

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	13
	Referências Bibliográficas	23
A	METODOLOGIA IBGE	26
B	EXEMPLO DE COLETA	31

Capítulo 1

INTRODUÇÃO/MOTIVAÇÃO

Firmas individuais não ajustam seus preços em contrapartida de choques relevantes na economia. Este fato não é controversia 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.

Os diferentes mecanismos propostos para incorporar a rigidez nominal dos preços podem ser divididos 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 microfundamentos.

Por conseguinte, recentes estudos 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).

Assim, pesquisas anteriores têm demonstrado a importância da avaliação da rigidez nominal dos preços usando microdados. Os fatos estilizados obtidos de microdados podem ajudar a examinar o comportamento dos preços em nível de firmas, onde as decisões de precificação são feitas. A informação individual sobre as definições de preços permite determinar em que medida as hipóteses usadas na derivação de modelos teóricos são atualmente realísticas, o que por sua vez refina as estratégias de modelagem. 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.

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 apresenta a revisão bibliográfica do tema 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 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 ensaio 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 seguintes questionamentos pretendem ser avaliados:

- É possível utilizar os dados coletados da internet como proxy para a inflação divulgada pelos órgãos públicos?
- Quanto frequente os preços se alteram?
- Existe heterogeneidade da rigidez nominal entre setores?
- Como podemos lidar com o problema de censura e amostragem quando a função risco é estimada a partir dos dados coletados da internet?
- A probabilidade de mudança dos preços pode variar ao longo da duração dos preços?
- Como podemos avaliar o efeito de variáveis explicativas sobre a taxa de risco?

Capítulo 2

REVISÃO BIBLIOGRÁFICA

MODELOS DE PRECIFICAÇÃO

Existem diversas abordagens teóricas na literatura de modelos de rigidez nominal em nível individual. Eles são baseados em várias hipóteses para os preços não se ajustarem: Contratos de Calvo/Taylor, Custo de Menu, Informação Rígida, Ira do Cliente. Na sequência, revisamos cada uma destas hipóteses e suas implicações para os modelos de precificação.

Contratos de Calvo/Taylor

Preços nominais de acordo com [Taylor \(1980\)](#) são fixos por um certo número de períodos. Se as alterações nos preços fossem perfeitamente escalonados ao longo do tempo, a duração dos preços nominais permaneceriam constantes para todas as firmas. No modelo de Taylor, os preços são fixos por N períodos e a taxa de risco é zero para todas as durações exceto N .

No modelo de [Calvo \(1983\)](#) a probabilidade de um preço mudar é constante. Em cada período, uma proporção fixa de firmas podem alterar seus preços, as firmas remanescentes mantêm seus preços nominais fixos. A probabilidade de ser capaz de mudar preços é a mesma para todas as firmas, independentemente de quando elas mudaram seus preços pela última vez. Isto significa que a função risco é constante.

Os modelos de Taylor e Calvo são insuficientes para gerar bastante persistência do produto e da inflação a choques de política monetária (Moore,1995;[Chari et al. \(2000\)](#); [Christiano et al. \(2005\)](#)) .

Uma popular justificativa teórica é adicionar indexação ao modelo de Calvo ([Smets and Wouters \(2003\)](#); [Woodford and WALSH \(2005\)](#); [Christiano et al. \(2005\)](#)). O preço é definido no começo do contrato e para a duração do contrato este é aumentado pela inflação do período. Embora seja possível o modelo de Calvo com indexação modelar a persistência da inflação e do produto, ele tem o custo de ter os preços alterando em todo o período.

Os modelos de Taylo e Calvo generalizados são introduzidos para explicar a persis-

tência da inflação e do produto sendo consistentes com a evidência micro de rigidez nominal. No modelo generalizado de Taylor, existem muitos setores com diferentes tamanho de preços e dentro de cada setor existem um simples processo de Taylor. No modelo generalizado de Calvo, a probabilidade de reposição é dependente da duração. Adicionalmente, podemos modelar a estratégia de definição dos preços como um modelo de Calvo com múltiplos setores, que é um caso especial do modelo generalizado de Calvo. Uma característica chave desses modelos generalizados é que eles refletem a grande heterogeneidade observada em dados individuais. A rigidez de preços varia entre setores. Na presença de complementariedade nos preços os setores de baixo ajuste tem um efeito maior e desproporcional sobre todos os ajustes de preços, retardando a resposta do preço e aumentando a resposta do produto a choques. Quando uma economia heterogênea é atingida por um choque, o ajustamento inicial ocorre principalmente em empresas de setores com ajuste rápido. Com o passar do tempo, uma grande proporção das firmas que ainda tem que ajustar são firmas de setores com ajustamento mais lento. Em outras palavras, o processo de ajuste é dominado inicialmente por ajustes de alta frequência e depois por ajustes de baixa frequência.

