Universidade de São Paulo

Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Departamento de Economia

Programa de Pós-Graduação em Economia Área: Economia Aplicada

ANA CAROLINA SANTANA MINIOLI

Modelagem de curvas de juros usando amostragem de frequências mistas

Orietador: Prof. Dr. Márcio Poletti Laurini

Ribeirão Preto - SP 2014

Prof. Dr. Marco Antonio Zago Reitor da Universidade de São Paulo

Prof. Dr. Sigismundo Bialoskorski Neto Diretor da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Prof. Dr. Sérgio Kannebley Júnior Chefe do Departamento de Economia

Profa. Dra. Elaine Toldo Pazello Coordenadora do Programa de Mestrado em Economia Aplicada

ANA CAROLINA SANTANA MINIOLI

Modelagem de curvas de juros usando amostragem de frequências mistas

Dissertação apresentada ao Programa de Pós-Graduação em Economia - Área: Economia Aplicada da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Mestre em Ciências. Versão corrigida. A original encontra-se disponível na FEA-RP/USP

Orietador: Prof. Dr. Márcio Poletti Laurini

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRA-BALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

FICHA CATALOGRÁFICA

Minioli, Ana Carolina Santana

Modelagem de curvas de juros usando amostragem de frequências mistas.

Ribeirão Preto, 2014.

86 p.: il.; 30cm

Dissertação de Mestrado apresentada à Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo.

Orientador: Laurini, Márcio Poletti

1. Modelo dinâmico Nelson-Siegel. 2. Inferência Bayesiana. 3. MIDAS.

FOLHA DE APROVAÇÃO

Nome: MINIOLI, Ana Carolina Santana	
Título: Modelagem de curvas de juros usa	ando amostragem de frequências mistas
	Dissertação apresentada ao Programa de Pós-Graduação em Economia - Área: Economia Aplicada da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.
Aprovada em:	
BANCA E	XAMINADORA
Prof. Dr	
Instituição:	
Prof. Dr	
Instituição:	Assinatura:

Instituição: _____ Assinatura: ____

Agradecimentos

Quando decidi fazer mestrado em Economia, muitos me perguntavam se meu objetivo final seria dar aula, pois, infelizmente, essa é a primeira associação que se faz a esse tipo de curso de pós-graduação. Uma pena, pois poucos sabem o que esse período de estudos pode proporcionar. Momentos de aprendizado que vão muito além da teoria econômica e seus modelos sofisticados, nada triviais; aprendizados que levarei comigo para o resto da vida, em que pude conhecer um pouquinho mais do que somos capazes de fazer quando se tem muito esforço e dedicação focados em único objetivo: aprender vivendo (ou vivendo para aprender, como preferirem). E o melhor de tudo, por mais que sejam momentos dificílimos, é saber que sempre haverá pessoas em quem podemos nos apoiar. É por essas pessoas que dedico parte do meu tempo buscando palavras de carinho e de reconhecimento, já que, sem dúvida alguma, sem elas nada disso seria concluído.

Primeiramente, gostaria de agradecer meu orientador, Márcio Laurini, com quem muito aprendi durante o desenvolvimento deste trabalho. Agradeço pela paciência a me ensinar, principalmente naqueles momentos em que eu aparecia em sua sala para tirar dúvidas com cara de quem pensa "Socorro, não estou entendendo nada!". Rs. É por isso que eu falo: Professor, muito obrigada! Sou super grata pela ajuda e orientação neste último ano e meio.

Dizem que a escola ensina e os pais educam! Com toda certeza, comigo foi exatamente assim. Portanto, um agradecimento mais que especial aos meus pais, Graça e José, pois sempre estiveram ali, guiando-me e mostrando quais caminhos eu poderia escolher ao invés de, simplesmente, falarem qual deveria ser escolhido. Obrigada por cada momento dedicado à minha educação. Sou quem sou hoje pois vocês me fizeram assim. Agradeço meus irmãos, Ana e Guilherme, por estarem ao meu lado em momentos bons e outros não tão bons assim; foi com eles que aprendi a dividir e compartilhar desde pequena.

Laurita, Roberts e Déniel, três pessoas super queridas que fizeram a diferença nessa jornada. Nosso primeiro ano juntos foi um ano intenso, de segunda a segunda, dentro daquela salinha de estudos. Reconheço que meu mestrado não seria o mesmo sem vocês ali comigo! Ao Paulo, um muito obrigada por nos acompanhar de vez em quando! Aos meus amigos que, de longe, sempre torceram por mim; em especial, à Karina e à Laís.

A todos os professores da Economia da FEARP que fizeram parte desta minha nova formação.

A CAPES pela bolsa de mestrado.

Sumário

1	Intr	dução	19
2	Rev	ão de Literatura	23
	2.1	Modelo dinâmico de Nelson-Siegel	23
		2.1.1 Fatos estilizados	25
		2.1.2 Relação entre os fatores latentes da curva de rendimentos com os	
		componentes principais	26
		2.1.3 Taxa de Juros e Dados Macroeconômicos	27
	2.2	Volatilidade Estocástica	29
	2.3	Inferência Bayesiana	30
	2.4	MIDAS - Mixed Data Sampling	32
3	Me	dologia	35
	3.1	Modelo Dinâmico Nelson-Siegel - DNS	35
	3.2	Volatilidade Estocástica	37
	3.3	Inferência Bayesiana	39
		3.3.1 Introdução à inferência bayesiana	39
		3.3.2 MCMC - Markov Chain Monte Carlo	43
		3.3.2.1 A Cadeia de Markov	43
		3.3.2.2 O método MCMC	45
	3.4	MIDAS - Mixed Data Sampling	50
	3.5	Descrição do modelo	52
		3.5.1 O modelo	52
		3.5.2 Procedimento de estimação	57
4	Dad	s	61
5	Res	tados Finais	65
6	Cor	lusão	81
Re	ferên	as	83

Lista de ilustrações

Figura 1	Dinâmica da estrutura a termo ao longo do tempo	19
Figura 2	Loading dos fatores	24
Figura 3	Dinâmica da estrutura a termo ao longo do tempo	61
Figura 4	Evolução da utilização da capacidade industrial ao longo do tempo	
	(meses)	63
Figura 5	Evolução da taxa básica de juros americana ao longo do tempo (dias) .	63
Figura 6	Evolução da taxa de inflação ao longo do tempo (meses)	64
Figura 7	SEM variáveis macroeconômicas: Fator nível $-N_t$	65
Figura 8	COM variáveis macroeconômicas: Fator nível $-N_t$	66
Figura 9	SEM variáveis macroeconômicas: Fator inclinação $-I_t$	66
Figura 10	COM variáveis macroeconômicas: Fator inclinação $-I_t$	66
Figura 11	SEM variáveis macroeconômicas: Fator curvatura $-C_t$	67
Figura 12	COM variáveis macroeconômicas: Fator curvatura $-C_t$	67
Figura 13	Fator Nível $-N_t$	68
Figura 14	Fator Inclinação $-I_t$	68
Figura 15	Fator curvatura $-C_t$	68
Figura 16	Volatilidade estocástica: Erro de Medida - σ_e	69
Figura 17	Volatilidade estocástica: Nível - $\sigma_{N,t}$	69
Figura 18	Volatilidade estocástica: Inclinação - $\sigma_{I,t}$	70
Figura 19	Volatilidade estocástica: Curvatura - $\sigma_{C,t}$	70
Figura 20	Volatilidade estocástica: Capacidade Utilizada - $\sigma_{UCI,t}$	70
Figura 21	Volatilidade estocástica: Forward Funds - Taxa Básica de Juros - $\sigma_{FFR,t}$	71
Figura 22	Volatilidade estocástica: Inflação - $\sigma_{Infl,t}$	71
Figura 23	Parâmetro de decaimento: λ_t	72
Figura 24	Peso dado pela regressão MIDAS para utilização da capacidade indus-	
	trial — fator nível	72
Figura 25	Peso dado pela regressão MIDAS para utilização da capacidade indus-	
	trial — fator inclinação	73
Figura 26	Peso dado pela regressão MIDAS para utilização da capacidade indus-	
	trial — fator curvatura	73
Figura 27	Peso dado pela regressão MIDAS para Inflação — fator nível	74
Figura 28	Peso dado pela regressão MIDAS para Inflação – fator inclinação	74

Figura 29	Peso dado pela regressão MIDAS para Inflação — fator curvatura $\ .\ .\ .$	75
Figura 30	Peso dado pela regressão MIDAS x coeficientes VAR(1)	75

Lista de tabelas

Tabela 1	Análise descritiva — taxa de juros	62
Tabela 2	Análise descritiva — informações macroeconômicas	63
Tabela 3	Modelo DNS SEM variáveis macroeconômicas	76
Tabela 4	Modelo DNS COM variáveis macroeconômicas	77
Tabela 5	Modelo DNS COM variáveis macroeconômicas - Inferência Bayesiana	
	- Ponderação MIDAS Pesos Iguais	78
Tabela 6	Modelo DNS SEM variáveis macroeconômicas - Filtro de Kalman	79
Tabela 7	Modelo DNS COM variáveis macroeconômicas - Filtro de Kalman	79

Resumo

MINIOLI, A. C. S. **Modelagem de curvas de juros usando amostragem de frequências mistas.** 2014. 25 p. Dissertação (Mestrado) - Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2014.

Neste trabalho, tínhamos por objetivo propor um modelo dinâmico de estrutura a termo de taxas de juros com variáveis macroeconômicas baseado na formulações de Diebold e Li (2006) e Nelson e Siegel (1987) (DNS). A estrutura de estimação proposta permite utilizar dados de frequências distintas, combinando observações diárias de curvas de juros e mensais de variáveis macroeconômicas de interesse através de uma estrutura MIDAS - Mixed Data Sampling. Também utilizamos uma estrutura de volatilidade estocástica multivariada para os fatores latentes e variáveis macroeconômicas e também permitimos que o parâmetro de decaimento do modelo DNS varie no tempo, permitindo capturar mudanças na estrutura de volatilidade condicional e no formato das curvas em períodos longos. O procedimento de estimação é baseado em métodos Bayesianos usando Markov Chain Monte Carlo. Aplicamos este modelos para a curva de juros de títulos do Tesouro Americano entre 1997 e 2011. Os resultados indicam que incorporação de informações diárias e mensais em um mesmo modelo permite ganhos significantes de ajuste, superando as estimativas usuais baseadas em modelos sem informações macroeconômicas e nos métodos usuais de estimação do modelo de Diebold e Li (2006).

Palavras-chaves: Modelo Dinâmico de Nelson-Siegel. Inferência Bayesiana. MI-DAS.

Abstract

MINIOLI, A. C. S. **The term structure of interest rates model using mixed data sampling.** 2014. 25 p. Dissertation (Master Degree) - Graduate School of Economics, Business and Accounting, University of São Paulo, Ribeirão Preto, 2014.

In this present work, we propose a dynamic model for the term structure of interest rates with macroeconomic variables based on Diebold e Li (2006)'s and Nelson e Siegel (1987)'s researches. The estimation procedure we intend to build allows time series data sampled at different frequencies, mixing daily observations of yield curves and monthly observations of macroeconomic variable through a Mixed Data Sampling (MIDAS) regression. We also make use of a multivariate stochastic volatility structure for the latent factors and allow the parameter that governs the exponential decay rate to vary trough time, which enables us to capture changes both in the conditional volatility structure and in the curve's shapes during long periods. The estimation procedure is based on Baeysian inference trough the usage of of Markov Chain Monte Carlo (MCMC) method. We applied these models to the U.S. Treasure bonds' yield curve from 1997 to 2011. The results denote that joining daily and monthly information into the same model allows significant gains on fitting these models to the term structure, overcoming the usual estimates based on models without macroeconomics information and on regular estimation methods of Diebold e Li (2006)'s model.

Key-words: Dynamic Nelson-Siegel Model. Bayesian Inference. MIDAS.

1 Introdução

Uma das variáveis mais observadas nos cenários financeiro e econômico é a taxa de juros. Há uma necessidade em poder antecipar algumas informações, incentivando muitos profissionais a desenvolver técnicas de previsão para a curva de juros. A estrutura a termo da taxa de juros é a relação entre taxas de juros de diferentes prazos até a maturidade em um determinado instante de tempo, considerando o mesmo nível de risco e liquidez; é também conhecida como curva de rendimento ou yield curve, que está representada na figura 1. Entender sua dinâmica de curto e longo prazos é de grande relevância, uma vez que tal variável apresenta valores diferentes a cada dia. Desta forma, precisamos descobrir como os rendimentos se comportam, de acordo com cada maturidade, ao longo do tempo.

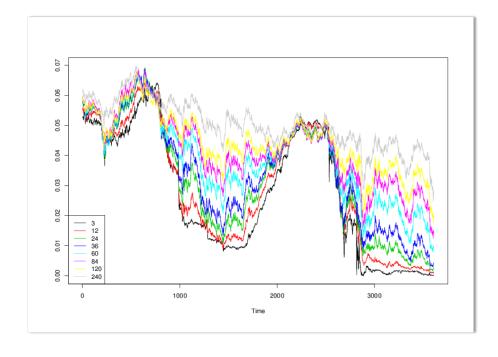


Figura 1 – Dinâmica da estrutura a termo ao longo do tempo

A figura 1 acima nos mostra a evolução no tempo do rendimento dos títulos do tesouro americano para diferentes maturidades. O eixo horizontal (*Time*) representa o tempo em dias e as maturidades estão representadas em meses.

No cenário financeiro, a taxa de juros é altamente influenciável no mercado de ações, uma vez que um aumento na taxa de juros, por exemplo, faz com que investimentos alocados no mercado de ações migrem para títulos do governo (lembrando que o inverso também ocorre quando há uma queda na taxa de juros). Ela é, também, muito

utilizada em precificação de derivativos, como, por exemplo, em contratos de opção ou contratos futuros - um dos estudos pioneiros na precificação de opção foi feito por Black e Scholes (1973). Além disso, no gerenciamento de risco, a taxa de juros exerce uma função importante, uma vez que o conhecido *risco de mercado* inclui o risco de perdas resultantes de movimentos de índices abrangentes de mercado, como, por exemplo, mudanças no nível geral de taxa de juros, no preço de ações, nas taxas de câmbios, no preço de *commoditties*.

Ao analisarmos a visão macroeconômica, fica clara a importância da taxa de juros para o andamento da economia. Pois sabe-se que bancos centrais de diversos países utilizam como ferramenta de política monetária a taxa de juros de curto prazo, que, por sua vez, acaba influenciando os indicadores de inflação e a taxa de longo prazo, que é comumente usada pelos participantes do mercado, os quais estão sempre preocupados em decidir entre consumir e poupar, em que o termo poupar pode ser interpretado como consumo futuro. Desta forma, sendo a taxa de juros uma informação importante na decisão de consumo intertemporal (ver Romer (2012)).

