



Instituto Superior de Economia e Gestão

UNIVERSIDADE TÉCNICA DE LISBOA

DESDE 1911

MESTRADO EM

ECONOMETRIA APLICADA E PREVISÃO

TRABALHO FINAL DE MESTRADO

DISSERTAÇÃO

**MODELAÇÃO DO INTERESSE DE VÍDEOS DE
MÚSICA MEDIDO PELO NÚMERO DE PROCURAS NA
INTERNET VIA GOOGLE TRENDS**

MARIA MABEL DE BARROS LIMA

Novembro/2014



Instituto Superior de Economia e Gestão

UNIVERSIDADE TÉCNICA DE LISBOA

DESDE 1911

MESTRADO EM
ECONOMETRIA APLICADA E PREVISÃO

TRABALHO FINAL DE MESTRADO
DISSERTAÇÃO

MODELAÇÃO DO INTERESSE DE VÍDEOS DE
MÚSICA MEDIDO PELO NÚMERO DE PROCURAS NA
INTERNET VIA GOOGLE TRENDS

MARIA MABEL DE BARROS LIMA

ORIENTAÇÃO:

PROFESSOR DOUTOR JOÃO CARLOS HENRIQUES
DA COSTA NICOLAU

Novembro/2014

Resumo

O mercado da música mundial continua a se expandir em novos mercados e criar novos negócios, atraindo cada vez mais usuários para os serviços de música sob o formato digital. A receita gerada pela indústria da música digital apresentou um crescimento de 4,3% de 2012 para 2013 (de US\$ 5.6 bi para US\$ 5.9 bi), já representa 39% da receita total gerada no mercado mundial. Para uma melhor compreensão da natureza do ciclo de vida do formato digital da música emergente, buscou-se estudar os vídeos de música da *internet*, dado a sua importância na indústria da música, por ser um dispositivo de *marketing* destinado, principalmente, a promover as vendas de gravações de música, por ser um importante contributo para a comercialização da música popular e dado a ausência de literatura de caráter qualitativo e quantitativo subjacente. Esta dissertação pretende propor um modelo capaz de descrever a dinâmica dos vídeos de música ao longo do tempo, ou seja, de como se dá o interesse coletivo por um determinado vídeo de música. A base empírica deste estudo consiste em séries temporais de vídeos de música (dados semanais) relacionando frequências de busca disponíveis a partir do Google Trends. Empiricamente avaliou-se o desempenho do modelo proposto, usando métodos de estimação não lineares de séries temporais. Os resultados obtidos permitem distinguir os vídeos de música de *internet* de curta duração de outros mais duradores.

Palavras-chave: Ciclo de vida do produto, Google Trends, Modelo de Séries Temporais Não Lineares, Vídeos de Música da *Internet*.

Abstract

The global music business continues to expand into new markets and create new business, attracting more and more users to digital format music services. The revenues generated by the digital music industry grew by 4.3% from 2012 to 2013 (US\$ 5.6 billion to US\$ 5.9 billion), already represents 39% of total revenues generated by the global music market. Internet music videos are a pervasive phenomenon on the Web, they typically consist in a short film made to advertise a popular song that spread through network. In order to contribute to a better understanding of the nature of the life cycle of internet music videos, given its importance in the music industry and in particular plausible models that would explain their temporal dynamics have not previously been reported. Our aim in this paper is thus to develop meaningful and interpretable model that describes the dynamics of music videos over time, i.e., how collective attention to internet music videos evolves over time, and how relate with their life cycle. The empirical basis of our study consists of time series of music videos relating frequencies available search from Google Trends. We conduct an empirical illustration to assess the performance of our model using nonlinear time series models. The results of the empirical illustration indicate to distinguish short and “long” life cycle’s internet music videos.

Keywords: Google Trends, Internet Music Video, Nonlinear Time Series Models, Product’s Life Cycle.

“The mind that opens to a new idea never returns to its original size.”

Albert Einstein

Agradecimentos

Ao *Professor Doutor João Carlos Henriques da Costa Nicolau*, expresso o meu profundo agradecimento pela orientação e pela imprescindível ajuda na definição do tema da dissertação, pela sua disponibilidade e suas sugestões, indispensáveis à conclusão deste trabalho.

Agradeço a todos os professores do Mestrado em Econometria Aplicada e Previsão, do ISEG/ULisboa, pelos valiosos ensinamentos no decorrer destes dois anos de curso.

Enfim, agradeço a todas as pessoas que, ao longo do Mestrado me ajudaram, directa ou indirectamente, a cumprir os meus objetivos e a realizar mais esta etapa da minha formação académica. Aos meus colegas, em específico *Samuel Arone* agradeço-lhe a sincera amizade, a ajuda e os estímulos nas alturas de desânimo.

À *Minha Família*, em especial à *Camilo Guimarães Lima*, a eles dedico todo este trabalho.

Sumário

1	Introdução	1
2	Modelo Proposto	10
3	Método e Análise Empírica	15
3.1	Introdução	15
3.2	Dados	15
3.3	Especificação e Estimação dos Modelos	17
3.4	Discussão dos Resultados	27
3.5	Análise dos parâmetros β_2 , β_3 , β_4 e α_i para diferentes vídeos de música	31
4	Conclusões	34

Lista de Figuras

1	Ciclo de vida do produto (<i>Standard</i>).	3
2	Ciclo de vida do produto (<i>Fads</i>).	5
3	Séries temporais de vídeos de música.	6
4	Simulação gráfica do modelo proposto.	13
5	Resíduos NLS - volume de consultas para o vídeo de música <i>No One</i> (Alicia Keys, 2007).	21
6	Resíduos FG-NLS - volume de consultas para o vídeo de música <i>No One</i> (Alicia Keys, 2007).	30
7	Dois diferentes vídeos de músicas, populares na <i>internet</i>	32

Lista de Tabelas

1	Modelo NLS estimado para o vídeo de música <i>No One</i> (Alicia Keys, 2007).	22
2	Modelo FGNLS estimado para o vídeo de música <i>No One</i> (Alicia Keys, 2007).	27
3	Modelo FGNLS estimado para os vídeos de música <i>No One</i> (Alicia Keys, 2007) e <i>Waka waka</i> (Shakira, 2010).	33

1 Introdução

Em 2013, a indústria da música foi responsável pela geração de uma receita mundial de US\$ 15 bilhões de dólares. Ano após ano, o cenário é de retração, tendo se observado uma redução de 3,9%, com relação ao ano de 2012, respectivos US\$ 15.6 bi, esta queda é justificada principalmente pela diminuição da demanda do formato físico (CD, DVD e disco de vinil), pelo segundo principal maior mercado da música, o japonês.

Em geral, as vendas em formato físico caíram 11,7% (de US\$ 8.75 bi em 2012 para 7.73 bi em 2013) e apesar do declínio observado ao longo dos anos, em vendas de formatos físicos, este ainda é responsável por pouco mais da metade, 51,4% da receita global, em comparação com 56% no ano anterior. O cenário é de mudança, a tendência mundial aponta para um impressionante crescimento do mercado digital (MP3, *downloads*) da música que tende a fazer a indústria fonográfica avançar e a compensar o declínio nas vendas de formato físico. O formato digital continua ascendente e já representa 39% da receita mundial gerada em 2013 pela indústria da música (*Digital Music Report, 2014*).

Com o rápido desenvolvimento da ciência e tecnologia, o processo de inovação do produto é acelerado e a preferência dos consumidores rapidamente é alterada, como consequência o ciclo de vida do produto tende a torna-se cada vez mais curto, repercutindo em mudanças nos mais diversos mercados.

