**Identificação de sistema dinâmico
em dados de estoque imobiliário**

Luiz Paulo Medina de Lima

TEXTO APRESENTADO
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
A OBTENÇÃO DO TÍTULO
DE
MESTRE EM CIÊNCIAS DA COMPUTAÇÃO

Programa: Ciência da Computação

Orientador: Prof. Dr. Junior Barrera

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro do CNPq

São Paulo, Agosto de 2018

Identificação de sistema dinâmico em dados de estoque imobiliário

Esta versão da dissertação contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 24/08/2018. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Junior Barrera - IME-USP (orientador)
- Prof. Dr. Edvaldo Angelo - Mackenzie
- Dr^a. Vanessa Gapriotti Nadalin - IPEA

Agradecimentos

Gostaria de expressar minha sincera gratidão ao meu orientador, Prof. Dr. Junior Barrera, que acreditou no meu potencial desde o início. Sua motivação, paciência e entusiasmo em me ensinar foram essenciais para a realização deste trabalho. Mais do que isso, este trabalho não teria sido possível caso o Prof. Barrera não tivesse me admitido como orientando, acreditando em minha capacidade mesmo dado minha formação em administração de empresas.

Aproveito aqui a oportunidade de agradecer também à Dra. Vanessa Gapriotti Nadalin, ao Prof. Dr. Walter Mascarenhas, e ao Prof. Dr. Edvaldo Angelo. À Vanessa, por ter me introduzido à área de mercados imobiliários e por ter me ajudado a encontrar meu problema de pesquisa. Embora nunca tenhamos nos conhecido pessoalmente, nossas conversas por Skype foram fundamentais. Ao Prof. Walter, não somente por ter alterado o semestre de uma de suas disciplinas para que eu pudesse participar da mesma, mas também por estar sempre disponível para me auxiliar com diversas dúvidas que surgiram no decorrer do trabalho. Ao Prof. Edvaldo, por ter altruisticamente me aconselhado no final desta pesquisa, me ajudando a polir a forma final deste texto. Todos os três me ajudaram muito além do que seria esperado, e eu nunca me esquecerei da ajuda que recebi de cada um individualmente.

Agradeço a minha mãe e a minha namorada, não só pelo apoio incondicional que recebi das duas durante esse longo e árduo trabalho, mas também pela leitura e revisão do texto final. Vocês duas com certeza tornaram o texto, algumas vezes complexo em demasia, algo mais agradável de se ler e compreender.

Não posso esquecer dois grandes amigos que fiz durante estes anos, entre outros que fizeram parte desta aventura, Rodrigo Faria e Rodrigo Teixeira. A motivação provida por estes dois indivíduos ao longo desses anos foi essencial para me manter focado no objetivo final. Nossas discussões de economia com certeza enriqueceram a minha compreensão sobre o funcionamento de mercados imobiliários.

Por último, mas não menos importante, gostaria de agradecer à F3 Capital, empresa em que trabalhei durante o último ano da realização deste trabalho. Eles foram extremamente solidários e compreensivos em todos os momentos difíceis que enfrentei ao longo deste projeto.

Vocês foram todos jogadores importantes nessa conquista!

Resumo

LIMA, L. P. M. **Identificação de sistema dinâmico em dados de estoque imobiliário**. Mestrado - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2018.

Modelos preditivos de mercado são ferramentas importantes para tomadores de decisões no âmbito público e privado. Devido à complexidade dinâmica do mercado imobiliário, composta pela interação de dois submercados distintos – mercado de ativos imobiliários e mercado de consumo de espaço – e pela limitação de dados disponíveis, o estudo analítico de mercados imobiliários requer a modelagem paramétrica de um sistema de equações que os descrevam, seguido pela identificação dos parâmetros deste sistema utilizando dados reais de uma região. Neste trabalho, estudamos o modelo dinâmico de mercado imobiliário proposto por [Wheaton \(1999\)](#), criado a partir do popular modelo de quatro quadrantes de autoria de [DiPasquale e Wheaton \(1996\)](#). Utilizamos técnicas de identificação de sistemas para elaborar um modelo de aprendizado para o estoque imobiliário, e o implementamos em Matlab. Aplicamos o método elaborado em dados simulados, para validá-lo, e então aplicamos o mesmo método, com adaptações, em dados reais do mercado imobiliário canadense. Os resultados obtidos validam o método de identificação de sistema dinâmico quando testado em dados simulados, e corroboram o modelo de [Wheaton \(1999\)](#) como modelo preditivo em dados reais. Ademais, os resultados indicam que um modelo que seja capaz de entender a evolução dinâmica dos parâmetros estáticos do modelo de [Wheaton \(1999\)](#), poderia melhorar os resultados deste como ferramenta preditiva.

Palavras-chave: mercado imobiliário, identificação de sistemas, sistema dinâmico, dinâmica de sistemas.

Abstract

LIMA, L. P. M. **Dynamical system identification in real estate stock data**. Thesis (Master's Degree) - Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2018.

Predictive market models are important tools for decision-makers in the public and private spheres. Due to the dynamic complexity of the real estate market, consisting of the interaction of two distinct submarkets - real estate asset market and space consumption market - and the lack of real estate data, the analytical study of real estate markets requires the parametric modeling of a system of equations describing them, followed by the identification of the parameters of this system using real data from a region. In this work, we study the dynamic real estate market model proposed by [Wheaton \(1999\)](#), created from the popular four-quadrant model of [DiPasquale e Wheaton \(1996\)](#). We use system identification techniques to develop a learning model for real estate inventory data, and implement it in Matlab. We apply the method devised in simulated data to validate it, and then apply the same method with adaptations in real data of the Canadian real estate market. The results validate the dynamic system identification method when tested in simulated data, and corroborate the [Wheaton \(1999\)](#) model as a predictive model in real data. In addition, the results indicate that a model that is able to understand the dynamic evolution of the static parameters of the [Wheaton \(1999\)](#) model, could improve its results as a predictive tool.

Keywords: real estate, system identification, dynamical system, system dynamics.

Sumário

Lista de Figuras	ix
Lista de Tabelas	xi
1 Introdução	1
1.1 Motivação	2
1.2 Definição do problema	3
1.3 Objetivo	4
1.4 Contribuição	4
1.5 Limitações da pesquisa	4
1.6 Organização do trabalho	5
2 Fundamentação Teórica	7
2.1 Conceitos econômicos	7
2.1.1 Mercados	7
2.1.2 Modelo de oferta e demanda	7
2.1.3 Elasticidades	10
2.2 Modelagem de mercados imobiliários	11
2.2.1 Modelo 4QM	13
2.2.2 Modelo dinâmico de Wheaton (1999)	18
2.3 Identificação de sistemas dinâmicos	21
2.3.1 Classificação de métodos e modelos	22
2.3.2 Métodos de Erro de Predição (PEM)	24
2.3.3 Validação, seleção e teste de modelos	28
3 Experimentos	33
3.1 Experimentos em dados simulados	33
3.1.1 Modelo e método	33
3.1.2 Dados simulados	37
3.1.3 Experimentos	38
3.2 Experimentos em dados reais	39
3.2.1 Modelo e método	40
3.2.2 Dados do mercado imobiliário canadense	44
3.2.3 Experimentos	47

4	Resultados e análise	51
4.1	Experimentos em dados simulados	51
4.1.1	Parâmetros obtidos	51
4.1.2	Predição em dados de teste	52
4.1.3	Simulação em dados de teste	53
4.2	Experimentos em dados reais	54
4.2.1	Parâmetros obtidos	54
4.2.2	Predição em dados de teste	55
4.2.3	Simulação em dados de teste	56
5	Conclusões	59
A	População	61
B	Estoque imobiliário	67
C	Simulações	73
	Referências Bibliográficas	79

Lista de Figuras

2.1	Curva de oferta e demanda.	8
2.2	Deslocamentos da curva de demanda.	9
2.3	Deslocamentos da curva de oferta.	10
2.4	Alterações no ponto de equilíbrio.	11
2.5	Ilustração do modelo 4QM	14
2.6	Modelo 4QM: deslocamento da curva de demanda pelo uso do espaço	16
2.7	Modelo 4QM: exemplo de alteração na taxa de juros da economia	17
2.8	Modelo 4QM: deslocamento da curva de oferta de novo espaço imobiliário	18
2.9	Simulação 1 do modelo de Wheaton (1999)	20
2.10	Simulação 2 do modelo de Wheaton (1999)	21
2.11	Simulação 3 do modelo de Wheaton (1999)	21
2.12	Simulação 4 do modelo de Wheaton (1999)	22
2.13	Exemplo de adequação de curvas	24
2.14	<i>Validação de modelos</i>	29
2.15	Seleção de modelos	30
2.16	<i>Leave-one-out cross-validation</i>	32
3.1	Gráficos das séries de dados de estoque simuladas	38
3.2	Experimento com dados simulados	39
3.3	Mapa político do Canadá	45
3.4	Séries temporais população de Newfoundland e Labrador	47
3.5	Séries anuais de estoque imobiliário – Canada e Províncias	48
3.6	Experimento com dados reais	49
4.1	Resultados experimentos – dados simulados	53
4.2	Erro percentual absoluto médio por ano de simulação	56
4.3	Erro percentual por ano de simulação por região	56
4.4	Gráfico simulação Canadá	57
4.5	Resultados experimentos – dados reais	58
A.1	População Alberta	61
A.2	População British Columbia	61
A.3	População Canadá	62
A.4	População Manitoba	62
A.5	População New Brunswick	63

A.6	População Newfoundland e Labrador	63
A.7	População Nova Scotia	64
A.8	População Ontario	64
A.9	População Prince Edward Island	65
A.10	População Quebec	65
A.11	População Saskatchewan	66
B.1	Estoque imobiliário Alberta	67
B.2	Estoque imobiliário British Columbia	67
B.3	Estoque imobiliário Brunswick	68
B.4	Estoque imobiliário Manitoba	68
B.5	Estoque imobiliário Newfoundland e Labrador	69
B.6	Estoque imobiliário Nova Scotia	69
B.7	Estoque imobiliário Ontario	70
B.8	Estoque imobiliário Prince Edward Islands	70
B.9	Estoque imobiliário Quebec	71
B.10	Estoque imobiliário Saskatchewan	71
C.1	Estoque simulado versus real – Alberta	73
C.2	Estoque simulado versus real – British Columbia	74
C.3	Estoque simulado versus real – Canadá	74
C.4	Estoque simulado versus real – Manitoba	75
C.5	Estoque simulado versus real – New Brunswick	75
C.6	Estoque simulado versus real – Newfoundland e Labrador	76
C.7	Estoque simulado versus real – Nova Scotia	76
C.8	Estoque simulado versus real – Ontário	77
C.9	Estoque simulado versus real – Prince Edward Islands	77
C.10	Estoque simulado versus real – Quebec	78
C.11	Estoque simulado versus real – Saskatchewan	78

Lista de Tabelas

3.1	Constantes, entradas e estados iniciais	37
3.2	Valores de parâmetros utilizados por simulação	37
3.3	Parâmetros convertidos por simulação	37
3.4	Restrição do espaço de busca	44
3.5	Dimensões das séries populacionais do Canadá e Províncias. Fonte: adaptado de Canada (2016c)	47
3.6	Dimensões das séries de estoque imobiliário do Canadá e Províncias. Fonte: adaptado de Canada (2016a)	48
3.7	Dimensões das séries de estoque e população utilizadas. Fonte: adaptado de Canada (2016a,c)	49
4.1	Constantes, entradas e estados iniciais	52
4.2	Resultados de experimentos em dados simulados	53
4.3	Resultados de experimentos em dados reais	54
4.4	Resultados de experimentos em dados reais	55

Capítulo 1

Introdução

Desde meados da década de 80, vários estudos foram conduzidos com o intuito de examinar os movimentos cíclicos presentes em alguns tipos de mercado imobiliário (Barras, 1994; Grenadier, 1995; Kaiser, 1997; King e McCue, 1987; Rosen e Smith, 1983; Voith e Crone, 1988; Wheaton, 1987; Wheaton *et al.*, 1997). Embora esta tenha sido uma área de interesse para alguns pesquisadores, nenhuma teoria geral de ciclos existe, com pesquisadores diferentes adotando perspectivas variadas para explicar, entender e prever ciclos (Rottke *et al.*, 2003). Um dos modos de se analisar e prever ciclos no mercado imobiliário é através do estudo da evolução do estoque imobiliário ao longo do tempo, de forma a identificar movimentos endógenos ao sistema imobiliário de excesso ou estagnação na construção de novas unidades, em contraste a movimentos em resposta a choques de demanda e oferta.

A análise dinâmica da economia é quase tão antiga quanto o próprio campo de economia (Schumpeter, 1954). Mesmo modelos matemáticos dinâmicos de fenômenos econômicos possuem uma longa história no desenvolvimento da área. O modelo de teia de aranha (*cobweb model*), por exemplo, que tenta explicar o ajuste de preços ao longo do tempo em determinados mercados, data de pelo menos 1887, embora o termo "teia de aranha" só tenha sido cunhado por Kaldor, em 1934 (Kaldor, 1934; Puu, 1991).

O surgimento da metodologia de dinâmica de sistemas, na década de 1950, deu ímpeto ao estudo de sistemas dinâmicos econômicos e empresariais através de modelos de estoque-fluxo (*stock-flow models*), que se destinam a estudar o comportamento não-linear de sistemas dinâmicos ao longo do tempo através da modelagem da interação de estoques, fluxos e ciclos de retroalimentação (*feedback loops*). Este método de estudo de sistemas complexos ganhou bastante popularidade nas décadas de 80 e 90, sendo a utilidade destes modelos sua capacidade de explicar padrões observáveis de comportamento a partir da análise da estrutura de micro-nível do fenômeno de interesse, e dos processos de tomada de decisão dos agentes do sistema (Oliva, 2003).

Em linha com o estudo de sistemas complexos através da modelagem e simulação matemática proposta pela escola de dinâmica de sistemas, Wheaton (1999) formulou um modelo de sistema estoque-fluxo para explicar a existência de ciclos imobiliários de construção em alguns mercados com base em um modelo estático conceitual, conhecido como o modelo de quatro-quadrantes, proposto, em sua concepção original, por DiPasquale e Wheaton (1992) (Eskinasi *et al.*, 2011). Enquanto o modelo de DiPasquale e Wheaton (1992) modelava estaticamente os equilíbrios obtidos no mercado imobiliário a partir de diferentes características da economia local, e permitia entender como a alteração de preferências dos consumidores ou de fatores econômicos poderia impactar estes equilíbrios, o modelo de Wheaton (1999) se preocupava em entender o comportamento do mercado obtido no período de transição entre um equilíbrio e outro.

O modelo proposto por Wheaton (1999) oferece uma possível explicação sobre o motivo pelo qual alguns mercados apresentam oscilações sem aparente conexão com a evolução estável da economia, enquanto outros parecem responder mais fidedignamente aos choques econômicos ao longo do tempo. Este modelo, devido sua simplicidade matemática e fundamentação teórica, é famoso entre pesquisadores do mercado imobiliário (Eskinasi *et al.*, 2011; Mashayekhi *et al.*, 2009; Zhang *et al.*,

2015).

As conclusões obtidas por Wheaton (1999) foram baseadas na simulação de dados com base em parâmetros considerados economicamente plausíveis para determinados tipos de mercado, prática também conhecida na literatura econômica e de dinâmica de sistemas como *calibração do modelo* (Oliva, 2003). A ideia por trás da prática de calibração de modelos, consiste em utilizar parâmetros razoáveis para simulação dos modelos, e observar se o comportamento obtido se assemelha ao comportamento do fenômeno objeto de estudo.

Embora a calibração de modelos econômicos tenha validade acadêmica no estudo de dinâmica de sistemas, diversas técnicas existem para se parametrizar modelos dinâmicos. Um estudo sistemático das principais técnicas disponíveis para identificação de sistemas pode ser encontrada em Åström e Eykhoff (1971). A família de técnicas mais utilizada para parametrização de sistemas dinâmicos é conhecida como *prediction-error methods* (métodos do erro de predição), ou PEM, e tem como finalidade parametrizar um modelo matemático de um sistema dinâmico a partir de um conjunto de mensurações de entrada/saída. Estas técnicas têm parentesco próximo com o *maximum likelihood method* (método de máxima-verossimilhança), ou ML, primeiro proposto por Fisher (1912), e aplicado ao contexto de sistemas dinâmicos por Åström e Torsten (1965) e Box *et al.* (2015).

A parametrização do sistema dinâmico implícito no modelo de Wheaton (1999), com base em dados reais do mercado imobiliário e através de técnicas de identificação de sistemas, poderia corroborar o modelo de Wheaton (1999), além de proporcionar ao usuário um método mais simples de utilizar este modelo para previsões e simulações do que o método de calibração das variáveis econômicas deste modelo.

Séries temporais de dados de estoque imobiliário são difíceis de encontrar, entretanto, e mesmo quando tais dados existem, nem sempre são de fácil acesso. Uma das agências internacionais que mantém este histórico de dados para o seu país é a Agência de Estatística do Canadá (*Statistics Canada*), que disponibiliza, através do seu *site*, a série de dados de estoque imobiliário habitacional, de 1941 até 2000, para o Canadá e províncias.

Apesar do modelo proposto por Wheaton (1999) ser um modelo genérico para o mercado imobiliário, sua construção foi realizada para explicar oscilações observadas em mercados imobiliários de áreas comerciais. Zhang *et al.* (2015) sugere que o modelo pode ser adaptado para outros tipos de mercados, embora não realize tal adaptação, optando por estudar o mercado imobiliário Chinês sob uma perspectiva de dinâmica de sistemas.

Neste trabalho, adaptamos o modelo de Wheaton (1999) para se adequar aos dados canadenses disponíveis, e utilizamos técnicas de identificação de sistemas, propostas por Ljung (2002, 2006, 2010), para parametrizar o modelo adaptado e verificar seu desempenho em dados reais.

1.1 Motivação

Modelos preditivos são ferramentas importantes para agentes econômicos, cujas decisões requerem conhecimento, tanto dos estados futuros de aspectos da economia onde atuam, quanto dos resultados que suas escolhas podem ter sobre estes mesmos mercados. Investidores, interessados em maximizar os retornos sobre seus capitais, necessitam de modelos que os auxiliem a decidir se a possibilidade de ganho em um empreendimento compensa o risco associado a este, enquanto organizações governamentais precisam entender os impactos de diferentes ações de controle e coordenação econômica para que possam atuar de maneira eficiente.

O setor imobiliário é um exemplo de mercado em que modelos preditivos confiáveis seriam de grande valor para tomadores de decisão, públicos e privados, e em que dados agregados são de difícil aquisição, como já apontado por Hendershott *et al.* (1999). Especificamente, a capacidade de se prever oscilações no setor imobiliário é importante tanto para investidores quanto para organizações estatais de controle, já que oscilações endógenas são decorrentes de má alocação de capital no setor imobiliário (Wheaton, 1999).

Em problemas marcados pela escassez de dados, a necessidade de modelar matematicamente o

fenômeno de interesse e delimitar o espaço de hipóteses da representação desenvolvida da melhor maneira possível é essencial para que modelos de aprendizado possam encontrar equações contidas dentro do espaço de hipóteses com boa generalização e baixo erro de predição (Abu-Mostafa *et al.*, 2012). Para tanto, um estudo aprofundado do problema e da literatura disponível sobre o mesmo é necessário de forma a delimitar o espaço de hipóteses através de uma modelagem matemática coerente.

Diversos modelos preditivos já foram propostos para estudar o comportamento de preços no mercado imobiliário (Cain e Janssen, 1995; Limsombunchao, 2004; Liu *et al.*, 2006; Rafiei e Adeli, 2015; Worzala *et al.*, 2009), com alguns modelos destinados especificamente a estudar o comportamento oscilatório do mercado imobiliário (Barlas *et al.*, 2007; Belej e Kulesza, 2015), mas pouca atenção é dada para o estudo do comportamento do estoque imobiliário, mesmo dado que os preços e estoque estão inter-relacionados através da oferta e demanda da economia. O desenvolvimento de modelos matemáticos de estoque, para o mercado imobiliário, permite a tomadores de decisões no âmbito de políticas públicas avaliar previamente os efeitos de possíveis políticas a serem adotadas, enquanto o mercado privado pode utilizar estes mesmos modelos para tentar evitar bolhas de excesso e escassez de oferta, tornando o mercado mais eficiente e alocando melhor os recursos na economia.

Além dos benefícios sociais do desenvolvimento de modelos matemáticos de estoque imobiliário, a parametrização do modelo de Wheaton (1999) corrobora as conclusões obtidas por este, na medida em que os resultados obtidos pelo modelo parametrizado coincidem com os resultados teóricos esperados. Esta metodologia de teste de sistemas dinâmicos pode também ser empregada em outros problemas estudados através da metodologia de dinâmica de sistemas, para a qual a literatura sobre validação de modelos é escassa (Barlas, 1996).

Por fim, um último ponto que motiva a realização deste trabalho é apontar para a deficiência de dados na área de estudos imobiliários, que dificulta o estudo aprofundado do mesmo. A criação e manutenção de mais séries de dados de estoque imobiliário, como as disponibilizadas pela Agência Nacional de Estatística do Canadá, permitiria que mais estudos como este pudessem ser feitos e mais modelos pudessem ser testados para entender o comportamento do estoque habitacional, comercial e industrial na economia.

1.2 Definição do problema

O estoque imobiliário em uma economia evolui de maneira dinâmica ao longo do tempo, sendo propenso a ciclos, devido à durabilidade de seus ativos e à grande janela de tempo existente entre a demanda por capital e sua subsequente entrega. Decisões de investimento são tomadas com base nas informações disponíveis no momento e nas perspectivas que os agentes tem em relação a como o mercado estará no fim da construção das unidades habitacionais (Wheaton, 1999).

Vários estudos já foram feitos para apontar o caráter dinâmico e cíclico do mercado de imóveis e para propor modelos para esses mercados (Case e Quigley, 1991; DiPasquale e Wheaton, 1992, 1996; Hendershott *et al.*, 2002a, 1999, 2002b; Pyhrh *et al.*, 2009; Wheaton, 1987, 1999; Wheaton *et al.*, 1997). O modelo de Wheaton (1999), normalmente usado como base para o desenvolvimento de outros modelos (Zhang *et al.*, 2015), no entanto, não foi desenvolvido com o propósito de realizar previsões quantitativas sobre os estados futuros de mercados reais, sendo sua finalidade entender o comportamento oscilatório teoricamente obtido pelo modelo caso este corresponda a realidade.

O presente estudo foi elaborado com a finalidade de utilizar o modelo de sistema dinâmico proposto por Wheaton (1999), desenvolvido para o mercado imobiliário de escritórios, e adaptá-lo para o mercado imobiliário de moradias urbanas, e, a partir deste modelo, desenvolver um algoritmo baseado em ferramentas de identificação de sistemas para encontrar os melhores parâmetros que descrevem um conjunto de dados de estoque habitacional. O modelo original de Wheaton (1999), descrito por uma série de equações relacionadas, foi adaptado de forma a torná-lo passível de ser aprendido através de técnicas de identificação de sistemas. Outras adaptações realizadas foram necessárias para adequar o modelo ao mercado imobiliário habitacional, e, por fim, para adequar o

modelo aos dados disponíveis.

O modelo foi inicialmente testado nos dados simulados pelo artigo de [Wheaton \(1999\)](#), para verificar a convergência dos parâmetros aprendidos para os parâmetros corretos. Em sequência, o modelo foi testado em dados de estoque de moradia do mercado imobiliário do Canadá, disponibilizados pela agência nacional de estatística do Canadá (*Statistics Canada*) ([Canada, 2016a](#)). A escolha dos dados se deve à completude e ao tamanho das séries temporais disponibilizadas pela *Statistics Canada*, características difíceis de se encontrar em dados de mercado imobiliário.

1.3 Objetivo

O objetivo deste estudo é desenvolver um algoritmo preditivo de estoque do mercado imobiliário, a partir do modelo dinâmico proposto por [Wheaton \(1999\)](#), capaz de parametrizar o modelo proposto pelo autor com base em dados de estoque imobiliário, e testar este mesmo algoritmo em dados reais do mercado imobiliário para verificar se as conclusões extraídas originalmente por [Wheaton \(1999\)](#) coincidem com os resultados obtidos.

1.4 Contribuição

A pesquisa desenvolvida contribui para a área de pesquisa de mercados imobiliários tanto do ponto de vista acadêmico, quanto do ponto de vista social e econômico.

