

Hudson Chaves Costa

Três Ensaaios em Comportamento dos
Preços na Economia Brasileira

Porto Alegre
2014

Hudson Chaves Costa

Três Ensaaios em Comportamento dos Preços na Economia Brasileira

Projeto de Tese apresentado ao Programa
de Pós-Graduação em Economia da Uni-
versidade Federal do Rio Grande do Sul na
Área de Economia Aplicada.

Orientador: Prof. Dr. Sabino Porto da Silva
Júnior

Porto Alegre
2014

Sumário

1	INTRODUÇÃO/MOTIVAÇÃO	1
2	REVISÃO BIBLIOGRÁFICA	5
3	METODOLOGIA	11
	Referências Bibliográficas	18

Capítulo 1

INTRODUÇÃO/MOTIVAÇÃO

Firmas individuais não ajustam seus preços em contrapartida de choques relevantes na economia. Este fato não é controvérsia e é uma hipótese padrão em modelagem macroeconômica que permite choques nominais influenciar as variáveis reais. Uma grande vertente da literatura analisou as implicações de alternativas formas de rigidez nominal sobre a dinâmica do comportamento da inflação e produto em níveis agregados. Não obstante, em resposta à relevância do tema, modelos macroeconômicos de preços rígidos têm sido desenvolvidos e fundamentam-se, em grande medida, no comportamento microeconômico de determinação de preços adotado pelos agentes.

É possível dividir os diferentes mecanismos propostos para incorporar a rigidez nominal dos preços em dois grupos: preços tempo-dependentes e preços estado-dependentes. Em um modelo de precificação tempo-dependente, a possibilidade de um preço mudar pode ser afetada apenas pelo tempo desde a mudança anterior e não pelo estado das vendas de uma firma, a economia ou outros fatores. Já em modelos estado-dependentes, a decisão de mudar os preços depende do estado da economia e o mercado enfrentado pela firma. Firmas enfrentam custos caso ajustem seus preços e exemplos destes custos incluem custos fixos (custo de menu, [Mankiw \(1985\)](#)) ou a desutilidade associada à fazer grandes alterações de preços se as firmas temem que tais mudanças podem contrariar seus clientes ([Rotemberg \(1982\)](#)).

Na mesma direção da preocupação teórica em relação ao comportamento dos preços, muitos Bancos Centrais têm adotado a política de metas de inflação como um fator relevante para a política monetária. Tipicamente, a meta é definida em termos de um índice de preço agregado, tal como o Índice Nacional de Preços ao Consumidor Amplo (IPCA) no Brasil. Dado que este índice agregado é uma soma ponderada dos preços individuais, mudanças nestes preços terão importantes implicações para o nível de preço geral e preços relativos. A partir disso, foi natural o surgimento de pesquisas com o objetivo de diferir a análise empírica da rigidez nominal dos preços baseada em dados agregados da avaliação do comportamento dos preços por meio de microfunda-mentos.

Por conseguinte, recentes estudos internacionais e nacionais têm usado de grandes bancos de dados de preços individuais (microdados) para examinar como os preços ao consumidor comportam em nível de produto. Em particular, [Bils and Klenow \(2004\)](#),

Nakamura and Steinsson (2008) e Klenow and Kryvtsov (2008) para os EUA, Dhyne et al. (2006) para a Zona do Euro, Gouvea (2007), Matos and Barros (2009) e LOPES (2008) para o Brasil e Bunn and Ellis (2012) para o Reino Unido. Apesar de pesquisas brasileiras terem utilizado de microdados para a avaliação empírica da rigidez dos preços, nenhuma delas conseguiu alta granularidade no que tange à periodicidade dos preços coletados bem como em relação às regiões em que a rigidez foi avaliada. Tal característica é oriunda das fontes de dados usadas (IBRE/FGV e IPC/FIPE).

A partir disso, este ensaio do projeto de tese, busca aplicar a tecnologia de *web scraping* na avaliação de como os preços individuais ao consumidor tipicamente se comportam e se existe rigidez de preços nos principais mercados brasileiros¹. Isto permitirá uma periodicidade diária na coleta de preços e maior quantidade de preços coletados em comparação com os tradicionais índices de preços.

Motivado por tais apontamentos, o presente ensaio explorará preços coletados de sites (supermercados, farmácias, lojas de eletrodomésticos, construção civil, ...) como uma alternativa à dificuldade em obter microdados de preços. A tecnologia de *web scraping* que será utilizada tem se tornado uma alternativa para o acesso a diversas fontes de dados para análise econômica empírica. Pretende-se contribuir para o estudo empírico da rigidez nominal nos preços por meio da estimação de funções de risco de alterações nos preços. Esta função é definida como a probabilidade condicional dos preços se alterarem em relação ao tempo, um conceito chave em entender o comportamento dos preços. Ela está relacionada aos modelos microfundamentados. Por exemplo, os modelos de Calvo (1983) e Taylor (1980) têm função de risco constante. A probabilidade de alterações nos preços depende apenas do tempo, independentemente de mudanças em variáveis de estado. Por outro lado, o modelo de Dotsey et al. (1999) tem uma função de risco crescente que muda dependendo das variáveis de estado. Ele é um dos mais populares modelos estado-dependentes em um framework de equilíbrio geral.

Este ensaio é organizado da seguinte forma: além deste primeiro capítulo que faz uma breve apresentação do tema, justificativa e objetivos do ensaio, o capítulo 2 mostra o referencial teórico sobre modelos de rigidez de preços e por fim no capítulo 3 a metodologia a ser utilizada bem como o cronograma da pesquisa são descritos.