Como proposto por [Dixon and Le Bihan \(2012\)](#), pode-se relacionar o modelo generalizado de Taylor com dados individuais e olhar a distribuição cross-sectional da duração entre firmas e atrelar o modelo generalizado de Calvo através da função de risco.

Custo de Menu

Os modelos de custo de menu assumem que a mudança no preço é custosa e esse custo impede que as firmas mudem seus preços continuamente. [Sheshinski and Weiss \(1977\)](#) mostraram que na presença de custo de alteração nos preços, a política ótima de preços é a do tipo (S, s) . S e s indicam o limite superior e inferior para o preço real, respectivamente. Uma vez que o preço real encontra-se dentro dos limites, o preço nominal será mantido constante. Durante o período de precificação, a política ótima é um tipo de política estado-contingente.

Os modelos de custo de menu usualmente são resolvidos usando métodos numéricos, assim não há expressão analítica para a taxa de risco. Na maioria das calibrações realizadas por investigações anteriores, o risco é crescente com a duração.

Informação Rígida

Assumem que é custoso para as firmas coletar informações sobre as condições econômicas correntes ([Mankiw and Reis \(2001\)](#)). Novas informações sobre o estado da economia tem sido adotadas e um novo padrão de preços ótimos são atualizadas em cada período. Informações desatualizadas são usadas para tomar decisões de preço pelo resto das empresas. A inflação, portanto, depende das expectativas anteriores da inflação e produto correntes.

Em função dos efeitos reais substancialmente maiores e persistentes serem resultados de choques monetários, o modelo de informação rígida ajusta fatores macroeconômicos melhores. Contudo, na ausência de outras fricções, alguma forma de indexação não

importando se é geral, setorial ou em nível de preços, é envolvida no planejamento do preço ótimo das firmas. Portanto, todas as firmas mudam seus preços em todo o tempo em modelos de rigidez de informação. Este argumento, contudo, é contraditório nas evidências empíricas baseadas sobre dados individuais. Estudos anteriores fazem tentativas para resolver este problema por combinar rigidez de informação com custo de menu (Klenow and Willis (2007); II and Edward (2010)). Uma economia em que as firmas se deparam tanto com custo de menu e o custo de conhecer as condições macroeconômicas precisa ser considerada. Dois métodos que as firmas podem adotar para obter informações é pagar os custos ou aprender as ações de outras empresas. Isto resulta em uma externalidade da informação e encoraja as firmas a atrasar o ajuste de preço.

Ira do Cliente

Rotemberg (2005) desenvolveu um modelo para explicar a rigidez de preços. Este modelo indica que clientes sempre analisam as decisões de precificação das firmas dependendo da percepção de justiça. Se os clientes estão convictos que os preços não são justos, eles terão reações adversas para produtos ou serviços relevantes. Assim, firmas podem abandonar alterações nos preços para evitar a ira do cliente.

Porém, consumidores não terão reação negativa e aceitam os ajustes de preços sobre as circunstâncias de rápida inflação. As empresas alterarão seus preços dentro de um calendário pertinente de forma que os clientes desenvolvam suas crenças.xvi

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 produto 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 produtos 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.

MODELOS TEMPO-DEPENDENTE E ESTADO-DEPENDENTE

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 \(2003\)](#) 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 Golo-
sov 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.

ESTUDOS EMPÍRICOS

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 compararam 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.