O desenvolvimento de um algoritmo preditivo capaz de aprender os parâmetros do modelo de sistema dinâmico proposto por [Wheaton \(1999\)](#) contribui, especificamente, das seguintes maneiras:

- Contribuição social
 - O desenvolvimento de um algoritmo capaz de parametrizar o modelo de [Wheaton \(1999\)](#) permite que agentes econômicos privados e públicos possam entender melhor as características econômicas da região em que atuam, desde que tenham dados sobre o mercado;
 - A capacidade de prever possíveis ciclos endógenos do mercado imobiliário de uma região possibilita que agentes econômicos tomem melhores decisões de alocação de capital, podendo, no limite, evitar que estas oscilações ocorram.
- Contribuição acadêmica
 - A adaptação do modelo para o mercado imobiliário habitacional permite estender as conclusões obtidas para o mercado imobiliário comercial para outros mercados;
 - O teste do modelo, através de métodos de identificação de sistema, confere maior confiança aos resultados obtidos por [Wheaton \(1999\)](#);
 - O procedimento metodológico pode ser utilizado para outros modelos de dinâmica de sistemas testados através da calibração e simulação de sistemas.

1.5 Limitações da pesquisa

Este trabalho é limitado pelas séries de dados disponíveis para sua realização. Idealmente, gostaríamos de testar o modelo de [Wheaton \(1999\)](#) em diferentes séries de dados, de diferentes indústrias e regiões, de forma a testar os resultados do modelo em séries temporais cíclicas e acíclicas, com diferentes comportamentos econômicos. Conforme mencionado anteriormente, estes tipos de dados são de difícil obtenção, e as únicas séries temporais adequadas para a realização deste trabalho que encontramos foram as séries temporais de estoque imobiliário residencial canadense, e, portanto, a pesquisa só poderá verificar os resultados do modelo de [Wheaton \(1999\)](#) parametrizado para as séries regionais específicas do Canadá e suas províncias, séries estas que não apresentam ciclos no período observado ([Fréchette, 2016](#); [Reynaud, 2015](#)).

1.6 Organização do trabalho

Neste capítulo, apresentamos uma breve introdução ao tema da pesquisa, explicando a sua inserção no contexto de pesquisas econômicas e de sistemas dinâmicos. São descritos aqui a motivação do estudo, a definição do problema a ser estudado, os objetivos, contribuições e limitações da pesquisa conduzida.

O Capítulo 2 é composto pela fundamentação teórica e pelos conceitos que serão utilizados nos demais capítulos. Inicialmente, inserimos uma breve descrição de conceitos econômicos básicos que podem não ser de conhecimento comum na área de Ciência da Computação, embora sejam importantes para a última seção deste Capítulo, que trata sobre a modelagem de sistemas habitacionais. A seguir, explicamos os dois modelos de mercado imobiliário relevantes para a condução deste trabalho, que são os modelos estático de DiPasquale e Wheaton (1992) e o modelo dinâmico de Wheaton *et al.* (1997). Em sequência, abordamos o assunto de identificação de sistemas dinâmicos, focando nas técnicas utilizadas nesta pesquisa, como métodos PEM, *cross-validation*, entre outros.

No Capítulo 3, descrevemos os experimentos realizados com base na análise de métodos PEM sugerida por Ljung (2006), tanto em dados simulados, para validar a capacidade do método desenvolvido de aprender corretamente os parâmetros de uma série temporal de estoques, quanto em dados reais, para verificar a capacidade do modelo em realizar previsões corretas sobre o estoque imobiliário futuro.

Por fim, nos capítulos 4 e 5, analisamos os resultados obtidos durante a etapa de testes e extraímos as conclusões com base nos mesmos.

Capítulo 2

Fundamentação Teórica

Este capítulo é dedicado a apresentar as bases teóricas necessárias ao desenvolvimento da pesquisa. Iniciamos o capítulo apresentando conceitos e terminologias básicos de economia empregados neste trabalho. Em seguida, apresentamos a modelagem de sistemas imobiliários, partindo do modelo estático de quatro-quadrantes de DiPasquale e Wheaton (1992), e evoluindo para o modelo dinâmico de Wheaton (1999).

Ao final do capítulo apresentamos as técnicas de identificação de sistemas utilizadas no método de parametrização do sistema dinâmico desenvolvido neste trabalho com base no modelo de Wheaton (1999).

2.1 Conceitos econômicos

No decorrer deste projeto, são empregados alguns termos e princípios econômicos que não ser de conhecimento comum para pesquisadores da área de Ciência da Computação. Introduzimos aqui, brevemente, conceitos básicos de economia que serão utilizados ao longo do texto. Sugerimos ao leitor interessado consultar as referências bibliográficas listadas, caso haja interesse.

2.1.1 Mercados

Em economia, mercados são definidos como um grupo de compradores e vendedores de um bem ou serviço em particular. Os compradores, em conjunto, definem a demanda por tal bem ou serviço, enquanto vendedores, também em conjunto, definem a oferta de tal bem ou serviço (Mankiw, 2014).

Mercados diferem em muitos aspectos, dependendo de suas características, como a forma em que são organizados, o valor médio dos bens ou serviços, o volume de operações que ocorrem, entre outros. Uma distinção útil de se traçar entre mercados é o quanto estes diferem em termos de competitividade.

Um mercado é dito perfeitamente competitivo quando as ações de qualquer comprador ou vendedor individual não afetam o preço pago pelos bens ou serviços deste mercado de maneira significativa. Este tipo de comportamento é comum em mercados compostos por um grande número de compradores e vendedores e nos quais os produtos vendidos não diferem em suas características, como é o caso do mercado de *commodities* (Krugman *et al.*, 2013; Mankiw, 2014).

Embora o mercado imobiliário não seja um mercado perfeitamente competitivo, é comum trabalhar-se com a premissa de que o mesmo é razoavelmente competitivo para ser modelado da mesma forma que mercados perfeitamente competitivos (DiPasquale e Wheaton, 1992, 1996).

2.1.2 Modelo de oferta e demanda

O modelo de oferta e demanda é usado para descrever o comportamento de mercados competitivos e os fatores que os afetam. Este modelo, segundo Krugman *et al.* (2013), é composto por cinco elementos:

1. a curva de demanda;
2. a curva de oferta;
3. fatores que levam ao deslocamento da curva de demanda e da curva de oferta;
4. o ponto de equilíbrio de mercado, composto pelo preço de equilíbrio e pela quantidade de equilíbrio;
5. a forma como o equilíbrio de mercado muda de acordo com o deslocamento das curvas de oferta e demanda.

Nas seções a seguir, explicamos cada um individualmente.

Curva de demanda

A demanda em um mercado representa o quanto os consumidores deste desejam consumir de um bem ou serviço para um determinado preço do mesmo, e costuma ser modelada através de uma função matemática, com sua representação gráfica denominada de *curva de demanda* (Krugman *et al.*, 2013; Mankiw, 2014).

A demanda por um bem ou serviço, no geral, está negativamente relacionada com o preço deste mesmo bem, resultando em uma curva de demanda decrescente em um gráfico de preço vs. demanda, como exibido na figura 2.1

Figura 2.1: Ilustração das curvas de oferta e demanda no mercado de algodão. As curvas se intersectam no ponto de equilíbrio do mercado, definido por um preço de equilíbrio e uma quantidade de equilíbrio. Fonte: adaptado de Krugman *et al.* (2013).

Curva de oferta

A curva de oferta representa o quanto os produtores de um bem ou serviço estão dispostos a vender dos mesmos por um determinado preço. Assim como ocorre na modelagem da demanda, a oferta é tipicamente modelada por uma equação matemática que estima a relação entre a oferta e o preço de um bem ou serviço (Krugman *et al.*, 2013; Mankiw, 2014).

Como produtores no geral são incentivados a produzir mais quanto maior o preço daquilo que vendem, tudo o mais constante, a curva de oferta apresenta uma inclinação positiva, como ilustrado na figura 2.1.

Deslocamento das curvas de demanda e oferta

A alteração na produção ou consumo de bem ou serviço, em virtude de eventos econômicos, pode levar à modificações nas curvas de oferta e demanda, normalmente denominadas de movimentos ou deslocamentos.

Um aumento na quantidade demandada para qualquer preço corresponde a um deslocamento da curva de demanda para a direita, enquanto uma diminuição na demanda corresponde a um deslocamento para a esquerda. Deslocamentos na curva de oferta são análogos a deslocamentos nas curvas de demanda, como exibido nas figuras 2.2 e 2.3.

Figura 2.2: Deslocamentos da curva de demanda para a direita – aumento da demanda – e para a esquerda – diminuição da demanda. Fonte: adaptado de *Krugman e Wells (2009)*.

Ponto de equilíbrio

O ponto de equilíbrio de um mercado é o ponto definido pelo preço de equilíbrio e pela quantidade de equilíbrio no qual a quantidade ofertada é a mesma que a quantidade demandada. Neste, todo comprador disposto a pagar o preço de equilíbrio pelo bem ou serviço, definido pelo mercado, consegue satisfazer sua demanda, da mesma forma, todo produtor consegue vender sua produção pelo preço de equilíbrio (*Krugman e Wells, 2009*).

O ponto de equilíbrio é encontrado igualando-se a função de demanda e a função de oferta, e é graficamente representado pelo ponto em que as duas curvas se tocam, como representado na figura 2.1.

Alterações no ponto de equilíbrio

Deslocamentos das curvas de oferta e demanda podem levar os mercados a estabelecerem novos pontos de equilíbrio, com possíveis alterações no preço de equilíbrio, na quantidade de equilíbrio, ou em ambos.

A figura 2.4 exibe duas alterações possíveis do ponto de equilíbrio de um mercado de algodão hipotético, devido a deslocamentos de magnitudes distintas nas curvas de oferta e demanda.

Figura 2.3: Deslocamentos da curva de oferta para a direita – aumento da oferta – e para a esquerda – diminuição da oferta. Fonte: adaptado de *Krugman e Wells (2009)*.

2.1.3 Elasticidades

Elasticidades medem a sensibilidade da quantidade demandada ou da quantidade ofertada em resposta à mudança de um de seus determinantes, normalmente o preço (*Robert e Bernanke, 2001*). Matematicamente, elasticidades são calculadas utilizando-se mudanças percentuais, de forma que a elasticidade-preço de uma função de demanda é dada por:

$$\beta_D = \frac{\partial Q}{\partial P} \frac{P}{Q} \quad (2.1)$$

onde β_D representa a elasticidade-preço da função de demanda, Q representa quantidade e P representa o preço.

Analogamente, a elasticidade-preço da oferta é dada por:

$$\beta_O = \frac{\partial Q}{\partial P} \frac{P}{Q} \quad (2.2)$$

Embora a elasticidade-preço da função de demanda seja um número tipicamente negativo, já que um aumento no preço de um bem ou serviço, normalmente, produz uma redução na quantidade demandada, a elasticidade-preço da demanda é normalmente expressa em módulo, por simplicidade.

Elasticidades-preço de oferta e demanda são classificadas em cinco tipos:

1. Perfeitamente elástica ($\beta \rightarrow \infty$): a elasticidade-preço é dita perfeitamente elástica quando a quantidade demandada ou ofertada varia infinitamente em resposta a uma mudança de preço;
2. Elástica ($1 < \beta < \infty$): uma curva é classificada como elástica quando uma variação no preço gera uma variação maior na quantidade demandada ou ofertada;
3. Elasticidade unitária ($\beta = 1$): a elasticidade é classificada como unitária quando quantidades demandadas e ofertadas variam igualmente em resposta a variações nos preços;
4. Inelástica ($0 < \beta < 1$): uma curva é classificada como inelástica quando uma variação no preço causa uma variação menor na quantidade ofertada ou demandada de um bem ou serviço;

Figura 2.4: Alterações no ponto de equilíbrio causadas por movimentos nas curvas de oferta e demanda. Em (a), uma pequena contração na oferta somada a um grande aumento na demanda resultam em quantidade e preço de equilíbrio superiores; em (b), um pequeno aumento na demanda somado a uma grande diminuição na oferta resultam em um ponto de equilíbrio com preços superiores e quantidades inferiores. Fonte: adaptado de *Krugman et al. (2013)*.

5. Perfeitamente inelástica ($\beta = 0$): a elasticidade-preço é dita perfeitamente inelástica quando a quantidade demandada ou ofertada não varia em resposta a uma variação do preço de um bem ou serviço.

2.2 Modelagem de mercados imobiliários

Propostas de modelagem matemática de mercados imobiliários existem há muito tempo na literatura, datando pelo menos de *Haig (1926)*. A literatura moderna sobre mercados imobiliários, no entanto, tem sido baseada em um modelo que descreve as inter-relações entre os mercados de ativos imobiliários e o mercado de uso de espaço (*Zhang et al., 2015*), conhecido como o modelo de quatro quadrantes (*four quadrant model - 4QM*) de autoria de *DiPasquale e Wheaton (1992)*.

O mercado imobiliário possui a peculiaridade de ser composto pela interação de dois sub-mercados distintos de compradores e vendedores. A distinção entre estes mercados fica clara quando moradores e proprietários de ativos imobiliários são pessoas distintas. Proprietários adquirem e vendem unidades habitacionais no mercado de ativos imobiliários, de acordo com sua expectativa de derivar renda de tal investimento ou transação, dando origem ao estoque de unidades habitacionais disponíveis. Este estoque é negociado pelo preço do metro quadrado na economia. A necessidade habitacional de moradores, por outro lado, e a qualidade e característica do espaço habitacional oferecido pelos proprietários, definem o aluguel de moradia no mercado de uso de espaço.

Embora a distinção seja melhor delineada quando moradores e proprietários são pessoas distintas, a mesma distinção ainda pode ser feita para o caso contrário de moradores habitando casas próprias, se considerarmos o aluguel do uso do espaço como sendo o valor obtido pelo serviço de habitação provido pelo ativo. Neste caso, um morador atua em ambos os mercados, tanto como investidor no mercado de ativos, quanto como consumidor do serviço de uso de espaço (*Zhang et al., 2015*).

Embasado na modelagem de oferta e demanda, discutida em 2.1.2, o modelo de quatro quadrantes, de autoria de *DiPasquale e Wheaton (1992)*, foi o primeiro a propor uma estrutura analítica destes dois mercados distintos, permitindo a usuários realizar estimativas sobre os possíveis impactos que diferentes mudanças no comportamento de investidores, nas condições macroeconô-

micas ou alterações na regulação imobiliária poderiam ter em ambos os mercados. Este modelo, no entanto, não permitia a compreensão dos estágios intermediários entre as mudanças de equilíbrio do sistema decorrentes de alterações nas condições macroeconômicas e financeiras do mercado. Para tanto, era necessário modelar o mercado através de um sistema de equações dinâmicas que o descrevessem.

A natureza dos estudos dinâmicos desenvolvidos sobre o mercado imobiliário tem origem na metodologia de modelagem de problemas dinâmicos conhecida como dinâmica de sistemas (Zhang *et al.*, 2015). Dinâmica de sistemas é uma abordagem auxiliada por computadores para construção de políticas, suporte de decisões estratégicas e desenvolvimento de teorias aplicada a sistemas dinâmicos marcados por interdependência, *feedback* informacional e causalidade circular (Richardson, 1991, 2009).

Os componentes utilizados para modelar a dinâmica de sistemas são (Radzicki, 2009):

- **Estoques:** acúmulos de informação ou material, que são afetados por fluxos positivos ou negativos ao longo do tempo, e são responsáveis pela memória do sistema, por criar atrasos e por alterar o comportamento dos fluxos;
- **Fluxos:** entradas ou saídas de informação ou materiais que afetam os estoques do sistema, de maneira dinâmica;
- **Feedback:** transmissão de informação a respeito dos níveis de acúmulo de estoque do sistema que afetam o comportamento dos seus fluxos. Podem ser de natureza positiva, reforçando o comportamento do sistema, ou negativa.

Estoques representam as variáveis de interesse do sistema, e são afetados por fluxos de entradas e saídas, que por sua vez são afetados por *feedbacks* de informação sobre os níveis dos estoques, estabelecendo uma dinâmica circular e endógena.

Apesar de estoques definirem as saídas do sistema, são os *feedbacks* de informação que determinam o comportamento dinâmico do sistema.

De acordo com Richardson (2009), a abordagem de dinâmica de sistemas envolve:

1. Definir problemas dinamicamente;
2. Procurar uma explicação endógena ao sistema para o seu comportamento;
3. Pensar em todos os componentes do sistema como quantidades contínuas interconectadas por *loops* de *feedback* de informação e causalidade circular;
4. Identificar estoques ou acumulações no sistema e suas entradas e saídas;
5. Formular um sistema dinâmico matemático capaz de reproduzir o comportamento do sistema real em questão;
6. Compreender o funcionamento do sistema e elaborar políticas para melhorá-lo.

Matematicamente, a estrutura formal de um modelo de dinâmica de sistemas é um sistema de equações diferenciais, ou equações de diferença, de primeira ordem, normalmente não-lineares, que governam o comportamento dos estoques do sistema (Radzicki, 2009). Dinâmica de sistemas, portanto, pode ser entendida como uma forma de se modelar matematicamente sistemas complexos presentes naturalmente no mundo real, através de sistemas dinâmicos de equações, com o objetivo de entender o comportamento dos mesmos de forma a alterá-los de maneira eficaz.

Em virtude da incapacidade do modelo 4QM de explicar o comportamento do mercado imobiliário no curto prazo, Wheaton (1999) propõe um modelo de equações dinâmicas inspirado no modelo estático 4QM para estudar a existência de ciclos de atividade e inatividade no setor de construção imobiliária comercial que não aparentavam apresentar conexão com o nível de atividade econômica, e que portanto pareciam endógenos ao sistema.

Em linha com a abordagem de dinâmica de sistemas, o autor desenvolve seu modelo estudando os ciclos do mercado imobiliário através do estudo do comportamento do estoque imobiliário comercial, identificando suas entradas e saídas, modelando o *feedback* informacional, e desenvolvendo um sistema dinâmico de equações capaz de simular seu comportamento. O autor então simula o modelo a partir de diferentes parâmetros para compreender seu comportamento, e, a partir dessas simulações, extrair suas conclusões.

Nas subseções seguintes, expomos o modelo 4QM de DiPasquale e Wheaton (1992) e o modelo dinâmico de Wheaton (1999), e explicamos seu funcionamento.

2.2.1 Modelo 4QM

O modelo de quatro-quadrantes utiliza o modelo de oferta e demanda, apresentado em 2.1.2, para unir conceitualmente as relações de equilíbrio presentes entre os mercados de ativos imobiliários e de uso do espaço.

Como construções imobiliárias são bens duráveis, o preço de unidades imobiliárias é determinado em um mercado de capitais, ou ativos, imobiliários. O preço de casas é determinado por quantas unidades familiares desejam possuir uma (ou mais) casas, e pelo volume de casas que compõem o estoque imobiliário deste tipo de moradia. Da mesma forma, o valor de escritórios comerciais depende de quantos investidores desejam possuir um (ou mais) escritórios, e de quantos escritórios existem no estoque imobiliário desta modalidade.

A oferta de novas unidades imobiliárias depende do preço que investidores acreditam que serão capazes de vender estas novas unidades, relativamente ao custo que incorrerão para construí-las. No longo prazo, o preço de novas unidades deve convergir para o custo¹ de produção das mesmas. No curto prazo, no entanto, com um estoque fixo para qualquer ponto no tempo, um aumento na demanda por novas unidades pode levar a preços superiores ao custo de construção, o que estimula produtores a construir novas unidades. Uma queda na demanda, por outro lado, pode levar o preço a cair abaixo do custo de produção, o que pode levar o mercado de construção a diminuir a produção de novas unidades.

No mercado pelo uso do espaço, a demanda é proveniente dos ocupantes do espaço, seja através da compra de um imóvel para uso próprio ou através do aluguel do mesmo. No setor comercial, o espaço é um dos diversos fatores de produção, e, portanto, seu consumo está relacionado com a receita da empresa relativamente ao custo do espaço. Para indivíduos e famílias, a decisão da área habitacional que desejam ter disponível para uso próprio é uma preferência que leva em conta o custo do espaço em relação ao custo de consumir outros bens e serviços.

O custo pelo uso do espaço, tanto para empresas quanto para pessoas, é o preço do aluguel do mesmo. Para inquilinos, este preço é determinado no contrato de aluguel. Para empresas e indivíduos que optam por comprar o imóvel, o preço pelo uso do espaço é o custo econômico de possuir o mesmo.

O preço do aluguel provém do mercado pelo uso do espaço, não do mercado de ativos imobiliários. A oferta de espaço para o mercado pelo uso do espaço, no entanto, provém do mercado de ativos imobiliários. A função do mercado pelo uso do espaço é determinar o nível de preços no qual a demanda por espaço é igual a oferta de espaço, fixa no curto prazo.

O modelo de quatro-quadrantes foi o primeiro modelo a ligar estes dois mercados em uma estrutura analítica coerente. O nome do modelo decorre da representação gráfica do mesmo, exibida na figura 2.5, que ilustra a curva de demanda pelo uso do espaço e a oferta de espaço imobiliário novo, ambas ocorrendo em mercados distintos, mas conectadas através de modelos de precificação e de formação de estoque;

A conexão entre os mercados pelo uso de espaço e de ativos imobiliários ocorre em dois pontos. Primeiramente, o preço do uso do espaço é um fator determinante na precificação de ativos imobiliários, já que o investimento em um ativo imobiliário representa um fluxo de caixa para o detentor

¹Nos referimos aqui ao conceito de custo econômico, que difere do conceito de custo contábil por levar em consideração o custo de oportunidade. Custo de oportunidade se refere a perda de ganho potencial de outras alternativas de investimento quando uma alternativa é escolhida.

do imóvel, e esse fluxo é determinado pelo preço de aluguel do mesmo. Uma alteração no nível de preços de aluguel, portanto, afeta imediatamente a demanda pelo mesmo. O segundo ponto de conexão entre estes dois mercados se dá através da construção de novas unidades habitacionais, que aumentam o estoque de espaço imobiliário na economia. Um aumento na construção de unidades imobiliárias aumenta o estoque de espaço disponível, reduzindo o preço pelo uso do mesmo.

Figura 2.5: Ilustração do modelo 4QM: o quadrante superior direito representa o mercado pelo uso do espaço, o quadrante superior esquerdo representa a avaliação dos ativos imobiliários a partir do preço do aluguel dos mesmos, o quadrante esquerdo inferior representa a construção de novo espaço no mercado de ativos imobiliários, e o quadrante inferior direito representa a composição do estoque imobiliário através da construção de novas unidades e a depreciação de unidades antigas. Fonte: adaptado de DiPasquale e Wheaton (1992).

Na figura 2.5, o quadrante superior direito representa o mercado pelo uso do espaço, o quadrante superior esquerdo representa a avaliação dos ativos imobiliário a partir do preço do aluguel dos mesmos, o quadrante esquerdo inferior representa a construção de novo espaço no mercado de ativos imobiliários, e o quadrante inferior direito representa a composição do estoque imobiliário através da construção de novas unidades e a depreciação de unidades antigas. No modelo, os quatro eixos definidores dos quadrantes emanam do centro, de forma que movimentos excêntricos representam aumentos positivos nas mensurações dos mesmos.

O quadrante superior direito, definido pelo eixo vertical de aluguel (R), medido em termos de valor por unidade de espaço por ano, e estoque (S), em unidade de espaço, contém a curva de demanda por espaço de um mercado imobiliário. A oferta de espaço é dada como fixa no curto prazo, e, como premissa, considera-se que a demanda iguale a oferta, de modo que o quadrante superior direito determina o valor do aluguel na economia. A demanda (D) é dada como uma função do aluguel e de condições econômicas do mercado:

$$D(R, Economia) = S \quad (2.3)$$

O quadrante superior esquerdo compartilha o eixo vertical de aluguel com o quadrante superior direito, e possui como eixo horizontal preço (P). A relação entre aluguel e preço (P) é dada pela

reta definida pela equação:

$$P = R/i \quad (2.4)$$

onde i representa a taxa de juros anual entre o aluguel por ano e o preço².

A taxa de juros i , em um mercado de agentes racionais, determina a escolha entre posse ou aluguel de uma propriedade, e é considerada no modelo como uma variável exógena, determinada em função de outros aspectos econômicos de onde o mercado imobiliário está inserido. Um aumento na taxa de juros é refletido no modelo como um movimento de rotação horário na reta determinada pela equação 2.4, enquanto sua diminuição tem o efeito contrário.

A função do quadrante superior esquerdo é ligar a demanda por espaço, no quadrante superior direito, com a oferta de novas unidades habitacionais, presente no quadrante inferior esquerdo.

O quadrante inferior esquerdo compartilha o eixo horizontal com seu quadrante superior, e possui como eixo vertical a construção de novo espaço habitacional (C). A curva de oferta é dada como função do preço:

$$C = f(P) \quad (2.5)$$

e sua angulação é determinada pela elasticidade da oferta de novo espaço em função do preço.

Uma maior elasticidade da oferta tornaria a curva mais vertical, com construtores oferecendo mais espaço em função de um mesmo preço, enquanto uma menor elasticidade da oferta tornaria a curva mais horizontal, com construtores menos responsivos a mudanças no preço do espaço.

A curva de oferta presente no quadrante inferior esquerdo toca o eixo horizontal em um valor positivo distinto de zero. Isso representa o fato de que quando o preço do espaço cai abaixo de um certo patamar, o custo de construção de novo espaço cai abaixo da renda que poderia ser obtida através da venda de novas unidades de espaço habitacional, e, nesse caso, a oferta de espaço novo passa a ser nula.