JUSTIFICATIVA

A dinâmica do comportamento dos preços individuais proporciona vários desdobramentos que são bastante debatidos na literatura dado os impactos que podem causar. Não compreender este tipo de comportamento levou a distintas abordagens para a análise da velocidade e intensidade de transmissão da política monetária. Além disso, compreender as estratégias de definição de preços das firmas levaria ao aprimoramento de modelos teóricos cujas abordagens e conclusões podem sofrer alterações expressivas na presença de fatos estilizados.

A falta de estudos que gerassem empiricamente um diagnóstico da definição e grau

¹A princípio serão coletados preços dos principais sites das regiões metropolitanas e cidades que participam do IPCA e INPC

de rigidez de preços individuais foi um limitador por diversas décadas em função da falta de informações estatísticas no nível de microdados que pudessem servir de base para estas análises. Porém, há alguns anos a disponibilização de preços coletados pelos órgãos governamentais tanto nacionais quanto internacionais, proporcionaram o surgimento de pesquisas que avaliassem o comportamento dos preços em nível de microdados ([Bils and Klenow \(2004\)](#); [Nakamura and Steinsson \(2008\)](#); [Klenow and Kryvtsov \(2008\)](#); [Dhyne et al. \(2006\)](#); [Gouvea \(2007\)](#); [Matos and Barros \(2009\)](#); [LOPES \(2008\)](#); [Bunn and Ellis \(2012\)](#)).

Porém, ainda existe um fator limitante nestes estudos, pois concentram-se em mercados específicos, não possibilitando análises generalizadas aos diversos setores da economia, pois as pesquisas são reféns das características dos dados utilizados. Também, dada a importância do tema para os tomadores de decisão em nível de política monetária, é preciso maior dinâmica na análise e não apenas um olhar para o passado.

Assim, o presente ensaio do projeto de tese apresenta o uso da tecnologia de *web scraping* para coletar preços diretamente das páginas das empresas que possuem sites de vendas e por conseguinte, contribuir para a avaliação da rigidez de preços de uma forma mais dinâmica dadas as características do processo de coleta. Estudos empíricos já mostraram a importância de dados coletados da *web* na avaliação dos pressupostos de rigidez de preços e proposição de medida de inflação oriunda de informações *online* ([Cavallo \(2010\)](#)).

OBJETIVOS

O objetivo geral deste projeto é avaliar empiricamente a rigidez nominal dos preços na economia brasileira por meio de dados coletados da *web*, bem como, propor um índice de inflação oriundo da mesma fonte de dados que seja estatisticamente significativa para o uso dos tomadores de decisões econômicas.

Dentro deste escopo, os objetivos específicos são: melhor compreensão do comportamento microeconômico dos preços e suas implicações para a economia.

Capítulo 2

REVISÃO BIBLIOGRÁFICA

Preços Rígidos e Preços Flexíveis

[Ball and Mankiw \(1994\)](#) argumentam que a política monetária afeta a atividade econômica real e que a principal razão para essa afirmação são as evidências históricas, especialmente os inúmeros episódios em que as contrações monetárias causaram recessões. Uma vez assumida a hipótese de não-neutralidade monetária, os macroeconomistas dividem-se quanto à melhor maneira de explicar flutuações econômicas de curto prazo. [Ball and Mankiw \(1994\)](#) acreditam que a rigidez nominal dos preços fornece a explicação mais natural para a não-neutralidade monetária, dadas as evidências microeconômicas de que muitos preços são, de fato, rígidos. Outros economistas, entretanto, desenvolveram modelos com preços flexíveis e substituíram a rigidez nominal dos preços por alguma outra imperfeição nominal para gerar o resultado de não-neutralidade monetária.

A alternativa mais famosa aos modelos de preços rígidos é o modelo de [Lucas \(1972\)](#). Nesse modelo os preços são flexíveis e a imperfeição nominal é informacional, sendo possível gerar o resultado de não-neutralidade monetária. O modelo de [Lucas \(1972\)](#) de informação imperfeita baseia-se na ideia de que quando um produtor observa uma mudança no preço de seu produto ele não sabe distinguir se isso é resultado de uma mudança no preço relativo ou se é resultado de uma mudança no nível agregado de preços. Uma mudança nos preços relativos altera a quantidade ótima a produzir, enquanto que uma mudança no nível de preço agregado deixa a quantidade ótima de produção inalterada. Considerando-se uma expansão monetária não-observada, o melhor que cada produtor pode fazer é admitir que uma parte do aumento da demanda por seu produto reflete um choque de preços relativos. Então, produtores elevam seus preços e a expansão monetária tem efeitos reais e não apenas efeitos nominais sobre os preços. A fragilidade do modelo de [Lucas \(1972\)](#) está no fato de que é difícil compreender como nas economias modernas produtores poderiam confundir movimentos nos preços relativos com movimentos no nível de preços agregado, dado o grande volume de informações. [Ball and Mankiw \(1994\)](#) conclui que o modelo de Lucas não é um substituto convincente para os modelos de preços rígidos.

A vertente novo-keynesiana estabelece a hipótese de existência de rigidez nominal

tanto nos preços quanto nos salários. Essas variáveis nominais teriam dificuldades em se ajustar dada a ocorrência de mudanças na política monetária o que, por conseguinte, provocaria impactos reais sobre o produto. Assim, a expansão monetária pode provocar diferentes impactos sobre cada preço da economia dependendo do grau de rigidez nominal de cada bem. Tal rigidez se for diversificada, resultará em alterações nos preços relativos provocando impactos reais.