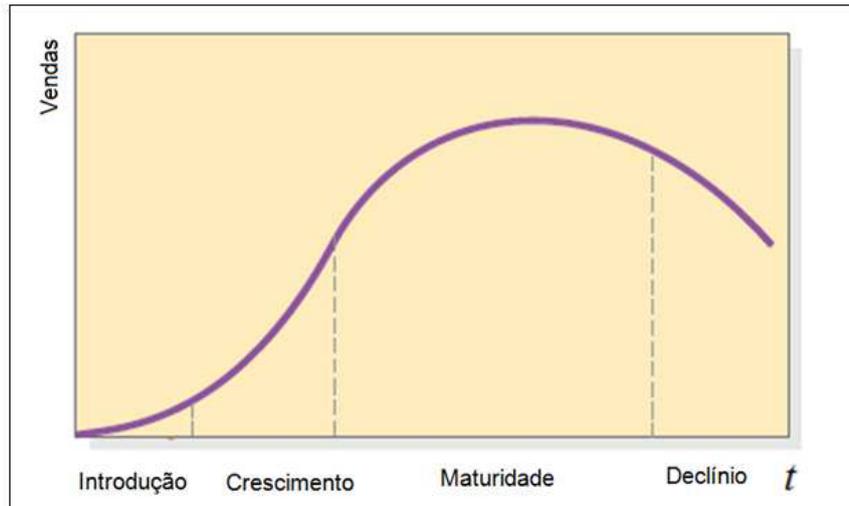
Segundo Kotler (2012) tudo que possa ser quantificado em valor monetário, pode ser considerado um produto¹, algo que pode ser oferecido em um mercado para satisfazer um desejo ou necessidade, incluindo bens físicos, serviços, experiências, eventos, pessoas, lugares, propriedades, organizações, informações e ideias. Assim como o ser humano, o produto tem um ciclo de vida, isto é o mesmo que afirmar que o produto tem vida limitada.

O ciclo de vida do produto é definido basicamente como a evolução de um produto, mensurado por suas vendas ao longo do tempo. Isto é, cada produto passa por uma série de estágios e o total destes é considerado como o ciclo de vida do produto, nos quais as vendas oscilam, enfrentam diferentes desafios, oportunidades e problemas, os lucros sobem e descem, portanto produtos necessitam de diferentes estratégias de produção, financeira, *marketing*, compras e recursos humanos de acordo com cada estágio do seu ciclo de vida (Cox, 1967). A versão clássica² da teoria do ciclo de vida do produto postula que o histórico de vendas de um novo produto segue uma curva em forma de sino ao longo do tempo e esta curva é normalmente dividida em quatro etapas, figura 1.

¹Pode ser classificados com base na durabilidade, tangibilidade e uso.

²Joel Dean foi precursor da Teoria do Ciclo de Vida do Produto, com o artigo intitulado “Pricing Policies for New Products”, 1950.

Figura 1: Ciclo de vida do produto (*Standard*).



Fonte: Kotler, 2012 (adaptado).

Inicialmente, as condições primárias para o crescimento das vendas (introdução/lançamento) e lucros de um novo produto estão associadas ao quão direcionado é este produto ao cliente, sendo este período caracterizado, em geral, por incerteza e risco, altos investimentos em tecnologia, propaganda, distribuição e embalagem/*design*, por vendas baixas, lucros em geral não existentes justificados pelos altos custos de inserção do produto no mercado.

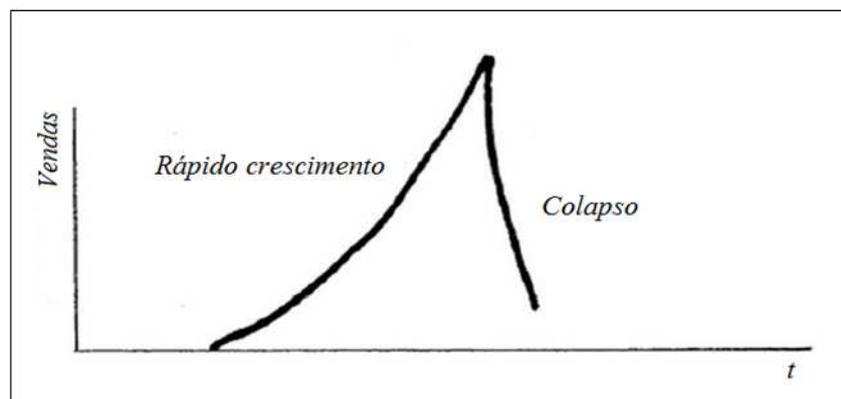
A usual característica do sucesso de um novo produto é o aumento gradual de suas vendas (crescimento) e, em algum ponto deste aumento, o mercado apresenta crescimento e acontece a demanda do consumidor, conseqüentemente as vendas. Quanto mais acelerada for a taxa de aceitabilidade do produto pelo consumidor, mais fácil é a abertura de novos canais de distribuição e há uma melhoria substancial no lucro.

A desaceleração no crescimento da demanda (maturidade) acontece, em geral, quando o produto alcança a aceitação pela maioria dos potenciais consumidores. A competição de preço se torna intensa e as tentativas competitivas de atingir e manter a preferência envolve uma fina diferenciação do produto, como em serviços ao cliente e em práticas promocionais, por exemplo. O estágio de maturidade pode ser passado rapidamente, como no caso da maioria dos *Fads*³ ou ele pode persistir durante gerações. Observa-se um decrescimento ou estabilização dos lucros nesta etapa. E por fim, o produto começa a perder consumidores (declínio) e as vendas se reduzem e, conseqüentemente, os lucros.

Alguns produtos, apesar de bem sucedidos na fase de introdução, são caracterizados por apresentarem um formato de curva que difere do formato clássico. Geralmente, são formas de comportamento relacionados às ideias, a atividades ou a produtos que são adotados por uma grande parcela de uma população durante um período de tempo, enquanto permanece sob a forma de novidade. Não conseguem sobreviver por muito tempo, por não satisfazerem uma forte necessidade da população, caracterizados por apresentarem um ciclo de vida curto (Mayerson e Katz, 1957).

³Produtos que são rapidamente introduzidos a um determinado público, com fase de aceitação curta, são adotados com grande zelo, apresentam um pico precoce, e declínio muito rápido. Geralmente atraem apenas um número limitado de pessoas que procuram diversão, por exemplo (Kotler, 2012).

Figura 2: Ciclo de vida do produto (*Fads*).



Fonte: Wasson, 1968 (adaptado).

Neste sentido, em geral produtos de entretenimento são caracterizados por apresentarem ciclos de vida curtos, apresentam uma demanda normalmente maior no momento do lançamento e diminui muito rapidamente nos meses subsequentes. Como colocado anteriormente, o mercado mundial da indústria da música passa por um período de transformação, observado a crescente substituição do formato físico pelo formato digital das músicas, com repercussão notória na receita gerada por este segmento.

Embora as origens de vídeos de música datem de muito longe, eles entraram em destaque na década de 1980⁴. Neste período, teve início a exploração e a expansão da forma e do estilo⁵ do gênero musical, usando efeitos mais sofisticados, misturando cinema e vídeo, e adicionando um enredo ou trama para o vídeo da música. Em meados da década de 2000,

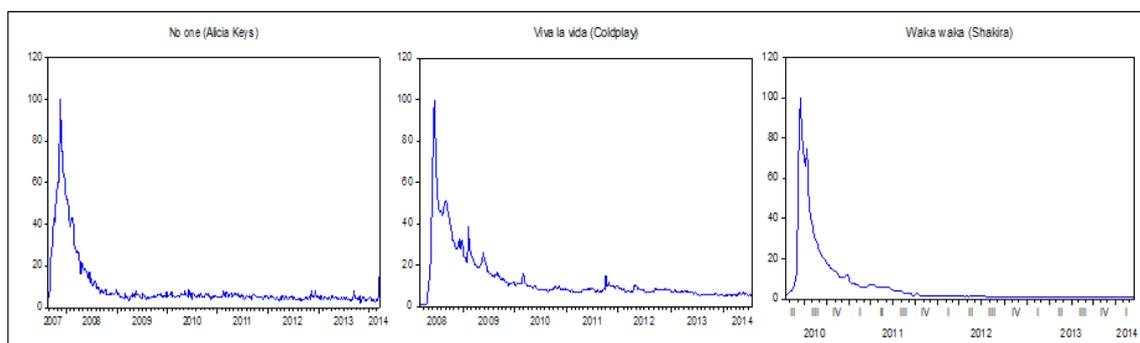
⁴Quando o canal de televisão a cabo MTV (*Music Television*) foi lançado com o propósito de divulgar vídeos de música.

⁵Por exemplo, ocasionalmente vídeos foram feitos em uma forma não representacional, em que o artista musical não era mostrado, mas como o intuito do vídeo é a promoção do artista este são relativamente raros.

sites da *internet*, como o iTunes (2001) e YouTube (2005), tiveram um efeito democratizante sobre a indústria de mídia digital, onde a visualização *online* tornou o acesso destes vídeos muito mais rápido e mais fácil para os diversos usuários. Cabe citar que o Yahoo! Video, Facebook e MySpace usam tecnologia similar. Estes *sites* tem um grande efeito sobre a visualização de vídeos de música, por exemplo, alguns artistas começam a ter sucesso como resultado de vídeos vistos na maior parte ou inteiramente *online*. Mais recentemente, a iTunes Store começou a vender vídeos de música para uso no iPod da Apple com capacidade de reprodução de vídeo.