Por último, o quadrante inferior direito divide o eixo vertical de Construção com o quadrante inferior esquerdo, e o eixo horizontal de estoque com o quadrante superior direito. O estoque disponível na economia, é, então, dado pela equação:

$$S = \frac{C}{\delta} \quad (2.6)$$

onde δ representa a taxa depreciação real do estoque, ou seja, o quanto do estoque é demolido a cada ano.

Em resumo, a partir de um estoque de espaço, o mercado pelo uso do mesmo irá determinar um preço que equacione a demanda com o estoque disponível no momento. O preço resultante desta interação, será convertido em valores de imóveis, que irá determinar o nível de construção de novas unidades, que, por fim, resultará em mudanças no nível de estoque imobiliário através de novas construções. Como o modelo 4QM é um modelo de equilíbrio, a construção de novo espaço deve ser igual à destruição de espaço dentro do mesmo mercado, pois caso contrário o modelo não se encontraria em equilíbrio. Desta forma, valores divergentes de estoque nesse estágio do modelo, determinam um mercado que não se encontra em equilíbrio.

O aspecto final do modelo é dado pelo retângulo formado pelos valores de estoque, aluguel, preço e construção, que tocam cada uma das curvas presentes em cada quadrante em apenas um ponto. O retângulo representa graficamente o equilíbrio obtido quando as equações 2.3, 2.4, 2.5 e 2.6, em conjunto com as variáveis consideradas exógenas pelo modelo, possuem R^* , P^* , C^* e S^* que

²A equação 2.4 decorre da equação financeira de perpetuidade com fluxos de caixa constantes. Uma variável, no caso, o preço do aluguel, com valor inicial R_0 , que cresce anualmente a uma taxa composta i , atingirá em t anos o valor de $R_0(1+i)^t$. Devido a existência de taxas de juros, portanto, podemos considerar que o *valor presente* do aluguel no ano t , no valor de $R_0(1+i)^t$, é de R reais hoje. Da mesma forma, se considerarmos que um investidor, ao comprar um imóvel, espera receber um fluxo de caixa decorrente de aluguéis de R reais por ano, temos que, o valor presente deste imóvel é de $\sum_{t=1}^{\infty} \frac{R}{(1+i)^t}$, o que equivale a um valor presente de $\frac{R}{i}$ reais hoje. Recomendamos a leitura de Capinski e Zastawniak (2006) para uma discussão completa de modelagem matemática de fluxos de caixa.

as satisfaçam.

O modelo 4QM ficou famoso entre participantes do mercado imobiliário devido a elegância da sua simplicidade de representação, além de sua utilidade pedagógica, por permitir avaliar de maneira estruturada o efeito de longo prazo de mudanças nas condições de mercado de uma determinada economia (Colwell, 2002).

Um deslocamento da curva de demanda no mercado residencial para a direita, causada por um aumento no número de habitantes de uma determinada região, por exemplo, pode ter seus efeitos de longo prazo analisados estudando os efeitos decorrentes no modelo 4QM, conforme a figura 2.6 leva a um aumento no nível de preços de alugueis. Este aumento do preço do aluguel, por sua vez, leva a um aumento do valor dos ativos imobiliários, tudo mais mantido constante. Um preço maior de ativos imobiliários, estimula a construção de novas unidades por parte de investidores, que no longo prazo leva o estoque imobiliário residencial para um patamar superior. A magnitude de impacto das variações depende da inclinação das curvas presentes nos quatro quadrantes. Se a curva de oferta de novas unidades tivesse alta elasticidade, por exemplo, um pequeno aumento no nível de preços levaria a uma construção de um número consideravelmente maior de novas unidades imobiliárias.

Figura 2.6: Modelo 4QM: exemplo de deslocamento da curva de demanda pelo uso do espaço. O deslocamento da curva de demanda no mercado residencial para a direita leva a um aumento no nível de preços de alugueis. Este aumento do preço do aluguel, por sua vez, leva a um aumento do valor dos ativos imobiliários, tudo mais mantido constante. Um preço maior de ativos imobiliários, estimula a construção de novas unidades por parte de investidores, que no longo prazo leva o estoque imobiliário residencial para um patamar superior. Fonte: adaptado de DiPasquale e Wheaton (1992).

Uma redução na taxa de juros, por outro lado, ou qualquer fator que leve investidores a requererem menor retorno em seus investimentos imobiliários, pode ser analisado através de uma rotação anti-horária na curva de avaliação de valor imobiliário, no quadrante superior esquerdo, como representado na figura 2.7. Este tipo de mudança, com tudo mais permanecendo constante na economia, levaria a um aumento no nível de construção de forma a aproveitar o preço mais elevado de imóveis em relação ao seu custo de construção. Com um nível maior de construção, e uma mesma taxa de depreciação, o nível de estoque no longo prazo seria elevado a um novo patamar, que para uma mesma curva de demanda levaria a uma redução no aluguel pelo uso do espaço.

Outra variação exógena capaz de afetar o mercado imobiliário levantada pelos autores seria a alteração no custo de construção de novo espaço imobiliário. A alteração, ilustrada na figura

Figura 2.7: Modelo 4QM: exemplo de alteração na taxa de juros da economia. Uma diminuição na taxa de juros leva a um aumento no nível de preços de imóveis, o que estimula a oferta de novas construções, que leva a um maior nível estoque no equilíbrio, diminuindo o nível de preços do aluguel. Fonte: adaptado de DiPasquale e Wheaton (1992).

2.8 como um deslocamento excêntrico na curva de oferta de novas unidades no quadrante inferior esquerdo, diminui o nível de construção de novo espaço. No longo prazo, com tudo mais mantido constante, o aumento de custos de construção leva a um estoque menor de espaço, o que acarreta um preço maior pelo uso do mesmo e um valor maior para as unidades imobiliárias existentes.

O modelo de quatro-quadrantes, embora elegante em sua análise, se destina a estudar as mudanças de longo prazo resultantes de alterações nas condições macroeconômicas e financeiras de uma economia. De fato, em dois trechos de seu artigo original, os autores procuram deixar claro que o modelo não é adequado para analisar as alterações de curto prazo decorrentes de alterações em condições econômicas, nem para analisar os passos intermediários que levam o mercado imobiliário de um equilíbrio para um novo. Primeiramente, os autores advertem contra o uso de seu modelo para compreender a dinâmica de curto prazo do mercado imobiliário:

"É importante perceber que o diagrama de quatro-quadrantes representa um equilíbrio de longo prazo nos mercados pelo uso de espaço e de ativos imobiliários. O diagrama não é adequado para descrever a dinâmica do mercado de curto prazo ou os desequilíbrios temporários que frequentemente ocorrem no setor imobiliário." (DiPasquale e Wheaton, 1992, p. 190)

Os autores encerram o artigo com a sugestão de que para compreender os estágios intermediários entre equilíbrios, um modelo de sistema dinâmico é necessário:

"Esta estrutura analítica simples funciona bem para ilustrar os novos equilíbrios que resultam à medida que o ambiente econômico exógeno muda. Uma desvantagem importante deste modelo é que não é fácil rastrear os estágios intermediários à medida que o mercado se desloca para o seu novo equilíbrio. Para descrever os ajustes intermediários do mercado seria necessário um sistema dinâmico de equações, o que poderia complicar significativamente nossa análise." (DiPasquale e Wheaton, 1992, p. 197)

Figura 2.8: Modelo 4QM: exemplo de deslocamento da curva de oferta de novo espaço imobiliário. Uma diminuição na oferta de espaço imobiliário causada por um aumento de custos na cadeia produtiva acarreta um menor nível de construção de novas unidades, que leva a um menor nível de estoque no equilíbrio, aumentando, por sua vez, o preço de aluguéis e o preço por ativos imobiliários. Fonte: adaptado de DiPasquale e Wheaton (1992).

2.2.2 Modelo dinâmico de Wheaton (1999)

Embora DiPasquale e Wheaton (1992) utilizem o modelo 4QM para representar o efeito que choques econômicos têm no equilíbrio de longo prazo do mercado, os próprios autores destacam a incapacidade do modelo de representar as dinâmicas que ocorreriam em uma economia entre um equilíbrio e um outro novo equilíbrio na eventualidade de mudanças nas condições econômicas de um mercado.

Para suprir a deficiência do modelo 4QM em explicar a dinâmica de curto prazo do mercado imobiliário decorrente de alterações das condições macroeconômicas e financeiras de uma região, Wheaton (1999) propõe um modelo urbano dinâmico baseado no 4QM para explicar o comportamento cíclico de mercados urbanos que pode ser observado após choques econômicos (Wheaton, 1987; Wheaton *et al.*, 1997). A análise de Wheaton se concentra na modelagem do estoque imobiliário, em linha com a método de modelagem de sistemas dinâmicos promulgada pela área de dinâmica de sistemas, em contraste com outros trabalhos destinados a entender o comportamento dinâmico de preços de aluguéis em um mercado (Hendershott, MacGregor, e White, 2002a; Hendershott, MacGregor, e Tse, 2002b; Hendershott, Lizieri, e MacGregor, 2010).

O modelo dinâmico proposto por Wheaton (1999) é um modelos de estoque-fluxo (*stock-flow models*). Modelos de estoque-fluxo, muito usados em trabalhos de dinâmica de sistemas, são modelos de sistemas dinâmicos destinados a modelar a variação de estoques no tempo através de fluxos físicos e de informação. Estes modelos são utilizados na literatura de dinâmica de sistemas para modelar matematicamente os modelos conceituais propostos (Smith e van Ackere, 2002).

O autor, em seu artigo original, desenvolve seu modelo de forma a explicar a presença de ciclos imobiliários de atividade e inatividade de construção, aparentemente endógenos, nos mercados de escritórios comerciais. Embora o autor não realize simulações com base em dados habitacionais, o mesmo sugere que este modelo pode ser utilizado para entender o comportamento de outros tipos de mercados imobiliários, como por exemplo o mercado de espaço habitacional, hoteleiro e industrial.

O modelo de [Wheaton \(1999\)](#) é baseado em cinco equações, que dão origem a um sistema dinâmico dado por uma equação de diferença. Inicialmente, a demanda é modelada como sendo proporcional a uma variável econômica exógena ao sistema – *ie.* uma variável de entrada – E_t , a uma constante comportamental α_1 e inversamente proporcional ao aluguel pelo uso do espaço R_t , sujeito a elasticidade de demanda constante β_1 .

$$D_t = \alpha_1 E_t R_t^{-\beta_1} \quad (2.7)$$

O modelo supõe que a oferta iguala perfeitamente a demanda, e, desta forma, a demanda na equação 2.7 pode ser substituída pelo estoque (S_t) de espaço para alugar já existente, permitindo que o aluguel pelo uso do espaço possa ser calculado como:

$$R_t = \left(\frac{S_t}{\alpha_1 E_t} \right)^{-\frac{1}{\beta_1}} \quad (2.8)$$

O estoque a cada período é formado pela destruição de parte do estoque, modelada através da taxa de depreciação³ constante δ , e pelo acréscimo ao estoque de construções C_t que foram iniciadas n períodos atrás. O parâmetro n reflete a janela de tempo contida entre a decisão de construir por parte de investidores e a conclusão do projeto. A evolução do estoque, portanto, é dada pela equação de diferença a seguir:

$$\frac{S_t}{S_{t-1}} = 1 - \delta + \frac{C_{t-n}}{S_{t-1}} \quad (2.9)$$

A taxa de novas construções entregues no período t , é proporcional a uma constante comportamental α_2 e à estimativa de preço P_t para o período t que o mercado tinha em $t - n$, sujeita a elasticidade de demanda β_2 .

$$\frac{C_{t-n}}{S_{t-1}} = \alpha_2 P_t^{\beta_2} \quad (2.10)$$

Como ativos imobiliários são bens privados e negociados em um mercado altamente ilíquido, [Wheaton \(1999\)](#) defende a irracionalidade do mercado na formação de preços como sendo responsável pelo comportamento cíclico entre bolhas de super e sub-construção, citando a baixa liquidez dos ativos e ao fato de que grande parte do estoque existente de moradias é de propriedade de famílias e pequenos investidores. O modelo de irracionalidade em formação de preços proposto por [Wheaton \(1999\)](#) é dado por:

$$P_t = \frac{R_{t-n}}{r} \quad (2.11)$$

onde r representa o custo de capital dos investidores. A equação acima presume que investidores formam suas expectativas de preços de imóveis projetando os aluguéis vigentes no momento presente para o futuro.

Fica clara a relação entre o modelo de [Wheaton \(1999\)](#) e o modelo 4QM, na medida em que temos um estoque comum de espaço, que é determinado pelas decisões de investimento e consumo que ocorrem em dois mercados distintos de ativos imobiliários e de uso do espaço. Estas decisões estão representadas no modelo através das relações de investir em novas construções, pelos equilíbrio de preços de aluguéis determinados pela oferta e demanda pelo uso do espaço, e pela relação entre preços de imóveis e seus respectivos aluguéis.

Embora a representação gráfica em quatro quadrantes não seja possível para o modelo dinâmico desenvolvido por [Wheaton \(1999\)](#), cada uma das equações que compõe o sistema dinâmico são diferentes modelagens de curvas presentes nos quadrantes do modelo 4QM. As equações 2.7 e 2.8 são decorrentes da modelagem da curva de demanda pelo uso do espaço, representadas graficamente

³ δ também pode ser considerada como a taxa de crescimento do mercado imobiliário, sendo que o autor em diversas partes do texto se refere a δ como a taxa de depreciação-crescimento.

no modelo 4QM no quadrante superior direito. A equação 2.11 refere-se ao quadrante superior esquerdo, que representa avaliação de valor de ativos imobiliários com base no retorno esperado por participantes do mercado de ativos imobiliários. A equação 2.10 modela a construção de novas unidades imobiliárias, presente no quadrante inferior esquerdo do modelo 4QM. A evolução do estoque imobiliário, no entanto, modelada pela equação de diferença 2.9 não tem representação nos quatro quadrantes do modelo 4QM, apesar de estar relacionada com o quadrante inferior direito, isto porque o modelo estático de DiPasquale e Wheaton (1992) modela o estoque imobiliário como um equilíbrio de longo prazo, enquanto o modelo de Wheaton (1999) se destina a estudar exatamente a dinâmica de curto prazo deste estoque.

A simplicidade do modelo reflete a escassez de dados sobre mercados imobiliários em qualquer economia. O autor admite a possibilidade de o modelo criado, devido sua simplicidade, possuir um viés grande o suficiente de forma a tornar suas conclusões inválidas, mas, ao mesmo tempo, ao optar por um modelo simples, diminui a variância dos parâmetros aprendidos durante o processo de aprendizado, o que torna mais provável que os resultados obtidos sejam generalizáveis (Abu-Mostafa *et al.*, 2012).

O modelo apresenta, durante simulações, diferentes comportamentos em resposta a alterações na variável de entrada de acordo com a escolha de seus parâmetros, com o sistema apresentando pouca ou nenhuma oscilação endógena enquanto se move de um equilíbrio para outro, ou grandes oscilações, que variam em intensidade e frequência, conforme ilustrado a seguir.

Nos gráficos 2.9 a 2.12, parâmetros diferentes foram utilizados para medir o comportamento do sistema em resposta a um único choque de demanda de 50%, i.e., a variável entrada aumenta em 50% a partir do período zero e se mantém constante até o final da simulação.

Figura 2.9: Simulação 1 do modelo de Wheaton (1999). Reação de mercado a um choque de demanda de 50% (atraso: $n = 5$; depreciação: $\delta = 0,05$; elasticidade de demanda: $\beta_2 = 1,0$; elasticidade de oferta: $\beta_1 = 1,0$). O primeiro gráfico mostra a reação do sistema ao longo do tempo medida por Estoque/Emprego, enquanto o segundo gráfico mostra a mesma reação do sistema, porém medida em termos de Preço. Fonte: adaptado de Wheaton (1999).

Wheaton (1999) define ciclos no mercado imobiliário como oscilações repetidas de mercado ao redor de um estado constante resultantes de um único choque econômico. O modelo dinâmico de Wheaton (1999) foi proposto originalmente com o intuito de mostrar que, sob determinadas condições de mercado, ciclos no mercado imobiliário são possíveis.

Alguns resultados são obtidos através da simulação do modelo de Wheaton (1999) no que diz respeito à presença de ciclos no mercado imobiliário. Em primeiro lugar, para que oscilações ocorram no modelo dinâmico, é necessário que a elasticidade de oferta seja maior que a elasticidade de demanda, ou seja, $\beta_1 \leq \beta_2$. Outro resultado que pode ser observado é que a variável de atraso, n , afeta a frequência do ciclo imobiliário, com a frequência diminuindo quanto maior o atraso. Por último, a variável de depreciação de ativos imobiliários, δ , também impacta o comportamento de ciclos, tornando a frequência e amplitude de ciclos maiores quanto maior o seu valor.

Em resumo, mercados imobiliários apresentam estabilidade (i.e., não são propensos a ciclos) contanto que a demanda seja mais elástica que a oferta, e, caso a oferta seja mais elástica que a demanda, o mercado ainda apresenta estabilidade contanto que o período de construção seja curto

Figura 2.10: Simulação 2 do modelo de *Wheaton (1999)*. Reação de mercado a um choque de demanda de 50% (atraso: $n = 5$; depreciação: $\delta = 0,05$; elasticidade de demanda: $\beta_2 = 0,4$; elasticidade de oferta: $\beta_1 = 2,0$). O primeiro gráfico mostra a reação do sistema ao longo do tempo medida por *Estoque/Emprego*, enquanto o segundo gráfico mostra a mesma reação do sistema, porém medida em termos de *Preço*. Fonte: adaptado de *Wheaton (1999)*.

Figura 2.11: Simulação 3 do modelo de *Wheaton (1999)*. Reação de mercado a um choque de demanda de 50% (atraso: $n = 8$; depreciação: $\delta = 0,05$; elasticidade de demanda: $\beta_2 = 0,4$; elasticidade de oferta: $\beta_1 = 2,0$). O primeiro gráfico mostra a reação do sistema ao longo do tempo medida por *Estoque/Emprego*, enquanto o segundo gráfico mostra a mesma reação do sistema, porém medida em termos de *Preço*. Fonte: adaptado de *Wheaton (1999)*.

ou a taxa de depreciação-crescimento pequena.

Mercados habitacionais, como citado pelo próprio autor no texto, em contraste a mercados de áreas comerciais, não são propensos a ciclos. Este comportamento está em linha com as previsões do modelo, já que o tempo de construção de unidades habitacionais é muito menor que o período de construção de áreas comerciais, assim como a elasticidade de oferta e demanda são ambas baixas para o mercado habitacional, em contraste com a alta elasticidade de oferta do mercado comercial. O crescimento e depreciação também são bem menores no mercado habitacional quando comparados ao mercado comercial *Wheaton (1999)*.

2.3 Identificação de sistemas dinâmicos

Um sistema dinâmico é uma abstração matemática criada para estudar o comportamento ou evolução do estado de um sistema, definido como um conjunto de partes que estão inter-relacionadas. A seguir, emprestamos a definição de sistemas dinâmicos dada por *Barrera (2016)*

Dado x e u funções de $\mathbb{R}$ para $\mathbb{R}^n$, isto é, $x : t \rightarrow x(t)$ e $u : t \rightarrow u(t)$, e sendo ϕ uma função de $\mathbb{R}^n \times \mathbb{R}^n$ para $\mathbb{R}^n$ e, para qualquer $t \in \mathbb{R}$, deixe ϕ_t ser uma função de $\mathbb{R}^n \times \mathbb{R}^n$ para $\mathbb{R}^n$, então um sistema dinâmico pode ser definido pela equação:

$$\dot{x}(t) = \phi_t(x(t), u(t)), \forall t \in \mathbb{R} \quad (2.12)$$

onde $x(t)$ e $u(t)$ são denominados, respectivamente, o estado e a entrada do sistema dinâmico em

Figura 2.12: Simulação 4 do modelo de [Wheaton \(1999\)](#). Reação de mercado a um choque de demanda de 50% (atraso: $n = 5$; depreciação: $\delta = 0,10$; elasticidade de demanda: $\beta_2 = 0,4$; elasticidade de oferta: $\beta_1 = 2,0$). O primeiro gráfico mostra a reação do sistema ao longo do tempo medida por *Estoque/Emprego*, enquanto o segundo gráfico mostra a mesma reação do sistema, porém medida em termos de *Preço*. Fonte: adaptado de [Wheaton \(1999\)](#).

t , e a função ϕ é chamada de função de transição. $x_i(t)$ e $u_i(t)$ são chamados, respectivamente, de componente i do estado $x(t)$ e da entrada $u(t)$ no período t .

O estado inicial de um sistema dinâmico é o conjunto numérico mínimo que precisa ser determinado no tempo $t = t_0$ de modo a ser capaz de compreender o comportamento do sistema para qualquer instante $t \geq t_0$, ou seja, é a descrição mínima do passado do sistema necessário para prever seu comportamento futuro.

Identificação de sistemas, de uma maneira ampla, pode ser entendida como a ciência de construir modelos matemáticos de sistemas dinâmicos a partir de dados de entrada-saída ([Ljung, 2010](#)). O problema de identificação de sistemas consiste em, dado uma amostra de sinais gerados a partir de mensurações de um fenômeno da natureza, encontrar um elemento de uma dada família de sistemas dinâmicos que melhor aproxime o fenômeno real com base na amostra de dados.

Iniciamos esta seção com uma breve caracterização dos diferentes tipos de métodos e modelos de identificação de sistemas, introduzindo parte da terminologia utilizada no trabalho. Em seguida, explicamos os métodos de erro de predição, uma classe de métodos de identificação de sistemas muito popular devido a algumas de suas propriedades, e que compõe as técnicas utilizadas nesta pesquisa para parametrizar o modelo adaptado de [Wheaton \(1999\)](#). Por fim, tratamos sobre a predição, validação, simulação e teste de modelos dinâmicos aprendidos.

2.3.1 Classificação de métodos e modelos

Existem vários métodos disponíveis na literatura para se identificar sistemas dinâmicos a partir de dados. Estes métodos podem ser classificados de acordo com diferentes critérios, que, enquanto não homogêneos nem exaustivos, provém uma forma útil de se entender as diferentes possibilidades disponíveis para o usuário ([Ljung, 2006](#)):

Modelos paramétricos vs. modelos não paramétricos

Modelos paramétricos são gerados pela determinação de uma classe de modelos,

$$G = \{g(x, \theta) | \theta \in \mathbb{R}^n\} \quad (2.13)$$

onde θ é um vetor n -dimensional de parâmetros de $g(x)$, e a escolha do modelo ótimo é dada pela busca pelo θ^* ótimo, que normalmente é realizada pela minimização de uma métrica de erro ou pela maximização de um critério de verossimilhança, resultando em um vetor de parâmetros estimados $\hat{\theta}$. A função $\hat{g}(x)$ é então determinada por:

$$\hat{g}(x) = g(x, \hat{\theta}) \quad (2.14)$$

Modelos não paramétricos, por outro lado, não são determinados por vetores de parâmetros n -dimensionais, e, tipicamente, produzem uma estimativa de função $\hat{g}(x)$ através de médias de observações y_n relevantes.

Modelos globais vs. modelos locais

Modelos globais utilizam todos os dados de entrada, X , e suas respectivas observações, Y , para encontrar a estimativa $\hat{g}(x)$. Modelos locais utilizam apenas um conjunto de pares de observações (x_n, y_n) em que x_n encontra-se localmente próximo de x segundo alguma medida de localidade determinada. Modelos não paramétricos são, usualmente, locais.

Problemas de regressão vs. problemas de classificação

Em um problema de regressão, o objetivo é determinar uma função $\hat{g}(x)$ que retorne estimativas pertencentes a um conjunto numérico contínuo dependendo do valor de entrada x enquanto em classificação o objetivo é determinar a qual grupo ou classe x pertence.

Modelos caixa-branca, caixa-cinza e caixa-preta

Os modelos dinâmicos implícitos em um método são comumente classificados de acordo com o grau de especificidade do desenvolvimento deste mesmo modelo para tratar o problema em questão para o qual o método está sendo empregado. Esta classificação é dada através do espectro de “opacidade” do modelo, limitada entre os denominados *white-box models* (modelos caixa-branca) e *black-box models* (modelos caixa-preta).

A classificação de opacidade está relacionada com o grau de informação existente a priori sobre o problema que foi utilizado no desenvolvimento do modelo. Existe um amplo espectro de tonalidades de cinza existentes entre modelos caixa-branca e modelos caixa-preta. Como as nuances de opacidade de modelos não são relevantes para o desenvolvimento deste trabalho, classificamos os modelos intermediários contidos entre os modelos caixa-branca e caixa-preta como modelos caixa-cinza, uma nomenclatura comum na literatura de identificação de sistemas.