Neste contexto, muitos modelos macroeconômicos de preços rígidos baseados em fundamentos microeconômicos têm sido desenvolvidos. As hipóteses desses modelos envolvem, em grande medida, características acerca do comportamento microeconômico de determinação de preços adotados pelos agentes. A busca da construção de modelos macroeconômicos cada vez mais apurados fez emergir a necessidade de verificar empiricamente alguns dos aspectos microeconômicos constantes nesses modelos.

Tempo-Dependentes e Estado-Dependentes

Em geral, modelos de precificação microfundamentados são classificados em dois tipos: modelos estado-dependentes e modelos tempo-dependentes. Nos modelos tempo-dependentes, a probabilidade dos preços mudarem depende apenas do período pelo qual o preço está fixo. Então, a função risco deste modelo tem uma certa forma constante em relação à duração dos preços. Por exemplo, o modelo [Calvo \(1983\)](#) tem uma função de risco plana e assume que a oportunidade de um preço variar segue um processo Poisson. A hipótese significa que um formador de preço tem uma oportunidade de alterar os preços com uma probabilidade constante em cada período. É bem conhecido que a curva de Phillips novo-keynesiana é derivada do modelo de Calvo com competição monopolística. Também, o modelo de [Taylor \(1980\)](#) tem uma função risco constante que assume 100% em certos períodos e 0% por outro lado. Ele assume que o formador de preços muda seus preços apenas no começo do contrato e não muda dentro do período de durabilidade do contrato. Então, sua taxa de risco toma o valor da unidade no começo do contrato e 0 por outro lado.

Em adição aos modelos de Calvo e Taylor, [Mash \(2004\)](#) e [Coenen et al. \(2007\)](#) generalizaram o modelo de Calvo. Eles atribuem diferentes probabilidades em distintos períodos, permitindo a função risco ter qualquer forma funcional incluindo riscos crescentes e decrescentes. Eles mostram que as curvas de Phillips derivadas de seus modelos dependem não apenas do corrente *gap* no produto e a inflação esperada para o próximo período como a curva de Phillips novo-keynesiana, mas também sobre as taxas de inflação esperadas em algum período passado e futuro.

Modelos estado-dependentes, começam com os modelos de custo de menu de [Barro \(1972\)](#) e [Sheshinski and Weiss \(1977\)](#) e mais recentemente [Dotsey et al. \(1999\)](#) e [Golosov and Lucas Jr \(2007\)](#), tendem a ter maior fundamentação microeconômica. Nestes modelos, a probabilidade condicional do preço alterar depende das variáveis de estado, preços relativos e taxas de inflação. Então, a função risco no modelo estado-dependente pode mudar sua forma em resposta à choques reais ou monetários em um transição, enquanto ela tem uma forma constante em *steady state*. Por exemplo, [Dotsey et al.](#)

(1999) desenvolveram um modelo de precificação estado-dependente ampliando o modelo de custo de menu de [Blanchard and Kiyotaki \(1987\)](#). Eles assumiram que o custo de menu segue um processo aleatório e varia entre formadores de preços. Neste caso, eles mostraram que uma função risco depende das taxas de inflação e a distribuição do processo aleatório. Em adição, a forma de uma função risco é crescente em um *steady state*. Em um *steady state*, quanto mais tempo o preço permanece fixo, mais o preço relativo desvia do preço relativo ótimo devido a choques de produtividade acumulados. Então, a probabilidade condicional de mudanças nos preços sobe equanto o preço permanece fixo.

Bakhshi et al. (2004) mostraram que a curva de Phillips derivada do modelo de [Dotsey et al. \(1999\)](#) tem uma forma complexa. Ela depende não apenas dos correntes *gaps* do produto e a taxa de inflação esperada no próximo período, mas também sobre as taxas de inflação esperada em alguns períodos passados e futuros. Assim, independentemente de modelos tempo-dependentes ou estado-dependentes, a curva de Phillips tem a forma complexa se uma função risco não é plana como no modelo de Calvo.

Trabalhos Utilizando Microdados para avaliação da rigidez de preços

Conforma salientado, existem trabalhos nacionais e internacionais que utilizaram microdados de preços ao consumidor para analisar a rigidez nominal nos preços. Um ponto forte deste tipo de trabalho é a viabilidade no acompanhamento do preço de um determinado produto vendido em uma loja específica ao longo do tempo. Assim, é possível comparar o grau de rigidez em vários níveis (setores, cidades, economia como um todo, produtos, ...).

[Bils and Klenow \(2004\)](#) examinaram a frequência das alterações nos preços mensais de 350 produtos e serviços que representavam em torno de 70% da cesta de consumo analisada pelo índice de preços do *Bureau of Labor Statistics* (BLS) no período de 1995 a 1997 nos EUA. Encontraram que metade dos preços não sobrevivem por menos do que 4 meses. [Taylor \(1980\)](#) concluiu que os preços alteravam tipicamente em torno de uma vez por ano. Além disso, examinaram se as séries temporais de inflação são consistentes com os modelos de rigidez de preços de Calvo e Taylor dada a frequência de alteração que os autores encontraram. Encontraram que, para a maioria dos produtos, esses modelos predizem taxas de inflação que são muito mais persistentes e muito menos voláteis do que as observadas pelos autores. Os modelos *over-predict* a persistência e *under-predict* a volatilidade para bens cujos preços se alteram em menor frequência.