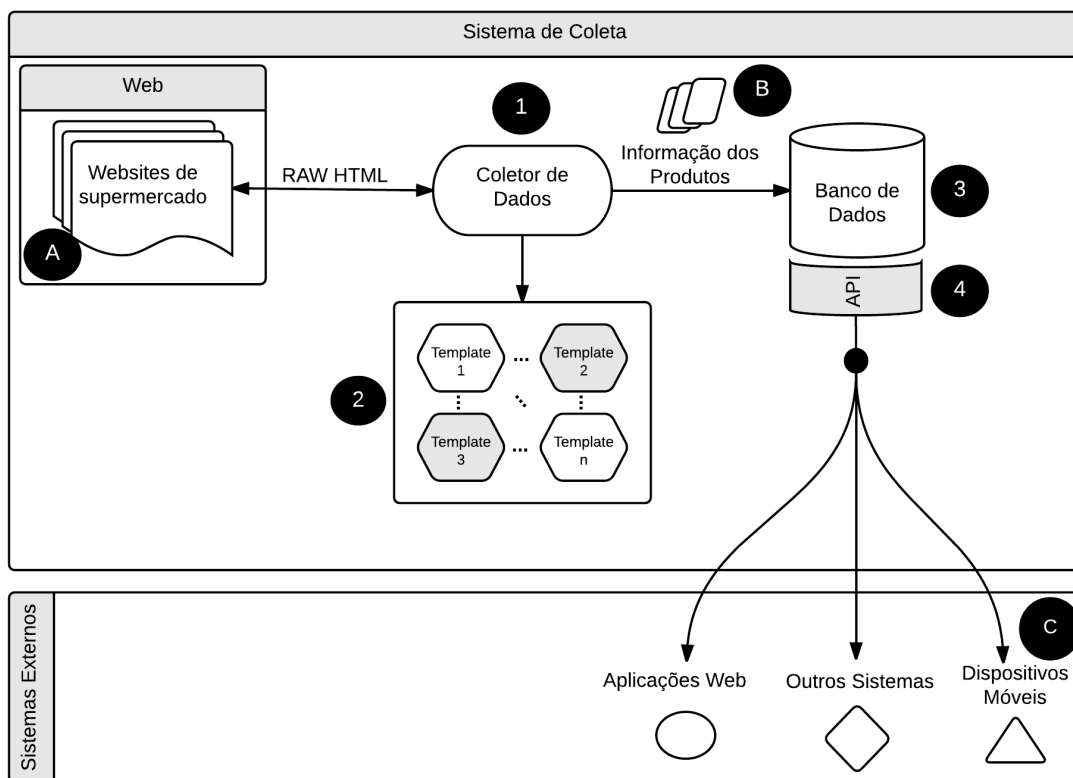


Figura 3.1: Arquitetura do Sistema de Coleta e Disponibilização dos Dados

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 [A](#). 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.

Modelo Logit de probabilidade dos preços alterarem

A metodologia descrita nesta seção é baseada sobre [Aucremanne and Dhyne \(2005\)](#) que usaram uma abordagem de dados em painel para encontrar os fatores determinantes da probabilidade de um preço se alterar na Bélgica. Abordagem similar também foi usada por [Lunnemann and Mathä \(2005\)](#) para [Baumgartner et al. \(2005\)](#) para a Austria e [Baudry et al. \(2004\)](#) para a França.

Para modelar a probabilidade de um preço mudar será preciso focar sobre os eventos de mudança nos preços enquanto ignoramos o tamanho da mudança nos preços. Assim, defina Y_{jkt} como uma variável binária:

$$Y_{jkt} = \begin{cases} 1 & P_{jkt} \neq P_{jk,t-1} \\ 0 & P_{jkt} = P_{jk,t-1} \end{cases} \quad (3.8)$$

onde Y_{jkt} indica se o preço do produto j vendido pela firma k foi alterado no começo do período t , e $P_{jk,t-1}$ é o preço do produto j vendido pela firma k no período t .

A escolha das variáveis explicativas para o modelo é dependente sobre as hipóteses sobre o mecanismo de formação de preços subjacente. Se assumimos que os definidores de preço aplicam a regra de precificação de [Calvo \(1983\)](#), então a probabilidade de ajuste dos preços não depende do preço decorrido desde a última alteração no preço ou sobre o estado da economia e a única variável explicativa será uma constante. Neste caso, o modelo logit de probabilidade da firma k alterar o preço do produto j no começo do período t é a seguinte:

$$Pr(Y_{jkt} = 1) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \quad (3.9)$$

Sobre a hipótese de uma regra de precificação conforme Calvo, a probabilidade do preço alterar é descrita apenas por β_0 . Quanto maior for β_0 menos rígidos são os preços. A equação 3.9 pode ser transformada para incluir também os elementos do modelo de [Taylor \(1980\)](#), que assume que as firmas ajustam seus preços depois de um número de períodos fixo desde a última mudança. Isto é feito afirmando que o truncamento ocorre depois de um número fixo de períodos.