Modelos caixa-cinza são desenvolvidos a partir princípios conhecidos *a priori* sobre o fenômeno estudado, mas ao contrário de modelos brancos (*white models*), que são elaborados inteiramente a partir do estudo de algum fenômeno da natureza, possuem alguns parâmetros para serem estimados a partir de dados observados do fenômeno em si (Ljung, 2010).

Modelos pré-determinados vs. modelos sob demanda

Modelos pré-determinados são estimados a partir de todos os dados e podem ser, então, usados para calcular o valor para qualquer x desejado. Modelos sob demanda, por outro lado, são estimados para pontos distintos dos dados de treinamento apenas quando a demanda por tal é feita.

A distinção entre modelos pré-determinados e modelos sob demanda também existe no campo de aprendizado de máquina na distinção entre *eager learning* (aprendizagem ansiosa) e *lazy learning* (aprendizagem preguiçosa) (Aha, 1998), com as mesmas respectivas características. Modelos não paramétricos costumam ser de natureza preguiçosa, enquanto modelos paramétricos costumam ser pré-determinados durante o aprendizado.

No presente trabalho, definimos um espaço de hipóteses através da elaboração de um sistema dinâmico paramétrico através do estudo do mercado imobiliário, e em seguida elaboramos um método de estimar os parâmetros desse sistema de forma a encontrar o elemento contido no espaço de hipóteses delimitado pelas diferentes parametrizações do sistema dinâmico elaborado que melhor aproxime a amostra de dados em mãos.

Nesse sentido, o modelo desenvolvido neste trabalho é um modelo paramétrico, contendo três parâmetros a serem estimados inteiramente a partir de dados, e, devido a isso, o método de identificação de sistemas utilizado é um método global, de regressão, caixa-cinza e pré-determinado. Nas

subseções seguintes, explicamos brevemente as técnicas de estimação de parâmetros e as técnicas de validação que serão utilizadas neste projeto.

2.3.2 Métodos de Erro de Predição (PEM)

A construção de modelos de sistema dinâmico a partir de dados de entrada-saída pode ser realizada de diversas formas, sendo os Métodos de Identificação de Subespaço (*Subspace Identification Methods – SIM*) e os Métodos de Erro de Predição (*Prediction-Error Methods – PEM*) as duas classes de técnicas mais populares no campo de identificação de sistemas atualmente, sendo o primeiro baseado em técnicas de projeção no espaço Euclidiano, enquanto o segundo é baseado na minimização de um critério dependente de parâmetros (Gevers, 2006).

Os algoritmos baseados em SIM, embora atraentes sob alguns aspectos, apresentam deficiências em relação a algoritmos baseados em PEM. Em geral, e de interesse para este trabalho, as estimativas de SIMs não são tão precisas quanto as obtidas através de PEMs, além de apresentarem vieses em condições de malha-fechada, nas quais as entradas são parcialmente determinadas como saídas do sistema durante a coleta de dados (Qin, 2006). Nos aprofundaremos nos métodos do erro de predição, que são utilizados no decorrer da pesquisa devido sua robustez e adequação para parametrização de modelos caixa-cinza.

O problema básico de identificação de sistemas através de métodos de erro de predição é muito semelhante ao problema de adequação de curvas (*curve fitting*) há muito estudado pela estatística. Existe um processo natural que pode ser descrito por uma função matemática $f : \mathcal{X} \rightarrow \mathcal{Y}$. Para um conjunto de amostras $\mathcal{D}$ composto de N dados de entrada, $\{x_1, x_2, \dots, x_N\}$, e N observações ruidosas para estes mesmos N dados de entrada, $\{y_1, y_2, \dots, y_N\}$, sendo $\mathcal{D} = \{(x_1, y_1), \dots, (x_N, y_N)\}$, tal que:

$$y_n = f(x_n) + r_n \quad (2.15)$$

onde r_n representa o ruído na observação n , para $n = 1, 2, \dots, N$, encontrar uma função $\hat{g}(x)$ que aproxime $f(x)$ bem segundo uma métrica de erro (Ljung, 2006), conforme ilustrado a figura 2.13.

Figura 2.13: Exemplo de adequação de curvas. Fonte: adaptado de Ljung (2006).

De forma análoga, em identificação de sistemas através de métodos PEM, temos um processo natural que desejamos modelar através de um sistema dinâmico paramétrico. O estado do sistema em t é dado por $y(t)$, e, da mesma forma, a entrada do sistema em t é dada por $u(t)$, sendo Z_N o conjunto de todos os dados de estado e entrada do sistema de $t = 1$ até $t = N$. O processo de

identificação de sistemas através do método PEM, então, consiste em⁴:

1. Escolher um modelo de sistema dinâmico paramétrico $\phi(Z, \theta)$, onde θ representa os parâmetros do modelo, capaz de aproximar o fenômeno natural objeto de estudo.
2. Descrever o modelo como um preditor do passo seguinte do modelo, *i.e.* utilizando toda ou parte da informação que temos até o momento $t - 1$, estimar o estado do sistema em t :

$$\hat{y}(t, \theta) = \phi(Z_{t-1}, \theta) \quad (2.16)$$

Aqui, $\hat{y}(t, \theta)$ representa a predição de um passo a frente do estado do sistema.

3. Escolher uma medida de distância, $\ell : \mathbb{R}^N \rightarrow \mathbb{R}$, adequada (também conhecida como função custo ou métrica de erro na literatura) entre a predição de um passo do estado do sistema e o estado observado para o mesmo período t . Para exemplificar, vamos utilizar como medida de custo o erro quadrático, também conhecido como norma- ℓ^2 .
4. Encontrar os parâmetros θ ótimos, denotados como $\hat{\theta}$, que parametrizem a função ϕ de forma a minimizar a média da métrica de distância definida no item anterior para todas as predições de um passo geradas pelo o conjunto de dados observados Z_N , representada aqui por V_N :

$$\hat{\theta}_N = \arg \min_{\theta} V_N(\theta) \quad (2.17)$$

$$V_N(\theta) = \frac{1}{N} \sum_{t=1}^N \ell(y_t - g(x_{t-1}, \theta)) \quad (2.18)$$

$$\ell(\varrho) = \varrho^2 \quad (2.19)$$

Em decorrência da estrutura de métodos PEM, em problemas de identificação de sistemas onde o espaço de hipóteses é dado pelas diferentes parametrizações possíveis de um sistema dinâmico, o termo identificação de sistemas é utilizado com o significado estrito de estimação de parâmetros de um sistema dinâmico a partir de dados de entrada-saída, onde, dado um modelo de sistema dinâmico com parâmetros estáticos desconhecidos, o objetivo é estimar estes parâmetros com base em dados de entrada-saída de forma a minimizar uma medida de erro (Andrieu *et al.*, 2004; Ding *et al.*, 2013).

Neste projeto, construiremos um modelo caixa-cinza, baseado no modelo dinâmico de Wheaton (1999), que será posteriormente parametrizado utilizando o método do erro de predição aqui descrito, utilizando para isso o conjunto de dados reais do mercado imobiliário canadense (Canada, 2016a).

Estrutura dos métodos PEM

A base dos métodos de identificação de sistemas é a teoria estatística (Ljung, 2010), empregando técnicas utilizadas para o problema de adequação de curvas (Ljung, 2006). Devido à semelhança entre os problemas de identificação de sistemas e aprendizado de máquina, alguns autores tomam emprestado técnicas tipicamente empregadas em problemas de aprendizado de máquina para auxiliar no processo de identificação de sistemas (Ljung, 2010; Nagumo e Noda, 1967; Pillonetto *et al.*, 2014).

É difícil quantificar o número de técnicas já propostas no campo de identificação de sistemas, mas os métodos PEM existentes são, no geral, combinações distintas de estruturas básicas que estão presentes em todos os métodos de forma explícita ou implícita. Conforme apontado por Ljung

⁴Utilizamos aqui o caso comum de um sistema dinâmico com tempo discreto, que correspondem ao modelo econômico sendo estudado, embora também seja possível trabalhar com modelos contínuos no tempo (Ljung, 2002).

(1978) e Domingos (2012), os diferentes métodos PEM propostos provêm de combinações diferentes de três componentes básicos. São eles:

1. **Representação:** corresponde ao formato do modelo escolhido, e implica a escolha de um espaço de hipóteses definido pelas diferentes representações que o modelo escolhido é capaz de apresentar. Um modelo, no caso, é uma simplificação matemática que busca descrever as propriedades ou comportamentos de um processo existente no mundo real.
2. **Avaliação:** uma função de avaliação, também conhecida como função erro, ou função custo, é necessária para distinguir bons modelos de modelos ruins. A função de avaliação utilizada nem sempre coincide com aquela que desejamos que seja otimizada na realidade, devido à dificuldades de implementação matemática ou custo computacional.
3. **Otimização:** Com base na representação definida e na função de avaliação escolhida, é necessário um método para se navegar o espaço de hipóteses em busca do elemento contido no mesmo que minimize métrica de avaliação escolhida.

É interessante notar que esta estrutura básica é comum tanto ao campo de identificação de sistemas por métodos PEM quanto ao campo de aprendizado de máquina supervisionado, sendo que a principal diferença recai sobre a escolha de representação. O campo de identificação de sistemas, como o próprio nome denota, utiliza diferentes classes de modelos de sistemas dinâmicos como formas de representação em seus métodos. O campo de aprendizado de máquina, por sua vez, utiliza normalmente representações estáticas como formas de representação, e é geralmente utilizado para modelar problemas de classificação. Não existe nenhuma linha clara que separe os dois campos, no entanto, com trabalhos publicados que poderiam ser classificados tanto como sendo membros de uma área quanto de outra (Drezet e Harrison, 1998; Jiang *et al.*, 2017; Martínez-Ramón *et al.*, 2006)

Generalização de modelos aprendidos

Para que o processo acima delineado possa ser aplicado de forma bem sucedida, algumas condições precisam ser atendidas. Primeiramente, presume-se que o fenômeno ou processo, objeto de estudo, existente no mundo real, pode ser modelado de maneira precisa por uma função matemática $f(x)$ desconhecida pelo pesquisador. Presume-se também que o pesquisador conheça ou possa encontrar uma classe de modelos G que tenha contido em si a função $f(x)$, ou uma boa aproximação $g^*(x)$ desta. Em último lugar, presume-se que é possível encontrar uma função $\hat{g}(x)$ que aproxime razoavelmente bem, de acordo com algum critério escolhido pelo pesquisador, a função $f(x)$ ou $g^*(x)$ (Abu-Mostafa *et al.*, 2012; Anderson *et al.*, 1978).

A representação adotada delimita a classe de modelos e o espaço de hipóteses que contém todas as possíveis representações passíveis de serem aprendidas com base em dados, sendo que a complexidade do espaço de hipóteses é uma medida do tamanho do mesmo. Representações mais flexíveis possuem complexidades maiores que funções menos flexíveis, e portanto a probabilidade de que algum elemento contido no espaço de hipóteses da representação mais flexível aproxime bem o fenômeno de interesse de acordo com a medida de avaliação é maior que a de algum elemento contido no espaço de hipóteses de uma função menos flexível. Entretanto, complexidades maiores também implicam uma maior necessidade de dados para encontrar um elemento que aproxime bem o fenômeno estudado (Abu-Mostafa *et al.*, 2012).

Tomando emprestado parte da nomenclatura do exemplo desenvolvido para ilustrar o problema de adequação de curvas, vamos supor novamente que exista uma função $f : \mathcal{X} \rightarrow \mathcal{Y}$ que descreva um fenômeno de interesse, onde $\mathcal{X}$ é o espaço de entrada (conjunto de todas as entradas x possíveis), e $\mathcal{Y}$ é o espaço de saída (conjunto de todas as saídas y possíveis) da função f . Vamos supor que o fenômeno de interesse que se deseja estudar seja um problema de classificação, onde $\mathcal{Y} = \{\text{sim}, \text{não}\}$. Assim como antes, existe um conjunto de dados de entrada-saída com N amostras,

$\mathcal{D} = \{(x_1, y_1), \dots, (x_N, y_N)\}$, mas nesse caso, supomos observações livres de ruído, de forma que $y_n = f(x_n)$ para $n = 1, 2, \dots, N$.

Por fim, consideremos que existe um algoritmo de aprendizado $\mathcal{A}$ que utilize o conjunto de dados $\mathcal{D}$ para escolher a função $g : \mathcal{X} \rightarrow \mathcal{Y}$ que aproxime f segundo uma métrica de erro. O algoritmo escolhe esta função g a partir de um espaço de hipóteses contendo todas as funções possíveis definidas pela representação escolhida pelo pesquisador. Consideremos que o espaço de hipóteses $\mathcal{H}$ seja finito, de forma que $\mathcal{H} = \{h_1, h_2, \dots, h_M\}$, onde M é o número de hipóteses contido neste espaço, de forma que $g \in \mathcal{H}$.

Existem duas medidas de erro que nos interessam neste problema, o erro dentro da amostra, E_{in} , e o erro fora da amostra, E_{out} . Neste caso, temos que (Abu-Mostafa *et al.*, 2012):

$$E_{in}(h) = \frac{1}{N} \sum_{n=1}^N \mathbb{I}[h(x_n) \neq f(x_n)] \quad (2.20)$$

$$E_{out}(h) = \mathbb{P}[h(x) \neq f(x)] \quad (2.21)$$

É possível demonstrar que, o problema descrito acima, possui a seguinte garantia probabilística (Abu-Mostafa *et al.*, 2012):

$$\mathbb{P}[|E_{in}(g) - E_{out}(g)| > \epsilon] \leq 2Me^{-2\epsilon^2 N} \quad (2.22)$$

para qualquer $\epsilon \in (0, 1)$.

A inequação 2.22, apesar de restrita em termos de suas premissas, confere a intuição necessária para se entender os *trade-offs* existentes em problemas de aprendizado em geral. As conclusões que podem ser extraídas deste simples exemplo já foram generalizadas para espaços de hipóteses infinitos e para premissas mais fracas (Blumer *et al.*, 1989; Domingos, 2012).

Duas questões são relevantes para o problemas de identificação de sistemas:

1. Quão certo podemos ficar de que $E_{in}(g)$ está próximo de $E_{out}(g)$?
2. Qual é o menor valor possível que $E_{in}(g)$ pode adotar?

A primeira questão trata do problema de generalização do resultado obtido através do processo de aprendizagem. A segunda questão se refere ao viés da representação escolhida para se modelar o problema.

Especificamente, para problemas marcados pela escassez de dados, como é o caso do mercado imobiliário, a utilização de representações muito flexíveis, marcadas por um espaço de hipóteses grande, torna o lado direito da inequação 2.22 demasiadamente frouxo, o que leva a uma pequena garantia de generalização da equação aprendida g . De fato, em problemas de identificação de sistemas para os quais não é possível modelar o fenômeno de interesse através de princípios teóricos, a aplicação de métodos caixa-preta requer cuidados especiais, como a combinação com métodos de subespaço, para obter resultados razoáveis (Ljung, 2002). Representações pouco flexíveis e genéricas, no entanto, correm o risco de não obterem uma boa minimização de E_{in} , apesar de, em princípio, melhorarem a generalização do resultado.

Outra estrutura de análise utilizada para guiar a escolha da complexidade de um modelo de sistema dinâmico para ser utilizada durante o processo de identificação de sistemas, e que se refere às mesmas questões tratadas acima, é a análise de viés-variância. Conforme descrito por Ljung (2006), o objetivo da escolha da classe de modelos a ser utilizada no processo de identificação de sistemas é encontrar o modelo $\hat{g}(x)$ que melhor se aproxime do modelo correto $f(x)$. Se pensarmos no ruído do modelo como uma variável aleatória, uma medida natural do tamanho do erro é o erro quadrático médio (*mean square error* – MSE):

$$W(x) = E(f(x) - \hat{g}(x))^2 \quad (2.23)$$

$$= B(x) + V(x) \quad (2.24)$$

$$B(x) = (f(x) - g^*(x))^2 \quad (2.25)$$

$$g^*(x) = E(\hat{g}(x)) \quad (2.26)$$

$$V(x) = E(g^*(x) - \hat{g}(x))^2 \quad (2.27)$$

onde E representa o símbolo matemático de esperança, e o MSE pode ser dividido no viés do erro ($B(x)$) e na variância do erro ($V(x)$). Um caso típico ocorre quando $g^*(x)$ não depende do número de observações N , e temos que $\hat{g}(x) \rightarrow g^*(x)$ conforme $N \rightarrow \infty$ [Ljung \(2006\)](#).

As conclusões decorrentes deste tipo de análise são condizentes com a análise da complexidade do espaço de hipóteses conduzida previamente. De forma a minimizar o MSE, é necessário escolher uma classe de modelos G que minimize tanto a expectativa do viés quanto a expectativa da variância. Conforme a complexidade C de G aumenta, temos que $B(x)$ diminui, enquanto $V(x)$ aumenta. Da mesma forma, conforme C diminui, $B(x)$ aumenta, e $V(x)$ diminui.

A conclusão obtida é a mesma para as duas análises. Dado uma quantidade fixa de dados, a melhor forma de se aumentar a garantia de generalização da equação aprendida g , e ao mesmo tempo utilizar uma representação com baixo viés, é através da delimitação do espaço de hipóteses decorrente de uma modelagem bem feita do problema em si, baseada no estudo da informação disponível previamente no corpo de conhecimento que tange a pesquisa sendo realizada.

2.3.3 Validação, seleção e teste de modelos

O processo de validação consiste na verificação da capacidade do modelo de generalizar seus resultados para conjuntos de dados não utilizados durante o treinamento. A técnica é normalmente empregada durante a etapa de seleção de modelos, onde deseja-se escolher um parâmetro que delimite o espaço de hipóteses a ser procurado durante a etapa de treinamento do modelo. O teste do modelo aprendido procura verificar a generalização do mesmo quando aplicado a novos conjuntos de dados não utilizados no processo de treino, diferindo da validação por não influenciar o processo de aprendizado, sendo, portanto, a última etapa em identificação de sistemas ([Abu-Mostafa et al., 2012](#); [Ljung, 2006, 2010](#)).

Validação

Embora o próprio processo de treinar o modelo resulte em uma medida de erro (E_{in}), a medida de erro obtida através do teste do modelo aprendido nos dados de treinamento é uma medida viesada do desempenho deste, já que os parâmetros do modelo foram aprendidos exatamente de forma a minimizar a função custo para aquele conjunto de dados, o que não implica necessariamente um resultado semelhante do modelo quando aplicado a conjuntos de dados diferentes. De fato, um dos maiores problemas resultantes do processo de identificação de sistemas guiado tão somente pela minimização da métrica de erro nos dados de treinamento (E_{in}) é o fenômeno conhecido como *overfitting*, que ocorre quando, durante a seleção de modelos, a melhora do desempenho do modelo em dados de treinamento (E_{in}) resulta em uma piora no desempenho em dados fora do treinamento (E_{out}) ([Abu-Mostafa et al., 2012](#)).

O método normalmente utilizado para obter uma estimativa não viesada do desempenho de modelos aprendidos através de técnicas PEM consiste, basicamente, na divisão do conjunto de dados disponível em dois conjuntos distintos. Um desses conjuntos, normalmente o maior destes, é usado para o treinamento do modelo, enquanto o segundo é utilizado para se estimar a generalização do modelo encontrado segundo a métrica de erro definida. Os nomes geralmente atribuídos a estes conjuntos de dados são *dados de treinamento* e *dados de teste* (também chamado as vezes de *dados de validação*) ([Murphy, 2012](#)), e uma ilustração do processo pode ser encontrada em [2.14](#)

Figura 2.14: Ilustração do processo de validação de modelos para obtenção de uma estimativa de E_{out} . O conjunto de dados D é dividido em um conjunto de dados de treino D_{treino} e conjunto de dados de validação D_{val} . O conjunto de dados D_{treino} é utilizado para produzir o modelo g^- , e este modelo é então utilizado para realizar previsões com base no conjunto de dados de validação com a finalidade de produzir uma estimativa do desempenho do modelo final g parametrizado através do conjunto de dados total D . Fonte: adaptado de Abu-Mostafa et al. (2012).

Em casos em que é necessário utilizar os dados para fazer algumas escolhas durante o processo de aprendizado, é comum dividir-se o conjunto de dados disponível em três conjuntos distintos: dados de treinamento, dados de validação e dados de teste. Para estes casos, é comum utilizar estimativas de erro de generalização, através do uso de dados de validação, para realizar estas escolhas. Isso permite ao pesquisador utilizar o conjunto de dados de validação para criar estimativas de erro do modelo de forma a fazer escolhas com base em estimativas de generalização do mesmo, e ainda manter um conjunto de dados não viesado (dados de teste) de forma a obter uma estimativa também não viesada do desempenho do modelo em dados diferentes dos dados de treinamento. Desta forma, o teste do modelo em dados de validação permite ao pesquisador fazer escolhas sobre a classe de modelos a ser utilizada no processo de aprendizado, evitando o risco de *overfitting* (Abu-Mostafa et al., 2012). Este processo é denominado de seleção de modelos, ilustrado pela figura 2.15.

Conforme explicado na seção 2.3.2, é natural esperar que modelos mais flexíveis obtenham melhores resultados em dados de treinamento que modelos menos flexíveis. No entanto, minimizar o desempenho em dados de treinamento (E_{in}) por si só raramente é o objetivo do processo de aprendizado, sendo o desempenho em dados fora do treinamento (E_{out}) o verdadeiro interesse no processo de aprendizado de um modelo. Deixe $\hat{m}$ ser o modelo aprendido a partir de um conjunto de dados de treinamento com N observações Z_{treino}^N , e deixe $\mathcal{F}$ representar a métrica de avaliação (*fitness*) do erro de um modelo em um determinado conjunto de dados. A expectativa probabilística que temos pode ser, então, representada por:

$$\bar{\mathcal{F}}(\hat{m}, Z_{val}) = \mathcal{F}(\hat{m}, Z_{treino}^N) + \psi(C(M), N) \quad (2.28)$$

onde o lado esquerdo da equação representa a adequação do modelo aprendido a dados de validação. A função ψ representa o prêmio de erro que obtemos ao partir da métrica de adequação medida em dados de treinamento para chegar na métrica de adequação esperada em dados de validação, dado como uma função da complexidade da classe de modelos M considerada, assim como da quantidade

Figura 2.15: Ilustração do processo de seleção de modelos através do uso de um conjunto de dados de validação. O conjunto de dados, D , é separado em um conjunto de dados de treino, D_{treino} , e um conjunto de dados de validação, D_{val} . M modelos são então treinados a partir dos dados de treinamento com base nas M hipóteses diferentes que se deseja testar. Estimativas do desempenho destes modelos, E_m , são obtidas a partir do teste dos mesmos com base em previsões realizadas no conjunto de validação. Escolhida a melhor hipótese, treina-se o modelo final, g_{m^*} , no conjunto de dados D . Fonte: adaptado de [Abu-Mostafa et al. \(2012\)](#).

de dados utilizada durante o treinamento ([Ljung, 2010](#)).

A delimitação das garantias probabilísticas que podemos obter quanto à generalização dos resultados obtidos em dados de treinamento para dados de validação é uma área de pesquisa ativa ainda hoje. Uma das garantias que podemos ter quanto a 2.29, decorrente da teoria de aprendizado de Vapnik (ver [Suykens et al. \(2002\)](#)), é que, com probabilidade de $1 - \delta$:

$$\bar{\mathcal{F}}(\hat{m}, Z_{val}) \leq \mathcal{F}(\hat{m}, Z_{treino}^N) \left(1 - \sqrt{\frac{C(\log N/C) + 1 - \log \delta}{N}} \right)^{-1} \quad (2.29)$$

onde a medida de complexidade C utilizada é a dimensão VC (*Vapnik-Chervonenkis*) do conjunto de modelos, que de maneira simplificada, pode ser entendida como uma medida de quão bem o conjunto de funções decorrentes do modelo podem separar pontos aleatórios.

Outras garantias na mesma linha da equação 2.29 já foram propostas por outros autores ([Akaike, 1974](#); [Bartlett et al., 2006](#); [Golub et al., 1979](#)). A intenção aqui, no entanto, não é dar um tratamento completo sobre as diferentes garantias já descobertas sobre o uso do erro de validação como uma estimativa do desempenho do modelo em dados não utilizados durante o treinamento, mas sim expor que estas garantias existem, mesmo porque as garantias propostas, exatamente por se aplicarem a quaisquer modelos, são demasiadamente frouxas. É comum se esperar, neste sentido, que os erros de validação obtidos na etapa de teste dos modelos aprendidos sejam uma medida mais fidedigna de E_{out} do que estas garantias levam a crer ([Abu-Mostafa et al., 2012](#); [Gevers, 2006](#)).

Cross-validation

A divisão do conjunto de dados disponíveis em dois subconjuntos de dados, de treinamento e teste, é uma abordagem eficaz quando uma grande quantidade de dados se encontra disponível para uso no processo de aprendizado. Quando a quantidade de dados disponível sobre o fenômeno

de interesse é limitada, no entanto, manter parte deles para o conjunto de teste reduz ainda mais o número de amostras disponíveis para treinamento, reduzindo a confiança de proximidade que obtemos entre E_{in} e E_{out} .