[Nakamura and Steinsson \(2008\)](#) avaliaram os preços mensais de 270 produtos que representavam 70% da cesta de consumo avaliada no período de 1998 a 2005 para os EUA. Apresentaram 5 fatos sobre os preços nos EUA: a frequência mediana de alteração nos preços de produtos que não estavam em promoção é aproximadamente a metade dos preços em relação aos promocionais (de 9% a 12% por mês contra 19% a 20% para itens idênticos), um terço das alterações nos preços são em relação a quedas, a frequência do

aumento nos preços está fortemente relacionada com a inflação enquanto que a frequência de queda nos preços e o tamanho do aumento e queda nos preços não está relacionada com a inflação e por fim, a frequência das alterações nos preços é altamente sazonal sendo maior no primeiro trimestre e declina a partir dele. Os autores encontraram evidências da função de risco ter inclinada ascendente para produtos individuais. Além disso, os fatos três primeiros fatos estão consistentes com o modelo de custo de menu enquanto os outros dois não.

Klenow and Kryvtsov (2008) usaram microdados coletados pelo *Bureau of Labor Statistics* (BLS) de 300 produtos que representavam 85% da cesta de consumo analisada para decompor a variação mensal da inflação no período de 1988 a 2004. Para resumir a medida em que as variações dos preços são sincronizados nos dados, os autores aproveitaram uma identidade da inflação. A inflação agregada em um dado mês é igual ao produto de dois termos: a fração de itens que mudaram seus preços (margem extensiva) e o tamanho médio das alterações nos preços (margem intensiva). Usando esta identidade, a variância da inflação ao longo do tempo pode ser decomposta na contribuição de cada margem. Assim, os autores implementaram esta decomposição nos dados e encontraram que aproximadamente 95% da variância da inflação mensal é dada pela margem intensiva, isto é, o termo que representa os modelos tempo-dependentes. Nestes modelos a fração de itens mudando seus preços é constante ao longo do tempo e desta forma. Os termos remanescentes estão relacionados com os modelos estado-dependentes dado que envolvem variação ao longo do tempo na fração de itens mudando seus preços e na amostra dos autores, esses termos representam apenas 5% da variação na inflação.

Dhyne et al. (2006) argumentam que os preços dos bens e serviços não se ajustam imediatamente em contrapartida de alterações de demanda ou oferta. O artigo caracteriza a frequência média e tamanho das alterações nos preços na zona do euro e seus países membros, investigando os determinantes da probabilidade dos preços alterarem. Os autores comparam as evidências para a zona do euro com resultados dos EUA e mostraram que os preços se alteram menos frequentemente na Zona do Euro do que nos EUA. Os dados utilizados são os registros mensais dos preços subjacentes ao cálculo nacional do *Consumer Price Indices* e Harmonized Consumer Price Indices. As evidências mostraram que existiu uma heterogeneidade no mercado na frequência com que os preços se ajustam (comidas, bens industriais e especialmente em serviços). Como Aoki (2001), de uma perspectiva de política monetária ótima, uma potencial implicação é que os Bancos Centrais deveriam ter maior atenção com a inflação em setores rígidos.

Gouvea (2007) investigou a padrão do ajustamento dos preços no Brasil. A autora derivou os principais fatos estilizados descrevendo o comportamento dos definidores de preços diretamente de um grande banco de dados do IBRE da Fundação Getulio Vargas e proporciona evidências microeconômicas do grau e características da rigidez de preços. Os dados representam 85% da cesta de consumo analisada com um total de 243 produtos no período de 1996 a 2006 totalizando aproximadamente 9 milhões de observações. Conclui que a média ponderada da frequência mensal de alterações nos preços é de 37% para todos os produtos e que a duração média dos preços é de

aproximadamente 2,7 meses enquanto a mediana é de 1,9 meses. Comparado com os EUA e Zona do Euro, esses números revelaram que os preços são mais flexíveis no Brasil. Além disso, documentou uma clara evidência de heterogeneidade no comportamento de definição dos preços em produtos e setores.

LOPES (2008) investigou o comportamento de determinação dos preços na cidade de São Paulo. Analisaram mais de 6 milhões de cotações do índice de preços ao consumidor da FIPE. Os principais resultados encontrados pelo autor são: a frequência média de mudança dos preços é de 32,35% ao mês; os preços duram em média 2,56 meses; há grande heterogeneidade entre produtos quanto ao comportamento de mudança dos preços; 40% das mudanças de preços são para baixo; as mudanças de preço possuem magnitude considerável; a frequência de mudança dos preços exhibe padrões sazonais em alguns grupos; a frequência de mudança dos preços respondeu às incertezas eleitorais de 2002 em alguns grupos; as funções de risco comum são decrescentes em apresentam picos na duração correspondente a doze meses para alguns subgrupos e o risco de mudança dos preços responde ao índice inflacionário para a aproximadamente 70% dos subgrupos.

É possível observar que todos estes estudos aplicaram metodologias semelhantes para a avaliação da rigidez de preços em seus mercados. Para tanto, utilizaram os principais índices de inflação como fonte para a cesta de produtos a serem avaliadas e assim, conseguiram usar grandes percentuais destes bens na avaliação da frequência das alterações nos preços. Porém, a quantidade de produtos presentes nestas cestas é pequena em relação aos comercializados diariamente apesar de serem estatisticamente representativos para a mensuração da inflação de acordo com as metodologia de cálculo de inflação. Ainda, para determinados setores ou produtos tal evidência pode ser ainda maior.

Utilizar a capacidade da tecnologia de *web scraping* para coletar preços de sites pode ser uma alternativa para acesso a mais produtos em periodicidade maior e assim, mensurar a rigidez de preços empiricamente com maior capilaridade em nível de produto, setores e centros específicos do país. Também, a partir do contínuo processo de coleta

Capítulo 3

METODOLOGIA

WEB SCRAPING

O surgimento da internet, em particular a Web (*Web Wide Word*) trouxe um crescimento exponencial nas disposições de informações. Embora muitas dessas informações sejam úteis, elas raramente estão de uma forma que podemos utilizá-las, pois ainda é comum pessoas gastarem horas na coleta manual de dados de páginas da web. Especificamente, apesar de disponibilidade, poucos trabalhos acadêmicos em Economia utilizam desta fonte de dados para análise econômica empírica. Tal característica pode ser dada pela dificuldade dos pesquisadores na área em lidar com linguagem de programação que demandam maior conhecimento de computação.