Na literatura, há algumas frentes de pesquisa para a previsão da taxa de juros que podem ser classificadas em três vertentes. A primeira considera modelos com livrearbitragem em que condições de consistência são impostas entre as taxas de juros de diversas maturidades para que se evite a existência de oportunidades de arbitragem (ver Heath et al. (1992), o qual considera, também, uma estrutura de volatilidade estocástica). A segunda vertente abrange modelos de equilíbrio de mercado, normalmente envolvem modelos para a taxa a termo instantânea e assume-se que o prêmio de risco dado por funções afim é utilizado para derivar as taxas de outras maturidades (ver Duffie e Kan (1996) e Dai e Singleton (2002) para exemplos). Por último, temos os modelos estatísticos, que não impõem condições de equilíbrio, nem de livre-arbitragem, e, tampouco, apresentam interpretação estrutural. Nelson e Siegel (1987) foi um dos primeiros estudos a considerar uma estrutura de variáveis latentes para a curva de rendimento, introduzindo um modelo paramétrico parcimonioso. Em Litterman e Scheinkman (1991), aplicou-se a análise de componentes principais para descrever a estrutura a termo da taxa de juros, atribuindo três fatores (nível, inclinação e curvatura) à curva de rendimento que pudessem explicar boa parte de sua variabilidade. Utilizando uma decomposição similar a dos componentes de nível, inclinação e curvatura, Diebold e Li (2006) formulou uma versão dinâmica do modelo de Nelson e Siegel (1987), que tem se mostrado extremamente bem sucedida na modelagem da curva de juros, como discutido em Diebold e Rudebusch (2013).

O passo seguinte na construção de modelos dinâmicos para a estrutura a termo de taxas foi a incorporação direta dos determinantes macroeconômicos, colocando variáveis macroeconômicas como determinantes de mudanças nos diversos vencimentos da curva. Um exemplo fundamental é o modelo proposto em Diebold et al. (2006), que formula um modelo Nelson-Siegel dinâmico colocando variáveis macroeconômicas como determinantes

da evolução temporal dos fatores de *nível*, *inclinação* e *curvatura*. No entanto, a utilização de variáveis macroeconômicas pode representar um problema de ineficiência do uso de informações nesta classe de modelos.

Com a inclusão de variáveis macroeconômicas no modelo de Nelson e Siegel (1987), como feito nos trabalhos de Diebold et al. (2006), precisamos nos atentar à presença de dados cujas frequências observadas são distintas entre algumas variáveis, uma vez que dados financeiros são obtidos diariamente e as informações macroeconômicas estão disponíveis numa frequência menor, por exemplo, a cada mês. Assim, nos modelos com variáveis macroeconômicas, é necessário agregar os dados diários de curvas de juros, o que representa uma perda efetiva de informações na modelagem da curva de juros, e impede a construção de modelos para a curva diária de juros que incorporem diretamente a informação presente nestas variáveis macroeconômicas.

Neste trabalho, propomos uma estrutura de amostragem com frequências distintas, usando métodos baseados em MIDAS (*Mixed Data Sampling*). Esta estrutura possibilita incorporar informações com diferentes frequências de observação usando uma estrutura ótima de ponderação que permite compatibilizar estas informações. Estes pesos são determinados usando uma estrutura baseada em um número restrito de parâmetros, e evitando uma estrutura *ad hoc* de ponderação, como a utilização de pesos iguais normalmente utilizada na agregação de dados diários. Esta estrutura MIDAS permite a formulação de um modelo Nelson-Siegel dinâmico que possibilita utilizar simultaneamente informações diárias para a curva de juros e dados macroeconômicas em frequências mensais, generalizando a estrutura proposta em Diebold et al. (2006).

Adicionalmente, também modificamos outras características da família de modelos dinâmicos de Nelson e Siegel (1987) com variáveis macroeconômicas para uma maior consistência com as características das curvas de juros observadas diariamente. Utilizamos uma estrutura de volatilidade estocástica para o erro de observação, fatores latentes e variáveis macroeconômicas do modelo, já que, quando usamos dados diários, as estruturas de volatilidade condicional são mais relevantes do que quando usamos dados mensais. Também assumimos que o padrão de decaimento da curva de Nelson-Siegel pode variar no tempo, transformando o parâmetro de decaimento em um fator latente adicional. Este procedimento é interessante por permitir capturar as mudanças no formato da curva de juros em períodos mais longos de tempo, e também é importante no período recente quando as curvas de juros se encontram próximas do limiar zero.

Devido à presença de volatilidades estocásticas e um parâmetro de decaimento variante no tempo, utilizamos uma estrutura de estimação Bayesiana usando métodos de Markov Chain Monte Carlo — MCMC. Neste procedimento, podemos estimar o modelo em um única etapa, evitando os procedimentos ineficientes de estimação em dois estágios normalmente utilizados para a estimação de modelos dinâmicos de Nelson-Siegel. Este

procedimento Bayesiano também incorpora a estimação da estrutura de ponderação ótima do procedimento MIDAS e permite obter intervalos de credibilidade exatos em amostras finitas, evitando as aproximações assintóticas nos métodos baseados em verossimilhança, além de ser robusto aos problemas existentes de máximos locais na estimação de modelos de estrutura a termo por máxima verossimilhança, como discutido, por exemplo, em Duffee (2002).

Nós aplicamos o modelo usando observações diárias de curvas de juros de títulos do tesouro americano para o período de 1997 a 2011, incorporando no modelo informações macroeconômicas como a utilização da capacidade industrial, taxa básica de juros americana (Federal Fund Rate) e inflação, de forma análoga ao modelo de Diebold et al. (2006), mas nossa estrutura de amostragem apresenta distintas frequências. O objetivo é avaliar o impacto que estas variáveis exercem sobre a taxa de juros, considerando, também, a relação inversa, ou seja, avaliando qual a influência da taxa de juros sobre variáveis macroeconômicas e tal relação será ajustada através da regressão MIDAS ao agregarmos as informações de maior frequência. Apenas a taxa básica de juros americana é obtida diariamente, as outras duas variáveis são capturadas mensalmente.

Este trabalho está estruturado da seguinte maneira. O Capítulo 2 é constituído de uma breve revisão da literatura sobre alguns modelos estatísticos para a curva de rendimentos, algumas vantagens do uso da estrutura de volatilidade estocástica, alguns ganhos de utilizar-se a inferência bayesiana ao invés do método clássico e, por fim, uma pequena introdução sobre a regressão MIDAS. No Capítulo 3, foi feita uma descrição mais detalhada sobre cada um dos métodos que serão utilizados neste trabalho. Já no Capítulo 4, temos uma visualização sobre os dados utilizados para a aplicação das metodologias. Por sua vez, no Capítulo 5, constam os resultados finais sobre a aplicação da estimação do modelo. Por fim, o Capítulo 6 faz um resumo geral sobre o trabalho.

2 Revisão de Literatura

Neste capítulo, será feita uma breve revisão da literatura sobre modelos estatísticos utilizados para descrever a dinâmica da estrutura a termo da taxa de juros, juntamente com estudos que acrescentaram informações macroeconômicas a alguns modelos. Posteriormente, serão apresentados estudos sobre a utilização da estrutura de volatilidade estocástica. Em seguida, faremos uma análise de trabalhos que aplicaram inferência bayesiana como alternativa aos métodos clássicos de modelagem. E, por fim, revisaremos estudos que mostraram a importância da regressão MIDAS ao agregar informações que são observadas com frequências distintas entre si.

2.1 Modelo dinâmico de Nelson-Siegel

Em seu artigo publicado em 1987, Nelson e Siegel (1987) sugeriram um modelo parcimonioso para a curva de rendimentos, considerando que a taxa a termo instantânea poderia ser obtida através da solução de uma equação diferencial de segunda ordem com raízes reais e iguais ¹ dada por

$$r(m) = \beta_0 + \beta_1 \cdot e^{-\frac{m}{\tau}} + \beta_2 \cdot \left[(m/\tau) \cdot e^{-\frac{m}{\tau}} \right]$$

Em que m representa maturidade e τ , o parâmetro arbitrário de decaimento. Desta forma, o rendimento médio até a maturidade seria dado por

$$R\left(m\right) = \frac{\int_{0}^{m} r(x)dx}{m}$$

$$R\left(m\right) = \beta_0 + \beta_1 \cdot \frac{1 - e^{-m/\tau}}{m/\tau} + \beta_2 \cdot \left[\frac{1 - e^{-m/\tau}}{m/\tau} - e^{-m/\tau}\right]$$

Os pesos (loadings) atribuídos a β_0 , β_1 e β_2 são, respectivamente, 1, $\frac{1-e^{-m/\tau}}{m/\tau}$ e $\frac{1-e^{-m/\tau}}{m/\tau} - e^{-m/\tau}$ e a interpretação econômica para cada um deles, assim como a do parâmetro de decaimento, pode ser feita da seguinte forma:

• Efeito de longo prazo ($\beta_0 \Rightarrow 1$): não depende de m, constante e igual a 1. Um

¹raízes reais e iguais para evitar o problema de super parametrização

- aumento em β_0 desloca a curva toda para cima (deslocamento paralelo). Em particular, temos que $R(\infty) = \beta_0$;
- Efeito de **curto** prazo $(\beta_1 \Rightarrow \frac{1-e^{-m/\tau}}{m/\tau})$: começa em 1, quando $m \to 0$; vai diminuindo, conforme m aumenta, até se aproximar de 0, quando $m \to \infty$. Um aumento em β_1 aumenta mais o rendimento de curto prazo do que o de longo prazo. Alguns autores, por exemplo Frankel e Lown (1994), definem *inclinação* da curva como sendo $R(\infty) R(0) = -\beta_1$;
- Efeito de **médio** prazo $(\beta_2 \Rightarrow \frac{1-e^{-m/\tau}}{m/\tau} e^{-m/\tau})$: começa em 0, quando $m \to 0$; vai aumentando até atingir seu ponto de máximo em um $m^{m\acute{a}x}$ específico e passa a diminuir, voltando a ser igual a 0, quando $m \to \infty$;
- Parâmetro arbitrário de decaimento (τ) : quanto menor τ , mais r'apido é o decaimento nos regressores (bom ajuste da curva com maturidades de horizonte pequeno); quanto maior τ , mais lento é o decaimento nos regressores (bom ajuste da curva com maturidades de maior horizonte).

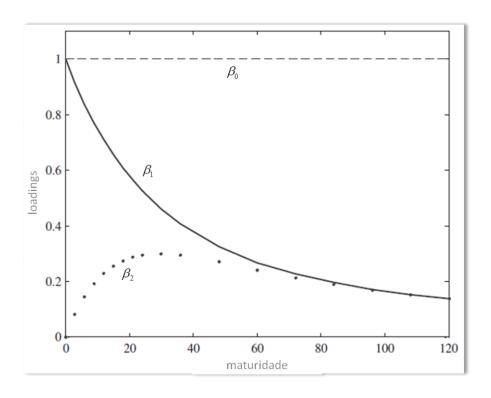


Figura 2 – Loading dos fatores

A figura 2 foi extraída de Diebold e Li (2006) e contém a evolução dos loadings (pesos) dos fatores explicados acima. Foi utilizado um valor fixo para τ igual a 0,0609. O objetivo desta figura é meramente exemplificativo para, assim, ilustrar o que foi descrito sobre os efeitos de curto, médio e longo prazo.

Neste estudo, eles mostraram que o modelo proposto apresenta, na maior parte das vezes, menor perda de precisão, obtendo um menor desvio-padrão para o modelo ajustado, em comparação com um modelo restrito em que $\beta_1 = 0$ (modelo monotônico). Mostraram, também, que o modelo não-restrito explica 96% da variabilidade total do rendimento dos títulos do Tesouro Americano para diferentes maturidades, ao longo do período entre 1981-83.

Litterman e Scheinkman (1991) introduziu a Análise de Componentes Principais para descrever a estrutura a termo da taxa de juros, em que os três primeiros fatores foram utilizados para descrever o comportamento da curva de juros. Primeiro, segundo e terceiro fatores foram nomeados, respectivamente, de nível, inclinação e curvatura. Mudanças no nível e na curvatura estão associados, respectivamente, a deslocamentos paralelos e a mudanças na volatilidade da curva de rendimentos. Neste artigo, eles analisaram, como variável dependente, o prêmio pelo risco; no caso, compararam títulos de renda fixa com títulos do governo americano e constataram que nível, inclinação e curvatura explicam, respectivamente, em média, 89,5%, 8,5% e 2% da variância total dos rendimentos.

Diebold e Li (2006) apresentou uma variação para o modelo de componentes exponenciais de Nelson e Siegel (1987), considerando que os fatores do modelo variam ao longo do tempo. Desta forma, os autores atribuíram modelos autoregressivos de primeira ordem para cada um dos fatores $-\beta' = \begin{pmatrix} \beta_{0t} & \beta_{1t} & \beta_{2t} \end{pmatrix}$. Apesar dessa alteração com relação ao modelo original, eles mantiveram o parâmetro arbitrário de decaimento (τ) estático, ou seja, não se altera com o tempo.

2.1.1 Fatos estilizados

Em seu estudo, Diebold e Li citam alguns fatos estilizados sobre a curva de rendimento. São eles:

- a curva de rendimento média é crescente e côncava e, além disso, assume diversos formatos ao longo do tempo, tais como inclinações positiva e negativa, formato de sino (em que rendimentos de médio prazo são maiores do que rendimentos de curto e longo prazos), formato de sino inverso podendo ser assimétrica ou não;
- as dinâmicas do rendimento são persistentes, enquanto que as dinâmicas do spread de rendimentos são menos persistentes (entenda por spread a diferença dos rendimentos de duas maturidades distintas, por exemplo, a diferença entre rendimentos de longo e curto prazos, $R(\infty) R(0)$;
- rendimento de curto prazo são mais voláteis do que de longo prazo, uma vez que o rendimento de curto prazo está associado tanto a β_{0t} , quanto a β_{1t} , e o de longo prazo, apenas a β_{0t}

• geralmente, taxas de juros de longo prazo apresentam uma persistência maior do que as de curto prazo.