Idealmente, o conjunto de dados de treinamento deve ser o menor possível, de forma a garantir que o modelo aprendido durante o treinamento utilize todos os dados para selecionar o melhor modelo dentre o espaço de hipóteses. Ao mesmo tempo, de forma ideal, o conjunto de validação deve ser o maior possível, de forma a garantir que a estimativa de erro produzida, E_{val} se aproxime da verdadeira estimativa de E_{out} do modelo sendo validado.

Uma outra abordagem, conhecida como *cross-validation*, pode ser utilizada para que todo o conjunto de dados seja usado tanto como conjunto de dados de treinamento quanto como conjunto de dados de validação. A técnica de *cross-validation* consiste na divisão do conjunto de dados de treino e validação em k partes, também conhecidas como *k-folds*. Cada um desses k subconjuntos de dados é utilizado como conjunto de validação contra os modelos aprendidos durante o treino dos modelos em seus respectivos conjuntos de dados.

Ilustramos aqui o caso mais simples, que utilizamos neste trabalho, conhecido como *leave-one-out cross-validation*, utilizando a terminologia empregada por Abu-Mostafa *et al.* (2012). Uma representação gráfica do processo pode ser encontrada na figura 2.16. Deixe D_n representar o conjunto de dados obtido pela exclusão da observação (x_n, y_n) , de forma que:

$$D_n = (x_1, y_1), \dots, (x_{n-1}, y_{n-1}), (\overline{x_n, y_n}), (x_{n+1}, y_{n+1}), \dots, (x_N, y_N) \quad (2.30)$$

e deixe g_n^- representar a hipótese escolhida após o treino em D_n . Chamemos a função erro de e , e de e_n o erro obtido pelo modelo g_n^- em seu respectivo conjunto de validação, que, por construção, é o dado (x_n, y_n) :

$$e_n = E_{val}(g_n^-) = e(g_n^-(x_n), y_n) \quad (2.31)$$

Temos, então, que a estimativa de erro obtida por *cross-validation* é dada pela média dos erros e_n :

$$E_{CV} = \frac{1}{N} \sum_{n=1}^N e_n \quad (2.32)$$

Embora a técnica de *cross-validation* tenha sido desenvolvida para lidar com problemas estáticos, o procedimento pode ser também empregado em séries temporais para seleção de modelos em identificação de sistemas (Stoica *et al.*, 1986). Diferentes métodos de validação de modelos já foram propostos para seleção de modelos em séries temporais. Bergmeir e Benítez (2012) testam empiricamente diferentes métodos propostos e encontram que a técnica de *cross-validation* obtém resultados melhores para seleção de modelos em diferentes séries temporais, tanto simuladas quanto reais, com diferentes características. De fato, *cross-validation* é especificamente vantajoso para problemas em que a série temporal disponível é limitada por natureza, pois permite ao pesquisador utilizar a totalidade de dados tanto para treinamento quanto para validação (Bergmeir e Benítez, 2012; Bergmeir *et al.*, 2014).

Teste

A última etapa em identificação de sistemas, é o teste do modelo produzido. Similar ao processo de validação, o teste de modelos de sistema dinâmico busca fornecer uma ideia de como o modelo aprendido durante a etapa de treinamento irá desempenhar sua função preditiva para outros conjuntos de dados. No caso de identificação de sistemas, duas formas básicas de teste de modelos existe, segundo Ljung (2006), já que estes podem ser usados para *simulação* e *predição*.

Por predição, Ljung (2006) refere-se à predição de um passo do sistema, conforme descrito na

Figura 2.16: Ilustração do processo de leave-one-out cross-validation para obtenção de uma estimativa de E_{out} em um conjunto de dados D com N observações. Separa-se o dado n , treina-se o modelo g_n no conjunto de dados restantes D_n , e então obtém-se uma estimativa de desempenho e_n do modelo g_n com base no dado n deixado de lado para validação. Após realizado este processo para todo n contido em D , calcula-se a estimativa de desempenho do modelo final g a partir da média do desempenho e_n dos N modelos g_n produzidos. Fonte: adaptado de [Abu-Mostafa et al. \(2012\)](#).

equação 2.16. Vamos supor um modelo de sistema dinâmico descrito por:

$$\hat{y}_t = \phi(y_{t-1}, x_{t-1}, \theta) \quad (2.33)$$

Onde $\hat{y}_t$ representa o estado previsto do sistema para o instante t , y_{t-1} e x_{t-1} representam o conjunto de dados de estado e entrada do sistema em $t - 1$, respectivamente, θ representa os parâmetros do sistema e ϕ representa a função de transição.

O teste da capacidade de predição do modelo busca, portanto, verificar se o modelo é capaz de produzir estimativas do estado do sistema para t com base no conjunto de dados disponível em $t - 1$.

O teste da capacidade de simulação do modelo é um teste mais rigoroso, e busca avaliar como o modelo se comportaria a partir de um determinado ponto caso suas predições fossem realizadas com base em outras predições. Utilizando a mesma notação empregada anteriormente para o caso de predição, temos que a simulação do modelo, a partir de um período arbitrário t , pode ser dada por:

$$\begin{aligned} \hat{y}_t &= \phi(y_{t-1}, x_{t-1}, \theta) \\ \hat{y}_{t+1} &= \phi(\hat{y}_t, x_t, \theta) \\ \hat{y}_{t+2} &= \phi(\hat{y}_{t+1}, x_{t+1}, \theta) \\ &\vdots \\ \hat{y}_{t+k} &= \phi(\hat{y}_{t+k}, x_{t+k}, \theta) \end{aligned}$$

onde k define o número de passos a ser simulado.

É natural que o desempenho do modelo de sistema dinâmico piore seu desempenho conforme o período de simulação utilizado, k , aumenta, já que o erro de predição, neste caso, é acumulado com cada passo do sistema ([Ljung, 2006](#)).

Capítulo 3

Experimentos

Este capítulo é dedicado a descrição dos experimentos elaborados para o teste do modelo dinâmico proposto por [Wheaton \(1999\)](#) como modelo preditivo em séries temporais de estoque imobiliário. Os algoritmos desenvolvidos para a realização dos experimentos descritos nesta sessão foram implementados em Matlab ([Mathworks, 2016a](#)).

Inicialmente adaptamos o modelo de forma a adequar seu formato ao propósito de predição e identificação de sistemas, e testamos sua validade em dados simulados. Posteriormente, utilizamos o mesmo método empregado com base em dados simulados, e adaptamos o modelo para verificar os resultados obtidos em dados reais provenientes de mercado habitacional canadense.

Em ambos os casos, o modelo adaptado é deduzido a partir das cinco equações que compõe o modelo de [Wheaton \(1999\)](#), configurando um modelo caixa-cinza com parâmetros ajustáveis para refletir as diferenças econômicas entre mercados imobiliários distintos. O método desenvolvido é elaborado com base em conceitos da área de identificação de sistemas, utilizando métodos de erro de predição para parametrizar o modelo proposto.

No total, são realizados 15 experimentos, divididos em 4 experimentos com base em dados simulados, e 11 experimentos com base nos dados reais disponíveis pela Agência de Estatística do Canadá ([Canada, 2016a,c](#)).

3.1 Experimentos em dados simulados

Iniciamos os experimentos do método de identificação de sistemas com base no modelo de [Wheaton \(1999\)](#) utilizando dados simulados por este mesmo modelo. As simulações utilizadas são provenientes do artigo original de [Wheaton \(1999\)](#), que busca parametrizar o modelo com parâmetros economicamente verossímeis.

A realização de testes do modelo em dados simulados antes de testá-lo em dados reais provém da necessidade de verificar a capacidade deste em aprender corretamente os parâmetros do modelo. Em dados simulados, possuímos duas garantias que facilitam na verificação do método proposto: primeiro, como os dados são provenientes do modelo em si, sabemos com certeza que o espaço de hipóteses contem o modelo real a ser identificado; segundo, sabemos com exatidão os valores corretos dos parâmetros ótimos. Desta forma, ao testarmos o método, podemos verificar se este está convergindo de maneira adequada para os valores corretos.

Nas subseções a seguir, elaboramos o método de identificação de sistemas, explicando cada passo de seu desenvolvimento, e em seguida descrevemos os testes realizados de forma a verificar seu desempenho.

3.1.1 Modelo e método

Conforme explicado na seção [2.3.2](#), os métodos de estimação de sistemas via minimização do erro de predição dependem dos seguintes passos para serem aplicados de forma adequada. De forma resumida, a boa aplicação dos métodos PEM depende da:

1. Escolha de um modelo de sistema dinâmico paramétrico;
2. Descrição do modelo como preditor de um passo a frente do estado do sistema;
3. Escolha de uma medida de erro entre predições *vs.* dados reais;
4. Busca pelos parâmetros que minimizem os erros de predição para o modelo desenvolvido.

Seguimos aqui esta estrutura para construir o método PEM empregado nos experimentos para dados simulados.

Modelo de sistema dinâmico

Como nos testes iniciais com base em dados simulados estamos testando um modelo de sistema dinâmico já disponível na literatura, temos que o passo 1 já se encontra atendido, com a descrição do modelo de sistema dinâmico paramétrico utilizado descrito em 2.2.2.

O modelo dinâmico proposto no artigo original de Wheaton (1999) é composto por cinco equações, elaboradas com base na teoria econômica disponível sobre as relações de oferta e demanda de estoque imobiliário. Enquanto, para o propósito original do artigo estas cinco equações são adequadas para o entendimento do fenômeno imobiliário que o autor se propunha a estudar, para fins preditivos, as mesmas podem ser unificadas em uma única equação de diferença realizando as substituições de variáveis corretamente. Esta equação de diferença pode, então, ser usada para expressar o modelo como preditor de um passo a frente do sistema, atendendo ao passo 2.

Além desta simplificação reduzir a complexidade da fórmula de erro de predição subsequentemente desenvolvida, a união das cinco equações em uma única permite que algumas constantes do modelo sejam unificadas, diminuindo o número de parâmetros a serem estimados a partir dos dados. Embora esta simplificação implique uma redução do número de variáveis com um significado econômico que podemos extrair do processo de aprendizado, a redução no número de parâmetros do modelo é benéfica do ponto de vista de generalização do modelo produzido, devido a diminuição da dimensionalidade do espaço de hipóteses (Friedman, 1997; Keogh e Mueen, 2011).

Conforme explicado na seção 2.2.2, o modelo de Wheaton (1999) é composto por cinco equações econômicas: a equação de demanda, a equação de oferta, a equação da taxa de novas construções, a equação de preço e a equação de diferença que determina a evolução do estoque de um período para o próximo. Enquanto Wheaton (1999) propõe a resolução sequencial destas mesmas equações, conectando as variáveis comuns entre as mesmas, utilizamos o mesmo processo sequencial de união de equações para unificá-las em uma única equação de diferença para a modelagem do estoque imobiliário de um mercado.

Inicialmente, a equação 2.7 de demanda pode ser substituída na equação 2.11 de preço, resultando em:

$$P_t = \frac{\left(\frac{S_{t-n}}{\alpha_1 E_{t-n}}\right)^{-\frac{1}{\beta_1}}}{r} \quad (3.1)$$

Em seguida, podemos substituir a equação expandida 3.1 de preço na equação 2.10 de construção:

$$\frac{C_{t-n}}{S_{t-1}} = \frac{\alpha_2}{r^{\beta_2}} \left(\frac{S_{t-n}}{\alpha_1 E_{t-n}}\right)^{-\frac{\beta_2}{\beta_1}} \quad (3.2)$$

Por fim, a equação 3.2 de construção obtida acima pode ser incorporada à equação 2.9, de forma que:

$$\frac{S_t}{S_{t-1}} = 1 - \delta + \frac{\alpha_2}{r^{\beta_2}} \left(\frac{S_{t-n}}{\alpha_1 E_{t-n}}\right)^{-\frac{\beta_2}{\beta_1}} \quad (3.3)$$

É importante notar que δ , α_1 , α_2 , r , β_1 e β_2 são parâmetros constantes do modelo. Desta forma, reescrevendo a equação 3.3 da seguinte forma:

$$\frac{S_t}{S_{t-1}} = 1 - \delta + \frac{\alpha_2 \alpha_1^{\frac{\beta_2}{\beta_1}}}{r^{\beta_2}} \left(\frac{E_{t-n}}{S_{t-n}} \right)^{\frac{\beta_2}{\beta_1}} \quad (3.4)$$

fica claro que o modelo depende, na verdade, de três parâmetros constantes¹:

$$\tau = 1 - \delta \quad (3.5)$$

$$\omega = \frac{\alpha_2 \alpha_1^{\frac{\beta_2}{\beta_1}}}{r^{\beta_2}} \quad (3.6)$$

$$\xi = \frac{\beta_2}{\beta_1} \quad (3.7)$$

onde τ e ξ representam, respectivamente, a porcentagem de estoque que se mantém de um período para outro e a relação entre a elasticidade de oferta e a elasticidade de demanda, enquanto ω não possui uma descrição econômica clara, resultando de uma conjunção de fatores de mercado.

Temos que o modelo, pode, então, ser descrito pela equação 3.8:

$$\frac{S_t}{S_{t-1}} = \tau + \omega \left(\frac{E_{t-n}}{S_{t-n}} \right)^{\xi} \quad (3.8)$$

Conforme mencionado anteriormente, seria muito mais rico do ponto de vista de uma análise econômica de um mercado se pudéssemos aprender com precisão todas as seis variáveis econômicas do modelo original a partir de uma série temporal de dados de estoque imobiliário. A combinação de algumas destas variáveis no modelo, porém, tornam as mesmas não identificáveis através do processo de minimização, já que várias combinações numéricas diferentes para variáveis combinadas podem resultar em um mesmo valor constante no modelo.

Se pegarmos ξ , por exemplo, e obtivermos que o valor que minimiza o erro para o modelo em uma série temporal de estoque imobiliário seja um valor constante de 2, esse mesmo ξ poderia, então, ser obtido tanto pelos parâmetros de elasticidade β_2 e β_1 de 2 e 1, respectivamente, quanto de 4 e 2. De fato, quaisquer valores no formato $2 \cdot x$ para β_2 e x para β_1 resultam em um parâmetro ξ com valor de 2.

Apesar da diminuição de variáveis obtida ao se agregar as diferentes equações reduzir o entendimento econômico do processo, o entendimento dos fluxos de estoque modelado pela equação torna-se mais simples de compreender. A partir da equação 3.9, podemos perceber que o comportamento do estoque em um mercado, segundo o modelo de [Wheaton \(1999\)](#), é regido por dois fatores: uma diminuição constante do estoque de período a período, devido a depreciação do mesmo ao longo do tempo, junto de um aumento de estoque, com atraso, gerado pela aumento da demanda em relação ao estoque disponível em um determinado período de tempo no passado.

¹Existe um quarto parâmetro relevante para o modelo, n , que representa o número de períodos que se leva entre o início de uma construção até seu término. Apesar de ser um parâmetro relevante, o mesmo é difícil de ser estimado segundo os mesmos métodos de otimização utilizados para se estimar os valores de τ , ω e ξ . De fato, parâmetros n diferentes implicam espaços de hipótese de modelos diferentes, sendo que o processo de escolha de um n adequado configura um processo de seleção de modelos. Conforme explicado na seção 2.3.3, utilizamos *cross-validation* para definir o parâmetro n ótimo para o conjunto de amostras.

Preditor de um passo do sistema

Partindo do modelo de sistema dinâmico descrito por 3.8, temos, então, que nosso modelo dinâmico de estoques pode ser escrito como:

$$\hat{S}_t = \left(\tau + \omega \left(\frac{E_{t-n}}{S_{t-n}} \right)^\xi \right) S_{t-1} \quad (3.9)$$

atendendo ao segundo passo de identificação de sistemas através de métodos PEM.

A saída da equação do modelo é a estimativa para o período t do estoque de área comercial em metros quadrados em um mercado. As entradas da equação são os estoques de áreas comerciais em metros quadrados em $t - 1$ e $t - n$ e a quantidade de trabalhadores neste mesmo mercado.

Os parâmetros do modelo são:

1. τ : porcentagem do estoque que se mantém de um período t para um período subsequente $t + 1$.
2. ω : variável decorrente da junção de diversas variáveis econômicas com significados distintos, mas que por si só não possui significado econômico claro.
3. ξ : elasticidade relativa de demanda por oferta.
4. n : atraso do sistema decorrente do número de períodos que leva entre a decisão de construção de uma nova unidade e a conclusão da mesma.

Função custo

Para atender ao passo 3, utilizamos como métrica de erro de predição do modelo a somatória dos erros quadráticos das predições realizadas em relação às observações reais, conforme a equação:

$$Erro = \sum_{t=1}^N (S_t - \hat{S}_t)^2 \quad (3.10)$$

Minimização

Por fim, de forma a parametrizar o modelo conforme o passo 4 da estrutura de métodos PEM, empregamos o algoritmo *GlobalSearch* (Mathworks, 2016b) para realizar a minimização do erro de predição de forma a determinar os parâmetros τ , ω e ξ ótimos para o conjunto de dados de treinamento utilizado. Para determinar o valor de n , utilizamos a técnica de *cross-validation* para seleção de modelos, conforme explicado na seção 2.3.3.

Para séries simuladas, o parâmetro ω adota valores elevados. Em virtude disso, sua minimização por métodos de gradiente não ocorre adequadamente, já que o mesmo é muito menos sensível a alterações que os demais parâmetros do modelo. Corrigimos esta deficiência substituindo o parâmetro ω na equação 3.8 por $10^{\omega'}$, onde $\omega' = \log_{10}^{\omega}$, e otimizamos o parâmetro ω' ao invés do parâmetro original ω , conforme exibido na equação 3.11:

$$\hat{S}_t = \left(\tau + 10^{\omega'} \left(\frac{E_{t-n}}{S_{t-n}} \right)^\xi \right) S_{t-1} \quad (3.11)$$

É fácil notar que, embora a alteração introduzida facilite a otimização dos parâmetros do modelo, a mesma não altera o mesmo de maneira qualitativa, sendo que tanto o modelo original quanto o modelo alterado são equivalentes. Como esta alteração se destina apenas a facilitar a otimização dos parâmetros, reportamos no decorrer do trabalho os valores obtidos para ω através da otimização do modelo para ω' , de forma a não poluir a análise dos experimentos conduzidos.

Como os parâmetros e comportamentos das séries simuladas apresentam valores bem divergentes por construção, optamos por tratar a delimitação dos limites dos valores dos conjuntos de busca dos parâmetros durante a seção 3.1.3.

3.1.2 Dados simulados

Conforme já mencionado, para compor as séries de dados de estoque simulados, realizamos 4 simulações com base nos parâmetros utilizados no artigo original de (Wheaton, 1999), de forma a replicar os comportamentos de estoque obtidos no mesmo.

As condições iniciais e parâmetros com valores constantes utilizados durante todas as simulações são conforme os valores exibidos na tabela 3.1. O subscrito *eq* denota valores do sistema em equilíbrio.

Parâmetro	Valor
$E_{t < n}$	10,0
$E_{t \geq n}$	15,0
S_{eq}	2.500,0
R_{eq}	20,0
r	0,05
P_{eq}	400,0

Tabela 3.1: Constantes, entradas e estados iniciais. Fonte: adaptado de Wheaton (1999).

Para todas as simulações, as entradas E_t possuem inicialmente o valor constante de 10,0, com o sistema encontrando-se em equilíbrio dinâmico, mas sofrem no período $t = n$ um choque positivo de 50%, passando a valer 15,0 nos períodos subsequentes. As simulações são feitas para um período de 100 unidades de tempo após a ocorrência do choque, sendo necessário um histórico de n períodos anteriores ao choque para atender à necessidade da variável de entrada do modelo.

Os parâmetros utilizados durante a simulação são listados na tabela 3.2, exibida abaixo.

Parâmetro	Sim 1	Sim 2	Sim 3	Sim 4
n	5	5	8	5
α_1	5.000,0	828,6	828,6	828,6
α_2	1.25×10^{-4}	3.125×10^{-7}	3.125×10^{-7}	3.125×10^{-7}
β_1	1,0	0,4	0,4	0,4
β_2	1,0	2,0	2,0	2,0
δ	0,05	0,05	0,05	0,1

Tabela 3.2: Valores de parâmetros utilizados por simulação. Fonte: adaptado de Wheaton (1999).

Para fazer a conexão entre o modelo original de Wheaton (1999) e o modelo concebido na seção 3.1.1, apresentamos na tabela 3.3 os respectivos valores dos parâmetros τ , ω e ξ por simulação.

Parâmetro	Sim 1	Sim 2	Sim 3	Sim 4
τ	0,95	0,95	0,95	0,9
ω	12,5	4.88×10^{10}	4.88×10^{10}	4.88×10^{10}
ξ	1,0	5,0	5,0	5,0

Tabela 3.3: Parâmetros convertidos por simulação. Fonte: adaptado de Wheaton (1999).

Os gráficos das séries simuladas de dados podem ser encontradas na figura 3.1. Todas as séries encontram-se inicialmente em equilíbrio dinâmico, mas apresentam comportamentos diferentes em resposta ao choque de demanda ocorrido no período n .

Figura 3.1: Gráficos das séries de dados de estoque simuladas. Fonte: proposto pelo autor.

3.1.3 Experimentos

Conforme explicado previamente, o experimento formulado tem como objetivo verificar a capacidade do método desenvolvido em identificar corretamente os parâmetros que resultem em um bom modelo de estimação dos dados utilizados. O experimento consiste em:

1. Simular dados de estoque utilizando o modelo e os parâmetros empregados por [Wheaton \(1999\)](#);
2. Utilizar o método desenvolvido para encontrar os parâmetros ótimos para a série de dados simulados;
3. Verificar se os parâmetros encontrados convergem para os parâmetros utilizados durante a simulação dos dados.

Assim como nos experimentos em dados reais, reservamos 35 anos de observações para realizar o treinamento do modelo. A escolha de separar aproximadamente 70 % dos dados para treinamento, deixando 30 % para testes, tenta balancear os ganhos de precisão obtidos por um maior número de dados separados para treinamento, com os ganhos obtidos de validação separando um conjunto maior para testes. Acreditamos, embora seja uma escolha arbitrária, que aproximadamente entre 10 e 15 anos de uma série para testes seja adequada para os propósitos deste trabalho.

As vantagens de se utilizar dados simulados para realizar o experimento estão na abundância de dados que podem ser obtidos através desse método, e também no fato de conhecermos exatamente os parâmetros implícitos na série gerada, permitindo-nos ter certeza de que os parâmetros estão convergindo para os valores corretos.

Uma ilustração do experimento com dados simulados pode ser encontrada na figura 3.2.

Figura 3.2: *Experimento com dados simulados: utilizamos o modelo de (Wheaton, 1999) em conjunto com parâmetros propostos pelo autor de forma a criar um conjunto de dados de estoque imobiliário simulado D_{sim} . Realizamos então o processo de seleção de modelos utilizando cross-validation em conjunto com o método de identificação de sistema proposto elaborado a partir de métodos PEM, restringindo o espaço de busca dos parâmetros n , τ , ω e ξ . Com base na estimativa de n obtida, empregamos novamente o método de identificação de sistemas proposto para encontrar os valores de parâmetros ótimos para τ , ω e ξ . Por fim, verificamos a convergência dos valores estimados para os valores corretos utilizados durante a simulação. Fonte: proposto pelo autor.*

3.2 Experimentos em dados reais

Até aqui, todos os dados utilizados e métodos desenvolvidos foram baseados em modelos teóricos e dados simulados, sobre os quais temos perfeito controle e podemos gerar uma quantidade ilimitada de matéria prima para análise. Dados reais, no entanto, sofrem normalmente de alguma limitação, seja ela econômica, física, temporal, entre outras.

Dados de estoque imobiliário são limitados tanto por motivos financeiros, devido ao custo de obtê-los, quanto por motivos temporais, já que a evolução do setor imobiliário ocorre lentamente. Em virtude destes fatores, as séries temporais de estoque imobiliário existente são escassas, já que poucos governos dedicam tempo e recursos para a coleta e manutenção de tais dados. Mesmo as séries existentes possuem poucos dados, já que a coleta dos mesmos se dá de forma anual, ou semestral, e possuem poucas décadas de existência.

Ademais, nem sempre os dados reais que temos disponíveis se ajustam aos modelos teóricos formulados, sendo necessário adaptar os modelos aos dados na incapacidade de alterarmos os dados para se adaptar aos modelos. No caso de dados de estoque imobiliários, esta é a realidade, já que não é possível voltar ao passado para obter séries temporais novas que não foram obtidas na época em que estavam disponíveis, assim como não é possível esperar para que novos dados sejam coletados, já que a obtenção de uma quantidade considerável de dados levaria tempo demais para se obter.