Na figura 3, são expostos alguns exemplos de vídeos de música de *internet*, que indicam como o interesse mundial por determinados vídeos de música (em termo da frequência relativa de pesquisa na *internet*) evolui ao longo do tempo.

Figura 3: Séries temporais de vídeos de música.



Fonte: www.google.com.br/trends (Acedido em 16/03/2014).

Embora os detalhes destas séries temporais pareçam caóticos, há características de uma tendência geral.

Saber como um produto se move de um estágio para o outro é importante para desenvolver estratégias de mercado para o produto e prever com precisão esta mudança é um dos temas mais debatidos quando se fala em ciclo de vida do produto. Belville (1966) usando o *Classical Normal Distribution Model* identificou o ponto de passagem do estágio de introdução ao de crescimento das vendas de TV em cores.

Poli e Cook (1969) analisaram a variação percentual nas vendas de bens duráveis, no momento t para o $t + 1$. As mudanças foram plotadas com uma distribuição normal com média zero, tendo concluído que se o produto apresentar uma mudança percentual nas vendas maior que $\frac{1}{2}\sigma$, ele se encontra no estágio de crescimento, se for menor que $\frac{1}{2}\sigma$, ele está no estágio de declínio e, se for $\pm\frac{1}{2}\sigma$, o produto é considerado estável.

Barclay (1963), Crawford (1966), King (1966), Learner (1968), Bass (1969), Massy (1969) desenvolveram modelos para novos produtos, que preveem os estágios de crescimento e de maturidade, com base nos dados do mercado. Os quais tem a limitação de não preverem com precisão a segunda metade da curva do ciclo de vida do produto. Se forem usados, os resultados acabam por subestimar o estágio da maturidade e, assim, entrarem na fase de declínio antecipadamente. Por outro lado, observou-se que a inclusão de certos indicadores⁶ da fase de maturidade às previsões tecnológicas, às análises de produtos similares, aos modelos epidemiológicos ou à previsão das vendas, ajuda a identificar o estágio

⁶Por exemplo, declínio do número de novos consumidores, lucros em queda, substituição do produto, diminuição na taxa de consumo pelos atuais usuários, etc.

de maturidade do ciclo de vida de um produto.

Para Dean (1950), a duração das fases do ciclo de vida de um determinado produto é função de uma taxa de mudança técnica, de uma taxa de aceitação do mercado e da facilidade de inserção no mercado. Cox (1967) examinou empiricamente a duração de cada fase do ciclo de vida de um medicamento controlado típico, bem como a sequência das fases. Os resultados indicaram: um mês (introdução), seis meses (crescimento), quinze meses (maturidade) e no total somando vinte e dois meses até o declínio.

Segundo Xu e Song (2007) os métodos de previsão tradicionais não são projetados para produtos de ciclo de vida curto, por não considerarem as características do ciclo de vida curto. Tendo proposto um modelo de Bass revisado, levando em consideração fatores sazonais e a informação passada das vendas de produtos similares, aplicado para previsão da demanda de um produto com ciclo de vida curto, enquanto que os parâmetros do modelo podem ser estimados pela informação de venda de um produto similar. Para validação do modelo de Bass com uma componente sazonal, foram considerados dados das vendas de cinco versões diferentes de computadores, tendo sido o modelo estimado por meio do NLS (*Nonlinear Least Squares*). Se comparado com outros modelos apresentados pela literatura subjacente, como por exemplo, quando se compara ao modelo de Bass tradicional para previsão de demanda do produto considerado, o modelo de Bass revisado apresentou melhores

resultados⁷.

Bauckhadge et al. (2013) estudaram um fenômeno da *internet*, os *memes*⁸, através do uso das distribuições estatísticas Weibull, Gompertz e Frechet como variantes de equações de crescimento. Com o objetivo principal de desenvolver modelos capazes de explicar a evolução do interesse coletivo em *memes* de *internet* ao longo do tempo. Os resultados empíricos mostraram quanto maior for o número de usuários de um determinado *meme*, significa menor novidade e menos atrativo se torna.

Conhecer as especificidades do produto, bem como compreender cada um dos estágios do ciclo de vida do produto é importante sobretudo do ponto de vista do *marketing*, pois as estratégias de mercado aplicadas a cada um destes estágios variam consoante as diferentes características do produto em análise. Portanto, a questão colocada é saber que tipo de modelo seria adequado para modelar este tipo de fenômeno.

Buscou-se estudar os vídeos de música da *internet*, dado a sua importância na indústria da música, por ser um dispositivo de *marketing* destinado, principalmente, a promover as vendas de gravações de música, por ser um importante contributo para a comercialização da música popular e dado a ausência de literatura de caráter qualitativo e quantitativo subjacente.

Esta dissertação tem como objetivo principal investigar a dinâmica

⁷DAM (Desvio absoluto médio), REQM (raiz do erro quadrático médio) e EPAM (erro percentual absoluto médio).

⁸São figuras, caricaturais que são acompanhadas por textos curtos, sob a forma de estilo de mensagens curtas, que coloca em pauta questões mundanas como a vida moderna e a cultura.

temporal dos vídeos de música ao longo do tempo. O capítulo seguinte tem como objetivo desenvolver e propor um modelo capaz de explicar a evolução do interesse coletivo em vídeos de música de *internet* ao longo do tempo. O método e a aplicação empírica são desenvolvidos no capítulo três e, tal como Xu e Song (2007), o modelo é estimado por meio de métodos de estimação não lineares. Finalmente, são sumarizadas as conclusões sobre o tema abordado.

2 Modelo Proposto

Neste ponto, será proposto um modelo para o estudo da dinâmica dos vídeos de música da *internet*, que considere a tendência geral ou a forma global do número de buscas por este produto na *internet* ao longo do tempo, como mostrado na figura 3, mas também analisa como esse interesse pode ser relacionado com o ciclo de vida de um produto, ou seja, lançamento, crescimento, maturidade e declínio. Ou seja, pretende-se desenhar um modelo que seja capaz de captar tal comportamento, isto é, como evolui o interesse por este determinado produto ao longo do tempo.

Um modelo simples, com dados de séries temporais, pode ser representado pela seguinte equação, por refletir o comportamento dos agentes econômicos,

$$y_t = \beta_1 + \beta_2 \phi(t; \beta_3; \beta_4) + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + u_t \quad (2.1)$$

Para simplificar escreve-se $\phi(t)$ em lugar de $\phi(t; \beta_3; \beta_4)$, com $t = 1, 2, \dots, T$.

$$y_t = \beta_1 + \beta_2\phi(t) + \alpha_1y_{t-1} + \alpha_2y_{t-2} + \dots + \alpha_py_{t-p} + u_t \quad (2.2)$$

Onde, y_t denota o índice do volume de consultas de usuários do GT (Google Trends) por um determinado vídeo de música, $\phi(\cdot)$ é a função de densidade de probabilidade (fdp) da distribuição normal. Os demais, $\alpha_1y_{t-1} + \alpha_2y_{t-2} + \dots + \alpha_py_{t-p}$ regressores (desfasamentos da variável dependente) e, u_t representa o termo de erro não observado.

A inclusão da função $\phi(\cdot)$, cujos respectivos parâmetros β_2 , $\beta_3 = \mu$ e $\beta_4 = \sigma^2$ são parâmetro de escla, à média e à variância, é justificada para captar o pico apresentado, em geral, nas séries de vídeos de música consideradas nesta dissertação, também está em função do tempo (t). Normalmente, espera-se que os sinais dos respectivos parâmetros β_2 , β_3 e β_4 sejam positivos.