Em Diebold e Li (2006), os fatores latentes $\hat{\beta}_{0t}$, $\hat{\beta}_{1t}$ e $\hat{\beta}_{2t}$ foram obtidos através do método de Mínimos Quadrados Ordinários em um procedimento de dois estágios. No primeiro estágio, assumindo que o parâmetro λ é fixo e conhecido, são estimados os parâmetros nível $(\hat{\beta}_{0t})$, inclinação $(\hat{\beta}_{1t})$ e curvatura $(\hat{\beta}_{2t})$ através de uma regressão cross-section para cada curva de juros, e no segundo estágio é estimado um processo autoregressivo para estes parâmetros estimados no primeiro estágio. Com dados empíricos, mostraram que o primeiro fator é o mais persistente $(\hat{\rho}_{\beta_{0t}}(30) = 0.454, \hat{\rho}_{\beta_{1t}}(30) = -0.082$ e $\hat{\rho}_{\beta_{2t}}(30) = -0.006$) e que o teste *Dickey-Fuller Aumentado* sugere que os fatores $\hat{\beta}_{0t}$ e $\hat{\beta}_{1t}$ apresentam raiz unitária (ou seja, sua dinâmica é bastante próxima a um passeio aleatório).

Diebold e Li (2006) comparou a previsão de 1, 6 e 12 meses à frente (para maturidades de 3 meses e 1, 3, 5 e 10 anos) do modelo proposto com vários outros modelos já existentes na literatura e concluíram que, para previsões com horizontes maiores (6, 12 meses), o modelo proposto apresentou erro quadrático médio (EQM) menor em relação a todos os outros modelos, para todas as cinco maturidades citadas acima. Já a previsão com um horizonte menor (1 mês), o modelo proposto se assemelha muito aos modelos já existentes (para comparação explícita dos modelos, ver Diebold e Li (2006)).

2.1.2 Relação entre os fatores latentes da curva de rendimentos com os componentes principais

No início do estudo apresentado por Diebold et al. (2006), os autores mostram a forte correlação dos componentes principais citados por Litterman e Scheinkman (1991) e o que Diebold e Li (2006) atribuiu como nível, inclinação e curvatura. Por conseguinte, explicitaram as seguintes relações:

- Fator *nível* e primeiro componente principal (média simples dos rendimentos com maturidades de 3, 24 e 120 meses): correlação de 80%;
- Fator *inclinação* e segundo componente principal (rendimento de maturidade de 3 meses menos o de 120 meses): correlação de 98%;
- Fator *curvatura* e terceiro componente principal (duas vezes rendimento de 24 meses menos a soma dos rendimentos de 3 e 120 meses): correlação de 96%.

2.1.3 Taxa de Juros e Dados Macroeconômicos

Há diversos estudos na literatura que envolvem informações macroeconômicas na previsão da taxa de juros. Em seu livro, Diebold e Rudebusch (2013), os autores expõem que, de um lado, pela perspectiva de finanças, a taxa de juros de curto prazo é fundamental para a construção das taxas de outras maturidades, uma vez que taxas de longo prazo são guiadas, em boa parte, pela expectativa futura da taxa de curto prazo. Por outro lado, pela perspectiva macroeconômica, a taxa de curto prazo é considerada peça fundamental dentre os instrumentos de política monetária, a qual é utilizada por bancos centrais com o objetivo de estabilizar a economia. Visto isso, relacionar dados macroeconômicos à curva de rendimentos pode nos fornecer informações relevantes na construção de modelos preditivos da estrutura a termo, e o contrário também é válido.

Ang e Piazzesi (2003) utilizou uma combinação de indicadores para inflação e para a atividade real da economia (taxa de crescimento do emprego, taxa de crescimento da produção industrial, entre outras) para que pudessem fazer previsões mensais para a curva de rendimentos. Eles mostraram que, para previsões com horizonte grande (mais de 60 meses à frente), 85% da variância é explicada por fatores macroeconômicos para maturidade de 12 meses e 64% para maturidade de 60 meses.

Hordahl et al. (2006) também incluiu variáveis macroeconômicas na previsão da curva de rendimentos. Alguns de seus resultados foram que choques na política monetária impactam bastante os rendimentos de curto prazo e, para os de longo prazo, o efeito é menor; choques na inflação e no produto afetam, na sua maior parte, a curvatura da curva de rendimentos para maturidades de médio prazo; mudanças na meta de inflação apresentam efeitos mais duradouros e tendem a ter um impacto mais longo em rendimentos de longo prazo.

Tanto Ang e Piazzesi (2003), quanto Hordahl et al. (2006), consideraram uma análise unidirecional, assumindo que produto e inflação são determinados independentemente do formato da curva de juros. Desta forma, utilizaram as informações macroeconômicas para prever a curva de rendimento, mas não o contrário.

Rudebusch e Wu (2008) apresentou uma análise bidirecional, explicitando a presença de dois fatores latentes junto com a taxa de inflação anual e uma medida do gap do produto (diferença entre o produto real e o produto potencial). Em que a taxa de juros é função dos fatores nível e inclinação. Tais fatores são explicados cada um de uma forma; o fator nível é expresso em função da inflação, já a inclinação é escrita em função tanto da inflação, quanto do gap do produto. A inflação, por sua vez, é função de inflações passadas, do fator nível e do gap do produto passado. Já o gap do produto é função do próprio gap do produto passado, da taxa de juros passada e do fator nível passado. Eles mostraram que a correlação entre a inflação esperada de um ano com o fator nível da

curva de rendimentos é de 73% e a do produto com o fator inclinação é de 66%.

Já Diebold et al. (2006) utilizou o mesmo modelo exposto em Diebold e Li (2006) para previsões mensais e incluiu informações macroeconômicas na estrutura do vetor autoregressivo de primeira ordem - VAR(1) - do modelo para os fatores latentes, porém manteve-se o parâmetro de decaimento (τ) fixo no tempo. As variáveis consideradas foram a utilização da capacidade industrial de manufaturados, a taxa básica de juros da economia e a inflação anual, que representam, respectivamente, a atividade econômica real em relação ao seu potencial, o instrumento de política monetária e a taxa de inflação.

Os autores observaram que as variáveis macro são bastante persistentes, pois apresentaram estimativas para o coeficiente autoregressivo de primeira ordem muito próximo de 1 (mais precisamente igual a 0.99) tanto para a taxa básica de juros, quanto para a inflação, e exatamente igual a 1 no caso da utilização da capacidade industrial.

Ao analisarem a função impulso-resposta, eles perceberam que as informações macroeconômicas são bastante sensíveis a choques no nível, em que, por exemplo, apesar do efeito inicial de um aumento deste fator ser quase nulo sobre a inflação, esta vai aumentando gradualmente ao longo do tempo. Em especial, um choque positivo no fator inclinação gera um aumento imediato de quase um para um na taxa básica de juros e, logo após o choque, inicia-se um decaimento até que a taxa volte ao mesmo valor que possuía anteriormente ao choque. Este efeito nos mostra, citam os autores, uma ligação entre o fator inclinação da curva de rendimentos e o instrumento de política monetária. Com isso, fornecem-nos duas interpretações. De um lado, pode ser que o Fed (Banco Central americano) esteja reagindo a tal mudança da estrutura a termo (lembrando que, como exposto anteriormente, o fator inclinação está diretamente ligado ao efeito de curto prazo) ao estipular qual será a taxa básica de juros da economia no início de cada mês. Por outro lado, dado o período entre a tomada de decisão de política monetária e a execução/aprovação de tal decisão, é provável que o mercado, consequentemente a curva de rendimentos, esteja reagindo em relação a informações macroeconômicas antecipadamente às ações do Fed. Por fim, ao analisarem os efeitos de choques das variáveis macro sobre os fatores do modelo, observaram que, por exemplo, um aumento na taxa básica de juros causa uma elevação temporária no fator inclinação, fazendo com que a curva de rendimentos fique mais negativamente inclinada e que choques positivos na inflação provocam um aumento no fator nível. A este último efeito, pode-se atribuir o fato de que os autores encontraram uma correlação de 43% entre inflação e o fator nível, implicando que uma elevação surpresa na inflação impulsiona uma maior expectativa de inflação futura. ²

Diebold et al. (2006) mostrou, também, a decomposição da variância para o modelo que inclui as variáveis macroeconômicas para os rendimentos de maturidades de 1, 12 e 60 meses, considerando um horizonte de previsão de 1, 12 e 60 meses. Observaram que,

²Para uma análise completa da função impulso resposta, ver Diebold et al. (2006)

para previsão de horizonte pequeno, tais variáveis pouco influenciam a variabilidade total da curva de rendimentos, inferindo que grande parte da variação idiossincrática de curto prazo não está atrelada a fundamentos macroeconômicos. Entretanto, para horizontes mais longos, as variáveis macro se mostraram mais influentes sobre a variabilidade total da curva de rendimentos, representando, na previsão de horizonte de 60 meses, 45%, 40% e 36% da variabilidade total dos rendimentos de maturidade de 1, 12 e 60 meses respectivamente. Ao analisarem o efeito inverso, ou seja, a decomposição da variância das variáveis macroeconômicas, constataram que os fatores da estrutura a termo pouco representam da variabilidade total tanto da inflação, quanto da utilização da capacidade industrial. Todavia, mostraram que os três fatores latentes juntos representam 46%, 53% e 51% da variância da taxa básica de juros, considerando os horizontes de previsão de 1, 12 e 60 meses respectivamente (para uma análise completa da decomposição da variância, ver Diebold et al. (2006)).

2.2 Volatilidade Estocástica

Séries temporais financeiras são caracterizadas por alguns fatos estilizados notáveis. Entre eles está a presença de estruturas de volatilidade condicional variantes no tempo. Esses efeitos foram estudados extensivamente para mercados de ações e moedas, mas também são bastante importantes, e existe uma linha de pesquisa em modelos de estrutura a termo considerando estruturas de volatilidade estocástica para os *yields* observados e fatores latentes da curva de juros. Koopman et al. (2010) introduz um modelo de estrutura a termo considerando uma estrutura fatorial usando um fator comum modelado por um processo GARCH na classe de Nelson-Siegel. Hautsch e Yang (2012) e Caldeira et al. (2010) introduzem estruturas de volatilidade estocástica na classe de modelos dinâmicos de Nelson-Siegel, mostrando que é fundamental considerar a volatilidade condicional em modelos da estrutura a termo.

Um dos fatos estilizados da estrutura a termo da taxa de juros citados por Diebold e Li (2006), em que a curva de rendimento apresenta diversos formatos ao longo do tempo, tais como inclinações positiva e negativa, formato de sino, formato de sino inverso — podendo ser assimétrica ou não, faz com que seja necessário considerar a presença de uma possível estrutura de volatilidade estocástica dos erros. Desta forma, com o aparecimento de mudanças no formato da curva de rendimentos, principalmente em momentos relacionados a determinadas incertezas sobre a taxa de juros futura e as expectativas em relação às políticas fiscal e monetária, faz-se necessário considerar não apenas que os erros sejam heteroscedásticos, mas que sua volatilidade seja estocástica.

Na literatura de finanças, existem duas classes fundamentais de modelos para capturar estruturas de volatilidade condicional variantes no tempo. A primeira classe está

baseada na chamada estrutura ARCH - Autoregressive Conditional Volatility proposta em Engle (1982) e depois generalizado por Bollerslev (1986). Nestes modelos a volatilidade é uma função determinística dos erros passados e da volatilidade passada. A segunda classe de modelos é de modelos de volatilidade estocástica, introduzida por Taylor (1986), que propôs modelar o logaritmo da volatilidade como um processo autoregressivo de primeira ordem.

Uma das principais características destes modelos, mencionada por Broto e Ruiz (2004), é que a volatilidade é modelada como uma variável latente não observada. Tais modelos são atrativos por estarem próximos a modelos frequentemente usados na teoria de finanças para representar o comportamento de preços de ativos financeiros. Além disso, suas propriedades estatísticas são fáceis de serem derivadas usando os resultados de distribuições log-normais. Ao serem comparados aos modelos GARCH (Generalized Autoregressive Conditional Heteroskedasticity), Carnero et al. (2004) mostraram que os modelos ARSV (Autoregressive Stochastic Volatility) são capazes de capturar as propriedades frequentes observadas nas séries diárias dos retornos financeiros.

2.3 Inferência Bayesiana

Em casos que encontramos dificuldades na implementação e aplicação de métodos de estimação para modelos não-lineares, é preciso que busquemos alternativas para solucionar a limitação dos métodos tradicionais, como, por exemplo, no caso de não linearidade ou, até mesmo, na existência de uma verossimilhança multimodal (que foi o caso encontrado por Chib e Ergashev (2009)). Portanto, utilizar um método clássico, que é o caso da máxima verossimilhança pela decomposição dos erros preditos através do filtro de Kalman, seria uma escolha errônea e teríamos estimativas inconsistentes. Uma vez que pode haver múltiplos máximos locais, gerando problemas de identificação dos parâmetros que maximizam a função. Uma alternativa que se enquadra muito bem a esse tipo de situação é a inferência Bayesiana, que, através de métodos computacionais baseados em métodos de *Markov Chain Monte Carlo*, podem ser aplicados a processos não-lineares, modelos em espaço de estado e processos não-Gaussianos sem as limitações usuais dos métodos frequentistas.

Em Barossi-Filho et al. (2010), os autores analisaram a série histórica do índice IBOVESPA, durante o período de julho de 1994 até fevereiro de 2008. Eles propuseram alguns modelos para o log do retorno deste índice em que o método de estimação utilizado foi a inferência bayesiana, uma vez que os modelos propostos apresentam uma função de verossimilhança muito complexa para que pudessem ser utilizadas técnicas de estatísticas clássicas usuais. São eles:

✓ Modelo de volatilidade estocástica em que a log-volatilidade do erro de medida segue

uma processo autoregressivo de ordem 1, cuja variância do processo ruído branco que compõe o erro de medida é conhecida e igual a 1;

- ✓ Modelo de volatilidade estocástica em que a log-volatilidade do erro de medida segue uma processo autoregressivo de ordem 1, cuja variância do processo ruído branco que compõe o erro de medida é desconhecida;
- \checkmark Modelo de volatilidade estocástica generalizado em que a log-volatilidade do erro de medida segue uma processo autoregressivo de ordem p, cuja variância do processo ruído branco que compõe o erro de medida é conhecida e igual a 1;
- ✓ Modelo de volatilidade estocástica generalizado em que a log-volatilidade do erro de medida segue uma processo autoregressivo de ordem p, cuja variância do processo ruído branco que compõe o erro de medida é desconhecida;
- ✓ Modelo ARCH(2), cuja variância do processo ruído branco que compõe o erro de medida é conhecida e igual a 1;
- ✓ Modelo GARCH(1,1)cuja variância do processo ruído branco que compõe o erro de medida é conhecida e igual a 1.