A principal série de dados disponível de estoques que temos para testar o método desenvolvido é uma série consistindo de dados de unidades habitacionais por área geográfica e por tipo (casas ou apartamentos) para o Canadá, abrangendo dados de 1941 até 2000 (Canada, 2016a). Esta série nos permite obter 13 séries temporais de áreas distintas, para as províncias e áreas canadenses e para o país como um todo.

Como descrito na seção 2.2.2, o modelo de Wheaton (1999), no entanto, lida com o estoque imobiliário em metros quadrados na economia. Isso decorre do fato deste ser baseado no modelo de quatro-quadrantes proposto por DiPasquale e Wheaton (1992), no qual a demanda por novo estoque em metros quadrados é dada pela entrada de novos trabalhadores em um mercado. O sistema, neste caso, sempre converge para uma relação de empregados por metro quadrado ótima, apresentando oscilações, às vezes, em torno desse ótimo quando perturbado.

Outro ponto no qual o modelo teórico difere dos dados disponíveis é quanto a natureza dos mercados imobiliários estudados. O modelo de Wheaton (1999) foi utilizado durante seu artigo para descrever o comportamento do mercado imobiliário de espaço comercial de uma economia, enquanto os dados a serem aqui analisados são referentes ao mercado habitacional canadense.

Nesta seção, formulamos experimentos semelhantes aos experimentos conduzidos com base em dados simulados, mas para serem empregados nos dados reais a serem analisados. Utilizamos novamente o processo de identificação de sistemas via PEM para guiar a construção do modelo e método, e para organizar os dados a serem utilizados durante o processo de parametrização do modelo. Adaptamos inicialmente o modelo para se adequar ao mercado imobiliário habitacional e aos dados disponíveis. Em sequência, definimos o modelo e o método empregado para a identificação de sistemas, explicamos a organização dos dados, e formulamos os experimentos a serem conduzidos.

3.2.1 Modelo e método

Assim como realizado na formulação de experimentos para dados simulados, seguimos aqui a estrutura em quatro passos para o desenvolvimento de métodos PEM, composta pela:

1. Escolha de um modelo de sistema dinâmico paramétrico;
2. Descrição do modelo como preditor de um passo a frente do estado do sistema;
3. Escolha de uma medida de erro entre previsões *vs.* dados reais;
4. Busca pelos parâmetros que minimizem os erros de previsão para o modelo desenvolvido.

Tratamos nas subseções a seguir de cada um destes passos de forma sistemática.

Modelo de sistema dinâmico

Para atender ao passo 1, duas modificações são necessárias para adaptar o modelo de sistema dinâmico desenvolvido para dados simulados em 3.1.1 de forma que o mesmo possa ser empregado para os dados reais do mercado habitacional que temos disponíveis. O modelo deve ser adaptado para as unidades de medida sendo utilizadas, e para as características específicas do mercado habitacional.

Primeiramente, o fato de não termos disponíveis o estoque imobiliário Canadense por áreas nos obriga a adequar o modelo aos dados por unidades. As características das séries de dados disponíveis são descritas em profundidade na seção 3.2.2.

A equação do modelo de Wheaton (1999) pode ser escrita da seguinte forma, como demonstrado na seção 3.1.1:

$$\frac{S_t}{S_{t-1}} = \tau + \omega \left(\frac{E_{t-n}}{S_{t-n}} \right)^\xi \quad (3.12)$$

Para contornar o descasamento entre dados e modelo, poderíamos utilizar o número total de unidades construídas em qualquer período t e transformá-las em áreas através da equação:

$$S_t = U_t \times a \quad (3.13)$$

onde u_t representa o número de unidades habitacionais disponíveis em t , e a representa a área média das unidades.

Ao fazer isso, obteríamos a seguinte equação:

$$\frac{U_t \times a}{U_{t-1} \times a} = \tau + \omega \left(\frac{E_{t-n}}{U_{t-n} \times a} \right)^\xi \quad (3.14)$$

Que pode ser simplificada para a seguinte forma:

$$\frac{U_t}{U_{t-1}} = \tau + \omega' \left(\frac{E_{t-n}}{U_{t-n}} \right)^\xi \quad (3.15)$$

onde ω' é uma nova constante obtida pela divisão de ω por a^ξ .

A premissa introduzida por este modelo é que a área média das unidades habitacionais ao longo do período é constante. Idealmente, utilizaríamos a área real de cada período no tempo, mas na ausência desta informação a introdução de alguma premissa no estimativa da área de cada período é inevitável.

Os dados disponíveis para estoque imobiliário, no entanto, são divididos entre casas e apartamentos, e seria uma premissa irrealística afirmar que casas e apartamentos possuem a mesma área média. É de se esperar que casas e apartamentos tenham áreas diferentes no geral, e a soma das unidades de casas e apartamentos, portanto, não levaria em consideração a mudança na composição do estoque imobiliário ao longo do tempo.

De forma a aproveitar da melhor forma os dados disponíveis, propomos utilizar as unidades de casas e apartamentos e acrescentar os parâmetros de área média de apartamentos (*multiples*), a_m , e área média de casas (*singles*), a_s , para permitir que o modelo faça uso dos dados sem demais modificações. A diferenciação entre unidades de casas e unidades de apartamentos no modelo resulta em:

$$\begin{cases} \frac{A'_t}{A'_{t-1}} = \tau + \omega \left(\frac{E_{t-n}}{A'_{t-n}} \right)^\xi \\ A'_t = U_{s,t} \times a_s + U_{m,t} \times a_m \end{cases} \quad (3.16)$$

onde A_t refere-se à área total de casas e apartamentos estimada para o período t , $U_{s,t}$ e $U_{m,t}$ referem-se, respectivamente, às unidades de casas (*singles*) e apartamentos (*multiples*) para o período t , e a_s e a_m são os parâmetros de área média de apartamentos e casas, respectivamente.

O modelo pode, ainda, ser simplificado para reduzir o número de parâmetros gerados pelo acréscimo dos dois novos parâmetros, a_s e a_m . De fato, tal simplificação é necessária para evitar um problema de identificabilidade destes parâmetros durante o processo de minimização do erro de predição a ser utilizada no decorrer deste trabalho.

A simplificação é obtida através da transformação dos parâmetros de áreas médias de casas em apartamentos em um único parâmetro de área relativa entre área média de casas e apartamentos. Para tal, dividimos ambos os lados das equações por a_m , assim como multiplicamos e dividimos o denominador da variável de entrada, A_t , mesmo parâmetro, resultando em:

$$\begin{cases} \frac{A'_t \left(\frac{1}{a_m} \right)}{A'_{t-1} \left(\frac{1}{a_m} \right)} = \tau + \omega \left(\frac{E_{t-n}}{A'_{t-n} \left(\frac{1}{a_m} \right) (a_m)} \right)^\xi \\ A'_t \left(\frac{1}{a_m} \right) = (U_{s,t} \times a_s + U_{m,t} \times a_m) \left(\frac{1}{a_m} \right) \end{cases} \quad (3.17)$$

Simplificando o sistema de equações, obtemos:

$$\begin{cases} \frac{A''_t}{A''_{t-1}} = \tau + \frac{\omega}{(a_m)^\xi} \left(\frac{E_{t-n}}{A''_{t-n}} \right)^\xi \\ A''_t = U_{s,t} \times \rho + U_{m,t} \end{cases} \quad (3.18)$$

onde ρ é um novo parâmetro resultante da divisão $\frac{a_s}{a_m}$, e $A'_t = A_t \left(\frac{1}{a_m} \right)$

Como ω , a_m e ξ são três parâmetros contantes, podemos realizar a substituição a seguir:

$$\begin{cases} \frac{A''_t}{A''_{t-1}} = \tau + \phi \left(\frac{E_{t-n}}{A''_{t-n}} \right)^\xi \\ A''_t = U_{s,t} \times \rho + U_{m,t} \end{cases} \quad (3.19)$$

onde ϕ é uma constante resultante dado pela divisão de ω por a_m^ξ .

A variável A''_t , no caso, substitui A_t como variável de estado do sistema e deixa de ser uma representação da área imobiliária habitacional disponível em qualquer período de tempo, passando a representar a quantidade equivalente de unidades de apartamentos do estoque imobiliário no período t caso todo o estoque fosse formado apenas por apartamentos. Esta transformação decorre da divisão da área total pela área de apartamentos, e resulta em um sistema análogo ao sistema 3.15.

Este sistema decorre naturalmente do modelo de Wheaton (1999) quando acrescido das premissas de que o estoque imobiliário é formado por duas classes diferentes de unidades habitacionais – casas e apartamentos – e que estas duas classes possuem áreas médias constantes durante o tempo. Apesar destas premissas serem restritivas, é o melhor que pode ser feito de acordo com os dados disponíveis, dado a impossibilidade de se enriquecer o conjunto de dados, conforme já discutido previamente.

Outra modificação realizada de forma adaptar o modelo original proposto por (Wheaton, 1999) para o mercado habitacional se refere ao comportamento da demanda por novo espaço imobiliário. Conforme descrito na equação 2.7, reescrita nesta seção por conveniência, temos que no modelo original formulado para o mercado comercial a equação de demanda é descrita por:

$$D_t = \alpha_1 E_t R_t^{-\beta_1}$$

com a demanda relacionada à quantidade de habitantes na economia.

A maioria dos modelos de previsão de demanda futura por espaço imobiliário comercial se baseia na relação entre a quantidade de trabalhadores por metro quadrado de uma economia, como apontado por Jacco e Robert (1996), um pressuposto que historicamente tem se mantido, já que a relação apresenta alguma oscilação histórica, e tende a apresentar um padrão de retorno à média (Miller, 2012).

Analogamente ao espaço comercial, a demanda por novo espaço imobiliário habitacional está tautologicamente relacionada ao número de habitantes de uma determinada área, afinal, para que exista demanda por novos espaços em uma certa região demográfica, é necessário que existam habitantes desejando residir em tal região (Glaeser *et al.*, 2005; Mankiw e Weil, 1989; Myers *et al.*, 2002; Thompson, 1937). De fato, a relação entre o número de habitantes de uma área e a oferta de espaço imobiliário é tão forte que o planejamento por novas áreas habitacionais deve decorrer de uma bom modelo preditivo habitacional (Alias *et al.*, 2016).

Em virtude desta relação entre habitantes e demanda por espaço habitacional, optamos por utilizar o número de habitantes de um mercado como variável de entrada substituta ao número de empregados em nossa adaptação do modelo proposto por Wheaton (1999) para o mercado habitacional. Analogamente ao modelo original, quanto maior o número de habitantes por metro quadrado de moradia em um mercado, maior será a demanda por espaço habitacional, enquanto, inversamente, quanto menor o número de habitantes, menor a demanda por espaço. Desta forma,

temos:

$$D'_t = \alpha_1 P_t R_t^{-\beta_1} \quad (3.20)$$

onde P_t representa a população do mercado no período t .

Realizando as devidas substituições, análogo ao demonstrado em 3.1.1, obtemos então o modelo de sistema dinâmico apresentado em 3.21:

$$\begin{cases} \frac{A''_t}{A''_{t-1}} = \tau + \phi \left(\frac{P_{t-n}}{A''_{t-n}} \right)^\xi \\ A''_t = U_{s,t} \times \rho + U_{m,t} \end{cases} \quad (3.21)$$

Preditor de um passo do sistema

Para atender ao passo 2 da criação de um método PEM, basta realizarmos a mesma transformação que aplicamos para o caso de dados simulados para a equação de diferença presente em 3.21. Obtemos assim:

$$\hat{A}''_t = \left(\tau + \phi \left(\frac{P_{t-n}}{A''_{t-n}} \right)^\xi \right) A''_{t-1} \quad (3.22)$$

A partir da equação 3.22, temos uma adaptação do modelo de Wheaton (1999) adequada para os dados imobiliários disponíveis sobre o mercado canadense.

A saída da equação do modelo é a estimativa de unidades equivalentes de apartamentos de um mercado em t (ie. o número equivalente de unidades habitacionais que seriam obtidas se dividíssemos a área habitacional total de um mercado pela área média de apartamentos nesse mesmo mercado).

As entradas da equação são as áreas equivalentes de apartamentos em $t-1$ e $t-n$, e o número de habitantes deste mesmo mercado em $t-n$.

Os cinco parâmetros a serem calibrados para os dados deste modelo são:

1. τ : porcentagem do estoque que se mantém de um período t para um período subsequente $t+1$.
2. ϕ : variável decorrente da junção de diversas variáveis econômicas com significados distintos, mas que por si só não possui significado econômico claro.
3. ρ : unidades equivalentes relativas a área média de apartamentos.
4. ξ : elasticidade relativa de demanda por oferta.
5. n : atraso do sistema decorrente do número de períodos que leva entre a decisão de construção de uma nova unidade e a conclusão da mesma.

Função custo

Devido às alterações aqui realizadas que alteram o modelo proposto para os experimentos realizados com base em dados simulados, também é necessário alterar a métrica de erro de forma a obtermos um método de parametrização do sistema coerente.

Uma das consequências de se adaptar o modelo de forma a utilizar toda a informação disponível pela separação do estoque entre casas e apartamentos é que a variável de área relativa, ρ , influencia tanto a área que queremos prever quanto sua previsão, conforme exibido na equação 3.23 de erro quadrático para o horizonte de um passo, exibida abaixo:

$$Erro = \sum_{t=1}^N \left(U_{s,t}\rho + U_{m,t} - \left(\tau + \phi \left(\frac{P_{t-n}}{U_{s,t-n}\rho + U_{m,t-n}} \right)^\xi \right) (U_{s,t-1}\rho + U_{m,t-1}) \right)^2 \quad (3.23)$$

Uma das consequências desta modificação, é que é possível minimizar a função erro tornando o valor da variável ρ o menor possível. Para evitar que a minimização da métrica de erro diminua o estoque total de forma a minimizar o erro total obtido, optamos por utilizar a diferença relativa entre observado e previsto, conforme a equação 3.24 abaixo:

$$Erro = \sum_{t=1}^N \left(\frac{A_t'' - \left(\tau + \phi \left(\frac{P_{t-n}}{A_{t-n}''} \right)^\xi \right) A_{t-1}''}{A_t''} \right)^2 \quad (3.24)$$

A equação 3.24 pode ser escrita de maneira simplificada como a soma dos quadrados das diferenças percentuais entre as áreas previstas e as áreas observadas, como ilustrado na equação 3.25 seguir:

$$Erro = \sum_{t=1}^N \left(1 - \frac{\hat{A}_t''}{A_t''} \right)^2 \quad (3.25)$$

Minimização

Assim como realizado nos experimentos conduzidos com dados simulados, utilizamos o algoritmo *GlobalSearch* (Mathworks, 2016b) para realizar a minimização do erro de predição de forma a determinar os parâmetros τ , ω e ξ ótimos para o conjunto de dados de treinamento utilizado. Para determinar o valor de n , utilizamos a técnica de *cross-validation* para seleção de modelos, conforme explicado na seção 2.3.3.

Restringimos os valores dos parâmetros de busca de forma a garantir que os valores obtidos sejam economicamente verossímeis. Os valores utilizados de forma a restringir o espaço de busca dos parâmetros são descritos na tabela 3.4.

Parâmetro	limite inferior ($\geq$)	limite superior ($\leq$)
n	2	5
τ	0,98	1,0
ω	0	50
ξ	0,1	10,0
ρ	1,0	3,0

Tabela 3.4: Restrição do espaço de busca. Fonte: proposto pelo autor.

3.2.2 Dados do mercado imobiliário canadense

Os dados obtidos para a realização deste trabalho foram fornecidos pela Agência Nacional de Estatística do Canadá (Canada, 2016a).

O conjunto de dados de estoque imobiliário histórico do Canada contém 405 séries, com dados de 1941 a 2000 (nem todas as combinações necessariamente possuem dados para todos os anos). O conjunto contém dados descritos pelas seguintes dimensões:

- **Geografia (15 itens):** Canadá; Newfoundland e Labrador; Prince Edward Island; Nova Scotia; New Brunswick; Quebec; Ontario; Manitoba; Saskatchewan; Alberta; British Columbia; Yukon e Northwest Territories incluindo Nunavut; Yukon; Northwest Territories; Nunavut;
- **Tipo de moradia (3 itens):** Total de apartamentos e casas; total de casas; total de apartamentos;
- **Tipo de posse (9 itens):** Estoque total líquido; estoque total líquido, comprado; estoque total líquido, alugado; total de unidades habitacionais vagas; unidades habitacionais vagas, para

venda; unidades habitacionais vagas, para alugar; total de unidades habitacionais ocupadas; unidades habitacionais ocupadas, compradas; unidades habitacionais ocupadas, alugadas.

As 405 séries disponíveis referem-se às diferentes combinações das três dimensões descritas anteriormente, e abrangem os territórios ilustrados na figura 3.3.

Figura 3.3: Mapa político do Canadá. Fonte: adaptado de *Worldatlas* (2018).

Os dados de estoque imobiliário, no entanto, referem-se a dados de unidades de moradias, e não a área habitável disponibilizada por tais moradias. Apesar de termos tentado entrar em contato com o órgão de estatística canadense para obter dados de estoque de área imobiliária, o contato revelou que esta série histórica não existe em sua base de dados. Na seção 3.2.1 adaptamos o modelo para contornar este descasamento de formatos.

Os conjuntos de dados imobiliários são, então, utilizados para extrair as séries de dados de estado do sistema imobiliário proposto.

Outra série necessária para testar o modelo de Wheaton (1999) é a série de entradas do sistema. No modelo original, destinado ao estudo do sistema imobiliário comercial, são utilizados como dados de entrada séries temporais de número de empregados. A escolha do autor é decorrente do modelo proposto por DiPasquale e Wheaton (1992, 1996), no qual a construção de novo espaço imobiliário comercial é decorrente da demanda por novo espaço comercial. A demanda por novo espaço comercial, por sua vez, é dada pela relação entre o estoque disponível de espaço comercial e o número de empregados da economia. Quanto maior o número de empregados por metro quadrado, maior a demanda por espaço, enquanto, inversamente, quanto menor o número de empregados, menor a demanda por espaço. Esta comunicação é realizada pelo preço por metro quadrado do estoque atual disponível.

Conforme explicado na seção 3.1.1, de forma a adaptar o modelo original para o mercado imobiliário habitacional, utilizamos como variável de entrada a quantidade de habitantes de um mercado. Novamente, obtemos as séries temporais de população para o Canadá, províncias e regiões diretamente através da base de dados disponível publicamente no sítio da Agência Nacional de Estatística do Canadá (Canada, 2016c). O conjunto de dados abrange um total de 15 séries descrevendo estimativas trimestrais de população para o Canadá, suas províncias e regiões entre o período compreendido entre 1946 até 2017. A tabela contém séries descritas pelas seguintes dimensões:

- **Geografia (15 itens):** Canadá; Newfoundland e Labrador; Prince Edward Island; Nova Scotia; New Brunswick; Quebec; Ontario; Manitoba; Saskatchewan; Alberta; British Columbia; Yukon e Northwest Territories incluindo Nunavut (terminado); Yukon (terminado); Northwest Territories; Nunavut (terminado);

O conjunto de dados encontra-se expresso em números de habitantes, e, portanto, não precisamos adaptar o modelo já que as unidades de medida da série de entrada encontram-se em acordo com o modelo original.

Organização dos dados

Para a execução dos experimentos com dados reais do mercado Canadense, os dois conjuntos de dados referentes às variáveis de estado e de entrada tiveram de ser organizados em grupos coesos de séries temporais com significado econômico.

Embora a descrição do conjunto de dados de estoque imobiliário (Canada, 2016a) e de população (Canada, 2016c) leve a crer que o conjunto possui o mesmo número de observações para todas as áreas geográficas presentes, algumas regiões possuem séries temporais mais longas do que outras. Ao mesmo tempo, os prazos entre os conjuntos de dados de estoque imobiliário e de população não são os mesmos. Organizamos os dados de forma a criar séries temporais que possuam ambas observações (de estoque imobiliário e de população) para cada período no tempo, eliminando assim as observações que não permitiriam realizar previsões segundo o modelo proposto.

Utilizamos como período base para criação das séries temporais de população os dados disponíveis para o último trimestre de cada ano, já que estes incorporam o número real de pessoas esperadas para cada ano base, embora a escolha de outros trimestres levariam a resultados muito semelhantes, já que as séries temporais para cada trimestre são muito semelhantes entre si, como é possível visualizar, por exemplo, nos gráficos contidos na figura 3.4, que mostra as séries temporais anuais, organizadas pelos trimestres de suas mensurações, da população canadense em nível nacional e por províncias.

Para a série de estoques, optamos por utilizar o estoque total líquido, que engloba tanto o estoque de casas e apartamentos ocupados quanto os vagos. Esta medida é mais condizente com o modelo proposto por Wheaton (1999), no qual o estoque total utilizável em uma economia é empregado. Para ilustrar, as séries temporais anuais de estoque imobiliário para Newfoundland e Labrador por casas e apartamentos são exibidas na figura 3.5.

Optamos por não incluir Yukon, territórios do noroeste e Navut no conjunto de dados a serem analisados segundo o modelo proposto. Esta opção decorre do fato da série destes territórios, mesmo em conjunto, possuírem pouca população quando comparados às demais regiões canadenses, por apresentarem uma série de estoques que se inicia apenas em 1956, e uma série de população estimada com pouca qualidade até 1971, reduzindo de fato a série histórica adequada para análise para um total de 29 observações, que é demasiadamente curta, especialmente levando em consideração que parte dos dados deve ser separada para a etapa de teste.

As séries finais utilizadas durante o teste do modelo proposto foram construídas a partir da junção dos conjuntos de dados de estoque e população. As dimensões das séries populacionais são exibidas na tabela 3.5, enquanto as dimensões das séries de estoque são exibidas na tabela 3.6. A junção dos conjuntos de dados populacionais e de estoque resultam nas séries com dimensões exibidas na tabela 3.7, onde também é mostrado o número total de observações em cada série.

Figura 3.4: Séries temporais população de Newfoundland e Labrador. Séries anuais por trimestre. Fonte: adaptado de *Canada (2016c)*.

Região	Início	Término
Canadá	1946	2017
Alberta	1951	2017
British Columbia	1951	2017
Manitoba	1951	2017
New Brunswick	1951	2017
Newfoundland e Labrador	1951	2017
Nova Scotia	1951	2017
Ontário	1951	2017
Prince Edward Island	1951	2017
Quebec	1951	2017

Tabela 3.5: Dimensões das séries populacionais do Canadá e Províncias. Fonte: adaptado de *Canada (2016c)*

3.2.3 Experimentos

Enquanto em experimentos conduzidos com base em dados simulados os valores corretos dos parâmetros do modelo são conhecidos previamente, em experimentos conduzidos em dados reais este não é o caso. Desta forma, em experimentos conduzidos com base em dados reais, não podemos utilizar a mesma estrutura de teste proposta na seção 3.1.3 formulada para testar modelos aprendidos com base em dados simulados.

Conforme explicado na seção 2.3.3, utilizamos as duas formas de teste de modelos propostas por *Ljung (2006)*. Dividimos inicialmente o conjunto de dados em dados de treinamento e dados de teste, reservando 35 anos para treinamento e deixando os demais dados para teste, e após obtidas as estimativas de parâmetros ótimos a partir da conclusão da etapa de treinamento, realizamos os procedimentos de predição e simulação nos dados de teste, e verificamos os resultados obtidos.

O experimento proposto, dessa forma, consiste em:

1. Organizar as séries de dados disponíveis sobre o mercado canadense, conforme descrito na seção 3.2.2, separando os dados em conjuntos de treinamento e teste;
2. Utilizar o modelo proposto na seção 3.2.1 e parametriza-lo segundo o método desenvolvido na mesma seção;

Figura 3.5: Séries anuais de estoque imobiliário – Canadá e Províncias. Fonte: adaptado de *Canada (2016a)*.

Região	Início	Término
Canadá	1941	2000
Alberta	1941	2000
British Columbia	1941	2000
Manitoba	1941	2000
New Brunswick	1941	2000
Newfoundland e Labrador	1949	2000
Nova Scotia	1941	2000
Ontario	1941	2000
Prince Edward Island	1941	2000
Quebec	1941	2000

Tabela 3.6: Dimensões das séries de estoque imobiliário do Canadá e Províncias. Fonte: adaptado de *Canada (2016a)*

3. Testar os modelos parametrizados obtidos durante a etapa de treinamento nos dados de teste correspondentes às respectivas séries de dados de treinamento.

A estrutura do experimento proposto encontra-se ilustrada na figura 3.6:

Região	Início	Término	Nº de Observações
Canadá	1946	2000	54
Alberta	1951	2000	49
British Columbia	1951	2000	49
Manitoba	1951	2000	49
New Brunswick	1951	2000	49
Newfoundland e Labrador	1951	2000	49
Nova Scotia	1951	2000	49
Ontario	1951	2000	49
Prince Edward Island	1951	2000	49
Quebec	1951	2000	49

Tabela 3.7: Dimensões das séries de estoque e população utilizadas. Fonte: adaptado de *Canada (2016a,c)*

Figura 3.6: Experimento com dados reais. Fonte: proposto pelo autor.