Se for assumido uma regra de precificação, então, seguindo [Cecchetti \(1986\)](#), a firma k mudará o preço do produto j apenas se a diferença entre o preço desejado P_{jkt}^* e o preço atual P_{jkt} excede uma constante limiar h_{jk}^* (especifica para cada produto e firma):

$$Pr(Y_{jkt} = 1) = Pr\left(\ln\left(\frac{P_{jkt}^*}{P_{jkt}}\right) \geq h_{jk}^*\right) \quad (3.10)$$

De acordo com [Cecchetti \(1986\)](#), a probabilidade de que a diferença entre o preço atual e desejado exceder um certo limiar pode ser expressa em termos de variáveis exploratórias: inflação acumulada até a última alteração do preço, tempo decorrido até a última alteração, tamanho da última alteração no preço e mudança acumulada na variável de demanda até o ajuste do preço anterior. Isso nos leva à seguinte representação logit do modelo de precificação estado-dependente:

$$Pr(Y_{jkt} = 1) = \frac{\exp\left(\beta_0 + \sum_{i=1}^N \beta_i X_{i,jkt}\right)}{1 + \exp\left(\beta_0 + \sum_{i=1}^N \beta_i X_{i,jkt}\right)} \quad (3.11)$$

onde $X_{i,jkt}$ denota uma variável exógena da listadas anteriormente.

A equação 3.11 pode ser vista como uma extensão da equação 3.9. Contudo, ela permite testar se todos os formadores de preços na economia são tempo-dependentes. Obviamente, se β_1, \dots, β_N não são significativamente diferentes de 0, podemos concluir que todas as firmas seguem um modelo de precificação de Calvo. Por outro lado, estimativas significantes para qualquer β_1, \dots, β_N poderia ser interpretada como uma rejeição ao modelo de Calvo. [Aucremanne and Dhyne \(2005\)](#) argumentam que as estimativas de β_1, \dots, β_N capturam tanto o impacto das variáveis sobre a probabilidade do preço alterar e a participação deste particular comportamento. Portanto, a rejeição do modelo de Calvo não significará que não existem formadores de preço que seguem esta regra na economia. Ao contrário, indicará que existe uma participação significativa de firmas seguindo o modelo de precificação estado-dependente.

Determinantes para a probabilidade do preço alterar

A seguir possíveis variáveis a serem utilizadas no modelo logit para investigar os fatores que afetam a frequência com que os preços aos consumidor se alteram:

1. Inflação: sobre as hipóteses dos modelos de precificação estado-dependentes de acordo com [Cecchetti \(1986\)](#), inflação total acumulada até a última alteração no preço estaria entre as variáveis. Maior a inflação acumulada está associado com duração curta entre mudanças nos preços. Em pesquisas empíricas, a abordagem para mensurar a inflação acumulada diverge. [Aucremanne and Dhyne \(2005\)](#) alteraram a especificação de [Cecchetti \(1986\)](#) através da substituição da inflação acumulada pela inflação acumulada mensurada ao nível setorial, enquanto as mudanças na inflação total foram consideradas por um conjunto de variáveis *dummy*. A mesma abordagem foi usada em [Baumgartner et al. \(2005\)](#).

2. Tempo desde a última alteração: O tempo passado até o último ajuste de preço é uma importante variável explicativa tanto em modelos estado-dependentes e modelos tempo-dependentes. Por um lado, usando o modelo de limite-alvo [Cecchetti \(1986\)](#) provaram, tanto teóricamente e empiricamente, que quanto maior o período desde a última alteração, maior a probabilidade de observar outra alteração no preço. Não obstante, o modelo de Taylor assume o truncamento de um preço depois de um período fixo no tempo. Um coeficiente positivo e estatisticamente significativo indicará que uma participação significativa das firmas segue uma regra de precificação tempo-dependente até alterar os preços depois de um certo número de meses, dias ou semanas.
3. Tamanho da alteração anterior: [Cecchetti \(1986\)](#) argumentam que o tamanho da alteração anterior nos preços pode conter informações sobre a próxima mudança no preço. Um ajuste passado grande poderia indicar que o limite para alterar os preços é alto e as firmas estão focadas a mudar os preços em frequência menor, embora que por montantes maiores. Da mesma forma, um ajuste passado pequeno poderia indicar que o limite é baixo e os preços podem mudar mais frequentemente.
4. Variável de demanda: O modelo teórico e empírico de [Cecchetti \(1986\)](#) mostrou a importância do fator demanda (representada pelo montante de vendas da indústria) para a frequência das alterações nos preços. De acordo com seu estudo para preços de revistas, o efeito da demanda é positivo e estatisticamente significativo. Assim, será preciso definir uma variável que represente a demanda dos produtos, pois na coleta de dados da internet não é possível mensurar a demanda sobre os itens disponíveis e apenas os preços.
5. Atratividade dos preços: A frequência com que os preços se alteram pode ser afetada por efeitos psicológicos e estratégias de marketing. Um dos efeitos que é usualmente incluído nos modelos logit de mudanças de preços é o efeito de um preço atrativo. Como em [Aucremanne and Dhyne \(2005\)](#), pode-se definir a atratividade dos preços como um preço finalizando com os dígitos 9, 5 ou 0. A variável pode ser inserida no modelo por meio de *dummies*.
6. Efeito Sazonal e anual: O ajuste nos preços pode mostrar padrões sazonais que serão capturados por variáveis *dummies* que dependerão da periodicidade dos preços coletados. Portanto, essas variáveis podem ser interpretadas como o efeito da omissão de condições macroeconômicas como, por exemplo, fatores de oferta e demanda.
7. Variáveis setoriais: Finalmente, os mecanismos de formação dos preços podem diferir entre firmas e estabelecimentos por setor de atuação. Este efeito pode ser capturado por um conjunto de variáveis que incluirão *dummies* para os principais setores da economia.