Ao compararem os seis modelos através do DIC (Deviance Information Criterion – Critério Desvio-Informação), concluíram que os modelos de volatilidade estocástica e de volatilidade estocástica generalizada induzem a melhores estimativas para a volatilidade da série analisada. A mesma conclusão foi obtida ao avaliarem os ajustes através do Fator de Bayes e através do critério BIC (Bayesian Information Criterion – Critério de Informação Bayesiana).

Em seu estudo de 2010, Laurini e Hotta (2010) propuseram algumas extensões ao modelo apresentado em Diebold e Li (2006), dentre as quais consideraram métodos Bayesianos fundamentados através do método de Monte Carlo em Cadeia de Markov (MCMC - Monte Carlo Markov Chain) em alternativa à estimação pelo método de máxima verossimilhança para situações cuja aplicação deste último método é considerada complicada ou, até mesmo, inviável de ser realizada. Os autores citam também que uma das vantagens de se utilizar inferência Bayesiana é que podemos considerar que tanto os parâmetros (β_{0t} , β_{1t} e β_{3t}), quanto os vetores de estado (que contêm a dinâmica dos fatores nível, inclinação e curvatura), sejam interpretados como variáveis latentes, visto que, com a formulação de Bayes, não há necessidade de se assumir linearidade e, desta maneira, o parâmetro de decaimento (τ) não precisa ser fixo no tempo, como é feito tanto por Nelson e Siegel (1987), quanto por Diebold e Li (2006). Atribuindo, então, uma estrutura autoregressiva de primeira ordem para a dinâmica de tal parâmetro.

Além disso, outra vantagem citada por Laurini e Hotta (2010) é de que as propriedades dos estimadores Bayesianos são atribuídas não somente a grandes amostras,

mas também para amostras finitas, o que nos permite calcular os intervalos de confiança para os hiper-parâmetros e fazer previsões para a estrutura a termo da taxa de juros, considerando a incerteza na estimação dos parâmetros.

Em seu artigo que relata uma discussão feita no Joint Statistical Meetings, em agosto de 1996, sobre o método MCMC, Kass et al. (1998) citam que o uso de métodos de simulação em geral, inclusive o MCMC, tem a vantagem de que, dado um conjunto de variáveis aleatórias θ^1 , θ^2 , ..., θ^G gerado a partir de uma distribuição posteriori, é possível estimar todas as informações-resumo de interesse, como, por exemplo, a média e a variância. Dois métodos básicos de MCMC são conhecidos como a amostragem de Gibbs e o algoritmo Metropolis-Hastings. O primeiro é utilizado para gerar amostras a partir das distribuições condicionais completas, isto é, $p(\theta_i|\theta_{j\neq i},y)$, com $i=1,2,\cdots,k$, e produz uma cadeia de Markov cuja densidade conjunta posteriori é a uma distribuição estacionária. Já o segundo método é utilizado quando há dificuldade em realizar as simulações a parir das distribuições condicionais completas. Neste caso, a amostra é feita a partir de uma distribuição candidata pré definida para cada parâmetro (ou grupo de parâmetros) e é feito um racional de aceitar ou rejeitar tal amostra até que se obtenha um número de amostras cuja sequência não apresente correlação ao longo do tempo.

2.4 MIDAS - Mixed Data Sampling

Nosso trabalho segue da contribuição de Diebold et al. (2006), mas permite incorporar diretamente as informações macroeconômicas na curva de juros observada diariamente. Nesse aspecto, o modelo ganha em eficiência, já que o número de observações da curva diária é maior que a curva mensal. Mas a contribuição mais relevante é evitar os problemas caudados pela agregação de informações, usando médias mensais da curva de juros ao invés dos dados diários. Nesse procedimento de agregação o impacto dos choques das informações macroeconômicas só é observado em nível mensal, impedindo o uso destas variáveis em estratégias de gestão de ativos e hedge de movimentos na curva, que devem ser realizados na frequência mais alta possível. Da mesma forma, os efeitos de agregação não são claros em um modelo não-linear como a versão dinâmica do modelo de Nelson-Siegel, e a agregação pode induzir a problemas de viés na estimação de parâmetros e impacto de choques.

Sabemos que informações macroeconômica e financeira são, normalmente, coletadas com frequências distintas. Por exemplo, enquanto a macroeconomia nos fornece dados, costumeiramente, a cada mês, a cada três meses ou, até mesmo, anualmente, é possível obtermos registros diários para variáveis financeiras. Com isso, é muito comum encontrarmos estudos que agreguem séries econômicas de tempo, cujas frequências se diferem de acordo com o tipo de variável, de forma a dar pesos iguais ao transformar a informação que

apresenta maior frequência para acompanhar as variáveis de menor frequência, ou seja, atribuindo apenas um cálculo, por exemplo, de média simples. Causando, desta forma, perda de informações, podendo, como dito anteriormente, gerar estimações viesadas e/ou ineficientes.

Estudos recentes, por exemplo Andreou et al. (2010), mostraram uma alternativa à agregação de variáveis que apresentam maior frequência, já que, anteriormente, eram atribuídos pesos iguais, e agora passam a estabelecer uma ponderação diferente ao longo do tempo de acordo com a influência que ela exerce na data específica t a ser analisada. É o caso da regressão MIDAS ($Mixed\ Data\ Sampling$). Tal análise consiste em estimar os parâmetros de um modelo, por exemplo, comumente chamado de ADL-MIDAS ($autoregressive\ distributed\ lag\ mixed\ data\ sampling\ -$ pertencente ao conjunto de modelos de regressão MIDAS), que explicita a relação autoregressiva da variável dependente (no nosso caso, as variáveis macroeconômicas) e a relação que esta variável pode apresentar com informações diárias da estrutura a termo da taxa de juros, estabelecendo a restrição de linearidade nos parâmetros. Adicionalmente a isso, ao analisarmos a evolução dos parâmetros ao longo do tempo, é esperado que haja um decaimento, seja linear, quadrático, cúbico ou, até mesmo, exponencial (a depender da velocidade de decaimento) conforme a data de observação dos regressores se afasta do tempo t analisado em questão. Exemplos da relação de decaimento dos parâmetros podem ser encontrados em Ghysels et al. (2007).

3 Metodologia

Neste capítulo, serão detalhadas, separadamente, cada uma das metodologias utilizas neste trabalho. São elas, o modelo Dinâmico Nelson-Siegel, a estrutura de volatilidade estocástica, a inferência Bayesiana com alguns algoritmos de simulação e, por fim, a regressão MIDAS. No final do capítulo, será feita a junção de cada uma das técnicas, explicitando, desta forma, como elas serão aplicadas.

3.1 Modelo Dinâmico Nelson-Siegel - DNS

O modelo Nelson-Siegel, como visto anteriormente, é uma forma parcimoniosa de ajustar a curva de rendimentos, pois captura boa parte dos fatos estilizados da estrutura a termo da taxa de juros, tais como os formatos exponenciais presentes na curva de rendimentos.

Primeiramente, vamos definir alguns conceitos básicos para a construção da estrutura a termo da taxa de juros¹.

1. Seja $P_t(m)$ o preço de um título que se compromete a pagar m cupons até o vencimento, cuja maturidade é igual a m, isto é, o valor presente no instante t de US\$1 a receber m períodos a frente e seja $y_t(m)$ o rendimento (yield) nominal de um título zero-cupom descontado continuamente até a maturidade. A partir da curva de rendimento, obtemos a curva de desconto a seguir

$$P_t(m) = exp\left\{-m_\times y_t(m)\right\}$$

2. A partir da curva de desconto, podemos obter a taxa forward (nominal) instantânea

$$f_t(m) = -\frac{P_t'(m)}{P_t(m)}$$

3. A relação entra a *yield to maturity* e a taxa *forward* é, portanto,

$$y_t(m) = \frac{1}{m} \int_0^m f_t(u) du$$

que implica que a *yield* zero-cupom é uma média igualmente ponderada das taxas *forward*.

¹Diebold e Li (2006)

Neste trabalho, utilizaremos a forma funcional de Nelson e Siegel (1987), estendido por Siegel e Nelson (1988)², que utiliza a curva da taxa forward como se segue

$$f_t(m) = \beta_{1t} + \beta_{2t}e^{-\lambda_t m} + \beta_{3t}\lambda_t e^{-\lambda_t m}$$

De acordo com Diebold e Li (2006), em que os autores mostram que o modelo de fator latente na representação Nelson-Siegel apresenta seus coeficientes variantes no tempo, a curva de rendimentos é dada por

$$y_t(m) = N_t + I_t \left(\frac{1 - e^{-\lambda_t m}}{\lambda_t m} \right) + C_t \left(\frac{1 - e^{-\lambda_t m}}{\lambda_t m} - e^{-\lambda_t m} \right)$$
(3.1)

As dinâmicas dos fatores $nivel(N_t)$, $inclinação(I_t)$ e $curvatura(C_t)$ seguem um processo autoregressivo vetorial (VAR) de primeira ordem. Desta maneira, o modelo forma um sistema chamado espaço de estado 3 , cuja equação de transição é dada por:

$$\begin{pmatrix}
N_t - \mu_N \\
I_t - \mu_I \\
C_t - \mu_C
\end{pmatrix} = \begin{pmatrix}
\alpha_{11} & \alpha_{12} & \alpha_{13} \\
\alpha_{21} & \alpha_{22} & \alpha_{32} \\
\alpha_{31} & \alpha_{32} & \alpha_{33}
\end{pmatrix} \begin{pmatrix}
N_{t-1} - \mu_N \\
I_{t-1} - \mu_I \\
C_{t-1} - \mu_C
\end{pmatrix} + \begin{pmatrix}
\eta_t^N \\
\eta_t^I \\
\eta_t^C
\end{pmatrix}$$
(3.2)

O modelo dinâmico de Nelson-Siegel (DNS) separa a curva de rendimentos em três fatores latentes (N_t, I_t, C_t) , cujas dinâmicas determinam a própria movimentação de y_t para qualquer maturidade m (neste estudo, vamos considerar que os fatores mudam ao longo do tempo), e os coeficientes (denominados de pesos dos fatores) determinam a evolução de y(m) para qualquer t. Este modelo permite trabalhar com um conjunto de variáveis de dimensão mais baixa ao criar os fatores latentes que explicam boa parte da variabilidade total da curva de rendimentos. Desta forma, evitam-se possíveis problemas de não identificação do modelo.

A proposta deste trabalho se assemelha ao que foi feito em Diebold et al. (2006) ao acrescentarem as variáveis macroeconômicas na dinâmica dos fatores. Desta forma, a representação do processo autoregressivo vetorial será dada por

$$f_t - \mu = A(f_{t-1} - \mu) + \eta_t \tag{3.3}$$

²Por Diebold e Li (2006)

³ver explicação em Hamilton (1994)

Em que

$$f_t = \begin{pmatrix} N_t \\ I_t \\ C_t \\ UCI_t \\ TBJ_t \\ INFL_t \end{pmatrix}$$

Sendo UCI_t a utilização da capacidade industrial, TBJ_t a taxa básica de juros e, por fim, $INFL_t$ representa a taxa de inflação do período.

Uma alteração a ser considerada em relação ao que foi apresentado por Diebold et al. (2006) é que aceitaremos que o parâmetro arbitrário de decaimento do modelo DNS varie no tempo.

Como o modelo a ser estudado não apresenta linearidade nos parâmetros, uma vez que será assumida a variação temporal do parâmetro de decaimento (λ_t) , e a estrutura a termo da taxa de juros requer que seja atribuída uma estrutura de volatilidade estocástica dos erros, não será possível utilizar métodos lineares como Filtro de Kalman para fazer a calibração do modelo. A proposta deste trabalho é utilizar inferência bayesiana através do método de simulação MCMC (Markov Chain Monte Carlo), que consiste em simular uma cadeia de Markov cuja distribuição estacionária converge para a distribuição $p(\beta|y)$. Além disso, como o modelo proposto procura não apenas gerar previsões para a curva de rendimentos, mas, também, para os dados macroeconômicos, será utilizado o método MIDAS para agregar as variáveis de diferentes frequências atribuindo um peso diferente ao longo do tempo, do que apenas uma média simples para o período.

3.2 Volatilidade Estocástica

Conforme citado anteriormente, modelos de volatilidade estocástica são muito utilizados em séries financeiras por serem capazes de capturar mudanças ao longo da curva *yield*, por exemplo. São muitas vezes utilizados para modelar preços de ações também. Este modelo foi introduzido por Taylor (1986) que propôs a seguinte estrutura:

$$\begin{cases} y_t = \alpha \,\varepsilon_t \sigma_t \\ \log \sigma_t^2 = \phi \log \sigma_{t-1}^2 + \eta_t \end{cases}$$

Tal estrutura é conhecida como ARSV(1) — autoregressive stochastic volatility de ordem 1, em que y_t é o retorno observado no instante t, σ_t é a volatilidade correspondente, α é um parâmetro escalar que remove a necessidade de se incluir uma constante na equação de log-volatilidade e ε_t é um processo ruído branco com variância unitária que representa as inovações no nível dos retornos. A perturbação da equação da volatilidade, η_t , é um processo ruído branco Gaussiano com variância σ_{η}^2 (em Andersen et al. (2001), os autores mostram que o processo da log-volatilidade pode muito bem ser aproximado por uma distribuição Normal⁴).

A variância do processo da log-volatilidade, σ_{η}^2 , mede a incerteza sobre a volatilidade futura. Se $\sigma_{\eta}^2 = 0$, o modelo ARSV(1) é não identificado. Já o parâmetro ϕ é, muitas vezes, considerado como uma medida de persistência dos choques na volatilidade. Note que, quando o valor de ϕ estiver próximo de 1 e o de σ_{η}^2 for próximo de 0, a evolução da volatilidade ao longo do tempo é bastante suave. No limite, se $\phi = 1$ e $\sigma_{\eta}^2 = 0$, a volatilidade é constante ao longo do tempo e, consequentemente, os retornos são homoscedásticos.

Se $|\phi| < 1$, então σ_t^2 é estacionário, consequentemente, y_t também será estacionário. Neste caso, a variância marginal dos retornos será dada por

$$\sigma_y^2 = \alpha^2 \exp\left\{\frac{1}{2} \times \frac{\sigma_\eta^2}{(1 - \phi^2)}\right\}$$

O modelo ARSV(1) pode ser linearizado ao aplicarmos o logaritmo no quadrado dos retornos, como mostrado em Broto e Ruiz (2004):

$$\log\left(y_{t}\right) = \mu + h_{t} + \xi_{t} \tag{3.4}$$

$$h_t = \phi h_{t-1} + \eta_t \tag{3.5}$$

Em que $\mu = \log(\alpha) + E[\log(\varepsilon_t^2)], h_t = \log(\sigma_t^2) \in \xi_t = \log(\varepsilon_t^2) - E[\log(\varepsilon_t^2)].$

O modelo composto pelas equações 3.4 e 3.5 é um modelo com representação em espaço de estado linear não-Gaussiano, em que a equação 3.4 é a equação de medições e a

⁴Por Broto e Ruiz (2004)

equação 3.5 é a equação de transição. As propriedades do ruído de medições, ξ_t , dependem da distribuição de ε_t .