Em recente publicação, [Varian \(2014\)](#) salienta que as técnicas utilizadas na Ciência da Computação e outras áreas correlatas para manipular e analisar dados, têm muito a oferecer. [Varian \(2014\)](#) defende que economistas deveriam conhecer melhor esses métodos e usá-los em seus trabalhos. Além disso, [Varian \(2014\)](#) cita a comumente colaboração entre os departamentos de Ciência da Computação e Estatística nas universidades dos EUA. Porém, o autor espera que em um futuro próximo os estudantes de econometria tenham maior colaboração com esses perfis e assim, contribuir para a pesquisa econômica empírica.

Uma metodologia que facilita o processo de coleta de dados da web é conhecida como *web scraping* que envolve escrever algoritmos que executam automaticamente o que nós fazemos manualmente quando navegamos por uma página de um site de e-commerce, por exemplo. Além disso, necessita-se de pouco conhecimento em programação para iniciar o processo de coleta de dados da internet. Segundo [Manning et al. \(2008\)](#) *web scraping* é o processo de tirar informações desestruturadas de páginas da web e transformá-las em informações estruturadas que podem ser usadas para análise.

A maior parte das páginas de sites são construídas usando uma linguagem de codificação estruturada chamada de *HyperText Markup Language* (HTML). Este código tem “tags”, tais como `< center >` e `< bold >`, que determinam o estilo e localização do texto em uma página. Estas tags tendem a permanecer constantes ao longo do tempo, uma vez que proporcionam um “look and feel” distinto para cada página. Por contraste,

a informação dentro dessas *tags*, tais como preço de produtos, mudam ao longo do tempo. O software de *scraping* pode ser ensinado a utilizar as *tags* em HTML para localizar informações relevantes sobre um produto e guardá-las em um banco de dados. A repetição desse processo todos os dias produz um banco de dados em formato de painel com um registro por produto por dia. Em adição, o endereço da página (URL) onde cada produto é localizado pode ser usado para classificar produtos em categorias padronizadas.

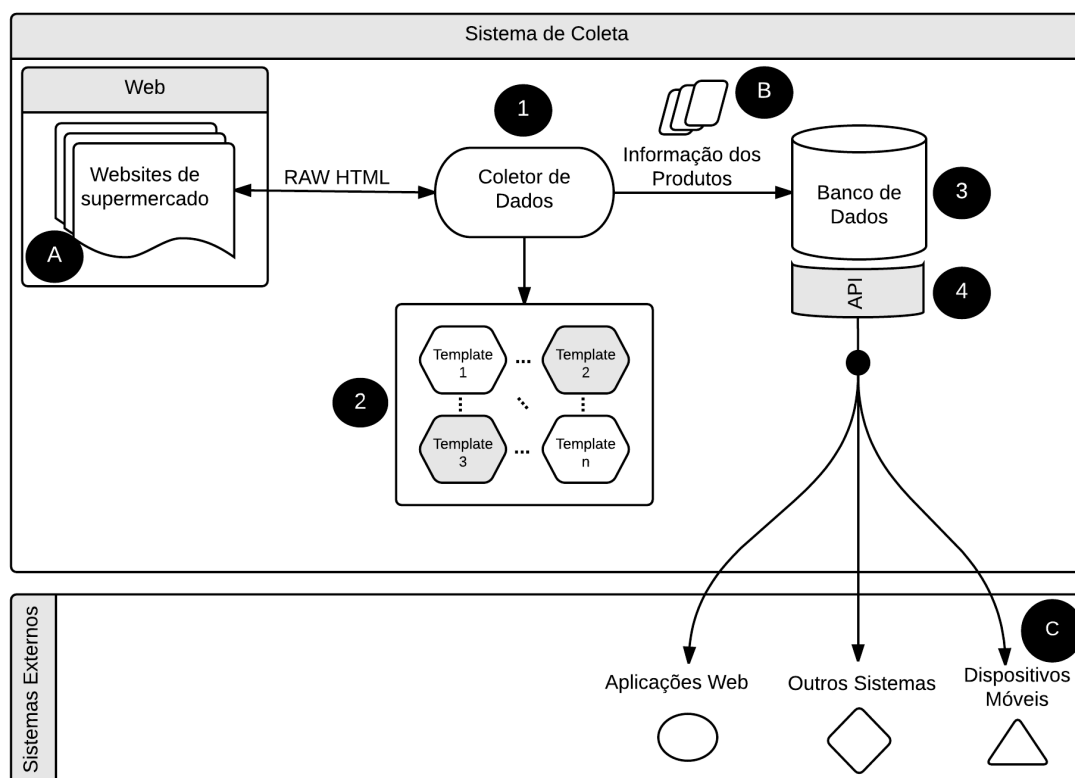


Figura 3.1: Arquitetura do Sistema de Coleta e Disponibilização dos Dados

Através de um coletor é possível arquitetar e executar de forma lógica e escalável todo esse processo. Para que um coletor seja funcional é necessário que o mesmo seja capaz de interagir com páginas da Web, extrair a informação de interesse e estruturar e armazenar os dados para futuras consultas. Em geral, exemplos corriqueiros de coletores podem ser citados como os desenvolvidos pelo Google e Microsoft para atuar na procura por páginas da internet ou outros mais específicos para coleta de preços de produtos como os portais de agregadores como Bom de Faro, Buscapé dentre outros. Assim, como [Cavallo \(2010\)](#), o presente projeto de tese busca de forma inovadora para a economia brasileira, explorar preços coletados de sites de supermercados, farmácias, companhia de energia elétrica, lojas de varejo online, lojas de roupas e calçados, entre outros, e propõe o uso de um sistema de coleta como o apresentado a seguir.