Assim, a representação do modelo logit considerando a abordagem de precificação estado-dependente da equação 3.11 é agora estendido para permitir efeitos aleatórios

u_{jk} que são específicos para todos os pares de produto-firmas:

$$Pr(Y_{jkt} = 1) = \frac{\exp(X_{jkt}\beta + u_{jk} + \varepsilon_{jkt})}{1 + \exp(X_{jkt}\beta + u_{jk} + \varepsilon_{jkt})} \quad (3.12)$$

onde X_{jkt} é um vetor linha de variáveis exógenas, β é um vetor coluna dos coeficientes do modelo logit e ε_{jkt} é um termo de erro. Por fim, pode-se distinguir a variável Y_{jkt} entre alterações em todos os preços ou excluir as promoções da análise.

Referências Bibliográficas

- Aucremanne, L. and Dhyne, E. (2005). Time-dependent versus state-dependent pricing: a panel data approach to the determinants of belgian consumer price changes.
- 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.
- Baudry, L., Le Bihan, H., Sevestre, P., and Tarrieu, S. (2004). Price rigidity in france—evidence from consumer price micro-data. Technical report, ECB Working Paper.
- Baumgartner, J., Glatzer, E., Rumler, F., and Stiglbauer, A. (2005). How frequently do consumer prices change in austria?: evidence from micro cpi data. Technical report, Citeseer.
- 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.
- Cecchetti, S. G. (1986). The frequency of price adjustment: A study of the newsstand prices of magazines. *Journal of Econometrics*, 31(3):255–274.
- Chari, V. V., Kehoe, P. J., and McGrattan, E. R. (2000). Sticky price models of the business cycle: can the contract multiplier solve the persistence problem? *Econometrica*, 68(5):1151–1179.
- Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*, 113(1):1–45.

- 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.
- Dixon, H. and Le Bihan, H. (2012). Generalised taylor and generalised calvo price and wage setting: Micro-evidence with macro implications*. *The Economic Journal*, 122(560):532–554.
- 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).
- II, K. and Edward, S. (2010). A tale of two rigidities: Sticky prices in a sticky-information environment. *Journal of Money, Credit and Banking*, 42(8):1543–1564.
- 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.
- Klenow, P. J. and Willis, J. L. (2007). Sticky information and sticky prices. *Journal of Monetary Economics*, 54:79–99.
- 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.
- Lunnemann, P. and Mathä, T. Y. (2005). Consumer price behaviour in luxembourg: evidence from micro cpi data.

- Mankiw, N. G. (1985). Small menu costs and large business cycles: A macroeconomic model of monopoly. *The Quarterly Journal of Economics*, pages 529–537.
- Mankiw, N. G. and Reis, R. (2001). Sticky information versus sticky prices: a proposal to replace the new keynesian phillips curve. Technical report, National Bureau of Economic Research.
- Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to information retrieval*, volume 1. Cambridge university press Cambridge.
- Mash, R. (2003). New keynesian microfoundations revisited: A calvo-taylor-rule-of-thumb model and optimal monetary policy delegation.
- 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.
- Rotemberg, J. J. (2005). Customer anger at price increases, changes in the frequency of price adjustment and monetary policy. *Journal of Monetary Economics*, 52(4):829–852.
- Sheshinski, E. and Weiss, Y. (1977). Inflation and costs of price adjustment. *The Review of Economic Studies*, pages 287–303.
- Smets, F. and Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European economic association*, 1(5):1123–1175.
- 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.
- Woodford, M. and WALSH, C. E. (2005). Interest and prices: Foundations of a theory of monetary policy.

Apêndice A

METODOLOGIA IBGE

O Sistema Nacional de Preços ao Consumidor (SNIPC) efetua a produção e sistemática de índices de preços ao consumidor tendo como unidade de coleta estabelecimentos comerciais e de prestação de serviços, concessionária de serviços públicos e domicílios (para levantamento de aluguel e condomínio). O sistema abrange as regiões metropolitanas do Rio de Janeiro, Porto Alegre, Belo Horizonte, Recife, São Paulo, Belém, Fortaleza, Salvador e Curitiba, além do Distrito Federal e do município de Goiânia. A partir de janeiro de 2014, o SNIPC passou a incorporar a Região Metropolitana de Vitória/ES e o município de Campo Grande/MS.