A função de autocorrelação será dada por:

$$\rho(\tau) = \frac{\sigma_h^2}{\left(\sigma_h^2 + \sigma_\xi^2\right)} \, \phi^{\tau}$$

Em que
$$\sigma_h^2 = \sigma_n^2 / (1 - \phi^2)$$
.

Neste trabalho, utilizaremos uma estrutura de volatilidade estocástica multivariada para todos os fatores latentes do modelo dinâmico de Nelson-Siegel (nível, inclinação e curvatura) e para as variáveis macroeconômicas utilização da capacidade industrial, taxa básica de juros americana e inflação. Lembrando que apenas a utilização da capacidade industrial e a inflação são capturadas mensalmente, já as demais informações são obtidas diariamente.

3.3 Inferência Bayesiana

3.3.1 Introdução à inferência bayesiana

Inferência bayesiana é baseada em algumas regras simples de probabilidade. Em Koop (2003), o autor comenta que essa é uma das vantagens de se utilizar uma abordagem Bayesiana, pois ao desejarmos estimar os parâmetros de um modelo, comparar diferentes modelos ou, até mesmo, obter previsões, é possível que utilizemos a mesma regra de probabilidade para diferentes fins. Ele complementa dizendo que métodos Bayesianos são universais e podem ser usados a qualquer momento que se tenha interesse em utilizar os dados para obter conhecimento de um determinado fenômeno.

Inicialmente, vamos considerar duas variáveis aleatórias, A e B. Pela definição de probabilidade condicional ⁵, temos que:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A,B)}{\mathbb{P}(B)} \; ; \; \mathbb{P}(B|A) = \frac{\mathbb{P}(A,B)}{\mathbb{P}(A)}$$

Com $\mathbb{P}(A)$ e $\mathbb{P}(A)$ extremamente positivos.

Em que:

• $\mathbb{P}(A, B)$ é a probabilidade conjunta de A e B;

⁵Ver Feller (1968)

- $\mathbb{P}(A|B)$ e $\mathbb{P}(B|A)$ são as probabilidades condicionais de A dado B e de B dado A respectivamente;
- $\mathbb{P}(A)$ e $\mathbb{P}(B)$ são as probabilidades marginais de A e B respectivamente.

Desta forma, temos que:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$
(3.6)

A expressão acima é conhecida como *Teorema de Bayes*⁶, que é considerado o núcleo da inferência Bayesiana.

Ao considerarmos, por exemplo, um modelo de regressão, desejamos estimar os coeficientes, que são chamados de parâmetros do modelo. Desta forma, sejam y a matriz de dados e θ o vetor de parâmetros a serem estimados para explicar o comportamento de y. Estamos interessados em aprender informações sobre θ baseado nos dados y. Considerando a formulação Bayesiana citada anteriormente, podemos substituir A por θ e B por y. Assim, temos que:

$$\mathbb{P}(\theta|y) = \frac{\mathbb{P}(y|\theta)\mathbb{P}(\theta)}{\mathbb{P}(y)}$$
(3.7)

Do ponto de vista bayesiano, o maior interesse se encontra em obter informações sobre a probabilidade de θ dado y. Ou seja, deseja-se saber o que se pode descobrir sobre alguma coisa desconhecida (neste caso, os parâmetros da regressão), dado algo já conhecido (no caso, os dados) e dado, também, a probabilidade da informação desconhecida condicionada às informações previamente conhecidas.

Podemos reescrever a equação 3.7, desconsiderando o termo que não envolve o parâmetro de interesse θ . Assim, temos que:

$$\mathbb{P}(\theta|y) \propto \mathbb{P}(y|\theta)\mathbb{P}(\theta) \tag{3.8}$$

Os termos da equação acima são definidos como:

⁶Ver Koop (2003)

- $\mathbb{P}(\theta|y)$ é conhecido como densidade *posteriori*;
- $\mathbb{P}(y|\theta)$ é conhecido como função densidade de probabilidade para os dados y, dado os parâmetros do modelo (também conhecido como função de verossimilhança);
- $\mathbb{P}(\theta)$ é conhecido como densidade a priori.

A densidade a priori, $\mathbb{P}(\theta)$, não depende das informações contidas nos dados, desta forma, ela resume tudo que se sabe sobre o vetor de parâmetros θ antes de observarmos os dados.

A função de verossimilhança, $\mathbb{P}(y|\theta)$, é a densidade de probabilidade dos dados condicionados ao vetor de parâmetros do modelo (θ) . São, normalmente, chamados de processo gerador de dados. Por exemplo, um modelo de regressão linear possui erros normalmente distribuídos, isso implica que $\mathbb{P}(y|\theta)$ segue uma distribuição Normal, que depende dos parâmetros (coeficientes da regressão e variância dos erros).

A distribuição posteriori, $\mathbb{P}(\theta|y)$, é a densidade de maior interesse na inferência bayesiana. Ela resume tudo o que sabemos sobre o vetor de parâmetros $\mathbb{P}(\theta|y)$ após observarmos os dados. A equação acima pode ser interpretada como uma regra de atualização, em que os dados podem atualizar nossas informações a priori de θ .

Outro ponto de interesse quando estamos criando modelos estatísticos é a comparação entre diferentes modelos. Um modelo é formalmente definido por uma função de verossimilhança e uma distribuição a priori. Suponha que tenhamos m modelos diferentes, M_i , para i=1,2,...,m, para explicar a matriz de dados y. Cada M_i depende de seus respectivos vetores de parâmetros θ^i . Assim, a distribuição posteriori para os parâmetros calculados usando o modelo M_i é obtida da seguinte forma:

$$\mathbb{P}(\theta^i|y, M_i) = \frac{\mathbb{P}(y|\theta^i, M_i)\mathbb{P}(\theta^i|M_i)}{\mathbb{P}(y|M_i)}$$
(3.9)

Usando a primeira equação do Teorema de Bayes (3.6), temos que:

$$\mathbb{P}(M_i|y) = \frac{\mathbb{P}(y|M_i)\mathbb{P}(M_i)}{\mathbb{P}(y)}$$
(3.10)

Assim, como o modelo de probabilidade a priori não envolve os dados, consideramos que $\mathbb{P}(M_i)$ é uma medida que quantifica qual o grau de crença que temos de que o modelo M_i esteja correto antes mesmo de termos acesso aos dados. $\mathbb{P}(y|M_i)$ é conhecida por função de verossimilhança marginal e é calculada utilizando a equação 3.9, ou seja, se integrarmos os dois lados da equação 3.9 em relação a θ_i e isolarmos $\mathbb{P}(y|M_i)$, temos que:

$$\mathbb{P}(y|M_i) = \int \mathbb{P}(y|\theta^i, M_i) \mathbb{P}(\theta^i) d\theta^i$$
(3.11)

Para chegarmos ao resultado acima, usamos o fato de que $\int \mathbb{P}(\theta^i|y, M_i)d\theta^i = 1$.

Note que a função de verossimilhança marginal depende apenas da distribuição a priori $-\mathbb{P}(\theta^i|M_i)$ — e da função de verossimilhança — $\mathbb{P}(y|\theta^i,M_i)$.

Como o denominador da equação 3.10 é difícil de ser obtido diretamente, normalmente, para compararmos dois modelos i e j, utilizamos a razão de chance de distribuições posteriori's. Desta forma, temos:

$$PO_{ij} = \frac{\mathbb{P}(M_i|y)}{\mathbb{P}(M_j|y)} = \frac{\mathbb{P}(y|M_i)\mathbb{P}(M_i)}{\mathbb{P}(y|M_j)\mathbb{P}(M_j)}$$

Considerando que $\mathbb{P}(M_i) = \mathbb{P}(M_j)$, considerando que se queira observar modelos que, mesmo antes de observarmos os próprios dados, acreditamos que possam explicar igualmente os dados. Neste caso, razão de chance de distribuições *posteriori*'s é simplesmente a razão entre as funções de verossimilhança marginais, que passa a ser chamada de *Fator de Bayes*, definido como:

$$FB_{ij} = \frac{\mathbb{P}(y|M_i)}{\mathbb{P}(y|M_j)}$$

Até aqui, mostramos, resumidamente, como construímos modelos através do uso do Teorema de Bayes e como podemos comparar os diversos modelos para um determinado conjunto de dados y. Agora, vamos falar, sucintamente, como podemos fazer previsões utilizando um determinado modelo. Basicamente, é outra maneira de utilizarmos o que foi visto até o momento, ou seja, as projeções devem ser baseadas na densidade predita $\mathbb{P}(\hat{y}|y)$, que pode ser escrita da seguinte forma:

$$\mathbb{P}(\widehat{y}|y) = \int \mathbb{P}(\widehat{y}, \theta) d\theta \stackrel{*}{=} \int \mathbb{P}(\widehat{y}|y, \theta) \mathbb{P}(\theta|y) d\theta^{7}$$

 $^{^{7}}$ * usando a regra de probabilidade apresentada em 3.11

3.3.2 MCMC - Markov Chain Monte Carlo

Métodos de Inferência Bayesiana representam um mecanismo simples de inferência, já que basicamente envolvem a especificação de uma função de verossimilhança e das distribuições a priori para os parâmetros de interesse e, desta forma, podemos obter a distribuição posterior usando apenas o Lema de Bayes. No entanto a existência de uma distribuição posterior em forma analítica só é possível em uma classe limitada de pares de verossimilhança e priores. Apesar disso, podemos utilizar métodos de aproximação numérica para avaliar a distribuição posterior. Um dos métodos mais utilizados neste caso é a classe de algoritmos de Markov Chain Monte Carlo, que simulam uma cadeia de Markov que tem como distribuição estacionária ou medida invariante a distribuição posterior de interesse, permitindo obter a distribuição posterior na ausência de soluções analíticas.

3.3.2.1 A Cadeia de Markov

O processo estocástico mais simples é a Cadeia de Markov, na qual a distribuição de X_t depende apenas de X_{t-1} . Desta forma, sejam χ o espaço amostral e $T=1,2,\cdots$, o conjunto de índices, temos:

Definição: o processo $\{X_n : n \in T\}$ é uma cadeia de Markov se:

$$\mathbb{P}(X_n|X_0, X_2, \cdots, X_{n-1}) = \mathbb{P}(X_n|X_{n-1})$$

Um conceito muito utilizado em Cadeia de Markov é a probabilidade de transição, que nos diz qual a probabilidade de passarmos de um estado para outro. Desta forma, temos a seguinte definição.

Definição: Seja a probabilidade de transição dada por

$$\mathbb{P}(X_{n+1} = j | X_n = i),$$

se a probabilidade de transição não mudar ao longo do tempo, dizemos que a cadeia é homogênea. Neste caso, definimos $p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i)$ e a matriz \mathbf{P} , cujo elemento (i,j) é p_{ij} , é conhecida como matriz de transição. Consideraremos apenas matrizes homogêneas.

Note que a matriz \mathbf{P} tem duas propriedades:

- $p_{ij} \geq 0$
- $\sum_i p_{ij} = 1$

Assim, cada linha da matriz ${f P}$ representa uma função de probabilidade e tal matriz será conhecida como matriz estoc ástica.

Seja $p_{ij}(n) = \mathbb{P}(X_{m+n} = j|X_m = i)$ a probabilidade de sair do estado i para o estado j em n passos, assim \mathbf{P}_n será a matriz cujo elemento (i,j) é $p_{ij}(n)$, que são chamados de probabilidade de transição em n passos.

Usando o teorema das equações de *Chapman-Komolgorov*, é possível provar que $\mathbf{P}_{m+n} = \mathbf{P}_m \mathbf{P}_n$ e que $\mathbf{P}_n = \mathbf{P}^n = \mathbf{P} \times \mathbf{P} \times \cdots \times \mathbf{P}$ (matrix \mathbf{P} multiplicada por ela mesma n vezes)⁸.

Um resultado muito útil ao utilizarmos a Cadeia de Markov é o seguinte:

$$p_n = p_0 \cdot \mathbf{P}^n$$

Em que:

- $p_n(i) = \mathbb{P}(X_n = i)$ denota a probabilidade marginal de que a cadeia, no período n, esteja no estado i;
- em particular, p_0 é chamado de distribuição inicial.

Para simularmos uma Cadeia de Markov, precisamos apenas conhecer a matriz de transição \mathbf{P} e a distribuição inicial p_0 .

As definições⁹ a seguir são importantes de ser citadas, pois serão pré-requisito para a definição do método MCMC.

Definições:

1. um estado é recorrente (ou persistente) se

$$\mathbb{P}(X_n = i | X_0 = i) = 1$$

⁸ver prova e mais propriedades da matriz de transição em Wasserman (2010)

⁹para maiores detalhes ver Wasserman (2010)

para qualquer $n \ge 1$; caso contrário, o estado é dito ser transiente;

- 2. periodicidade de um estado i é definido como o número mínimo de passos (n) necessários para sair deste estado i e voltar para ele mesmo. Um estado é dito aperiódico quando n=1. Uma cadeia é aperiódica quando todos os seus estados são aperiódicos;
- 3. tempo médio de recorrência de um estado recorrente i é dado por

$$m_i = E(T_{ii}) = \sum_n n f_{ii}(n)$$

Em que $f_{ii}(n) = \mathbb{P}(X_1 \neq j, X_2 \neq j, \dots, X_{n-1} \neq j, X_n = j | X_0 = i)$. Um estado recorrente é dito nulo se $m_i = \infty$, caso contrário o estado é $n\tilde{a}o$ nulo;

- 4. um estado é *ergódico* se ele for recorrente, não nulo e aperiódico. Uma cadeia é ergódica se todos os estados forem ergódicos;
- 5. a função de distribuição f é uma distribuição estacion ária se

$$f = f \mathbb{P}$$
.

3.3.2.2 O método MCMC

Na literatura, infelizmente, só encontramos fórmulas analíticas para poucos casos em inferência bayesiana. Desta maneira, uma forma alternativa para casos mais gerais como o que gostaríamos de propor com este estudo é recorrermos a métodos numéricos, sendo o método de simulação Monte Carlo em Cadeia de Markov (ou *Markov Chain Monte Carlo - MCMC*) o mais utilizado.