O sistema de coleta de preços está apresentado na Figura 3.1. O coletor recebe como entrada os templates dos sites que se deseja coletar e produz como resposta informações estruturadas com os atributos dos produtos. Para que o coletor seja capaz de realizar a tarefa de extração de informação, o mesmo deve apresentar os seguintes componentes: um componente centralizado capaz de ler instruções e aplicar regras para extração (1 - módulo coletor de dados); ter um conjunto de regras que descreva de forma não ambígua como realizar a coleta dos dados e os atributos de interesse (2 - templates dos webistes de supermercado, por exemplo); ter um banco de dados capaz de lidar com as características dos dados armazenados (3 - módulo de banco de dados); ter uma interface para facilitar o acesso aos dados por meio de outras aplicações ou sistemas web (4 - módulo de disponibilização da informação). Dessa forma, o coletor (1) é o centralizador do processo de coleta de dados, fazendo a interação com os templates (2), os webistes (A) e o módulo de banco de dados (3). Em resumo, o coletor através de um algoritmo inicia o processo de coleta carregando em uma lista os templates de coleta dos websites (2) e em seguida através de um processo iterativo visita o website, coleta os documentos de interesse e realiza a extração das informações indicadas pelo template. Ao final da coleta os dados são estruturados em um formato de documento denominado JSON e armazena-os em um banco de dados NoSQL adequado a essa estrutura de dados. O processo se repete para cada template até que todos os templates sejam avaliados. O algoritmo que descreve esse processo é apresentado em Algoritmo 1. Por fim, além da coleta em si, há um módulo para disponibilizar o acesso a informação coletada. Esse módulo (4) é responsável por permitir de forma segura e racional o uso dos dados coletados por diferentes sistemas e aplicações web existentes (C).

Segundo Cavallo (2010), preços coletados da internet possuem duas desvantagens: Primeiro, percentual menor de empresas disponibiliza seus produtos e preços na internet em comparação com as lojas físicas. Tal limitação pode ser minimizada ao longo do tempo com uma maior oferta de produtos e serviços na internet. Segundo, os preços coletados da internet não incluem informações sobre as quantidades vendidas o que impede de obter market share e estimativas de elasticidade.

Avaliações futuras precisarão ser feitas de forma que seja capaz explorar se os preços online e off-line se comportam similarmente. Preocupado com tal validação, Cavallo (2010) fez pesquisa de preços nas lojas físicas dos supermercados utilizados para coleta de dados da internet. Desta forma, o autor examinou se os preços dos produtos nas lojas físicas eram similares aos preços nos sites. Uma importante característica é que um alto percentual de produtos vendidos nas lojas físicas também era comercializado nos sites em todos os países. O autor comparou os preços tanto em termo de nível quanto em tamanho e intervalo de tempo de alterações nos preços. Tal comportamento é muito importante para a avaliação de rigidez nos preços. Para tanto, o autor criou uma série de mudança de preços para cada produto que recebe valor 1 se o preço aumentou, 0 se o preço permaneceu constante e -1, caso contrário. Assim, foi possível avaliar se os preços dos produtos nas lojas físicas são semelhantes em nível e em direção de mudança para cada produto e supermercado dos países avaliados por Cavallo (2010).

Não obstante, Cavallo (2010) apresenta algumas vantagens dos preços coletados da

internet que os fazem uma fonte única de informação para análise de rigidez nos preços. Primeiro, pode-se obter preços diários para os produtos e serviços e por conseguinte, reduzir medidas de erro em relação à frequência de cálculo da inflação, analisar promoções de produtos, controles e sincronização nos preços. Segundo, os dados estão disponíveis para vários países, com maior facilidade de acesso e possibilidade de comparação entre países. Terceiro, existem informações detalhadas sobre cada produto e não há substituições forçada de itens como ocorre em estatísticas oficiais de inflação. Por fim, preços coletados da internet estão viáveis em tempo real, sem qualquer atraso para acessá-los. Isto pode ser usado para providenciar estimativas de rigidez nos preços em tempo real.

ÍNDICE DE PREÇOS ONLINE

Para calcular o índice de preços online que será comparado com o Índice Nacional de Preços ao Consumidor Amplo (IPCA) e Índice Nacional de Preços ao Consumidor (INPC) divulgados pelo Instituto Brasileiro de Geografia e Estatística (IBGE), utilizaremos a abordagem proposta por [Cavallo \(2010\)](#).

Assim, o índice de preços usa a combinação de dados online e as estruturas de ponderação oficiais do IBGE para as categorias da “cesta de mercadorias”¹ de cada índice de inflação. Maiores detalhes sobre a metodologia de coleta e cálculo do IPCA e INPC podem ser obtidas no ???. Dados diários serão utilizados para construir o índice de preços online o que é útil para observar padrões de curto prazo nos dados que ajudam a validar as informações online.