As motivações para criação do Índice Nacional de Preços ao Consumidor Amplo (IPCA) e Índice Nacional de Preços ao Consumidor (INPC) foram a obtenção de medida geral de inflação e a indexação salarial, respectivamente.

As etapas para a construção dos índices de preços são elencadas abaixo. Para maiores detalhes, consultar [IBGE \(2013\)](#).

1. Definição da população objetivo:
 - (a) Para o INPC são as famílias residentes nas áreas urbanas das regiões de abrangência do SNIPC com rendimentos de 1 a 6 salários-mínimos e cujos chefes são assalariados;
 - (b) Para o IPCA, as famílias residentes nas áreas urbanas das regiões de abrangência do SNIPC com rendimentos de 1 a 40 salários-mínimos, qualquer que seja a fonte de rendimentos.
2. Obter estruturas de ponderação: O conjunto de bens e serviços representativos do consumo dos grupos e os valores de despesa que lhes são associados.
 - (a) Pode ser diferente para uma determinada população-objetivo;
 - (b) São resultado da consolidação dos orçamentos familiares levantados pela POF;

(c) São montadas de forma que categorias de consumo de mesma natureza fiquem juntas. Hierarquicamente ²:

- i. Grupo
- ii. Subgrupo
- iii. Item
- iv. Subitem

3. Calcular os Pesos:

- (a) Anualisar os valores de despesa com consumo oriúndas da POF que são coletados em diferentes períodos de referência;
- (b) Colocar as despesas anuais em preços constantes de 15 de janeiro de 2009;
- (c) Somar para cada subitem, despesas das famílias pertencentes à população-objetivo;
- (d) A razão da soma anterior e a despesa total de todas as famílias da região em questão gera o índice.

4. Definir estruturas de consumo: A partir da participação dos subitens, define-se quais permanecerão para o cálculo do índice. Para tanto, utiliza-se o seguinte critério:

- (a) subitens com participação igual ou superior a 0,07% fazem parte das estruturas;
- (b) subitens com participação inferior a 0,01% em hipótese alguma fazem parte das estruturas;
- (c) subitens com ponderação igual ou superior a 0,01% e inferior a 0,07% podem fazer parte para assegurar que o item do qual fazem parte tenha cobertura de 70% dos gastos realizados com os componentes do item.

5. Cadastrar informantes: Por meio da Pesquisa de Locais de Compra (PLC) faz-se o cadastro dos estabelecimentos.

6. Cadastrar produtos: Por meio da Pesquisa de Especificação de Produtos e Serviços (PEPS) obtém-se os produtos.

7. Coletar preços: Tarefa contínua, realizada mensalmente, nas áreas de cobertura da pesquisa, ao longo do mês. Para viabilizá-la existem pesquisadores de campo dedicados à coleta de informações necessárias à produção dos índices. Questionário eletrônico de coleta instalado em computador de mão, no qual estão descritas as características dos produtos ou serviços nele investigados.

²Por exemplo, Laranja-pera é um subitem do item "Frutas" que conjuntamente com outros itens formam o subgrupo "Alimentação no Domicílio", o qual, unido ao subgrupo "Alimentação Fora do Domicílio" compõe o grupo "Alimentação e Bebidas". Retirado de [IBGE \(2013\)](#)

A tabela A.1, apresenta um resumo das fontes de informações relevantes para os índices de preços: Pesquisa de Orçamentos Familiares (POF), Pesquisa de Locais de Compra (PLC) e Pesquisa de Especificação de Produtos e Serviços (PEPS).

PESQUISAS BÁSICAS	
POF	Fornece as estruturas de ponderação para cada grupo de bens e serviços).
PLC	Fornece o cadastro de informantes da pesquisa que tem manutenção constante.
PEPS	Fornece o cadastro de produtos e serviços a serem pesquisados.

Tabela A.1: Principais Pesquisas Utilizadas na Metodologia

Por fim, temos a metodologia de cálculo dos índices de preços. Sinteticamente, partindo-se de milhares de preços coletados mensalmente, obtêm-se no primeiro processo-síntese, as estimativas dos movimentos de preços referentes a cada produto pesquisado. Estes resultados são agregados por uma fórmula elementar de cálculo e geram a estimativa para variação de preços de cada subitem; essas estimativas, por sua vez, por outro processo agregativo, produzem os índices referentes a itens, que, por fim, geram os índices regionais e nacional mensais de cada população-objetivo.