Capítulo 4

Resultados e análise

Reportamos neste capítulo os resultados obtidos na condução dos experimentos descritos no capítulo 3. Inicialmente, descrevemos os resultados obtidos nos experimentos conduzidos em dados simulados, de forma a mostrar a convergência do método proposto aos parâmetros corretos. Em sequência, reportamos os resultados obtidos nos experimentos realizados em dados reais.

Com o intuito de tornar os resultados obtidos de fácil entendimento, ilustramos neste capítulo apenas alguns exemplos de resultados necessários à análise do método proposto. As tabelas completas com a descrição dos dados, parâmetros obtidos e gráficos gerados, que se apresentadas em sua totalidade atrapalhariam a análise dos resultados devido ao volume de séries analisadas, podem ser encontradas no apêndice deste trabalho.

Conforme explicado no capítulo 3, dividimos a análise dos resultados dos experimentos em três partes. Primeiro, analisamos os resultados obtidos para os parâmetros resultantes da aplicação do método de identificação de sistemas desenvolvido, interpretando o significado econômico dos mesmos. Em sequência, e seguindo a análise sugerida por [Ljung \(2006\)](#), avaliamos a qualidade dos modelos produzidos através dos métodos de predição e simulação, comparando os resultados obtidos com o conjunto de dados reservado para testes. Por fim, avaliamos as características dos mercados analisados e como estes se relacionam aos resultados obtidos.

4.1 Experimentos em dados simulados

Os resultados obtidos através dos experimentos em dados simulados corroboram a validade do método desenvolvido em identificar corretamente o sistema dinâmico implícito dentro dos dados. A seguir, analisamos os detalhes dos resultados obtidos durante os experimentos.

4.1.1 Parâmetros obtidos

Ao todo, realizamos quatro experimentos com base em séries temporais simuladas distintas, seguindo o método proposto na seção 3.1. Em todos os casos, os parâmetros convergem para os valores corretos, conforme esperado. A tabela 4.1 resume os resultados obtidos durante os experimentos conduzidos por simulação para os quatro parâmetros de interesse.

Comparando-se os valores utilizados durante a simulação com os valores obtidos pelo processo de identificação de sistemas, temos que em todos os experimentos os parâmetros convergem corretamente para os valores alvo, sendo que resultados mais próximos poderiam ser obtidos alterando-se os parâmetros de convergência do algoritmo. Utilizamos parâmetros iniciais com valores economicamente plausíveis para iniciar a busca pelos parâmetros corretos, e embora os valores iniciais escolhidos possam ser arbitrários, estes apresentam pouca relevância para as conclusões extraídas, já que valores próximos aos valores iniciais de busca poderiam ser obtidos através do uso de um algoritmo de minimização com multi-início.

Outro fato importante a se observar, é que delimitamos espaços de busca diferentes para a identificação dos parâmetros ω . Esta escolha decorre não da incapacidade do método de identificação

Parâmetro	Simulado	Resultado	limite inferior ($\geq$)	limite superior ($\leq$)	Valor inicial
n_1	5	5	2	10	-
n_2	5	5	2	10	-
n_3	8	8	2	10	-
n_4	5	5	2	10	-
τ_1	0,95	0,9497	0,9	1,0	0,98
τ_2	0,95	0,9500	0,9	1,0	0,98
τ_3	0,95	0,9500	0,9	1,0	0,98
τ_4	0,90	0,9000	0,9	1,0	0,98
ω_1	12,5	12,2123	5,0	30,0	20,0
ω_2	4.88×10^{10}	4.87×10^{10}	4.0×10^{10}	5.0×10^{10}	4.5×10^{10}
ω_3	4.88×10^{10}	4.88×10^{10}	4.0×10^{10}	5.0×10^{10}	4.5×10^{10}
ω_4	9.77×10^{10}	9.76×10^{10}	9.0×10^{10}	10.0×10^{10}	9.5×10^{10}
ξ_1	1,0	0,9946	0,1	10,0	4,0
ξ_2	5,0	4,9995	0,1	10,0	4,0
ξ_3	5,0	5,0000	0,1	10,0	4,0
ξ_4	5,0	5,0000	0,1	10,0	4,0

Tabela 4.1: Constantes, entradas e estados iniciais. Fonte: proposto pelo autor.

de sistemas em encontrar corretamente o valor do parâmetro ω utilizado, mas sim da finalidade das séries simuladas com o intuito de produzir diferentes comportamentos no sistema. O parâmetro ω , em dados simulados, decorre da equação:

$$\omega = \frac{\alpha_2 \alpha_1^{\frac{\beta_2}{\beta_1}}}{r^{\beta_2}} \quad (4.1)$$

Ou seja, como r é constante e α_1 e α_2 tem seus valores calculados a partir de β_1 e β_2 , o valor de ω opera em patamares diferentes durante as simulações exatamente devido à intenção do autor de verificar como oscilações bruscas nos valores das elasticidades afetam o comportamento do sistema. No entanto, se isolarmos o parâmetro ω da equação do modelo, verificamos que o mesmo pode ser expresso segundo a equação 4.2:

$$\omega = \frac{\frac{S_t}{S_{t-1}} - \tau}{\left(\frac{P_{t-n}}{S_{t-n}}\right)^\xi} \quad (4.2)$$

Temos que o valor de ω é decorrente de τ , ξ e do comportamento das curvas de estoque e entrada. Como não esperamos que os dois primeiros oscilem muito em seus valores, e como os comportamentos das curvas de estoque e população são similares entre as diversas séries temporais que temos disponíveis, o valor de ω não deveria apresentar muita oscilação entre mercados diferentes, e, portanto, a delimitação de espaços de busca distintos para cada série não seria necessária nos experimentos realizados nos dados canadenses.

Durante a análise dos resultados obtidos com dados reais, ficará claro que os valores obtidos para o parâmetro ω convergem para valores próximos ao longo de todas as séries analisadas, assim como esperado.

Ademais, as diferenças obtidas entre os parâmetros utilizados durante a simulação dos dados e os parâmetros resultantes do processo de identificação de sistemas são economicamente irrelevantes.

4.1.2 Predição em dados de teste

Conforme esperado em dados simulados, as métricas de erro de previsões de um passo a frente em dados de treinamento apresentam valores bem baixos, sendo um pouco inferiores à mesma métrica em dados de teste. A tabela 4.2 exibe os resultados obtidos para ambas as métricas citadas.

Série	E_{in}	E_{out}
Simulação 1	1.60×10^{-6}	1.84×10^{-6}
Simulação 2	1.83×10^{-6}	2.08×10^{-6}
Simulação 3	7.35×10^{-7}	7.90×10^{-7}
Simulação 4	8.82×10^{-7}	1.13×10^{-6}

Tabela 4.2: Resultados de experimentos em dados simulados – métricas de erro de predição. Fonte: proposto pelo autor.

4.1.3 Simulação em dados de teste

Embora a simulação de dados seja um teste mais rigoroso de validade do modelo obtido, como construímos o espaço de busca e geramos os dados de forma que os mesmos contêm o sistema real, e sabemos que os valores dos parâmetros convergem adequadamente para os valores alvo, é de se esperar que os resultados obtidos através da simulação do modelo identificado no conjunto de dados de teste produza resultados com quase perfeita sobreposição.

Assim como esperado, a simulação do modelo em dados não utilizados durante o treinamento mostra que as duas séries não apresentam diferenças significativas, como é possível observar na figura 4.1.

Figura 4.1: Resultados experimentos – dados simulados. Fonte: proposto pelo autor.

Como é possível notar nos gráficos da figura 4.1, as séries de referência, utilizadas para compor os conjuntos de treinamento e teste, são sobrepostas pelas séries simuladas a partir dos parâmetros aprendidos. A simulação se inicia no início do conjunto de teste, e se mantém até a sua conclusão. De fato, as séries de referência e as séries simuladas se sobrepõem tão bem para dados simulados, que não é possível distingui-las visualmente nos gráficos, ficando visível apenas as séries simuladas a partir do fim do conjunto de dados de treinamento.

Como os erros percentuais para as quatro séries não ultrapassam a ordem de grandeza de 10^{-6} , e

como a finalidade dos testes realizados em dados simulados eram tão somente verificar a convergência dos parâmetros aprendidos aos reais, não apresentamos aqui os gráficos de erro obtidos, já que estes não contribuem para a análise desenvolvida neste trabalho.

Um dos motivos pelos quais os resultados por simulação se mantêm tão próximos das séries de referência utilizadas no processo de aprendizado é que os parâmetros durante a simulação se mantêm constantes durante o tempo.

Desta forma, como sugerido por [Ljung \(2002\)](#), podemos utilizar o método desenvolvido em dados reais com a confiança de que todos os parâmetros do modelo convergem para suas melhores aproximações caso haja dados o suficiente e seja empregado um método de otimização adequado.

4.2 Experimentos em dados reais

Realizamos, conforme descrito na seção 3.2 os experimentos utilizando método similar ao método utilizado para dados simulados, mas adaptado para as características das informações disponíveis para dados do mercado imobiliário canadense. Como o conjunto de séries de dados canadenses é composto por um total de onze séries distintas, apresentamos aqui apenas um conjunto de exemplos e dados agregados de forma a realizar a análise dos resultados obtidos.

4.2.1 Parâmetros obtidos

Os valores obtidos durante o processo de aprendizado para os parâmetros do modelo desenvolvido são conforme a tabela 4.3.

Região	n	τ (%)	ϕ (10^{-2})	ρ	ξ
Alberta	5	98,01	2,2	3,0	2,4
British Columbia	3	98,10	2,7	3,0	2,1
Canadá	4	98,17	2,7	1,7	0,5
Manitoba	2	98,36	2,6	2,7	0,5
New Brunswick	5	99,97	1,8	1,6	0,1
Newfoundland e Labrador	5	98,45	2,5	1,4	0,3
Nova Scotia	5	99,98	1,6	1,8	0,1
Ontário	2	99,96	9,1	2,0	1,5
Prince Edward Islands	5	99,99	1,5	1,0	0,1
Quebec	5	99,93	7,4	1,0	1,1
Saskatchewan	4	99,98	1,2	1,0	0,1

Tabela 4.3: Resultados de experimentos em dados reais – parâmetros e métricas de erro. Fonte: proposto pelo autor.

Como em modelos caixa-cinza não possuímos a garantia de que o verdadeiro modelo encontra-se contido no conjunto de hipóteses delimitado pelo mesmo, os valores obtidos durante o processo de aprendizado para os parâmetros do modelo não são particularmente informativos, já que os mesmos representam a melhor aproximação do modelo real contida dentro do espaço de hipóteses ([Gevers, 2006](#)). Além disso, em contraste aos experimentos realizado com dados simulados, não temos o benefício de conhecer os valores corretos dos parâmetros para os quais os valores devem convergir. Devido a isso, conforme explicado, seguimos a sugestão de [Ljung \(2002\)](#) e verificamos o resultado obtido pelo método através da comparação entre as séries de referência contra as simulações obtidas com a aplicação dos parâmetros obtidos durante a fase de aprendizado.

Embora não possamos extrair conclusões com base tão somente nos parâmetros aprendidos, é interessante notar alguns padrões nos valores obtidos.

- O parâmetro τ , por exemplo, converge para o limite superior de 100% para algumas regiões, enquanto outras convergem para valores próximos ao limite inferior de 98%, sem encostar em nenhum dos limites. Como o valor de τ representa o quanto do estoque imobiliário se mantém

de um período para outro, e como o estoque imobiliário é um bem durável, esperaríamos que seu valor naturalmente fosse próximo de 1,0 como sugerido por [Wheaton \(1987\)](#). Como nenhum dos valores obtidos encosta no limite inferior de 98%, consideramos que os valores obtidos retornam dentro do esperado;

- Os valores obtidos para a variável de atraso do sistema, n , não apresentam um padrão claro que possamos identificar;
- O parâmetro ϕ , apesar de não possuir um significado econômico claro, converge para valores baixos em todos os casos. Outro ponto a notar, é que os valores não divergem de maneira significativa entre si, assim como apontado em 4.1. Acreditamos que a proximidade dos valores de ϕ ocorra em virtude da similaridade entre as taxas de crescimento médias dos estoques entre todos os conjuntos de dados, enquanto o baixo valor médio deste decorra devido ao lento crescimento percentual do estoque ao longo do tempo;
- Os resultados obtidos para o parâmetro ρ também não apresentam um padrão claro, com valores contidos entre 1 e 3.
- O valor de ξ varia consideravelmente entre as regiões, mas fica abaixo do valor estimado por [Wheaton \(1999\)](#) de 5 para mercados comerciais, como é de se esperar. Como o parâmetro resulta da divisão de dois parâmetros, é economicamente aceitável uma variação maior no valor deste, já que as elasticidades de oferta e demanda podem variar bastante de uma região para outra, dependendo das características do mercado local. É interessante notar que Alberta, British Columbia, Ontário e Quebec, as quatro regiões mais desenvolvidas economicamente ([Canada, 2016b](#)), obtêm valores de ξ superiores a 1,0, enquanto as demais províncias obtêm valores inferiores a 1,0. Isto indica a possibilidade de uma relação entre o desenvolvimento econômico de uma região e a relação entre sua elasticidade de oferta e demanda, já que faz sentido esperar que regiões economicamente desenvolvidas possuam uma responsividade melhor na oferta de habitações do que regiões menos desenvolvidas. Esta relação, no entanto, deveria ser estudada mais a fundo através de técnicas econométricas;

4.2.2 Predição em dados de teste

Avaliamos aqui os resultados obtidos através das métricas de erro de predição de um passo a frente no sistema em dados de treinamento (E_{in}) e dados de teste (E_{out}). Os resultados obtidos são exibidos na tabela 4.4.

Name	E_{in} (%)	E_{out} (%)
Alberta	0,88	0,94
British Columbia	0,40	0,42
Canadá	0,39	0,42
Manitoba	0,34	0,36
New Brunswick	0,61	0,65
Newfoundland e Labrador	0,54	0,60
Nova Scotia	0,57	0,60
Ontário	0,45	0,51
Prince Edward Islands	0,86	0,90
Quebec	0,40	0,43
Saskatchewan	0,60	0,61

Tabela 4.4: Resultados de experimentos em dados reais – métricas de erro de predição. Fonte: proposto pelo autor.

A tabela exhibe as métricas de erro por região e por conjunto de dados das métricas de erro de um passo a frente do sistema medidas em termos percentuais médios. Em linha com o esperado, os erros obtidos são menores para os conjuntos de dados de treinamento, piorando seu resultado

quando aplicados a dados de teste. Os erros, no geral, são bem baixos para ambas as amostras, sendo em média 0,55% para dados de treinamento e 0,59% para dados de teste. Os erros obtidos são superiores aos erros obtidos nos experimentos realizados em dados simulados, o que é natural, já que, por construção, para dados simulados existe uma combinação de parâmetros que resulta em erro 0, o que não é o caso para séries de dados reais.

4.2.3 Simulação em dados de teste

Os resultados apresentados durante a simulação de modelos estão em linha com nossa expectativa inicial. Em nenhuma das simulações os resultados apresentam oscilações endógenas, assim como esperado para mercados residenciais. Além disso, as simulações obtêm bons resultados durante os primeiros anos, com os erros aumentando com a passagem do tempo. Em alguns casos, as simulações voltam a convergir aos valores reais após alguns anos, mas o fazem por mero acaso. O erro absoluto médio obtido por horizonte de previsão é exibido na figura 4.2.

Figura 4.2: Erro percentual absoluto médio por ano de simulação. Fonte: proposto pelo autor.

Dependendo da finalidade do modelo de previsão, portanto, o aumento do erro médio dos modelos obtidos por horizonte de previsão pode tornar seus resultados não confiáveis.

Os erros percentuais obtidos por região por passos são exibidos na figura 4.3.

Figura 4.3: Erro percentual por ano de simulação por região. Fonte: proposto pelo autor.

É possível notar que em todas as regiões, exceto Alberta, o modelo superestima o estoque imobiliário alguns anos após 1990. A década de 90 é marcada por uma desaceleração no mercado de construção habitacional canadense, influenciado tanto por uma crise econômica e de excesso de oferta (Fréchette, 2016; Reynaud, 2015), quanto pela adoção de uma agenda neoconservativa de descentralização de programas sociais e redução de custos do governo canadense, que buscava reduzir o déficit fiscal pelo qual passava (CMHC, 2011; Dreier e Hulchanski, 1993; Sousa e Quarter, 2003; Van Dyk, 1995).

Embora cada província tenha sido afetada de maneira diferente, devido à estrutura provincial de incentivo e subsídios ao setor de construção, é possível notar uma desaceleração do setor imobiliário como um todo. Províncias como Saskatchewan, Prince Edward Islands e Quebec apresentam as maiores divergências negativas em relação ao modelo. British Columbia passa a apresentar desaceleração na taxa de crescimento do estoque imobiliário com um pouco de defasagem, enquanto a região de Alberta aparenta não ser afetada pela desaceleração.

O melhor resultado obtido pelo modelo se refere aos dados do agregados do Canadá. Por representar o mercado imobiliário do país como um todo, esta série não apresenta as variações decorrentes de choques de mercado específicos de províncias, mantendo uma taxa de crescimento mais constante que as demais séries de dados provinciais, como é possível observar na figura 4.4. Um dos possíveis fatores que contribui para o crescimento estável do mercado canadense é por este não ser afetado por movimentos migratórios internos, já que a movimentação entre províncias pode justificar o crescimento ou diminuição nas taxas de crescimento dos mercados imobiliários locais, mas não o crescimento ou diminuição nas taxas de crescimento do mercado imobiliário nacional. Como é possível notar, também, o mercado canadense como um todo demora um pouco mais a ser afetado pela desaceleração econômica no setor habitacional e pelo fim do incentivo federal ao setor imobiliário.

Figura 4.4: Gráfico de resultados simulados para os dados do Canadá. Fonte: proposto pelo autor.

No nível provincial, os conjuntos de dados apresentam maiores oscilações nas taxas de crescimento ao longo do tempo. As províncias de Saskatchewan e Quebec são as séries com maior oscilação nas taxas de crescimento, e, conforme esperado, o modelo apresenta seu pior desempenho nessas duas regiões. Em contraste, mercados imobiliários mais estáveis apresentam desempenhos melhores de simulação, apresentando erros relativamente baixos por horizontes de previsões maiores, como

é possível observar nas simulações para as províncias de British Columbia e Nova Scotia. Gráficos referentes às simulações dos modelos aprendidos com base nos dados dessas províncias podem ser encontrados na figura 4.5.

Figura 4.5: Resultados experimentos – dados reais. Fonte: proposto pelo autor.

Podemos notar que a estabilidade de mercados, que em princípio se traduzem para parâmetros econômicos mais estáveis, afetam positivamente o resultado dos modelos obtidos. O fato do modelo proposto por Wheaton (1999) partir da premissa de parâmetros econômicos constantes ao longo do tempo, impede que o mesmo se adapte, ou até mesmo preveja a alteração dos mesmos, fazendo com que resultado dos modelos obtidos deteriore com o advento de alterações nas características do mercado.

Ao agregar diversos fatores econômicos em uma mesma variável constante ϕ , o modelo torna-se incapaz de prever a desaceleração no ritmo de crescimento do mercado no período pós 1985. Como seria de se esperar, as constantes econômicas do modelo oscilam de acordo com o tempo decorrente de outros fatores de mercado além da densidade populacional da área habitável de uma região.

Para melhorar a performance do modelo, a forma com a qual os parâmetros econômicos se alteram ao longo do tempo no sistema dinâmico imobiliário deveriam ser estudadas. Em virtude dos resultados obtidos, é possível conjecturar-se que o enriquecimento do modelo com dados de atividade econômica e de estímulos para o setor possam melhorar o seu desempenho preditivo, já que ambos afetam diretamente o setor imobiliário como um todo. Um modelo que possua parâmetros variáveis, e que consiga modelar corretamente as alterações dos parâmetros econômicos de Wheaton (1999) ao longo do tempo, pode melhorar significativamente o resultado do mesmo em séries com mercados imobiliários menos estáveis.

Capítulo 5

Conclusões

Como ressaltado durante a introdução, a importância de modelos preditivos em um segmento da economia que representa grande parte do capital econômico acumulado de um país, torna os modelos preditivos neste setor econômico algo relevante para os agentes que coordenam as alocações de recursos e a criação de políticas que afetam a construção e destruição de áreas habitacionais, comerciais e industriais. A capacidade de prever o estado do mercado imobiliário em períodos futuros tem o potencial de tornar as decisões tomadas hoje por estes agentes melhor fundamentadas, que por sua vez pode evitar recursos de serem alocados de maneira ineficiente na economia. O capital gasto durante a construção de uma casa construída em um período de excesso de produção habitacional poderia ser melhor alocado em outra atividade econômica que possua melhor demanda futura, já que o intuito da produção em uma economia não é apenas produzir bens e serviços, mas sim prover aquilo que indivíduos desejam consumir (Friedman e Friedman, 1990).

O modelo de quatro-quadrantes de DiPasquale e Wheaton (1996) é um modelo de grande popularidade tanto no meio acadêmico quanto entre participantes do mercado imobiliário devido sua simplicidade e elegância, e a versão de sistema dinâmico do mesmo proposto por Wheaton (1999) fornece os meios para utilizar o mesmo para fazer previsões sobre futuros estados do estoque imobiliário.

Este trabalho foi escrito com o intuito de se testar o modelo de Wheaton (1999) como ferramenta preditiva do mercado imobiliário através de sua aplicação em um conjunto de dados reais. Para tanto, obtivemos séries históricas de dados reais e modificamos o modelo, tanto para se adequar matematicamente às técnicas preditivas de identificação de sistemas, quanto para se adequar aos dados disponíveis.

Os resultados obtidos em dados simulados validam a capacidade do método proposto em identificar corretamente o sistema dinâmico gerador das séries, dado que o mesmo encontra-se contido no espaço de busca do mesmo. Em todos os casos, os parâmetros aprendidos convergiram corretamente para os valores alvo.

Tomados sozinhos, os resultados obtidos com experimentos em dados simulados corroboram o método preditivo criado, mas não corroboram o modelo dinâmico de Wheaton (1999) em si. A confiança no método conferida pelos resultados positivos obtidos em dados simulados, no entanto, motiva sua aplicação em dados reais.

Os resultados obtidos em dados reais, por sua vez, retornaram dentro de nossas expectativas *a priori*. Os erros obtidos por predição de um passo a frente em dados de treinamento apresentam valores pequenos, com o modelo apresentando uma deterioração em desempenho preditivo quando testado em dados não utilizados na etapa de treinamento.

O teste por simulação do sistema em dados de teste, como sugerido por Ljung (2006), também retorna dentro do esperado, com o modelo apresentando bons resultados em períodos mais próximos, e piorando sua performance com a passagem do tempo.

A piora dos resultados com a passagem do tempo durante a simulação de modelos é algo esperado. Uma das fontes da piora de resultados é o próprio acúmulo de erros decorrente de gerar previsões com base em números estimados para períodos anteriores. Em algumas séries, no entanto,

é possível perceber-se modificações no próprio mercado imobiliário não explicadas pela entrada do sistema, o que sugere uma alteração do mercado imobiliário que não é passível de ser capturada ou prevista pelo sistema de parâmetros constantes.

Desta forma, uma das limitações que podemos identificar no modelo está na premissa de que os parâmetros econômicos de um mercado imobiliário se mantêm constantes ao longo do tempo. Em mercados onde estes parâmetros se modificarem ao longo do tempo, obviamente, o método criado não apresentará bons resultados, a não ser que seja proposto um método de prever as alterações dos mesmos ao longo do tempo.

A análise econômica dos motivos que levam o modelo a superestimar o estoque imobiliário por simulação a partir da década de 90 também leva a crer que o modelo poderia ser enriquecido com a agregação de outros fatores econômicos, como a atividade econômica e o montante de subsídios conferidos ao setor imobiliário, embora a aquisição destes dados possa tornar difícil a aplicação do mesmo.

Acreditamos que futuros trabalhos dedicados ao estudo da evolução dos parâmetros do modelo de [Wheaton \(1999\)](#) ao longo do tempo podem representar uma grande contribuição ao tema. Muitos dos parâmetros dados como constantes, podem, possivelmente, ser melhor modelados como entradas do sistema, e não como parâmetros. Isto permitiria ao método melhor se adaptar a alterações no ambiente econômico. Ademais, outros trabalhos que sejam desenvolvidos com base nesta pesquisa e que apliquem o método aqui desenvolvido em outros conjuntos de dados, como por exemplo dados imobiliários de outros países ou de mercados distintos (*e.g.* comercial, industrial, hoteleiro, entre outros) poderão estudar o comportamento do modelo em séries históricas com outras características, ajudando a compreender o desempenho preditivo do modelo de [Wheaton \(1999\)](#).