O índice de preço online será calculado utilizando os preços de todos os produtos disponíveis para compra em cada site. Isto implica que a cesta de bens muda dinamicamente ao longo do tempo podendo um produto aparecer ou desaparecer da cesta a qualquer momento devido à disponibilidade ou indisponibilidade no site. Além disso, o número de preços por produto tende a ser muito maior o que os coletados usualmente pelos órgãos governamentais. Para construir o índice, mudanças de preço são calculadas em nível de produto, então as médias dentro das categorias usando média geométrica ponderada e finalmente agregado entre categorias com uma média aritmética ponderada. Em particular, o primeiro passo é obter a média geométrica ponderada das mudanças nos preços na categoria j para cada dia t :

$$R_{t,t-1}^j = \prod_i \left(\frac{p_t^i}{p_{t-1}^i} \right)^{1/n_{j,t}} \quad (3.1)$$

onde p_t^i é o preço do bem i no tempo t , $n_{j,t}$ é o número de produtos na categoria j que estão presentes na amostra neste dia.

O segundo passo é computar o índice em nível de categoria em t :

¹Segundo o [IBGE \(2013\)](#) os índices constituem uma medida síntese de movimento de preços de um conjunto de bens e serviços, chamado “cesta de mercadorias”, representativo de um determinado grupo populacional, em certo período de tempo

$$I_t^j = R_{1,0}^j * R_{2,1}^j * \dots * R_{t,t-1}^j \quad (3.2)$$

Finalmente, o índice de preços no tempo t é a média aritmética ponderada de todos os índices das categorias:

$$IPO_t = \sum_j \frac{w_j}{W} I_t^j \quad (3.3)$$

onde w^j é o peso oficial utilizado pelo IBGE para tal categoria e W a soma de todos os pesos incluídos na amostra.

A classificação de produtos e pesos de categorias é uma das partes mais complexas deste processo. Nos dados originais, cada produto é atrelado à um endereço de web (URL) que corresponde à página onde o produto é localizado.

RIGIDEZ DE PREÇOS

Diversas estatísticas poderão ser utilizadas para a análise da rigidez de preços coletados da internet, como por exemplo: frequência de produtos com alterações diárias e frequências de alta e baixa em relação ao total de alterações nos preços em um dia. Assim, teremos um parâmetro que reflete a probabilidade incondicional de mudança no preço de uma firma ao longo de um dado período de tempo. Porém, a análise da frequência apresenta uma visão parcial do comportamento dos preços e faz-se necessário avaliar o tamanho da mudança dos preços por meio do valor absoluto da alteração no preço de um determinado produto (também o tamanho das mudanças positivas e negativas) e avaliação da sua distribuição de probabilidade. Desta forma, poder-se-á comparar o comportamento da inflação em determinadas regiões, municípios e estados em relação ao tamanho da mudança nos preços nestes locais.

[Cavallo \(2010\)](#) encontrou uma característica bimodal na distribuição do tamanho das alterações nos preços e uma forte queda da densidade das alterações próximo a 0% em alguns países, o que é consistente com os modelos de custo de menu que mudanças muito pequenas não são ótimas na presença de custo de ajuste. Outro tipo de análise é a avaliação da assimetria na densidade das alterações que pode refletir maior quantidade de preços crescendo que diminuindo.

Análise de Sobrevida

A análise de frequência ajudará na avaliação da rigidez de preços, mas ela sugere que a probabilidade de um preço se alterar é independente do tempo que uma mudança ocorre em relação à última alteração no preço. Ainda, a taxa de risco do preço se alterar é constante ao longo do tempo durante todo o período amostral. Embora esse método seja simples e efetivo para a comparação do grau de rigidez entre setores, regiões, cidades e países, um importante ponto reside sobre a forma da função risco.

Para avaliar a função risco utilizaremos a Análise de Sobrevida assim como [Cavallo \(2010\)](#). Conforme [Colosimo and Giolo \(2006\)](#), em análise de sobrevivência, a variável

resposta é, geralmente, o tempo até a ocorrência de um evento de interesse. Tal tempo é comumente conhecido como tempo de falha. Em medicina é comum o uso do método para a avaliação do tempo até a morte, transplante, doença, cura entre outros. No contexto de preços, estamos interessados no tempo até o ajuste do preço. Assim, tanto o aparecimento do risco e o evento de falha ocorrem quando uma firma muda seus preços.

A principal característica de dados de sobrevivência é a presença de censura que é a observação parcial da resposta. Isto se refere a situações em que por alguma razão, o acompanhamento do preço foi interrompido, seja porque a firma não vende mais um produto ou este não é produzido. A variável aleatória não-negativa T , usualmente contínua, que representa o tempo de falha, é geralmente especificada em análise de sobrevivência pela sua função de sobrevivência ou pela função de risco (tempo de falha). Estas duas funções são extensivamente usadas na análise de dados de sobrevivência.

Segundo [Colosimo and Giolo \(2006\)](#), a função de sobrevivência é definida como a probabilidade de uma observação não falhar até um certo tempo t , ou seja, a probabilidade de uma observação sobreviver (preço não se alterar) ao tempo t . Por outro lado, se T é a variável aleatória que mede a duração do preço, com função densidade $f(t)$ e densidade acumulada $F(t)$, o risco $h(t)$ é a probabilidade limite de que a mudança no preço ocorra em t , condicional ao preço não se alterar até este momento.

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{Pr(t < T < t + \Delta t | t < T)}{\Delta t} = \frac{f(t)}{1 - F(t)} \quad (3.4)$$

Esta função risco mede o risco instantâneo de um preço se alterar, condicionado à sobrevida. Podemos adicionar todas as taxas de risco ao longo do tempo e obter o risco total de um preço alterar acumulado até o tempo t . Isto é representado pelo função risco acumulado, $H(t)$:

$$H(t) = \int_0^t h(u) du = -\ln(1 - F(t)) \quad (3.5)$$

$H(t)$ é um aumento, função ilimitada de t , que acumula a probabilidade condicional do preço mudar ao longo do tempo. No contexto de repetidas "falhas"(preço se alterar), ela pode ser interpretada com o número esperado de ajustamento nos preços de 0 à t . O risco acumulado recebe grande atenção na Análise de Sobrevida porque ele é mais fácil de estimar do que a função risco sózinha.