Cálculo no nível de produto

Primeiro, calcula-se mensalmente o relativo de preços referentes a dois meses e temos a estimativa da variação mensal dos preços do produto j , ou relativo do produto j , conforme:

$$R_{t-1,t}^j = \frac{\bar{P}_t^j}{\bar{P}_{t-1}^j} = \frac{\frac{1}{n_t} \sum_{L=1}^{n_t} p_t^{j,L}}{\frac{1}{n_{t-1}} \sum_{L=1}^{n_{t-1}} p_{t-1}^{j,L}} \quad (\text{A.1})$$

onde $R_{t-1,t}^j$ é a medida da variação de preços do produto j entre os meses $t-1$ e t , \bar{P}_t^j e \bar{P}_{t-1}^j os preços médio do produto j nos meses t e $t-1$, respectivamente, assim como $p_t^{j,L}$ e $p_{t-1}^{j,L}$ são os preços do produto j no local L nos meses t e $t-1$. Por fim, n_t e n_{t-1} são os números de locais que compõem a amostra do produto nos meses t e $t-1$.

Por conseguinte, o próximo passo é a agregação no nível de subitem. Para tanto, calcula-se a média geométrica dos resultados obtidos para cada produto que compõe o subitem. Assim,

$$R_{t-1,t}^k = \sqrt[m_k]{\prod_{j=1}^{m_k} R_{t-1,t}^j} \quad (\text{A.2})$$

onde $R_{t-1,t}^k$ é a variação média de preços entre os meses $t-1$ e t dos produtos que compõem o subitem k , $R_{t-1,t}^j$ é a variação do preço do produto j entre os meses $t-1$ e t (fórmula

No que diz respeito aos resultados ao longo do tempo, evidencia-se a importância de manter-se o painel de produtos fixos, a exemplo do que ocorre com o painel de locais, sob pena de incorporar falsas variações de preços. Desta forma, o IBGE imputa o preço de um produto para determinado local ou subitem. Para maiores informações de como é feito esse processo, consultar [IBGE \(2013\)](#).

Cálculo no nível de item

Usa-se a fórmula de Laspeyres que expressa a razão entre o gasto efetuado no momento t , necessário para consumir as mesmas quantidades do momento 0, e o gasto efetuado no momento 0. A fórmula [A.3](#) representa o índice:

$$L_{0,t} = \sum_{i=1}^n \left(\frac{p_0^i q_0^i}{\sum_{j=1}^n p_0^j q_0^j} \right) \left(\frac{p_t^i}{p_0^i} \right) \quad (\text{A.3})$$

onde $\frac{p_t^i}{p_0^i} = R_{0,t}^i$ é o estimador da variação de preços do subitem i entre os momentos 0 e t e $\frac{p_0^i q_0^i}{\sum_{j=1}^n p_0^j q_0^j} = W_0^i$ é o peso do subitem i obtido a partir da POF. Observe-se que tanto $R_{0,t}^i$ como W_0^i referem-se, na prática, a pequenos agregados de produtos. Para se conhecer a variação de preços do item m para uma determinada área e faixa de rendimento em ciclos mensais utiliza-se a fórmula [A.4](#).

$$I_{t-1,t}^m = \frac{\sum_{i=1}^n W_{t-1}^i R_{t-1,t}^i}{\sum_{i=1}^n W_{t-1}^i} \quad (\text{A.4})$$

onde $I_{t-1,t}^m$ é o índice do item m entre os meses $t-1$ e t , W_{t-1}^i é o peso do subitem i em $t-1$ e $R_{t-1,t}^i$ é o relativo do subitem i entre $t-1$ e t . Além disso, o peso W_{t-1}^i a partir de $t=2$ é dado por:

$$W_{t-1}^i = W_0^i \prod_{j=0}^{t-2} \frac{R_{j,j+1}^i}{I_{j,j+1}} \quad (\text{A.5})$$

onde W_0^i é o peso do subitem i obtido a partir da POF, $R_{j,j+1}^i$ é o relativo do subitem i entre os meses j e $j+1$ e $I_{j,j+1}$ é o resultado do índice geral entre os meses j e $j+1$.

Cálculo dos índices regionais

O resultado mensal para a área A e população-objetivo F é dado por:

$$IPC_{t-1,t}^{A,F} = \sum_m^M W_{t-1}^m I_{t-1,t}^m \quad (\text{A.6})$$

onde $I_{t-1,t}^m$ é o índice do item m obtido conforme a equação [A.4](#) e W_{t-1}^m corresponde ao peso de cada item e é obtido somando-se os pesos no período $t-1$ por meio da equação [A.5](#) utilizando todos os subitens que compõem o respectivo item m .