Resolvendo (2.2) com relação à y_t , considerando uma componente autoregressiva AR(1), sem perda de generalidade e escrevendo-a nos polinômios L ,

$$A(L)y_t = \beta_1 + \beta_2\phi(t) + u_t \quad (2.3)$$

$A(L) = 1 - \alpha_1L$ o polinômio autoregressivo de ordem 1,

$$\iff y_t(1 - \alpha_1 L) = \beta_1 + \beta_2 \phi(t) + u_t$$

$$\iff y_t = \frac{\beta_1}{(1 - \alpha_1 L)} + \frac{\beta_2 \phi(t)}{(1 - \alpha_1 L)} + \frac{u_t}{(1 - \alpha_1 L)}$$

Efetutando a inversão do polinômio,

$$\begin{aligned} \iff \frac{\phi(t)}{(1 - \alpha_1 L)} &= \phi(t) + \alpha_1 L \phi(t) + \alpha_1^2 L^2 \phi(t) + \alpha_1^3 L^3 \phi(t) + \dots \\ &= \phi(t) + \alpha_1 \phi(t - 1) + \alpha_1^2 \phi(t - 2) + \alpha_1^3 \phi(t - 3) + \dots \end{aligned} \quad (2.4)$$

Para que esta série geométrica seja convergente, a razão em módulo deve ser menor do que um, $|\alpha_1| < 1$, então a soma da série geométrica existe e é finita. Substituir (2.4) em (2.3), o que é equivalente a reescrever (2.3) como,

$$\iff y_t \approx \beta_1 + \beta_2 \sum_{i=1}^{30} \alpha_1^i \phi(t - i) \quad (2.5)$$

Ou seja, vai somar densidades,

$$i = 0 \implies \alpha_1 \phi(t)$$

$$i = 1 \implies \beta_2 \alpha_1^2 \phi(t - 1)$$

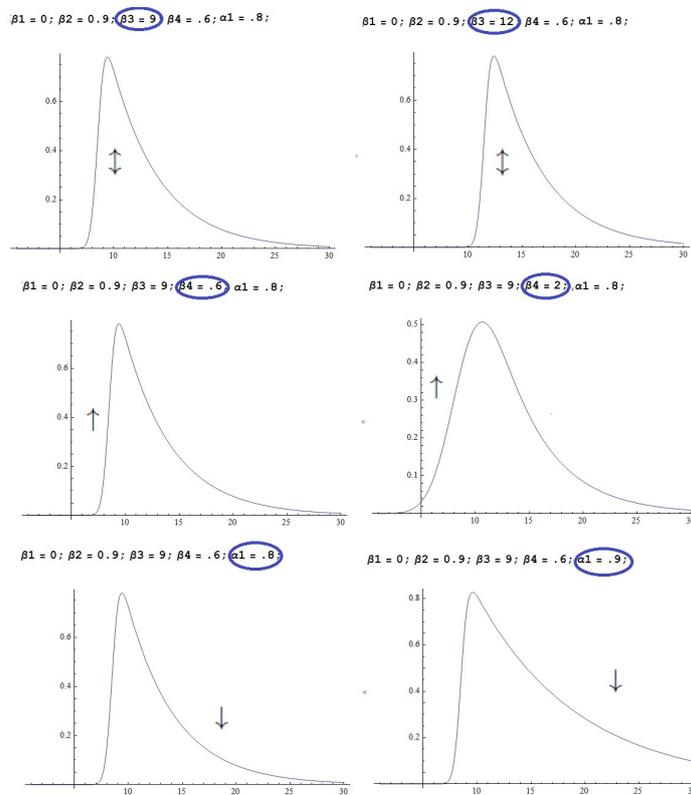
...

$$i = 30 \implies \beta_2 \alpha_1^{30} \phi(t - 30)$$

O que resultará no formato de curva almejado, o que mostra o com-

portamento desta equação 2.5 para diferentes valores dos parâmetros do modelo proposto. Para melhor compreender o funcionamento do modelo proposto, uma simulação foi realizada com o auxílio do *software Mathematica* e foram atribuídos diferentes valores para os respectivos parâmetros.

Figura 4: Simulação gráfica do modelo proposto.



Isto para $i = 0$ até 30, tendo sido escolhido este valor arbitrariamente, a título de exemplo. Como o modelo é não linear em um dos parâmetros, é importante obter bons valores iniciais dos parâmetros, para facilitar a convergência do algoritmo de minimização. Desta forma são atribuídos valores de inicialização aos respectivos parâmetros.

β_2 (parâmetro de escala) quanto maior, maior o volume de acessos

por um determinado vídeo de música, ou seja, mais alta é a curva.

O parâmetro β_3 referente à localização do centro da curva, ou seja, indica o momento no tempo onde ocorre o maior número de acessos. Se o máximo de acessos é atingido ao fim de nove semanas (dados semanais, por exemplo) após o lançamento, ou seja, quando $t = 9$, logo β_3 deve estar próximo deste valor.

Com relação ao parâmetro β_4 , tem haver com a velocidade da procura pelos usuários por um determinado vídeo de música após o lançamento e, também ele tem haver com o período de maturidade. Quando maior, significa que o número de procura na *internet* aumenta lentamente. Por fim, os parâmetros autoregressivos $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_p$, com $\alpha_i \neq 0$, tem haver com a velocidade do decaimento da procura após o período de maturidade, quanto maior mais lento é o decaimento da procura.

As séries analisadas apresentam características similares. Em geral, elas são leptocúrticas, cujo pico é mais agudo, com caudas mais longas e pesadas, tendo maior probabilidade de apresentar valores próximos da média e maior probabilidade de apresentar valores extremos, isto se comparando com a distribuição normal. Também, possui uma assimetria positiva a direita, pois se observa uma maior frequência para valores menores e a cauda mais longa à direita. Assim, majoritariamente a magnitude dos desvios à direita é maior que a magnitude à esquerda.

3 Método e Análise Empírica

3.1 Introdução

O objetivo do presente capítulo é descrever os dados, aspectos metodológicos e analisar empiricamente a aplicabilidade do modelo proposto, no capítulo anterior, com intuito de modelar o interesse por vídeos de música da *internet*, a partir do volume de acessos (y_t), de milhões de usuários da Web no tempo (t).

3.2 Dados

As bases empíricas da análise são séries temporais (dados semanais) para um determinado vídeo de música típico da *internet*, obtidas junto ao site GT, que sumarizam o comportamento de busca de milhões de usuários da Web em todo o mundo. Usadas na análise da plausibilidade do modelo proposto e para entender como a atenção coletiva se dá pelos vídeos de música da *internet* ao longo do tempo.

Com advento da *internet*, a obtenção de informação muitas vezes consiste na busca de fontes *online*. Desde que foi disponibilizada para uso em 2004, a ferramenta estatística GT tem se destacado por fornecer um índice do volume de consultas de usuários do Google, que é baseado no volume total de consultas para o termo de busca, na palavra-chave, numa determinada região geográfica dividida pelo número total de consultas naquela região e num determinado ponto no tempo. Quanto à

frequência dos dados, estes podem ser obtidos semanalmente ou mensalmente. A partir de 2008, tornou-se possível fazer o *download* dos dados em arquivo CVS, também uma representação gráfica é fornecida, onde os números do gráfico indicam quantas pesquisas foram feitas para um termo específico, em relação ao número total de pesquisas feitas no Google ao longo do tempo. Devido à normalização dos dados, estes são apresentados em uma escala de 0 a 100, quando não há dados suficientes, zero é exibido.

Choi e Varian (2009) publicaram o artigo “*Predicting the Present with Google Trends*” com o objetivo principal de familiarizar os leitores com os dados do GT, tendo ilustrado alguns métodos de previsão que usam esses dados, e incentivar os leitores a realizar suas próprias análises. Servindo esses exemplos de linhas de base para ajudar os analistas a começarem a trabalhar com os seus próprios esforços de modelagem e que esta pode, posteriormente, ser refinada para aplicações específicas.

Em duas contribuições recentes, comprovou-se a utilidade e a validade dessa metodologia, isto é, a significância de dados de pesquisa da *internet*, ou seja, a partir do GT. Para Mellon (2011) os dados fornecidos pelo GT fornecem proxies precisas sobre a dinâmica de opiniões públicas. Segundo Preis et al. (2013), os dados obtidos a partir do GT são úteis para prever o movimento dos preços diários do *Dow Jones industrial average*. Sendo seus resultados consistentes com a sugestão de que os dados do GT não apenas refletem aspectos do estado atual da eco-

nomia, mas também podem fornecer algumas dicas sobre as tendências futuras no comportamento dos agentes econômicos.

No que diz respeito a esta dissertação, estas descobertas sugerem que os dados do GT, que agregam informações sobre a atividade de milhões de usuários, são de facto indicativos do interesse coletivo e da atenção e, assim formam *proxies* adequados para as variáveis não observáveis, para a pesquisa sobre a dinâmica dos vídeos de música da *internet*.