Para estimar $H(t)$ e $h(t)$ empiricamente, usaremos conforme [Cavallo \(2010\)](#) uma abordagem não paramétrica dada por Nelson (1972) e Aalen (1978), que não requer hipóteses de distribuição de probabilidade. Métodos semi-paramétricos como o modelo Cox podem ser utilizados futuramente uma vez que permitem a incorporação de variáveis explicativas e a consideração da heterogeneidade não observável em nível de categoria de preços. Uma estimativa simples da função risco acumulado, $H(t)$, é dado por:

$$\hat{H}(t) = \sum_{j|t_j \leq t} \frac{c_j}{n_j} \quad (3.6)$$

onde c_j é o número de preços que mudaram em t_j e n_j é o número de preços sob risco em t_j , ou seja, os preços que não alteraram e não foram censurados até o instante imediatamente anterior a t_j . O passo incremental $\frac{c_j}{n_j}$ é uma estimativa para a probabilidade do preço mudar em t_j , levando em consideração apenas aqueles preços que sobreviveram até este ponto no tempo.

Para obter a função risco suavizada $\hat{h}(t)$, pode-se usar a seguinte equação:

$$\hat{h}(t) = \frac{1}{b} \sum_{j \in D} K\left(\frac{t - t_j}{b}\right) \Delta \hat{H}(t_j) \quad (3.7)$$

onde K é um kernel com densidade simétrica, b é a *bandwidth* de suavização e D é o conjunto de vezes com mudança de preços.

Referências Bibliográficas

- Ball, L. and Mankiw, N. G. (1994). A sticky-price manifesto.
- Barro, R. J. (1972). A theory of monopolistic price adjustment. *The Review of Economic Studies*, pages 17–26.
- Bils, M. and Klenow, P. J. (2004). Some evidence on the importance of sticky prices. *Journal of Political Economy*, 112(5):947–985.
- Blanchard, O. J. and Kiyotaki, N. (1987). Monopolistic competition and the effects of aggregate demand. *The American Economic Review*, pages 647–666.
- Bunn, P. and Ellis, C. (2012). Examining the behaviour of individual uk consumer prices*. *The Economic Journal*, 122(558):F35–F55.
- Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of monetary Economics*, 12(3):383–398.
- Cavallo, A. F. (2010). *Scraped data and prices in macroeconomics*. Harvard University.
- Coenen, G., Levin, A. T., and Christoffel, K. (2007). Identifying the influences of nominal and real rigidities in aggregate price-setting behavior. *Journal of Monetary Economics*, 54(8):2439–2466.
- Colosimo, E. A. and Giolo, S. R. (2006). Análise de sobrevivência aplicada. In *ABE-Projeto Fisher*. Edgard Blücher.
- Dhyne, E., Alvarez, L. J., Le Bihan, H., Veronese, G., Dias, D., Hoffmann, J., Jonker, N., Lünneemann, P., Rumler, F., and Vilmunen, J. (2006). Price changes in the euro area and the united states: Some facts from individual consumer price data. *The Journal of Economic Perspectives*, pages 171–192.
- Dotsey, M., King, R. G., and Wolman, A. L. (1999). State-dependent pricing and the general equilibrium dynamics of money and output. *Quarterly journal of Economics*, pages 655–690.
- Golosov, M. and Lucas Jr, R. E. (2007). Menu costs and phillips curves. *Journal of Political Economy*, 115(2).

- Gouvea, S. (2007). Nominal price rigidity in brazil: A micro evidence approach. Technical report, Mimeo. Banco Central do Brasil.
- IBGE (2013). *Sistema Nacional de Índices de Preços ao Consumidor - Métodos de Cálculo*. Diretoria de Pesquisas - Instituto Brasileiro de Geografia e Estatística (IBGE).
- Klenow, P. J. and Kryvtsov, O. (2008). State-dependent or time-dependent pricing: Does it matter for recent us inflation? *The Quarterly Journal of Economics*, 123(3):863–904.
- LOPES, L. (2008). *A rigidez nominal de preços na cidade de São Paulo: evidências baseadas em microdados do índice de preços ao consumidor da Fipe. 2008*. PhD thesis, Dissertação (Mestrado)–Universidade de São Paulo, São Paulo.
- Lucas, R. E. (1972). Expectations and the neutrality of money. *Journal of economic theory*, 4(2):103–124.
- Mankiw, N. G. (1985). Small menu costs and large business cycles: A macroeconomic model of monopoly. *The Quarterly Journal of Economics*, pages 529–537.
- Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to information retrieval*, volume 1. Cambridge university press Cambridge.
- Matos, S. and Barros, R. (2009). Comportamento dos preços no brasil: evidências utilizando microdados de preços ao consumidor. *Pesquisa e Planejamento Econômico*, 39(3).
- Nakamura, E. and Steinsson, J. (2008). Five facts about prices: A reevaluation of menu cost models. *The Quarterly Journal of Economics*, 123(4):1415–1464.
- Rotemberg, J. J. (1982). Sticky prices in the united states. *The Journal of Political Economy*, pages 1187–1211.
- Sheshinski, E. and Weiss, Y. (1977). Inflation and costs of price adjustment. *The Review of Economic Studies*, pages 287–303.
- Taylor, J. B. (1980). Aggregate dynamics and staggered contracts. *The Journal of Political Economy*, pages 1–23.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *The Journal of Economic Perspectives*, 28(2):3–27.