Cálculo dos índices nacionais

Os índices nacionais são obtidos a partir dos índices regionais. O método empregado para obtenção dos índices nacionais consiste no cálculo de uma média aritmética ponderada dos índices regionais mensais, conforme:

$$INPC_{t-1,t} = \sum_{A=1}^{11} W^{A,F} IPC_{t-1,t}^{A,F} \quad (\text{A.7})$$

onde $INPC_{t-1,t}$ é o índice nacional referente à variação de preços entre os meses $t-1$ e t , $IPC_{t-1,t}^{A,F}$ é o índice da área A , população-objetivo F , obtido via [A.6](#). Além disso, $W^{A,F}$ é o peso da área A , população-objetivo F . Na mais recente atualização, tendo como fonte a POF 2008-2009, os pesos das regiões foram obtidos com base nas estimativas da população urbana para os estados, Grandes Regiões e Brasil. A tabela [A.2](#), apresenta os índices regionais antes e após a alteração.

Regiões	IPCA	INPC
Belém	4,65	7,03
Fortaleza	3,49	6,61
Recife	5,05	7,17
Salvador	7,35	10,67
Belo Horizonte	10,86	10,60
Vitória	1,78	1,83
Rio de Janeiro	12,06	9,51
São Paulo	30,67	24,24
Curitiba	7,79	7,29
Porto Alegre	8,40	7,38
Campo Grande	1,51	1,64
Goiânia	3,59	4,15
Brasília	2,80	1,88

Tabela A.2: Participação do índice regional no agregado nacional

Apêndice B

EXEMPLO DE COLETA

```
> # library(devtools)
> # install_github("hadley/rvest")
> #
> #
> # library(rvest)
> #
> # # biscoitos <- html("http://www.paodeacucar.com.br/secoes/C4228_C4371/biscoitos-salgados")
> # # preco <- biscoitos %>%
> # #   html_nodes("h3 a") %>%
> # #   html_text() # aqui, o valor final da tag a dentro de h3
> # # preco
> #
> # paoacucar <- html("http://www.paodeacucar.com.br/") # pegar o código html da página
> #
> # secoes <- paoacucar %>%
> #   html_nodes("li a") %>% # dentro do nó "li" pego a tag "a"
> #   html_attr("href") # aqui extração de atributo da tag "a"
> # secoes
> #
> # # Porém, dentro do nó "li" pode existir outras coisas além das seções de produtos
> # do supermercado. Assim, utilizo o código abaixo para extrair de "secoes" apenas as
> # strings que têm a palavra secoes. Para tanto utilizo a função grepl.
> #
> # filtered_secoes <- secoes[grepl(".*secoes.*", secoes)]
> #
> #
> # # Agora, utilizo as seções filtradas para coletar os preços e os nomes dos produtos.
> # O for abaixo realiza essa tarefa. Cada seção pode ter um número x de páginas que contêm
> # os preços e nomes dos produtos. O código abaixo utiliza o "while" para realizar a
> # extração enquanto houver nova página, caso contrário, ele para.
> #
> # dataframe <- data.frame()
> # line_index <- 0
> #
```

```

> # for(url in filtered_secoes) {
> #   page_number <- 0
> #   while(TRUE) {
>     # url's que serão utilizads na coleta
> #   url_formatted <- paste(paste(paste(url, '?p=', sep=''), page_number, sep=''), '&qt=36', sep=
> #   print (url_formatted)
> #   html_page <- html(url_formatted)
> #
> #   preco <- html_page %>%
> #     # Navegação pelos nós "p" e "a" até a tag "span"
> #     html_nodes("p a span") %>%
> #     html_text() # extração do valor da "tag"
> #   print(preco) # mostrar os preços da página
> #
> #   name <- html_page %>%
> #     html_nodes("h3 a") %>% # navegação pelo nó "h3"
> #     html_text() # extrair os nomes de cada seção
> #   print(name) # mostrar os nomes dos produtos
> #
> #   if (length(preco) == 0) { # condição para parar a coleta da seção
> #     break
> #   }
> #
> #   for(i in 1:length(preco)) {
> #     line_index <- line_index + 1
> #     dataframe[line_index,1] <- name[i]
> #     dataframe[line_index,2] <- preco[i]
> #   }
> #
> #   page_number <- page_number + 1
> #
> #   # Save data at MongoDB
> #   # Start mongodb
> #
> # }
> # }
> #
> # write.csv2(dataframe, file='dataframe2.csv')

```