3.3 Especificação e Estimação dos Modelos

A abordagem utilizada baseia-se na teoria assintótica com regressores pré-determinados. Com relação à estacionaridade das séries estudadas, pode-se dizer de maneira preliminar que os processos de geração de dados, subjacente às séries em estudo, são os TSP (*Trend Stationary Process*). Portanto, o processo $\{y_t\}$ não é estacionário precisamente porque apresenta uma tendência, ou seja, a média é função do tempo. Para que processos deste tipo sejam transformados em estacionários deve-se remover a parte da tendência determinística da média. Também, não são $I(1)$ por estas séries serem limitadas entre 0 e 100 e no limite tendem para zero, isto é, $y_t \rightarrow 0$ quando $t \rightarrow \infty$.

Para os propósitos do objetivo, com um modelo linear nas variáveis e não linear em um dos parâmetros, os parâmetros serão estimados utilizando o método de estimação NLS, o qual desempenha papel similar ao OLS (*Ordinary Least Squares*), para o caso linear. Para modelos não

lineares, porém com o mesmo número de parâmetros e observações, é sempre mais difícil à estimação pelo método NLS.

$$\mathbf{y} = \mathbf{x}(\boldsymbol{\beta}) + \mathbf{u} \quad (3.1)$$

Onde \mathbf{y} e $\mathbf{u} \sim IID(\mathbf{0}, \sigma^2 \mathbf{I})$ são vectores de dimensão T com elementos típicos y_t e u_t , respectivamente, e $\mathbf{x}(\boldsymbol{\beta})$ é um vector de dimensão T cujo t -ésimo elemento é $x_t(\boldsymbol{\beta})$. Assim, $\mathbf{x}(\boldsymbol{\beta})$ é o análogo não linear do $\mathbf{X}(\boldsymbol{\beta})$ do caso linear. $\boldsymbol{\beta}$ é o vector de parâmetros de dimensão k a ser estimado (Davidson e Mackinnon, 1999).

Não há uma solução fechada para o problema não linear, em vez disso é preciso recorrer à utilização de um algoritmo para minimização da SQR (soma dos quadrados dos resíduos) com respeito ao vector dos parâmetros ($\boldsymbol{\beta}$), para encontrar as estimativas NLS, onde, em sua grande maioria, os algoritmos operam essencialmente da mesma forma. Sendo importante obter bons valores iniciais (*initial guess*) dos parâmetros, na busca para facilitar a convergência do algoritmo, os quais são aqueles que minimizam a SQR (Davidson e Mackinnon, 1999).

Em economia, frequentemente, para modela-se adequadamente os fenômenos, são incluídas na mesma equação variáveis referidas a períodos diferentes de tempo. Estes são os chamados modelos dinâmicos, com o propósito de tomar em consideração os defasamentos nas relações entre variáveis. Seja pela persistência dos hábitos dos indivíduos

ou porque a economia não se encontra em estado de equilíbrio, mas pelo contrário em cada momento a economia está ajustando-se. Seja devido às restrições tecnológicas, por exemplo, porque os ajustamentos não são feitos imediatamente quando há alterações, portanto esta dificuldade dos agentes adaptarem-se às novas situações demora algum tempo, entre outros.

Neste sentido, fala-se em modelo dinamicamente completo, significando que toda informação extra, que venha a adicionar ao conjunto de informação, será irrelevante, ou seja, já foram incluídos defasamentos suficientes, regressores adicionais não explicam a variável dependente. Por outro lado, se o modelo estimado é dinamicamente completo, corretamente especificado, então os erros são não autocorrelacionados, significa que vai haver uma maneira indireta de analisar a dinâmica do modelo, ou seja, usando teste de autocorrelação para analisar indiretamente se o modelo é dinamicamente completo. Os quais por sua vez são baseados na hipótese de que os termos de erro são homoscedásticos. Quando esta hipótese crucial é violada, as distribuições assintóticas das estatísticas de teste deixam de ser confiáveis, pondo em causa toda a inferência estatística.

Em modelos estimados com dados de séries temporais frequentemente são encontradas autocorrelação nos erros, portanto se o modelo estimado apresenta fortes sintomas de autocorrelação residual é possível tomar isto como sendo indícios de modelo não dinamicamente completo e esta auto-

correlação residual ser a parte visível de uma má especificação dinâmica do modelo, de ter uma dinâmica insuficiente para explicar o comportamento da variável dependente. O modelo é sujeito a uma bateria de testes para detecção de erros de especificação, para saber se é congruente com os dados, no caso de encontrar sintomas de autocorrelação dos erros do modelo, o que se deve fazer é aumentar a especificação dinâmica do modelo, com inclusão de defasamentos da variável dependente como regressores e não fazer correções de autocorrelação propriamente dita. Porém, convém ver o comportamento do ajustamento do modelo para que não ocorra uma deterioração deste.

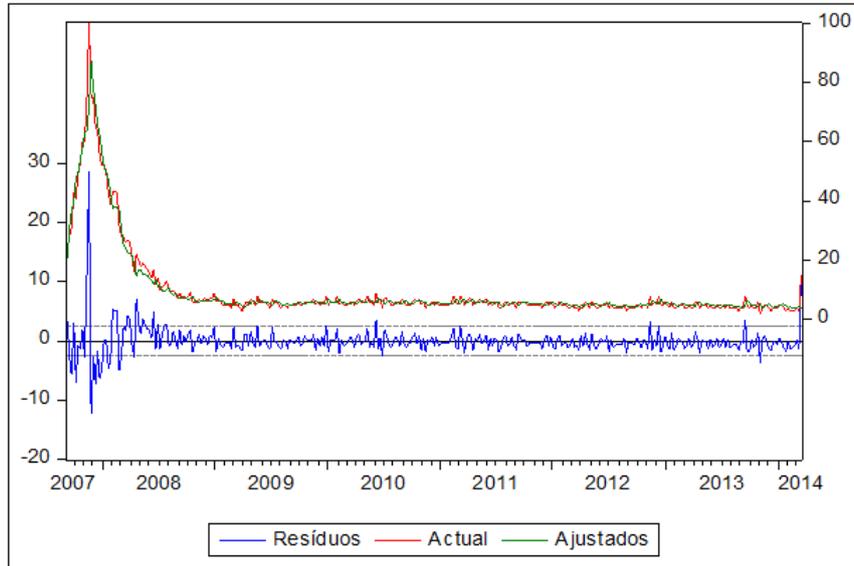
Basta que exista heteroscedasticidade condicional para que a inferência estatística seja invalidada, o mesmo se aplica ao caso de presença de autocorrelação. Neste trabalho, testa-se a presença de autocorrelação dos erros usando um teste robusto a heteroscedasticidade e depois testa-se a presença de heteroscedasticidade (Wooldridge, 2009).

Quando são encontradas evidências estatísticas de presença de heteroscedasticidade condicional é possível escolher a opção de erros padrão robustos a heteroscedasticidade ou modelar a heteroscedasticidade pelo GNLS (*Generalized Nonlinear Least Squares*).

Por ter-se observado previamente uma grande variação nos respectivos resíduos, indícios de um padrão pronunciado de heteroscedasticidade observado informalmente, Figura 5, optou-se, primeiramente, por estimar o modelo pelo método NLS com respectivos SE^* (*standard errors*)

robustos a heteroscedasticidade.

Figura 5: Resíduos NLS - volume de consultas para o vídeo de música *No One* (Alicia Keys, 2007).



E de forma a eliminar, respectivamente, os indícios de autocorrelação, foram incluídos dois desfasamentos no modelo estático, o que resultou em um modelo supostamente dinamicamente completo.

$$y_t = \beta_1 + \beta_2\phi(t) + \alpha_1y_{t-1} + \alpha_2y_{t-2} + u_t \quad (3.2)$$

Tendo sido complementados com testes de autocorrelação de ordem 1, com recurso à estatística de teste de Breusch-Godfrey. Na Tabela 1, foram compilados os resultados da estimação do modelo dinâmico, para o estimador NLS.

Informalmente, é possível fazer um teste em directo de heteroscedasticidade baseado na seguinte lógica, se a heteroscedasticidade for forte

Tabela 1: Modelo NLS estimado para o vídeo de música *No One* (Alicia Keys, 2007).

Coeficientes estimados	NLS
β_1	1.624*** (0.292) [0.486]
β_2	47.865*** (5.956) [11.267]
β_3	10.932*** (1.191) [2.240]
β_4	3.058*** (0.116) [0.339]
α_1	0.534** (0.054) [0.219]
α_2	0.186 (0.054) [0.169]
<i>Observações</i>	341
R^2	0.964

Erros-padrão robustos a heteroscedasticidade (SE*) em parênteses retos.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

então as correcções aos SE usuais há de ser grandes, ou seja, comparar os dois tipos de SE e se as diferenças forem muito grandes, devem-se usar estas grandes diferenças como sintomas adicionais da presença de heteroscedasticidade. Que é o que se observa analisando os erros padrão usuais e os SE*, da estimação NLS do modelo, Tabela 1.