Um método de simulação MCMC de uma distribuição f, por definição 10 , é qualquer método que produz uma Cadeia de Markov ergódica cuja distribuição estacionária é a f. Ele é um método genérico para simulações de amostragens de distribuições posterior's. É utilizado para gerar sucessivas amostras a partir de uma distribuição pré definida. Cada amostragem depende da amostragem anterior, por isso que está atrelado à Cadeia de Markov, uma vez que esta última é uma sequência de variáveis aleatórias θ^1 , θ^2 , θ^3 , \cdots , em que θ^t depende única e exclusivamente de θ^{t-1} , como visto anteriormente. Desta forma, uma Cadeia de Markov é dita sem memória, dado que a variável aleatória em questão depende apenas do que aconteceu no instante imediatamente anterior.

¹⁰Ver Robert e Casella (2005)

* Algoritmo Metropolis-Hastings

Suponhamos que estejamos interessados em obter T amostras de uma distribuição univariada cuja função densidade de probabilidade (f.d.p.) seja representada por $f(\theta|y)$ e que θ^t seja a t-ésima amostra gerada por f. Para utilizarmos o algoritmo Metropolis-Hastings, é preciso que tenhamos um valor inicial θ^0 e uma densidade $q(\theta^{t+1}|\theta^t)^{11}$, conhecida como distribuição instrumental ou distribuição proposta. Para a (t+1)-ésima iteração, o algoritmo gera uma amostra a partir de $q(\cdot|\cdot)$ baseado na amostra atual θ^t e, assim, faz-se a escolha entre aceitar ou rejeitar a nova amostra. Se a nova amostra for aceita, o algoritmo será repetido, mas o ponto inicial passa a ser a nova amostra (neste exemplo, seria θ^{t+1}). Caso contrário, o algoritmo também será repetido, porém o ponto de partida será a amostra atual (θ^t) . O algoritmo é auto-repetitivo e, na prática, precisamos decidir, inicialmente, qual o número total de amostras que será necessário e a simulação parará somente depois que todas as iterações forem concluídas.

Resumidamente, temos o seguinte raciocínio:

1. Definir um valor inicial θ^0 , que pode ser um valor arbitrário contanto que:

$$f(\theta^0|y) > 0;$$

- 2. Gerar uma nova amostra θ_{novo} usando a distribuição pré definida $q(\cdot|\theta^t)$;
- 3. Calcular o seguinte valor:

$$r = \min \left\{ \frac{f(\theta_{novo}|y)}{f(\theta^t|y)} \frac{q(\theta^t|\theta^{novo})}{q(\theta^{novo}|\theta^t)}, 1 \right\}$$

- 4. Gerar uma variável aleatória u, tal que esta variável seja uniformemente distribuída entre 0 e 1, ou seja, $u \sim U(0,1)$;
- 5. Definir $\theta^{t+1} = \theta_{novo}$, se u < r. Caso contrário, definir $\theta^{t+1} = \theta^t$;
- 6. Atribuir t = t + 1 e voltar ao passo 2 enquanto t < T. Parar quando t = T.

Uma condição mínima necessária é que seja válida a seguinte expressão

$$\bigcup_{\theta \in \varepsilon} suportes \ de \ q(\cdot|\theta) \supset \varepsilon$$

Em que ε representa o suporte da distribuição f. Ainda assim, é necessário verificar se houve convergência na simulação¹².

 $^{^{11}}$ Se $q(\theta^i|\theta^j)=q(\theta^j|\theta^i)$, dizemos que $q(\cdot|\cdot)$ é simétrica e o algoritmo é conhecido como algoritmo *Metropolis*

¹²Para maiores detalhes, ver Robert e Casella (2005)

* Amostragem de Gibbs

A amostragem de *Gibbs* é um caso especial do método de *Metropolis*. Segundo Casella e George (1992), a amostragem de *Gibbs* é uma técnica para gerar variáveis aleatórias, indiretamente, a partir de uma distribuição marginal, sem que precisemos calcular a função densidade. Assim, com esta técnica, é possível evitar cálculos mais complicados, substituindo-os por cálculos mais simples.

Sejam o vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$, $p(y|\boldsymbol{\theta})$ a função de verossimilhança e $f(\boldsymbol{\theta})$ a distribuição a priori. A distribuição condicional posteriori completa $f(\theta_i|\theta_j, i \neq j, y)$ será proporcional à densidade conjunta posteriori, ou seja,

$$f(\theta_i|\theta_j, i \neq j, y) \propto p(y|\boldsymbol{\theta}) f(\boldsymbol{\theta})$$

Uma característica particular sobre este método é que a função densidade $f(\theta_i|\theta_j, i \neq j, y)$ é a única densidade utilizada para fazer a simulação. Assim, a vantagem é de que, mesmo em um problema que apresente uma alta dimensão, todas as simulações serão univariadas.

A amostragem de Gibbs pode ser descrita da seguinte maneira:

- 1. Definir um valor inicial arbitrário $\boldsymbol{\theta}^{(0)} = \left\{\theta_1^{(0)}, \theta_2^{(0)}, \cdots, \theta_k^{(0)}\right\}$
- 2. Gerar cada componente de θ através de:
 - $\theta_1^{(t+1)} \sim f_1(\theta_1|\theta_2^{(t)}, \theta_3^{(t)}, \cdots, \theta_k^{(t)}, y)$
 - $\theta_2^{(t+1)} \sim f_2(\theta_2|\theta_1^{(t)}, \theta_3^{(t)}, \cdots, \theta_k^{(t)}, y)$

• •

- $\theta_k^{(t+1)} \sim f_k(\theta_k | \theta_1^{(t)}, \theta_2^{(t)}, \cdots, \theta_{k-1}^{(t)}, y)$
- 3. Atribuir t = t + 1 e voltar ao passo 2 enquanto t < T. Parar quando t = T.

Uma outra maneira que podemos escrever uma amostragem de Gibbs é considerar a seguinte transformação:

Definição: Dada uma função densidade de probabilidade f, a densidade que satisfizer a seguinte relação é chamada de $realização^{13}$ de f.

$$\int_{\mathbf{Z}} g(\theta, z) dz = f(\theta)$$

¹³Ver Robert e Casella (2005)

Desta forma, o algoritmo da amostragem de Gibbs será feito com a utilização de g quando, por exemplo, cada densidade condicional associada a f não for explícita. Consideraremos, então, o seguinte algoritmo, em que $w^{(t)} = (\theta^{(t)}, z^{(t)})$ e $g(w|y) = g(w_1, w_2, \dots, w_k|y)$, cujas densidades conjuntas são $W_1|w_2, \dots, w_k \sim g_1(w_1|w_2, \dots, w_k, y)$, $\dots, W_k|w_1, \dots, w_{k-1} \sim g_k(w_1|w_1, \dots, w_{k-1}, y)$

- 1. Definir um valor inicial arbitrário $\boldsymbol{w}^{(0)} = \left\{w_1^{(0)}, w_2^{(0)}, \cdots, w_k^{(0)}\right\}$
- 2. Gerar cada componente de \boldsymbol{w} através de:
 - $W_1^{(t+1)} \sim g_1(w_1|w_2^{(t)}, w_3^{(t)}, \cdots, w_k^{(t)}, y)$
 - $W_2^{(t+1)} \sim g_2(w_2|w_1^{(t)}, w_3^{(t)}, \cdots, w_k^{(t)}, y)$
 - $W_k^{(t+1)} \sim g_k(w_k|w_1^{(t)}, w_2^{(t)}, \cdots, w_{k-1}^{(t)}, y)$
- 3. Atribuir t = t + 1 e voltar ao passo 2 enquanto t < T. Parar quando t = T.

Teorema: para o algoritmo apresentado acima, se (W^t) for ergódico, então a distribuição g é uma distribuição estacionária para a cadeia (W^t) e f é a distribuição limite da sub-cadeia $(\theta^t)^{14}$.

O método de amostragem de Gibbs mostrado acima, utilizando a densidade g (que é a função de realização de f), é equivalente à composição de p algoritmos Metropolis-Hastings, cujas probabilidades de aceite são uniformemente iguais a 1^{15} .

* MCMC híbrido

Robert e Casella (2005) citam que as desvantagens do algoritmo Metropolis-Hastings são diferentes daquelas da amostragem de Gibbs, uma vez que aquele é mais frequentemente relacionado a uma possível má concordância entre a densidade f e a distribuição $instrumental, q(\cdot|\cdot)$. Apesar disso, a liberdade apresentada no algoritmo de Metropolis-Hastings permite, às vezes, solucionar tal desvantagem através da modificação de algumas escalas (parâmetros ou hiperparâmetros são particularmente úteis).

Ao analisarmos a amostragem de Gibbs, ela apresenta uma desvantagem em relação ao algoritmo Metropolis-Hastings ao atribuir uma nova distribuição g (realização de f) que não necessariamente apresenta um sistema de coordenadas que combina com a forma da distribuição f. Em Hills e Smith (1992) e Hills e Smith (1993) 16 , os autores propõem exemplos em que parametrizações incorretas do modelo aumentam significativamente o tempo de convergência para a amostragem de Gibbs.

¹⁴Ver prova em Robert e Casella (2005)

¹⁵Ver prova em Robert e Casella (2005)

¹⁶Por Robert e Casella (2005)

Todavia, seguindo Tierney (1994)¹⁷, uma maneira de tirarmos vantagens de ambos algoritmos é implementar uma abordagem híbrida, na qual usa-se tanto o algoritmo *Metropolis-Hasting* quanto a amostragem de *Gibbs*, conhecida como MCMC híbrido.

O algoritmo, com modificações propostas por Muller $(1991)^{18}$, é dado da seguinte maneira:

- 1. Para $i=1,2,\cdots,k$, definir um valor inicial arbitrário $\boldsymbol{y}^{(0)} = \left\{ y_1^{(1)}, y_2^{(1)}, \cdots, y_{i-1}^{(1)}, y_i^{(0)}, y_{i+1}^{(0)}, \cdots, y_k^{(0)} \right\}$
- 2. Gerar \tilde{y}_i a partir de

$$\widetilde{y}_i \sim q_i \left(y_i | y_1^{(t+1)}, \cdots, y_{i-1}^{(t+1)}, y_i^{(t)}, y_{i+1}^{(t)}, \cdots, y_k^{(t)} \right)$$

3. Usar a seguinte regra para obter $y_i^{(t+1)}$

$$y_i^{(t+1)} = \begin{cases} y_i^{(t)} & com \ probabilidade \ 1 - \rho \\ \widetilde{y}_i & com \ probabilidade \ \rho \end{cases}$$

em que

$$\rho = \min \left\{ 1; \frac{\left(\frac{g_i(\widetilde{y}_i|y_1^{(t+1)}, \cdots, y_{i-1}^{(t+1)}, y_{i+1}^{(t)}, \cdots, y_k^{(t)})}{q_i(\widetilde{y}_i|y_1^{(t+1)}, \cdots, y_{i-1}^{t+1}, y_i^{(t)}, y_{i+1}^{(t)}, \cdots, y_p^{(t)})}\right)}{\left(\frac{g_i(y_i^{(t)}|y_1^{(t+1)}, \cdots, y_{i-1}^{(t+1)}, y_{i+1}^{(t)}, \cdots, y_p^{(t)})}{q_i(y_i^{(t)}|y_1^{(t+1)}, \cdots, y_{i-1}^{(t+1)}, \widetilde{y}_i, y_{i+1}^{(t)}, \cdots, y_p^{(t)})}\right)}\right\}$$

Robert e Casella (2005) citam que um ponto importante sobre esta substituição é que o passo Metropolis-Hastings acima é apenas usado em uma iteração da simulação pela amostragem de Gibbs com a modificação que utiliza a função g (realização de f). Assim, o passo modificado utiliza apenas uma única simulação de \tilde{y}_i , ao invés de tentar aproximar $g_i(y_i|y_j, j \neq i)$ mais precisamente através de T simulações a partir de q_i . Os autores nos dão dois motivos para esta escolha, são elas:

- 1. o algoritmo híbrido resultante é valido uma vez que g é a distribuição estacionária da cadeia;
- 2. a amostragem de Gibbs também conduz a uma aproximação de T.

Ao se escolher uma aproximação mais precisa de $g_i(y_i|y_j, j \neq i)$ neste último algoritmo apresentado não necessariamente nos leva a uma melhor aproximação de g e a substituição de g_i por q_i deve ser ainda mais benéfica para a velocidade de convergência da cadeia.

¹⁷Por Robert e Casella (2005)

¹⁸Por Robert e Casella (2005)

3.4 MIDAS - Mixed Data Sampling

A disponibilidade de dados coletados em diferentes frequências é sempre um dilema para pesquisadores que trabalham com dados de séries temporais, conforme citado em Ghysels et al. (2007). Por um lado, variáveis que estão disponíveis com uma frequência alta contêm informações potencialmente importantes. Por outro lado, os pesquisadores acabam não podendo utilizar tais informações diretamente se algumas variáveis são disponibilizadas com uma frequência mais baixa, uma vez que regressões de séries de tempo envolvem dados amostrados em um mesmo intervalo. A solução mais comum, nestes casos, é fazer uma pré seleção dos dados que faça com que todas as informações analisadas apresentem a mesma frequência, no entanto, desta forma, muita informação que poderia ser útil seria descartada, tornando a relação entre as variáveis difícil de ser encontrada. Uma alternativa a esse tipo de problema é a utilização do método MIDAS, que pode ser descrito da seguinte forma.