Um estudo em particular que acreditamos que seria de grande relevância para o tema seria a aplicação do método aqui desenvolvido em dados do mercado imobiliário comercial dos Estados Unidos, uma vez que este motivou a criação do modelo de [Wheaton \(1999\)](#) em primeiro lugar. Para [Wheaton \(1999\)](#), mercados imobiliários comerciais são propensos a oscilações endógenas, e o teste do método aqui desenvolvido em dados com este comportamento poderiam validar a capacidade do modelo em identificar movimentos oscilatórios em mercados comerciais, de forma a melhorar a alocação de recursos nos mesmos.

Esperamos que outros trabalhos na área se beneficiem dos resultados aqui obtidos como base de comparação para seus próprios resultados, de forma a avançar ainda mais o conhecimento neste campo de estudo.

No geral, os resultados obtidos corroboram o modelo de [Wheaton \(1999\)](#) como ferramenta preditiva no mercado imobiliário em períodos de estabilidade do setor, e aumentam a confiança que usuários do mesmo podem ter neste como ferramenta para guiar decisões no mercado imobiliário, com a cautela de considerar possíveis impactos que alterações macroeconômicas possam ter sobre suas conclusões.

Embora as limitações do estudo não permitam realizar afirmações conclusivas quanto à validade do modelo de [Wheaton \(1999\)](#), seus resultados positivos motivam futuros estudos no tema. É importante notar que ao mesmo tempo em que a limitação de dados restringiu as conclusões do estudo, ela também permitiu que o mesmo modelo desenvolvido com o mercado imobiliário comercial em mente pudesse ser adaptado a outro mercado de características distintas, obtendo resultados coerentes.

Ademais, os passos aqui tomados para elaborar a metodologia de validação do sistema dinâmico teórico podem ser utilizados para testar outros modelos econômicos de naturezas similares.

Acreditamos que o trabalho atingiu os objetivos que o mesmo se propunha a atingir, e esperamos que o mesmo motive futuras investigações nos temas de mercados imobiliários e na avaliação de sistemas dinâmicos na economia, promovendo maior interação entre pesquisadores de disciplinas distintas.

Apêndice A

População

Figura A.1: *População Alberta.* Fonte: adaptado de *Canada (2016c)*.

Figura A.2: *População British Columbia.* Fonte: adaptado de *Canada (2016c)*.

Figura A.3: População Canadá. Fonte: adaptado de *Canada (2016c)*.

Figura A.4: População Manitoba. Fonte: adaptado de *Canada (2016c)*.

Figura A.5: População New Brunswick. Fonte: adaptado de *Canada (2016c)*.

Figura A.6: População Newfoundland e Labrador. Fonte: adaptado de *Canada (2016c)*.

Figura A.7: *População Nova Scotia. Fonte: adaptado de [Canada \(2016c\)](#).*

Figura A.8: *População Ontario. Fonte: adaptado de [Canada \(2016c\)](#).*

Figura A.9: População Prince Edward Island. Fonte: adaptado de *Canada (2016c)*.

Figura A.10: População Quebec. Fonte: adaptado de *Canada (2016c)*.

Figura A.11: *População Saskatchewan. Fonte: adaptado de [Canada \(2016c\)](#).*

Apêndice B

Estoque imobiliário

Figura B.1: *Estoque imobiliário Alberta.* Fonte: adaptado de *Canada (2016a)*.

Figura B.2: *Estoque imobiliário British Columbia.* Fonte: adaptado de *Canada (2016a)*.

Figura B.3: *Estoque imobiliário Brunswick. Fonte: adaptado de [Canada \(2016a\)](#).*

Figura B.4: *Estoque imobiliário Manitoba. Fonte: adaptado de [Canada \(2016a\)](#).*

Figura B.5: *Estoque imobiliário Newfoundland e Labrador. Fonte: adaptado de [Canada \(2016a\)](#).*

Figura B.6: *Estoque imobiliário Nova Scotia. Fonte: adaptado de [Canada \(2016a\)](#).*

Figura B.7: *Estoque imobiliário Ontario. Fonte: adaptado de [Canada \(2016a\)](#).*

Figura B.8: *Estoque imobiliário Prince Edward Islands. Fonte: adaptado de [Canada \(2016a\)](#).*

Figura B.9: *Estoque imobiliário Quebec. Fonte: adaptado de [Canada \(2016a\)](#).*

Figura B.10: *Estoque imobiliário Saskatchewan. Fonte: adaptado de [Canada \(2016a\)](#).*

Apêndice C

Simulações

Figura C.1: *Estoque simulado versus real – Alberta. Fonte: proposto pelo autor.*

Figura C.2: *Estoque simulado versus real – British Columbia. Fonte: proposto pelo autor.*

Figura C.3: *Estoque simulado versus real – Canadá. Fonte: proposto pelo autor.*

Figura C.4: *Estoque simulado versus real – Manitoba. Fonte: proposto pelo autor.*

Figura C.5: *Estoque simulado versus real – New Brunswick. Fonte: proposto pelo autor.*

Figura C.6: *Estoque simulado versus real – Newfoundland e Labrador. Fonte: proposto pelo autor.*

Figura C.7: *Estoque simulado versus real – Nova Scotia. Fonte: proposto pelo autor.*

Figura C.8: *Estoque simulado versus real – Ontário. Fonte: proposto pelo autor.*

Figura C.9: *Estoque simulado versus real – Prince Edward Islands. Fonte: proposto pelo autor.*

Figura C.10: *Estoque simulado versus real – Quebec. Fonte: proposto pelo autor.*

Figura C.11: *Estoque simulado versus real – Saskatchewan. Fonte: proposto pelo autor.*

Referências Bibliográficas

- Abu-Mostafa et al.(2012)** Yaser S Abu-Mostafa, Malik Magdon-Ismael e Hsuan-Tien Lin. *Learning from data*, volume 4. AMLBook Singapore. Citado na pág. [3](#), [20](#), [26](#), [27](#), [28](#), [29](#), [30](#), [31](#), [32](#)
- Aha(1998)** David W Aha. Feature weighting for lazy learning algorithms. Em *Feature Extraction, Construction and Selection*, páginas 13–32. Springer. Citado na pág. [23](#)
- Akaike(1974)** Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723. Citado na pág. [30](#)
- Alias et al.(2016)** Ahmad Rizal Alias, Noor Yasmin Zainun e Ismail Abdul Rahman. Comparison between arima and des methods of forecasting population for housing demand in johor. Em *MATEC Web of Conferences*, volume 81, página 07002. EDP Sciences. Citado na pág. [42](#)
- Anderson et al.(1978)** Brian DO Anderson, John B Moore e RM Hawkes. Model approximations via prediction error identification. *Automatica*, 14(6):615–622. Citado na pág. [26](#)
- Andrieu et al.(2004)** Christophe Andrieu, Arnaud Doucet, Sumeetpal S Singh e Vladislav B Tadic. Particle methods for change detection, system identification, and control. *Proceedings of the IEEE*, 92(3):423–438. Citado na pág. [25](#)
- Åström e Eykhoff(1971)** Karl Johan Åström e Peter Eykhoff. System identification—a survey. *Automatica*, 7(2):123–162. Citado na pág. [2](#)
- Åström e Torsten(1965)** Karl-Johan Åström e Bohlin Torsten. Numerical identification of linear dynamic systems from normal operating records. *IFAC Proceedings Volumes*, 2(2):96–111. Citado na pág. [2](#)
- Barlas(1996)** Yaman Barlas. Formal aspects of model validity and validation in system dynamics. *System dynamics review*, 12(3):183–210. Citado na pág. [3](#)
- Barlas et al.(2007)** Yaman Barlas, B Özbas e Onur Özgün. Modeling of real estate price oscillations in istanbul. Em *Proceedings of the 2007 System Dynamics Conference*. Citado na pág. [3](#)
- Barras(1994)** Richard Barras. Property and the economic cycle: building cycles revisited. *Journal of Property Research*, 11(3):183–197. Citado na pág. [1](#)
- Barrera(2016)** Junior Barrera. System identification. Working Paper. Citado na pág. [21](#)
- Bartlett et al.(2006)** Peter L Bartlett, Michael I Jordan e Jon D McAuliffe. Convexity, classification, and risk bounds. *Journal of the American Statistical Association*, 101(473):138–156. Citado na pág. [30](#)
- Belej e Kulesza(2015)** Mirosław Belej e Sławomir Kulesza. The dynamics of time series of real estate prices. *Real Estate Management and Valuation*, 23(4):35–43. Citado na pág. [3](#)
- Bergmeir e Benítez(2012)** Christoph Bergmeir e José M Benítez. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192–213. Citado na pág. [31](#)

- Bergmeir et al.(2014)** Christoph Bergmeir, Mauro Costantini e José M Benítez. On the usefulness of cross-validation for directional forecast evaluation. *Computational Statistics & Data Analysis*, 76:132–143. Citado na pág. 31
- Blumer et al.(1989)** Anselm Blumer, Andrzej Ehrenfeucht, David Haussler e Manfred K Warmuth. Learnability and the vapnik-chervonenkis dimension. *Journal of the ACM (JACM)*, 36(4):929–965. Citado na pág. 27
- Box et al.(2015)** George EP Box, Gwilym M Jenkins, Gregory C Reinsel e Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons. Citado na pág. 2
- Cain e Janssen(1995)** Michael Cain e Christian Janssen. Real estate price prediction under asymmetric loss. *Annals of the Institute of Statistical Mathematics*, 47(3):401–414. Citado na pág. 3
- Canada(2016a)** Statistics Canada. Housing stock dataset 1941-2000. <http://www5.statcan.gc.ca/cansim/a26?lang=eng&retrLang=eng&id=0300001&pattern=housing+stock&tabMode=dataTable&srchLan=-1&p1=1&p2=-1>, 2016a. Último acesso em 20/11/2016. Citado na pág. xi, 4, 25, 33, 40, 44, 46, 48, 49, 67, 68, 69, 70, 71
- Canada(2016b)** Statistics Canada. Gross domestic product dataset. <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610022201>, 2016b. Último acesso em 20/11/2016. Citado na pág. 55
- Canada(2016c)** Statistics Canada. Population dataset, 1946-2017. <http://www5.statcan.gc.ca/cansim/a26?lang=eng&id=510005>, 2016c. Último acesso em 20/11/2016. Citado na pág. xi, 33, 46, 47, 49, 61, 62, 63, 64, 65, 66
- Capinski e Zastawniak(2006)** Marek Capinski e Tomasz Zastawniak. *Mathematics for finance: an introduction to financial engineering*. Springer. Citado na pág. 15
- Case e Quigley(1991)** Bradford Case e John M Quigley. The dynamics of real estate prices. *The Review of Economics and Statistics*, páginas 50–58. Citado na pág. 3
- CMHC(2011)** CMHC. The evolution of social housing in canada, 2011. Citado na pág. 57
- Colwell(2002)** Peter F Colwell. Tweaking the dipasquale-wheaton model. *Journal of Housing Economics*, 11(1):24–39. Citado na pág. 16
- Ding et al.(2013)** Feng Ding, XG Liu e Jian Chu. Gradient-based and least-squares-based iterative algorithms for hammerstein systems using the hierarchical identification principle. *IET Control Theory Appl*, 7(2):176–184. Citado na pág. 25
- DiPasquale e Wheaton(1992)** Denise DiPasquale e William C Wheaton. The markets for real estate assets and space: A conceptual framework. *Real Estate Economics*, 20(2):181–198. Citado na pág. 1, 3, 5, 7, 11, 13, 14, 16, 17, 18, 20, 40, 45
- DiPasquale e Wheaton(1996)** Denise DiPasquale e William C Wheaton. *Urban economics and real estate markets*, volume 23. Prentice Hall Englewood Cliffs, NJ. Citado na pág. iii, v, 3, 7, 45, 59
- Domingos(2012)** Pedro Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87. Citado na pág. 26, 27
- Dreier e Hulchanski(1993)** Peter Dreier e J David Hulchanski. The role of nonprofit housing in canada and the united states: some comparisons. *Housing Policy Debate*, 4(1):43–80. Citado na pág. 57
- Drezet e Harrison(1998)** PML Drezet e RF Harrison. Support vector machines for system identification. Citado na pág. 26

- Eskinasi et al.(2011)** Martijn Eskinasi, EAJA Rouwette e JAM Vennix. Houdini: a system dynamics model for housing market reforms. Citado na pág. [1](#)
- Fisher(1912)** Ronald A Fisher. On an absolute criterion for fitting frequency curves. *Statistical Science*, 12(1):39–41. Citado na pág. [2](#)
- Friedman(1997)** Jerome H Friedman. On bias, variance, 0/1—loss, and the curse-of-dimensionality. *Data mining and knowledge discovery*, 1(1):55–77. Citado na pág. [34](#)
- Friedman e Friedman(1990)** Milton Friedman e Rose Friedman. *Free to choose: A personal statement*. Houghton Mifflin Harcourt. Citado na pág. [59](#)
- Fréchette(2016)** Jean-Denis Fréchette. Household formation and the housing stock: a stock-flow perspective. Citado na pág. [4](#), [57](#)
- Gevers(2006)** Michel Gevers. A personal view of the development of system identification: A 30-year journey through an exciting field. *IEEE Control Systems*, 26(6):93–105. Citado na pág. [24](#), [30](#), [54](#)
- Glaeser et al.(2005)** Edward L Glaeser, Joseph Gyourko e Raven E Saks. Urban growth and housing supply. *Journal of Economic Geography*, 6(1):71–89. Citado na pág. [42](#)
- Golub et al.(1979)** Gene H Golub, Michael Heath e Grace Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223. Citado na pág. [30](#)
- Grenadier(1995)** Steven R Grenadier. The persistence of real estate cycles. *The Journal of Real Estate Finance and Economics*, 10(2):95–119. Citado na pág. [1](#)
- Haig(1926)** Robert Murray Haig. Toward an understanding of the metropolis. *The Quarterly Journal of Economics*, páginas 402–434. Citado na pág. [11](#)
- Hendershott et al.(2002a)** Patric Hendershott, Bryan MacGregor e Michael White. Explaining real commercial rents using an error correction model with panel data. *The Journal of Real Estate Finance and Economics*, 24(1-2):59–87. Citado na pág. [3](#), [18](#)
- Hendershott et al.(1999)** Patric H Hendershott, Colin M Lizieri e George A Matysiak. The workings of the london office market. *Real Estate Economics*, 27(2):365. Citado na pág. [2](#), [3](#)
- Hendershott et al.(2002b)** Patric H Hendershott, Bryan D MacGregor e Raymond YC Tse. Estimation of the rental adjustment process. *Real Estate Economics*, 30(2):165–183. Citado na pág. [3](#), [18](#)
- Hendershott et al.(2010)** Patric H Hendershott, Colin M Lizieri e Bryan D MacGregor. Asymmetric adjustment in the city of london office market. *The Journal of Real Estate Finance and Economics*, 41(1):80–101. Citado na pág. [18](#)
- Jacco e Robert(1996)** Hakfoort Jacco e Lie Robert. Office space per worker: evidence from four european markets. *Journal of Real Estate Research*, 11(2):183–196. Citado na pág. [42](#)
- Jiang et al.(2017)** Xiaomo Jiang, Sankaran Mahadevan e Yong Yuan. Fuzzy stochastic neural network model for structural system identification. *Mechanical Systems and Signal Processing*, 82:394–411. Citado na pág. [26](#)
- Kaiser(1997)** Ronald Kaiser. The long cycle in real estate. *Journal of Real Estate Research*, 14(3):233–257. Citado na pág. [1](#)
- Kaldor(1934)** Nicholas Kaldor. A classificatory note on the determinateness of equilibrium. *The review of economic studies*, 1(2):122–136. Citado na pág. [1](#)

- Keogh e Mueen(2011)** Eamonn Keogh e Abdullah Mueen. Curse of dimensionality. Em *Encyclopedia of Machine Learning*, páginas 257–258. Springer. Citado na pág. [34](#)
- King e McCue(1987)** John L King e Thomas E McCue. Office building investment and the macroeconomy: empirical evidence 1973-85. *Journal of the American Real Estate and Urban Economics Association*, 15(3):234–55. Citado na pág. [1](#)
- Krugman e Wells(2009)** Paul Krugman e Robin Wells. *Macroeconomics*. Worth Publishers. Citado na pág. [9](#), [10](#)
- Krugman et al.(2013)** Paul Krugman, Robin Wells e Martha Olney. *Essentials of economics*. Macmillan. Citado na pág. [7](#), [8](#), [11](#)
- Limsombunchao(2004)** V Limsombunchao. House price prediction: Hedonic price model vs. artificial neural network. Citado na pág. [3](#)
- Liu et al.(2006)** Jian-Guo Liu, Xiao-Li Zhang e Wei-Ping Wu. Application of fuzzy neural network for real estate prediction. Em *International Symposium on Neural Networks*, páginas 1187–1191. Springer. Citado na pág. [3](#)
- Ljung(1978)** Lennart Ljung. Convergence analysis of parametric identification methods. *IEEE transactions on automatic control*, 23(5):770–783. Citado na pág. [25](#)
- Ljung(2002)** Lennart Ljung. Prediction error estimation methods. *Circuits, Systems and Signal Processing*, 21(1):11–21. Citado na pág. [2](#), [25](#), [27](#), [54](#)
- Ljung(2006)** Lennart Ljung. Some aspects on nonlinear system identification. *IFAC Proceedings Volumes*, 39(1):110–121. Citado na pág. [2](#), [5](#), [22](#), [24](#), [25](#), [27](#), [28](#), [31](#), [32](#), [47](#), [51](#), [59](#)
- Ljung(2010)** Lennart Ljung. Perspectives on system identification. *Annual Reviews in Control*, 34(1):1–12. Citado na pág. [2](#), [22](#), [23](#), [25](#), [28](#), [30](#)
- Mankiw(2014)** N Gregory Mankiw. *Essentials of economics*. Cengage learning. Citado na pág. [7](#), [8](#)
- Mankiw e Weil(1989)** N Gregory Mankiw e David N Weil. The baby boom, the baby bust, and the housing market. *Regional science and urban economics*, 19(2):235–258. Citado na pág. [42](#)
- Martínez-Ramón et al.(2006)** Manel Martínez-Ramón, Jos Luis Rojo-Alvarez, Gustavo Camps-Valls, Jordi Muñoz-Marí, Emilio Soria-Olivas, Anbal R Figueiras-Vidal et al. Support vector machines for nonlinear kernel arma system identification. *IEEE Transactions on Neural Networks*, 17(6):1617–1622. Citado na pág. [26](#)
- Mashayekhi et al.(2009)** Ali N Mashayekhi, Soheil Ghili e Arash Pourhabib. Real estate cycles: A theory based on stock-flow structure of durable goods markets. Em *Proceedings of the 2009 System Dynamics Conference*. Citado na pág. [1](#)
- Mathworks(2016a)** Mathworks. Matlab, 2016a. Natick, MA, USA. Citado na pág. [33](#)
- Mathworks(2016b)** Mathworks. Matlab optimization toolbox, 2016b. Natick, MA, USA. Citado na pág. [36](#), [44](#)
- Miller(2012)** Norm Miller. Estimating office space per worker. *Burnham-Moores Center for Real Estate University of San Diego*. Citado na pág. [42](#)
- Murphy(2012)** Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press. Citado na pág. [28](#)
- Myers et al.(2002)** Dowell Myers, John Pitkin e Julie Park. Estimation of housing needs amid population growth and change. *Housing Policy Debate*, 13(3):567–596. Citado na pág. [42](#)

- Nagumo e Noda(1967)** Jin-Ichi Nagumo e Atsuhiko Noda. A learning method for system identification. *IEEE Transactions on Automatic Control*, 12(3):282–287. Citado na pág. 25
- Oliva(2003)** Rogelio Oliva. Model calibration as a testing strategy for system dynamics models. *European Journal of Operational Research*, 151(3):552–568. Citado na pág. 1, 2
- Pillonetto et al.(2014)** Gianluigi Pillonetto, Francesco Dinuzzo, Tianshi Chen, Giuseppe De Nicolao e Lennart Ljung. Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, 50(3):657–682. Citado na pág. 25
- Puu(1991)** Tõnu Puu. Nonlinear economic dynamics. Em *Nonlinear Economic Dynamics*, páginas 1–7. Springer. Citado na pág. 1
- Pyhrr et al.(2009)** Stephen Pyhrr, Stephen Roulac e Waldo Born. Real estate cycles and their strategic implications for investors and portfolio managers in the global economy. *Journal of Real Estate Research*. Citado na pág. 3
- Qin(2006)** S Joe Qin. An overview of subspace identification. *Computers & chemical engineering*, 30(10-12):1502–1513. Citado na pág. 24
- Radzicki(2009)** Michael J Radzicki. System dynamics and its contribution to economics and economic modeling. Em *Complex Systems in Finance and Econometrics*, páginas 727–737. Springer. Citado na pág. 12
- Rafiei e Adeli(2015)** Mohammad Hossein Rafiei e Hojjat Adeli. A novel machine learning model for estimation of sale prices of real estate units. *Journal of Construction Engineering and Management*, 142(2):04015066. Citado na pág. 3
- Reynaud(2015)** Julien Reynaud. *Gauging Housing Supply in Canada: A Stock Approach*. Number 15-128. International Monetary Fund. Citado na pág. 4, 57
- Richardson(1991)** George P Richardson. System dynamics: Simulation for policy analysis from a feedback perspective. Em *Qualitative simulation modeling and analysis*, páginas 144–169. Springer. Citado na pág. 12
- Richardson(2009)** George P Richardson. System dynamics, the basic elements of. Em *Complex Systems in Finance and Econometrics*, páginas 856–862. Springer. Citado na pág. 12
- Robert e Bernanke(2001)** Frank Robert e Ben Bernanke. Principles of economics, 2001. Citado na pág. 10
- Rosen e Smith(1983)** Kenneth T Rosen e Lawrence B Smith. The price-adjustment process for rental housing and the natural vacancy rate. *The American Economic Review*, 73(4):779–786. Citado na pág. 1
- Rottke et al.(2003)** Nico Rottke, Martin Wernecke e Arthur Schwartz Jr. Real estate cycles in germany—causes, empirical analysis and recommendations for the management decision process. *Journal of Real Estate Literature*, 11(3):325–346. Citado na pág. 1
- Schumpeter(1954)** Joseph A Schumpeter. *History of economic analysis*. Psychology Press. Citado na pág. 1
- Smith e van Ackere(2002)** Peter C Smith e Ann van Ackere. A note on the integration of system dynamics and economic models. *Journal of Economic Dynamics and Control*, 26(1):1–10. Citado na pág. 18
- Sousa e Quarter(2003)** Jorge Sousa e Jack Quarter. The convergence of nonequity housing models in canada: Changes to housing policy since 1990. *Housing Policy Debate*, 14(4):591–620. Citado na pág. 57

- Stoica et al.(1986)** Petre Stoica, Pieter Eykhoff, Peter Janssen e Torsten Söderström. Model-structure selection by cross-validation. *International Journal of Control*, 43(6):1841–1878. Citado na pág. 31
- Suykens et al.(2002)** Johan AK Suykens, Tony Van Gestel e Jos De Brabanter. *Least squares support vector machines*. World Scientific. Citado na pág. 30
- Thompson(1937)** Warren S Thompson. Population growth and housing demand. *The ANNALS of the American Academy of Political and Social Science*, 190(1):131–137. Citado na pág. 42
- Van Dyk(1995)** Nick Van Dyk. Financing social housing in canada. *Housing Policy Debate*, 6(4): 815–848. Citado na pág. 57
- Voith e Crone(1988)** Richard Voith e Theodore Crone. National vacancy rates and the persistence of shocks in us office markets. *Real Estate Economics*, 16(4):437–458. Citado na pág. 1
- Wheaton(1987)** William C Wheaton. The cyclic behavior of the national office market. *Areuea Journal*, 15(4):281–99. Citado na pág. 1, 3, 18, 55
- Wheaton(1999)** William C Wheaton. Real estate “cycles”: some fundamentals. *Real estate economics*, 27(2):209–230. Citado na pág. iii, v, ix, 1, 2, 3, 4, 7, 12, 13, 18, 19, 20, 21, 22, 25, 33, 34, 35, 37, 38, 39, 40, 42, 43, 45, 46, 55, 58, 59, 60
- Wheaton et al.(1997)** William C Wheaton, Raymond G Torto e Peter Evans. The cyclic behavior of the greater london office market. *The Journal of Real Estate Finance and Economics*, 15(1): 77–92. Citado na pág. 1, 3, 5, 18
- Worldatlas(2018)** Worldatlas. Canada political map, 2018. URL <https://www.worldatlas.com/r/w728-h425-c728x425/upload/86/37/4a/694px-political-map-of-canada.png>. Citado na pág. 45
- Worzala et al.(2009)** Elaine Worzala, Margarita Lenk e Ana Silva. An exploration of neural networks and its application to real estate valuation. *Journal of Real Estate Research*. Citado na pág. 3
- Zhang et al.(2015)** Xin Zhang, David Geltner e Richard de Neufville. System dynamics modeling of chinese urban housing markets for pedagogical and policy analysis purposes. *Available at SSRN 2600777*. Citado na pág. 1, 2, 3, 11, 12