Caso existam fortes sintomas de heteroscedasticidade, se a perda de eficiência for grande, convém considerar outro método de estimação onde essa heteroscedasticidade seja acomodada, estimar pelo GNLS para tentar melhorar a eficiência de estimação. Neste contexto, a matriz de

covariâncias dos erros deixa de ser a usual. Como consequência desta última implicação, a inferência usual é inválida na presença de heteroscedasticidade. Então, o método GNLS visa corrigir estes problemas.

Uma hipótese bastante forte é imposta, que é a da matriz de covariâncias dos erros (Ω) ser conhecida. Neste contexto, a transformação consiste em dividir a variável dependente e cada um dos regressores do modelo pelos desvios padrão (w_t) das respectivas observações.

$$\frac{y_t}{w_t} = \frac{\beta_1}{w_t} + \frac{\beta_2\phi(t)}{w_t} + \frac{\alpha_1 y_{t-1}}{w_t} + \frac{\alpha_2 y_{t-2}}{w_t} + \frac{u_t}{w_t} \quad (3.3)$$

Ou seja, esta transformação permite que os erros do modelo sejam não heteroscedásticos e aplicar o NLS ao modelo transformado, é obter o estimador GNLS. O método consiste em minimizar a SQR do modelo transformado, na prática, o que vai acontecer é que as observações com maior desvio padrão vão pesar menos na soma dos quadrados dos resíduos e as observações com menor desvio padrão vão ter uma importância maior e isto faz sentido, pois se estas observações têm maior variância são menos credíveis, portanto, deve-se dar menor importância. Na prática, para que o estimador GNLS seja factível, é preciso assumir uma forma funcional para a heteroscedasticidade e esta forma que se assume leva a estimar a matriz Ω , não conhecida, mas, quando estimada, passa a ser aleatória e o estimador passa a ser chamado FGNLS (*Feasible Generalized Nonlinear Least Squares*), cujas propriedades em pequenas

amostras são desconhecidas, sendo apenas possível a análise em termos assintóticos.

Desta forma, alternativamente à estimação pelo método NLS, foi aplicado à metodologia citada no parágrafo anterior, os referidos resultados encontram-se na Tabela 2. Como visto, na Figura 5, a estimação pelo método NLS é mais imprecisa quando se chega ao pico das observações (quando é maior o volume de acessos, isto é, o interesse por um determinado vídeo atinge o máximo), porque a variância dos resíduos é maior e apresenta maior variabilidade, podendo ser dito que esta variabilidade é uma função do nível de interesse ou do nível de interesse no período anterior.

A primeira constatação que se pode chegar é que a variância condicional é uma função de y_{t-1} , do gênero $Var(y_t|\mathfrak{S}_{t-1}) \propto f(y_{t-1})$, ou seja, a variância desta série, y_t , dado os valores passados, \mathfrak{S}_{t-1} , poderá ser proporcional a uma função de y_{t-1} , onde esta função é desconhecida e tem a propriedade $f : \mathbb{R} \rightarrow \mathbb{R}^+$, que é uma função estritamente positiva.

A questão é saber como essa variabilidade responde a y_t .

1. $Var(y_t|\mathfrak{S}_{t-1}) \propto y_{t-1}^2$, então $\sqrt{Var(y_t|\mathfrak{S}_{t-1})} \propto y_{t-1}$;
2. $Var(y_t|\mathfrak{S}_{t-1}) \propto y_{t-1}$, então $\sqrt{Var(y_t|\mathfrak{S}_{t-1})} \propto \sqrt{y_{t-1}}$;
3. Ou supor que o padrão de heteroscedasticidade seja do tipo exponencial, $\sigma_t^2 = E[u_t^2|\mathbf{z}_t(\gamma)] = \exp(\mathbf{z}_t'(\gamma))$, onde algumas das variáveis que estão em X podem pertencer a Z e $Z = X$, mas não

necessariamente.

A forma funcional, da alínea (3), exponencial é muito utilizada para modelar a heteroscedasticidade condicional. O que será modelado é uma variância e a função exponencial só assume valores positivos para quaisquer valores que estejam no vector \mathbf{z}'_t e em γ . O método requer os seguintes passos: estimação do modelo original com o método NLS com o objetivo de obter os resíduos (e_t), aproximar os erros (u_t) com os resíduos, ou seja, os quadrados dos resíduos (e_t^2) vão servir para aproximar os quadrados dos erros (u_t^2). A ideia é considerar que as variâncias condicionais seguem uma função exponencial, ou seja, as variâncias são os segundos momentos condicionais, ora se os segundos momentos condicionais seguem uma função exponencial, então, as próprias variáveis (os próprios quadrados dos erros), por analogia com os modelos de regressão, são dadas pela função mais um erro. Isto é,

$$E [u_t^2 | \mathbf{z}_t(\gamma)] = \exp(\mathbf{z}'_t(\gamma)) \iff u_t^2 = \exp(\mathbf{z}'_t(\gamma)) \exp(\text{erro}) \quad (3.4)$$

Próximo passo é a linearização dessa função, logo passar aos logaritmos,

$$\log(u_t^2) = \mathbf{z}'_t(\gamma) + \text{erro}$$

Como pode haver uma diferença entre os erros e os resíduos, a re-

gressão auxiliar estimada é para encontrar as estimativas NLS $\hat{\gamma}$,

$$(\approx)\log(e_t^2) = \mathbf{z}'_t(\gamma) + erro^*$$

Onde essas estimativas são usadas para calcular,

$$\hat{w}_t = exp(\mathbf{z}'_t(\hat{\gamma}))^{1/2}, t = 1, 2, \dots, T \quad (3.5)$$

As variáveis originais são transformadas com as estimativas dos desvios padrão (\hat{w}), por fim, o estimador FGNLS é obtido usando NLS para estimar a equação em (3.3), substituindo o (w) desconhecido (Davidson e MacKinnon, 1999). Este é apenas um exemplo, portanto não é verdade em geral.

O método FGNLS foi usado para estimar os parâmetros do modelo não linear e como não há uma solução de forma fechada para este caso, são usados algoritmos numéricos para encontrar o valor dos parâmetros que minimizam os resíduos de forma iterativa, isto é, os valores são obtidos por aproximação sucessiva (Davidson e MacKinnon, 1999). Os resultados, resumidos na Tabela 2, foram obtidos utilizando FGNLS e considerando o padrão de heteroscedasticidade da alínea (1).

Tabela 2: Modelo FGNLS estimado para o vídeo de música *No One* (Alicia Keys, 2007).

Coeficientes estimados	FGNLS
β_1	2.528*** (0.363)
β_2	51.444*** (7.568)
β_3	10.306** (4.755)
β_4	4.215*** (0.350)
α_1	0.246*** (0.072)
α_2	0.280*** (0.071)
<i>Observações</i>	341
R^2	0.325
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$	

3.4 Discussão dos Resultados

Nesta seção, irá analisar-se a validade de todas as estimativas mostradas na seção anterior e seguir-se-á a mesma sequência.