Considere um processo MIDAS $\left\{y_t,\mathbf{x}_{t/m}^{(m)}\right\}$ em que y_t é observado em $t=1,\cdots,T$ e $\mathbf{x}_{t/m}^{(m)}=\left(1,x_{1,t/m}^{(m)},x_{2,t/m}^{(m)},\cdots,x_{p,t/m}^{(m)}\right)'$ é um vetor p-dimensional dos dados que apresentam a maior frequência, observados m vezes entre o instante t e t-1. Vamos assumir aqui que os dados são fracamente dependentes. Uma abordagem convencional utiliza a regressão linear sujeita a restrições de agregação temporal dos dados considerando uma ponderação igual para todos os tempos. Mas a proposta é considerar uma ponderação diferente, como a sugerida em Andreou et al. (2010), em que eles, primeiramente, fazem a projeção dos dados de maior frequência sobre \mathbf{y} . Tal projeção é caracterizada por um polinômio defasagem de alta frequência, $W\left(L^{1/m}; \boldsymbol{\theta}\right)$, que será parametrizado por um vetor de parâmetros $\theta_{r\times 1}$. Assim, a abordagem sugerida é representada a seguir.

$$y_t = \boldsymbol{\beta}' \mathbf{x}_t(\boldsymbol{\theta}) + u_t$$

tal que

$$oldsymbol{eta} = egin{bmatrix} eta_0 \ eta_{1,1} \ dots \ eta_{1,p} \end{bmatrix},$$

$$\mathbf{x}_{t}(\boldsymbol{\theta}) = \begin{bmatrix} 1 \\ x_{1,t}^{(m)}(\theta_{1}) \\ \vdots \\ x_{p,t}^{(m)}(\theta_{p}) \end{bmatrix}$$

e

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_n)$$
 tal que $\boldsymbol{\theta}_k = (\theta_{k,1}, \theta_{k,2})$

Em que $t=1,2,\cdots,T,\ u_t$ é um processo martingale diferença em relação ao espaço-sigma $\mathbb{F}_{t-1},\ E\left(u_t^2\right)=\sigma^2<\infty$ e $\mathbf{x}_t(\boldsymbol{\theta})$ é uma função não linear que mapeia as informações de maior frequência para dentro de uma frequência menor que seja comum a todas as variáveis, tal que:

$$x_{k,t}^{(m)} = W\left(L^{1/m}; \boldsymbol{\theta}_k\right) x_{k,t}^{(m)} = \sum_{j=1}^q w_{j,k}(\boldsymbol{\theta}_k) L^{j/m} x_{k,t/m}^{(m)}, \quad k = 1, 2, 3, \dots, p$$

E assumiremos que

 $\checkmark L^{j/m}$ é o operador defasagem tal que

$$L^{j/m} x_t^{(m)} = x_{t - \frac{j}{m}}^{(m)}$$

$$\checkmark w_{j,k}(\boldsymbol{\theta_k}) \in (0,1)$$

$$\checkmark \sum_{j=1}^{q} w_{j,k}(\boldsymbol{\theta}_k) = 1$$

As duas últimas condições citadas acima permitem a identificação do vetor coeficiente de inclinação, β .

Seguindo Ghysels et al. (2007), os coeficientes de defasagem, $w_{j,k}(\boldsymbol{\theta}_k)$, os quais determinam a ponderação que será utilizada na agregação temporal, serão obtidos através de dois parâmetros $\theta_{k,1}$ e $\theta_{k,2}$, isto é,

$$w_{j,k}(\theta_{k,1}, \theta_{k,2}) = \frac{f\left(\frac{j}{m}; \theta_{k,1}, \theta_{k,2}\right)}{\sum_{i=1}^{m} f\left(\frac{j}{m}; \theta_{k,1}, \theta_{k,2}\right)}$$

Tal que

$$f(x, a, b) = \frac{x^{a-1}(1-x)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

e

$$\Gamma(a) = \int_0^\infty e^{-x} x^{a-1} dx$$

Note que, quando $\theta_{k,1} = \theta_{k,2} = 1$, temos a agregação temporal que considera uma ponderação igual para todos os tempos.

3.5 Descrição do modelo

3.5.1 O modelo

O objetivo desta seção é agregar todas as metodologias descritas até aqui e estruturar de forma completa o modelo proposto por este trabalho.

Sejam

$$y_t(\tau) = N_t + I_t \left(\frac{1 - e^{\lambda_t \tau}}{\lambda_t \tau} \right) + C_t \left(\frac{1 - e^{\lambda_t \tau}}{\lambda_t \tau} - e^{\lambda_t \tau} \right) + \sigma_t \epsilon_t$$
 (3.12)

$$\ln \sigma_t^2 = \mu_\sigma + \phi_\sigma \ln \sigma_{t-1}^2 + e_t \tag{3.13}$$

$$\begin{bmatrix} N_{t} \\ I_{t} \\ C_{t} \\ UCI_{t} \\ TBJ_{t} \\ Infl_{t} \end{bmatrix} = \begin{bmatrix} \mu_{N} \\ \mu_{I} \\ \mu_{C} \\ \mu_{UCI} \\ \mu_{TBJ} \\ \mu_{Infl} \end{bmatrix} + \mathbf{\Phi} \begin{bmatrix} N_{t-1} \\ I_{t-1} \\ C_{t-1} \\ UCI_{t-1} \\ TBJ_{t-1} \\ Infl_{t-1} \end{bmatrix} + \begin{bmatrix} v_{N,t} \\ v_{I,t} \\ v_{C,t} \\ v_{UCI,t} \\ v_{TBJ,t} \\ v_{Infl,t} \end{bmatrix}$$
(3.14)

tal que

$$\begin{bmatrix} v_{N,t} \\ v_{I,t} \\ v_{C,t} \\ v_{UCI,t} \\ v_{TBJ,t} \\ v_{Infl,t} \end{bmatrix} = \begin{bmatrix} \zeta_{1,t} \, \sigma_{N,t} \\ \zeta_{2,t} \, \sigma_{I,t} \\ \zeta_{3,t} \, \sigma_{C,t} \\ \zeta_{4,t} \, \sigma_{UCI,t} \\ \zeta_{5,t} \, \sigma_{TBJ,t} \\ \zeta_{6,t} \, \sigma_{Infl,t} \end{bmatrix}$$

$$(3.15)$$

e

$$\begin{bmatrix} \ln \sigma_{N,t}^{2} \\ \ln \sigma_{I,t}^{2} \\ \ln \sigma_{C,t}^{2} \\ \ln \sigma_{C,t}^{2} \\ \ln \sigma_{TBJ,t}^{2} \\ \ln \sigma_{Infl}^{2} \end{bmatrix} = \begin{bmatrix} \mu_{\sigma_{N}} \\ \mu_{\sigma_{I}} \\ \mu_{\sigma_{C}} \\ \mu_{\sigma_{UCI}} \\ \mu_{\sigma_{TBJ}} \\ \mu_{\sigma_{Infl}} \end{bmatrix} + \mathbf{\Psi} \begin{bmatrix} \ln \sigma_{N,t-1}^{2} \\ \ln \sigma_{I,t-1}^{2} \\ \ln \sigma_{C,t-1}^{2} \\ \ln \sigma_{C,t-1}^{2} \\ \ln \sigma_{TBJ,t-1}^{2} \\ \ln \sigma_{TBJ,t-1}^{2} \\ \ln \sigma_{Infl,t-1}^{2} \end{bmatrix} + \begin{bmatrix} \gamma_{t,\sigma_{N}} \\ \gamma_{t,\sigma_{I}} \\ \gamma_{t,\sigma_{C}} \\ \gamma_{t,\sigma_{UCI}} \\ \gamma_{t,\sigma_{TBJ}} \\ \gamma_{t,\sigma_{Infl}} \end{bmatrix}$$

$$(3.16)$$

Em que

$$\epsilon_t \sim Normal(0,1)$$
 e $\epsilon_t \perp \epsilon_s, \ \forall t \neq s$

$$e_t \sim Normal(0, \sigma_e^2)$$
 e $e_t \perp e_s, \ \forall t \neq s$

$$\begin{bmatrix} \zeta_{1,t} \\ \zeta_{2,t} \\ \zeta_{3,t} \\ \zeta_{4,t} \\ \zeta_{5,t} \\ \zeta_{6,t} \end{bmatrix} \stackrel{iid}{\sim} Normal \begin{pmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \mathbb{I}_{(6 \times 6)} \\ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

com $\mathbb{I}_{(6\times 6)}$ representando uma matriz identidade de dimensão 6×6 .

$$\begin{bmatrix} \gamma_{t,\sigma_{N}} \\ \gamma_{t,\sigma_{I}} \\ \gamma_{t,\sigma_{C}} \\ \gamma_{t,\sigma_{UCI}} \\ \gamma_{t,\sigma_{TBJ}} \\ \gamma_{t,\sigma_{Infl}} \end{bmatrix} \sim Normal \begin{pmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}_{\gamma} \\ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

tal que

$$oldsymbol{\Sigma}_{\gamma} = egin{bmatrix} \sigma_{\gamma}^{N} & 0 & 0 & 0 & 0 & 0 & 0 \ & \sigma_{\gamma}^{I} & 0 & 0 & 0 & 0 \ & & \sigma_{\gamma}^{C} & 0 & 0 & 0 \ & & & \sigma_{\gamma}^{UCI} & 0 & 0 \ & & & \sigma_{\gamma}^{TBJ} & 0 \ & & & & \sigma_{\gamma}^{Infl} \end{bmatrix}$$

em que a matriz Σ_{γ} é uma matriz simétrica e representa a matriz variância-covariância dos resíduos γ , em que a diagonal principal apresenta as variâncias de cada uma das log-volatilidades estocásticas de suas respectivas variáveis e os termos fora da diagonal principal representam as covariâncias entre todas as volatilidades estocásticas, em que, neste estudo, vamos considerar que não há correlação entre as volatilidades.

Desta forma, na equação 3.14, UCI_t representa a utilização da capacidade industrial, TBJ_t representa a taxa básica de juros americana ($Fed\ Funds\ Rate$), $Infl_t$ representa a mudança no índice de preço ao consumidor. Por fim, Φ é a matriz de parâmetros a serem estimados do VAR(1). Apenas a taxa básica de juros é medida diariamente, as outras duas variáveis são capturadas mensalmente. De acordo com Laurini e Caldeira (2013), as três variáveis UCI, TBJ e Infl representam, repectivamente, o nível real de atividade em relação a seu potencial, o instrumento de política monetária e a inflação. Adicionalmente, será através da equação 3.14 que poderemos avaliar o quanto da variabilidade das informações macroeconômicas será explicada pelos fatores do modelo dinâmico de Nelson-Siegel.

Em 3.15, temos que os termos $\zeta_{h,t}$, com $h=1,2,\cdots,6$, representam um processo ruído branco normalmente distribuídos com variância unitária. E, adicionalmente, as respectivas volatilidades apresentam uma estrutura tal que contenha um componente

estocástico $(\gamma_{t,\sigma_l}, \text{ com } l = (N, I, C, UCI, TBJ, Infl))$ como mostrado em 3.16. Como será mostrado abaixo, a matriz Ψ é representada por uma matriz diagonal para garantir que não haverá correlação entre as diferentes volatilidades de cada um dos fatores latentes.

A inclusão das variáveis macroeconômicas no modelo nos permite verificar o quanto da dinâmica da taxa de juros é explicada por informações da economia, entretanto, o modelo não se restringe a apenas isso, podendo avaliar a influência da estrutura a termo da taxa de juros sobre as variáveis macroeconômicas.

As matrizes Φ e Ψ podem ser representadas por:

$$\mathbf{\Phi} = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} & \phi_{14} & \phi_{15} & \phi_{16} \\ \phi_{21} & \phi_{22} & \phi_{23} & \phi_{24} & \phi_{25} & \phi_{26} \\ \phi_{31} & \phi_{32} & \phi_{33} & \phi_{34} & \phi_{35} & \phi_{36} \\ \phi_{41} & \phi_{42} & \phi_{43} & \phi_{44} & \phi_{45} & \phi_{46} \\ \phi_{51} & \phi_{52} & \phi_{53} & \phi_{54} & \phi_{55} & \phi_{56} \\ \phi_{61} & \phi_{62} & \phi_{63} & \phi_{64} & \phi_{65} & \phi_{66} \end{bmatrix}$$

e

$$\mathbf{\Psi} = \begin{bmatrix} \psi_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & \psi_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & \psi_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & \psi_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \psi_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & \psi_{66} \end{bmatrix}$$

Conforme Laurini e Hotta (2010), a especificação de primeira ordem para o modelo autoregressivo não apresenta nenhuma limitação teórica, podendo ser utilizada alguma outra ordem superior.

Adicionalmente, temos que o parâmetro de decaimento segue um processo autoregressivo de primeira ordem com a seguinte estrutura:

$$\ln \lambda_t = \mu_\lambda + \phi_\lambda \ln \lambda_{t-1} + \eta_t \tag{3.17}$$

$$\eta_t \sim Normal(0, \sigma_n)$$
 e $\eta_t \perp \eta_s, \ \forall t \neq s$

A aplicação da metodologia MIDAS é feita simultaneamente com a estimação dos parâmetros do VAR(1), tal que, por exemplo, para a variável UCI, temos que:

$$UCI_{t} = \mu_{UCI} + \phi_{41} \sum_{i=1}^{dias} {}^{\acute{u}teis} \omega_{i}^{N} N_{t-\frac{i}{dias}} + \phi_{42} \sum_{i=1}^{dias} {}^{\acute{u}teis} \omega_{i}^{I} I_{t-\frac{i}{dias}} +$$

$$+ \phi_{43} \sum_{i=1}^{dias} {}^{\acute{u}teis} \omega_{i}^{C} C_{t-\frac{i}{dias}} + \phi_{44} UCI_{t-1} +$$

$$+ \phi_{45} \sum_{i=1}^{dias} {}^{\acute{u}teis} \omega_{i}^{TBJ} TBJ_{t-\frac{i}{dias}} + \phi_{46} Infl_{t-1} + \zeta_{4,t} \sigma_{UCI,t}$$

$$(3.18)$$

em que ϕ_{41} , ϕ_{42} , ϕ_{43} , ϕ_{44} , ϕ_{45} , ϕ_{46} são elementos da matriz $\mathbf{\Phi}$ e ω_i^N , ω_i^I , ω_i^C e ω_i^{TBJ} são os pesos atribuídos pela regressão MIDAS referentes aos fatores nível, inclinação e curvatura e à variável taxa básica de juros americana respectivamente, com $i=1,2,\cdots,dias$ úteis, para a variável utilização da capacidade industrial. O mesmo racional também é feito para a variável $Infl_t$ e, assim, temos a seguinte especificação:

$$Infl_{t} = \mu_{Infl} + \phi_{61} \sum_{i=1}^{dias \, \acute{u}teis} \varphi_{i}^{N} N_{t-\frac{i}{dias}} + \phi_{62} \sum_{i=1}^{dias \, \acute{u}teis} \varphi_{i}^{I} I_{t-\frac{i}{dias}} +$$

$$+ \phi_{63} \sum_{i=1}^{dias \, \acute{u}teis} \varphi_{i}^{C} C_{t-\frac{i}{dias}} + \phi_{64} UCI_{t-1} +$$

$$+ \phi_{65} \sum_{i=1}^{dias \, \acute{u}teis} \varphi_{i}^{TBJ} TBJ_{t-\frac{i}{dias}} + \phi_{66} Infl_{t-1} + \zeta_{6,t} \, \sigma_{Infl,t}$$

$$(3.19)$$

em que ϕ_{61} , ϕ_{62} , ϕ_{63} , ϕ_{64} , ϕ_{65} , ϕ_{66} são elementos da matriz $\mathbf{\Phi}$ e φ_i^N , φ_i^I , φ_i^C e φ_i^{TBJ} são os pesos atribuídos pela regressão MIDAS referentes aos fatores nível, inclinação e curvatura e à variável taxa básica de juros americana respectivamente, com $i=1,2,\cdots,dias$ úteis, para a variável inflação.