Para o caso do modelo estimado pelo método NLS, os resultados dos testes de especificação são os testes para detecção da presença de autocorrelação e de heteroscedasticidade condicional. A inclusão de dois defasamentos ao modelo original estático acabou por mitigar os indícios de autocorrelação residual. De acordo com o teste Breusch-Godfrey ($BG_{(1)} = 0.070, (0.079)$), não foram encontrados indícios da presença de autocorrelação residual de ordem 1, considerando os usuais níveis de significância estatística (1%, 5% e 10%), não se rejeita a hipótese nula

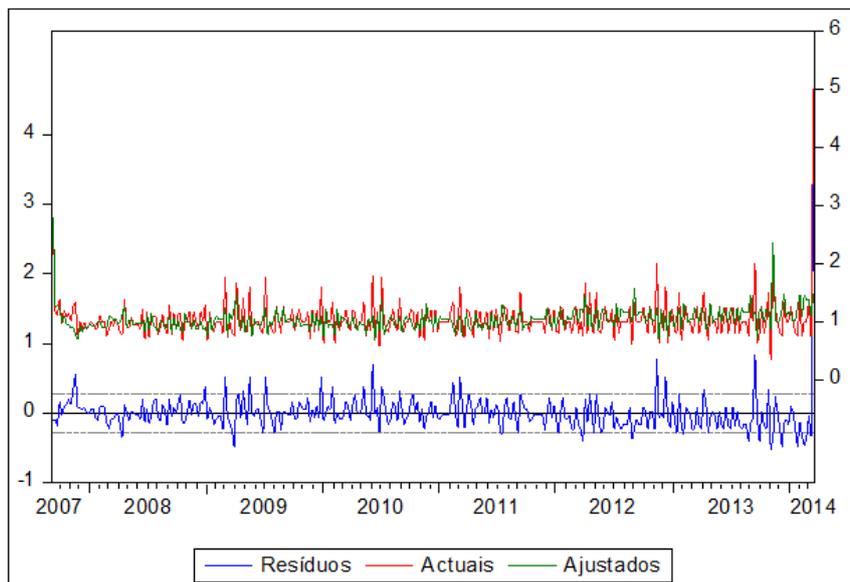
de ausência de autocorrelação residual. Confirmando que os sintomas de autocorrelação, detectados no modelo estático, estavam a refletir o problema de uma má especificação do modelo, pois bastou incluir dois defasamentos da variável dependente como regressores para que as autocorrelações desaparecessem.

Num segundo momento, tendo sido controlado o problema de especificação do modelo, testou-se a presença de heteroscedasticidade através dos testes ARCH (*Autoregressive Conditional Heteroscedasticity*) (Engle, 1982) e de Breusch-Pagan-Godfrey (Breusch-Pagan, 1979, e Godfrey, 1978), por ser visto em geral como um teste mais específico, baseando-se em informação que esteja disponível sobre as variáveis que podem explicar a variabilidade dos segundos momentos condicionais. Dir-se-ia que é um teste para uma situação em que se suspeita que haja heteroscedasticidade e existindo, se suspeita que ela seja explicada ou relacionada com uma determinada variável. Em qualquer que seja a versão do teste, respectiva estatística $\chi_{(7)} = 36.469(0.000)$ para o teste ARCH, a hipótese nula é fortemente rejeitada com base nos valores-p das estatísticas de teste e são fornecidas evidências empíricas da presença de heteroscedasticidade nos resíduos, considerando um nível de significância de 5%. Com base na estatística do teste de Breusch-Pagan-Godfrey, $LM = 9.960(0.000)$, também se rejeita fortemente a hipótese nula de variância homoscedástica. Nesse contexto o estimador NLS continua consistente, porém deixa de ser o mais eficiente.

Frente aos fortes indícios de heteroscedasticidade e buscando a melhoria da eficiência da estimação, foram considerados três diferentes padrões de heteroscedasticidade na estimação do modelo, capítulo 2, pelo método FGNLS, considerados nas alíneas (1), (2) e (3). A avaliação do modelo foi realizada através da análise empírica à estimação do modelo e de testes de especificação. Os testes de especificação utilizados neste tipo de modelo são o teste de ausência de autocorrelação e o de heteroscedasticidade, citados anteriormente.

Dentre os modelos estimados, o que considerou o padrão descrito na alínea (1), apresentou melhor resultado, no sentido de não violar a hipótese de ausência de autocorrelação ($BG_{(1)} = 2.137(0.143)$) e heteroscedasticidade. A estatística LM de Breusch-Pagan-Godfrey é 4.697, sendo significativa para todos os usuais níveis de significância estatística, dado que, com base no valor-p (0.195) da estatística de teste, não foram encontradas evidências estatísticas contra a hipótese nula de ausência de heteroscedasticidade. Ainda, os resíduos podem ser visto graficamente na figura 6.

Figura 6: Resíduos FGNLS - volume de consultas para o vídeo de música *No One* (Alicia Keys, 2007).



O modelo estimado, considerando o padrão de heteroscedasticidade descrito na alínea (2), não apresentou indícios de autocorrelação residual ($BG_{(1)} = 0.500(0.479)$), quando considerados os usuais níveis de significância estatísticos. Quando aplicados os testes de Breusch-Pagan-Godfrey e ARCH, qualquer que seja a versão do teste a hipótese nula de homoscedasticidade condicional é claramente rejeitada, portanto os testes fornecem evidências empíricas fortes de rejeição da homoscedasticidade. Quando considerado o padrão de heteroscedasticidade descrito na alínea (3), o recurso da inclusão de defasamentos da variável dependente como regressor não conseguiu eliminar, respectivamente, os indícios de autocorrelação residual do modelo.

É possível comparar os resultados obtidos via estimação NLS e FGNLS,

Tabelas 1 e 2 respectivamente. Com relação aos coeficientes estimados dos regressores, estes mudaram bastante, com exceção do coeficiente $\hat{\beta}_3$ que quase não apresentou mudança. Comparando os erros padrão em ambos os modelos, percebe-se que há uma melhoria, em geral, para o caso FGNLS, contudo o SE associado variável determinística parece ter menos precisão, ou seja, maior do que o SE* obtido a partir da estimação NLS relativamente.

Com relação aos métodos de estimação considerados, estudos de simulação são referidos (Wooldridge, 2009) e dizem que se existir heteroscedasticidade condicional forte, mesmo que o padrão de heteroscedasticidade condicional considerado seja incorrecto, é melhor usar o método FGNLS a ignorar e usar o NLS.

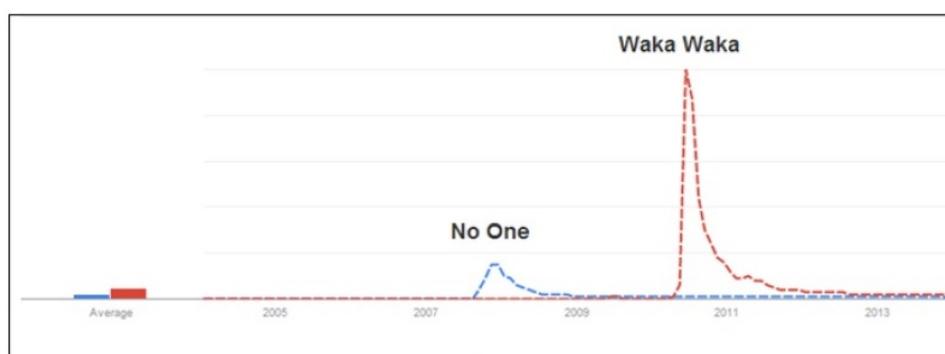
Este é o caso do modelo analisado neste trabalho, onde no caso NLS, tanto as estatísticas de Breusch-Godfrey-Pagan e ARCH rejeitam muito claramente a hipótese nula e apresentam fortes indícios de heteroscedasticidade condicional. Portanto, nestes casos, é melhor usar o FGNLS e não ignorar o problema, mesmo que a forma de heteroscedasticidade condicional não seja a correta.

3.5 Análise dos parâmetros β_2 , β_3 , β_4 e α_i para diferentes vídeos de música

A análise apresentada concentra-se nos modelos dinâmicos, mas especificamente nas estimações a partir do FGNLS e para facilitar a comparação

dos parâmetros estimados, são apresentados dois conjuntos diferentes de gráficos na figura 7, como representantes de diferentes vídeos de música (típicos), dado sua popularidade na *internet*. Respectivamente, são referentes às músicas *No One* (Alicia Keys, 2007) e *Waka waka* (Shakira, 2010).

Figura 7: Dois diferentes vídeos de músicas, populares na *internet*.



Fonte: www.google.com.br/trends (Acedido em 26/04/2014).

Primeiramente, a interpretação dos sinais das estimativas. Naturalmente, se a estimativa não apresenta o sinal esperado pela teoria econômica, isso deve ser tomado como uma indicação de não confiabilidade dos resultados da estimação. De acordo com o apresentado, no capítulo 2, espera-se que os sinais dos respectivos parâmetros β_2 , β_3 e β_4 sejam positivos.

Depois de estimados os modelos dinâmicos (FGNLS, padrão de heteroscedasticidade da alínea 1), analisada a especificação, considerando ambas as séries temporais expostas graficamente na figura 7, as comparações entre os parâmetros estimados podem ser realizadas a partir dos resultados explicitados na Tabela 3.

Tabela 3: Modelo FGNLS estimado para os vídeos de música *No One* (Alicia Keys, 2007) e *Waka waka* (Shakira, 2010).

Coeficientes estimados	<i>No One</i> (2007)	<i>Waka waka</i> (2010)
β_1	2.528*** (0.363)	0.076*** (0.024)
β_2	51.444*** (7.568)	192.440*** (0.072)
β_3	10.306** (4.755)	8.465*** (0.024)
β_4	4.215*** (0.350)	0.799*** (0.065)
α_1	0.246*** (0.072)	0.584*** (0.063)
α_2	0.280*** (0.071)	0.353*** (0.060)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

O β_2 é um factor de escala e quanto maior, mais alta é a curva, o que leva a concluir que em média o volume de pesquisa ao longo do tempo, em valor absoluto ou total, pelo vídeo de música *Waka waka* é superior, se comparado ao longo do tempo pelo vídeo de música *No One*.

De acordo com o parâmetro β_3 estimado, estima-se que o auge de acessos ao vídeo de música *No One* deve acontecer em média ao final da décima semana, após ter ocorrido o lançamento (10/09/2007). Enquanto que para o vídeo de música *Waka waka*, estima-se que o auge de acessos ao respectivo vídeo de música acontece em média oito semanas e meia após o respectivo lançamento (07/06/2010).

Com relação ao parâmetro estimado β_4 , quanto mais alto o seu valor, significa que o número de procura na *internet* aumenta lentamente. O que leva a concluir que em média o número de buscas da *internet*

pelo o vídeo de música *Waka waka* acontece com uma velocidade muito superior ao que se observa pelo vídeo de música *No One*.

Os resultados baseados no α_i estimado, indicam uma maior permanência de interesse (após o auge de acessos) pelo vídeo de música *Waka waka* do que pelo vídeo de música *No One*.

4 Conclusões

O mercado da música mundial continua a se expandir em novos mercados e a criar novos negócios, atraindo cada vez mais usuários para os serviços de música sob o formato digital. Segundo o *Digital Music Report* (2014), a receita gerada pela indústria da música digital apresentou um crescimento de 4,3% de 2012 para 2013.

Os vídeos de música se tornaram parte integrante da moderna cultura da *internet*. Dado sua popularidade, diversidade, proliferação e importância para o mercado mundial da música, por outro lado a pesquisa deste tipo de produto é ainda surpreendentemente escassa. Em particular, modelos plausíveis que explicariam sua dinâmica temporal anteriormente não foram relatados na literatura subjacente.

Nesse trabalho, questionou-se sobre a dinâmica deste tipo de produto e sobre o ciclo de vida destes tipos de séries temporais, a fim de caracterizar como a atenção coletiva, para um determinado vídeo de música, evolui ao longo do tempo, tendo sido lançada uma proposta de modelo no capítulo 2. As séries temporais consideradas, para análise empírica do

modelo, foram obtidas junto ao GT, tendo sido considerados os métodos de estimação (dinâmica) NLS e FGNLS.

A abordagem proposta nesta dissertação permite analisar a velocidade com que os usuários buscam o vídeo de música após o lançamento, indicar o momento no tempo onde ocorre o máximo de acessos por um determinado vídeo de música. Finalmente, permite distinguir os vídeos de música de *internet* de curta duração de outros mais duradores. O que é importante sobretudo do ponto de vista do *marketing*, pois em cada um dos estágios (lançamento, crescimento, maturidade e declínio) as estratégias de mercado variam.

O modelo proposto é simples, interpretável e aplicável. Um possível caminho de investigação adicional consiste em utilizar o modelo proposto para além dos vídeos de música da *internet* e testar a aplicabilidade para outros tipos de produtos de ciclo de vida de curta duração, como por exemplo, para analisar a dinâmica temporal das tendências de produtos tecnológicos. Seria também interessante, frente aos fortes indícios de heteroscedasticidade detectados pelos testes considerados neste trabalho, realizar um estudo de simulação como forma de corroborar os resultados obtidos com a estimação NLS e FGNLS.

Referências

- [1] Amemiya, T. (1985). *Advanced Econometrics*, Cambridge: Harvard University Press, Mass. Bauckhage, B. Kersting, K. e Hadiji, F. (2013). Mathematical Models of Fads Explain the Temporal Dynamics of Internet Memes. ICWSM, The AAAI Press.
- [2] Bauckhage, Christian., Kersting, Kristian., and Hadiji, Fabian. (2013). *Mathematical Models of Fads Explain the Temporal Dynamics of Internet Memes*. Proc. AAAI Int. Conf. on Weblogs and Social Media.
- [3] Belville, H. Jr. (1966). *The Product Life Cycle Theory Applied to Color Television*. Master thesis. New York City: New York University.
- [4] Choi, H., and Varian, H. (2012). *Predicting the Present with Google Trends*. Economic Record 88(S1):2–9.
- [5] Cox, Jr. W. (1967). *Product Life Cycles as Marketing Models*. J. Bus. 40 (October, 1967): 375-384.
- [6] Davidson, R. e MacKinnon, J. G. (1999). *Estimation and Inference in Econometrics*, New York: Oxford University Press.
- [7] Davidson, R. e MacKinnon, J. G. (2004). *Econometric Theory and Methods*, New York: Oxford University Press.

- [8] Dean, J. (1950). *Pricing Policies for New Products*. Harvard Bus. Rev. 28 (November-December, 1950): 45-53.
- [9] Digital Music Report (2014). United Kingdom. Disponível em: <http://www.ifpi.org/downloads/Digital-Music-Report-2014.pdf>. [Acesso em: 2014/7/6].
- [10] Enders, W. (2010). *Applied Econometric Time Series*, 3^a Ed. New York: John Wiley & Sons.
- [11] Forrester, J. W. (2013). *Industrial Dynamics*. Vol. 14, No. 7, Theory Series (Mar., 1968), pp. 398-415. JSTOR. Disponível em: <http://www.jstor.org/stable/2628888>. [Acesso em: 2014/3/8].
- [12] Grant, R. M. (2012). *Contemporary Strategy Analysis: Text and Cases*. 8^a Ed. John Wiley and Sons.
- [13] Greene, W. H. (2012). *Econometric Analysis*, 11^a Ed. New York: Pearson.
- [14] Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton.
- [15] Hansen, B. (2012). *Econometrics*. Disponível em: www.ssc.wisc.edu/bhansen.
- [16] Hayashi, F. (2000). *Econometrics*, Princeton University Press, Princeton.

- [17] Kotler, P. and Keller. K.L. (2012). *Marketing Management*, 14^a Ed. Upper Saddle River, N.J. Pearson Prentice Hall.
- [18] Johnston. J. and DiNardo, J. E. (1997). *Econometric Methods*, 4^a Ed. McGraw-Hill.
- [19] Levitt, T. (1965). *Exploit the Product Life Cycle*, HARV. BUS. REV., Nov.–Dec. 1965, at 81
- [20] Lopes, A. S. (2009). *Estimação OLS do modelo de regressão linear com séries temporais*, mimeo, ISEG-UTL.
- [21] Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer.
- [22] Meyerson, R., and Katz, E. (1957). *Notes on a Natural History of Fads*. American J. of Sociology 62(6):594–601.
- [23] Mellon, J. (2011). *Search Indices and Issue Salience: the Properties of Google Trends as a Measure of Issue Salience*. Sociology Working Papers 2011-01, University of Oxford.
- [24] Wasson, C.R. (1968). *How Predictable Are Fashion and Other Product Life Cycles?* *Journal of Marketing*. Vol. 32, No. 3 (Jul., 1968), pp. 36-43. JSTOR. Disponível em: <http://www.jstor.org/stable/1249760>. [Acesso em: 2014/3/11].
- [25] Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge Ma.

- [26] Wooldridge, J. M. (2006). *Introductory Econometrics, a Modern Approach*, 3^a Ed. Thomson South-Western, Mason.
- [27] Polli, R. and Cook, V. (1969). *Validity of the Product Life Cycle*. J. Bus. 42 (October, 1969): 385-400. JSTOR. Disponível em: <http://www.jstor.org/stable/2351877>. [Acesso em: 2014/3/6].
- [28] Preis, T, Moat, H. S. & Stanley, H. E. (2013). *Quantifying Trading Behavior in Financial Markets Using Google Trends*. Scientific reports, vol.3.
- [29] Teevan, J. Liebling, D. and Geetha, G. (2011). *Understanding and Predicting Personal Navigation*. In Proc. WSDM.
- [30] Xu, X. e Song, Q. (2007). *Forecasting for products with shot cycle based on improved Bass model*. 19^a International Conference on Production Research, China.