A matriz de variância e covariância é representada por:

$$\sum_{e,\,\gamma,\,\eta} = egin{bmatrix} \sigma_e^2 & \mathbf{0} & 0 \ \mathbf{0} & \mathbf{\Sigma}_\gamma & \mathbf{0} \ 0 & \mathbf{0} & \sigma_\eta^2 \end{bmatrix}$$

Nesta especificação, que deve ser considerada como um modelo de espaço de estado não linear, a equação 3.12 corresponde à equação de medição, conectando as taxas observadas y_t que descrevem a taxa de juros como funções da maturidade do título, m.

A matriz de variância-covariância, $\sum_{e,\gamma,\eta}$, contém:

- \checkmark a variância escalar, $\sigma_e^2,$ da equação de medida descrita em 3.13;
- \checkmark a matriz de variância-covariância, Σ_{γ} , da estrutura de log-volatilidade descrita em 3.16;
- \checkmark a variância escalar, σ_{η}^2 , do processo autoregressivo de ordem 1 referente ao parâmetro de decaimento λ_t descrita em 3.17.

Por fim, a equação 3.13 descreve os componentes da volatilidade estocástica para os erros da equação de medida. Esta estrutura segue a especificação original do modelo de volatilidade estocástica proposta por Taylor (1986). A presença da heteroscedasticidade é um fato estilizado na modelagem das séries de taxa de juros e, conforme Laurini e Hotta (2010), a adição de componentes de volatilidade estocástica é especialmente importante em momentos de mudanças no formato da curva *yield*, especialmente porque estes momentos estão ligados a grandes incertezas sobre a taxa de juros futura e as expectativas sobre as políticas fiscal e monetária.

3.5.2 Procedimento de estimação

O procedimento de estimação do modelo deste trabalho está muito em linha com o que foi feito no texto de Laurini e Caldeira (2013). Em que considera-se que o modelo pode ser escrito através de uma estrutura hierárquica com três níveis e 11 processos latentes, são eles:

✓ 3 fatores para média condicional

$$F_t \equiv \{N_t, I_t, C_t\}$$

 \checkmark 1 fator de decaimento

 λ_t

✓ 7 volatilidades latentes

$$V_{t} \equiv \left\{ \ln \sigma_{t}^{2}, \, \ln \sigma_{t,N}^{2}, \, \ln \sigma_{t,I}^{2}, \, \ln \sigma_{t,C}^{2}, \, \ln \sigma_{t,UCI}^{2}, \, \ln \sigma_{t,TBJ}^{2}, \, \ln \sigma_{t,Infl}^{2} \right\}$$

 \checkmark Adicionalmente, a matriz de parâmetros será dada por Θ , que contém

$$\mu_{\lambda}, \ \phi_{\lambda}, \ \mu_{N}, \ \mu_{I}, \ \mu_{C}, \ \mu_{UCI}, \ \mu_{TBJ}, \ \mu_{Infl},$$

$$\omega_{i}^{N}, \ \omega_{i}^{I}, \ \omega_{i}^{C}, \ \omega_{i}^{TBJ}, \ \varphi_{i}^{N}, \ \varphi_{i}^{I}, \ \varphi_{i}^{C}, \ \varphi_{i}^{TBJ} \ \mathbf{e} \ \mathbf{\Phi}$$

Desta forma, a função de verossimilhança será dada por:

$$p\left(\Theta|y\right) = \int_{F_1} \int_{F_2} \cdots \int_{F_T} p\left(Y|\Theta, F_1, F_2, \cdots, F_t\right) \cdot p\left(F_1, F_2, \cdots, F_t|\Theta\right) dF_1 dF_2 \cdots dF_T$$

Em que $p(Y|\Theta, F_1, F_2, \dots, F_t)$ é a densidade condicional de Y como uma função dos parâmetros Θ e dos fatores latentes e $p(F_1, F_2, \dots, F_t|\Theta)$ é a densidade condicional dos fatores latentes dada a matriz de parâmetros Θ . Como citado pelos autores, uma vez que os fatores latentes não são observados, é preciso que eles sejam marginalizados, conduzindo-nos a uma integral multidimensional mostrada a seguir.

$$p(F_1, F_2, \cdots, F_T | \Theta) =$$

$$\int_{V_1} \int_{V_2} \cdots \int_{V_T} p\left(F_1, F_2, \cdots, F_T \middle| \Theta, V_1, V_2, \cdots, V_T\right) \cdot p\left(V_1, V_2, \cdots, V_T \middle| \Theta\right) dV_1 dV_2 \cdots dV_T$$

Conforme Laurini e Caldeira (2013), a estrutura de priori's de Φ é dada por uma distribuição Normal multivariada, cuja matriz de variância e covariância segue uma distribuição de Wishart para o vetor autoregressivo dos parâmetros definindo a evolução de $n\'{i}vel~(N_t)$, $inclinação~(I_t)$ e curvatura (C_t) . Os parâmetros da equação 3.17 seguem uma distribuição normal, cuja variância segue uma distribuição Gama Inversa

Ainda de acordo com Laurini e Caldeira (2013), as distribuições dos parâmetros N_t , I_t e C_t seguem uma estrutura hierárquica de priori's Gaussianas, cujas variâncias são descritas por uma distribuição Gama Inversa, tal que:

$$p(x|\alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\beta/x} \quad \forall x > 0$$

$$E(X) = \frac{\beta}{\alpha - 1}$$
 e $Var(X) = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)}$

Os parâmetros da equação 3.16 (μ_{σ} e ϕ_{σ}) são obtidos através de uma distribuição Normal-Gama inversa.

Conforme citado anteriormente, o método MCMC será aplicado para obter a estimativa dos parâmetros, que pode ser exemplificado da seguinte forma:

$$p(F_1, F_2, \dots, F_T, V_1, V_2, \dots, V_T, \Theta | Y) \propto$$

$$p(Y | F_1, F_2, \dots, F_T, V_1, V_2, \dots, V_T, \Theta) \times$$

$$p(F_1, F_2, \dots, F_T | V_1, V_2, \dots, V_T, \Theta) \times$$

$$p(V_1, V_2, \dots, V_T, \Theta) \times$$

$$p(\Theta)$$

Para os fatores latentes e os parâmetros do processo do parâmetro de decaimento (λ_t) , será aplicado o algoritmo de *Metropolis-Hastings*. Assim, serão descartadas as primeiras 50 mil simulações e, a partir de então, serão feitas mais 600 mil iterações, das quais será mantida 1 simulação a cada 10 iterações relizadas. Já a amostragem de *Gibbs* será utilizada para estimar as *posteriori's* conjugadas.

4 Dados

Os dados a serem analisados são referentes aos retornos de títulos do tesouro americano com maturidades de 3, 12, 24, 36, 60, 84, 120 e 240 meses, durante o período de outubro de 1997 a novembro de 2011, totalizando em 3.610 observações¹. Note que, no banco de dados, temos informações ausentes em algumas datas, mas essas observações não são retiradas do modelo e podem ser analisadas diretamente pelo método de estimação Bayesiano que considera as observações ausentes como parâmetros adicionais no modelo.

A evolução da estrutura a termo da taxa de juros encontra-se no gráfico abaixo. Em que o eixo horizontal representa o tempo em dias e cada curva representa uma maturidade, em meses. Podemos perceber que, conforme a maturidade aumenta, em um determinado período de tempo, maior é o retorno do título. Podemos observar, nesta curva, a grande variedade de formatos e padrões na curva de juros Americana neste período.

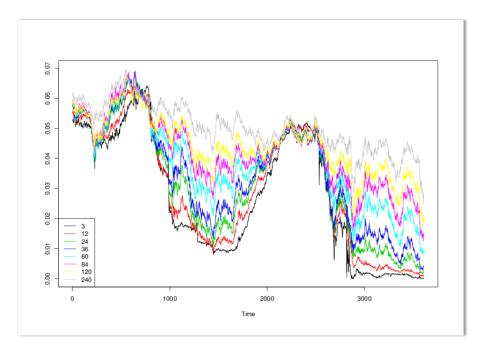


Figura 3 – Dinâmica da estrutura a termo ao longo do tempo

Nota: Esta figura descreve a evolução da estrutura a termo da taxa de juros no período de outubro 1997 a novembro de 2011.

Através da tabela 1, em que constam as estatísticas descritivas da curva de juros neste período, podemos notar, novamente, que quanto mais longo é o período até o

¹Dados obtidos em http://research.stlouisfed.org/fred2/

62 Capítulo 4. Dados

Tabela 1 – Análise descritiva – taxa de juro	Tabela 1 –	Análise	descritiva -	- taxa	de	juros
--	------------	---------	--------------	--------	----	-------

Maturidade (meses)	Média	Desvio-padrão	Mínimo	Máximo
3	2.74	2.06	0	6.42
12	2.98	2.00	0.08	6.44
24	3.25	1.88	0.16	6.93
36	3.46	1.72	0.29	6.88
60	3.87	1.41	0.79	6.83
84	4.20	1.22	1.24	6.87
120	4.44	0.98	1.72	6.79
240	5.03	0.82	2.48	6.97

Nota: Dados diários de outubro 1997 a novembro 2011.

vencimento, maior será o retorno do título. Basta compararmos os valores médios de maturidade de 3 e 240 meses, em que a primeira apresentou uma média de 2.74 e a segunda, de 5.03.

Um dos fatos estilizados citados anteriormente é de que a volatilidade dos retornos de curto prazo apresenta um valor maior que a dos de longo prazo, que também pode ser visto pela tabela 1, em que o desvio-padrão de 3 meses é um pouco mais de 2 vezes e meia ao de 240 meses.

A tabela 2 nos mostra os valores médios para as variáveis macroeconômicas que serão utilizadas no modelo deste trabalho. Nós utilizamos o mesmo conjunto de variáveis macroeconômicas utilizada em Diebold et al. (2006), que representam o núcleo básico de informação sobre o estado da economia. A utilização de capacidade é um indicador de atividade econômica e serve como uma proxy mensal para o PIB que só é mensurado anualmente. A taxa de inflação é um objetivo de política econômica, especialmente relacionado a condução da política monetária, que utiliza a taxa de juros como instrumento de combate a inflação nos períodos de inflação acelerada ou então a redução de juros como incentivo a atividade econômica nos períodos de recessão. Finalmente a última variável observada é a Forward Funds Rate, que é o target de juros de empréstimos federais para bancos comerciais nos EUA, e pode ser compreendida como a taxa básica de juros da Economia. Note que esta variável é observada diariamente. A utilização média da indústria está em torno de 78%, a taxa básica de juros apresentou um valor médio de 2.94 e a inflação média para o período analisado foi de 0.21.

Nas figuras 4, 5 e 6, constam a evolução do tempo das três variáveis macroeconômicas utilização da capacidade industrial, taxa básica de juros americana e inflação respectivamente.

Tabela 2 – Análise descritiva – informações macroeconômicas

Variável	Média	Desvio-padrão	Mínimo	Máximo
Utilização da capacidade industrial	78.0	3.84	67.3	84.9
Taxa Básica de Juros Inflação	$2.94 \\ 0.21$	$2.18 \\ 0.32$	$0.05 \\ -1.81$	$7.06 \\ 1.38$

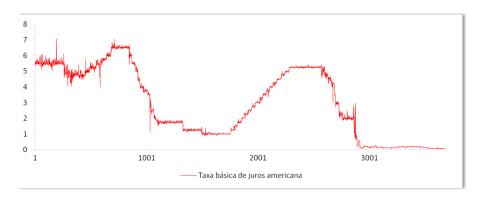
Nota: Dados mensais (utilização da capacidade industrial e Inflação) e diários (taxa básica de juros) de outubro 1997 a novembro 2011.

Figura 4 – Evolução da utilização da capacidade industrial ao longo do tempo (meses)



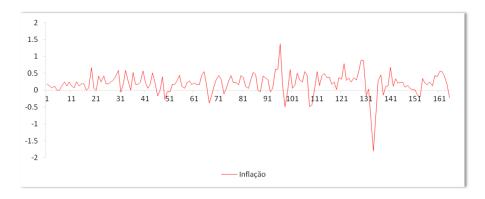
Nota: Dados mensais de outubro 1997 a novembro 2011.

Figura 5 – Evolução da taxa básica de juros americana ao longo do tempo (dias)



Capítulo 4. Dados

Figura 6 – Evolução da taxa de inflação ao longo do tempo (meses)



5 Resultados Finais

Nesta capítulo, temos os resultados obtidos com a estimação do modelo proposto na Seção 3.5. Apresentamos uma análise comparativa dos fatores latentes estimados com o modelo usando informações macroeconômicas e MIDAS em relação a um modelo com apenas os *yields* observados diariamente, mostrando os fatores latentes e volatilidades estocásticas obtidos nestas duas especificações. Temos, também, uma análise comparativa do ajuste obtido entre estas especificações, além do ajuste desenvolvido com uma especificação usual deste modelo obtida pela estimação com o Filtro de Kalman sem a estrutura MIDAS.

As estimações dos fatores nível, inclinação e curvatura, que foram obtidas a partir da estimação pelo método MCMC híbrido para os dois modelos SEM e COM variáveis macroeconômicas (MIDAS), encontram-se nas figuras 7, 8, 9, 10, 11 e 12 a seguir. Embora os padrões obtidos sejam visualmente similares, é possível observar que a estimação usando informações macroeconômicas elimina os saltos nos intervalos de credibilidade para o fator Nível $-N_t$, uma vez que tais saltos ocorrem nas datas ausentes da curva de juros e o uso destas informações faz com que os saltos sejam reduzidos devido à informação condicional obtida destas variáveis. Apesar do método de estimação conseguir estimar os fatores ausentes nestas datas, ele atribui uma incerteza grande quando não são usadas as variáveis macroeconômicas.

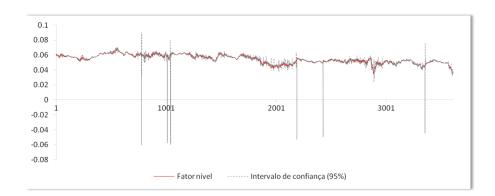


Figura 7 – SEM variáveis macroeconômicas: Fator nível – N_t

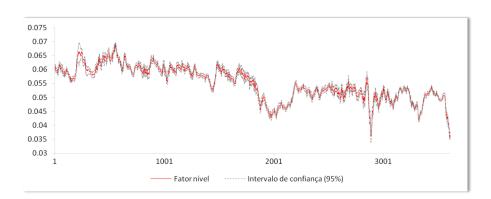


Figura 8 – COM variáveis macroeconômicas: Fator nível – N_t

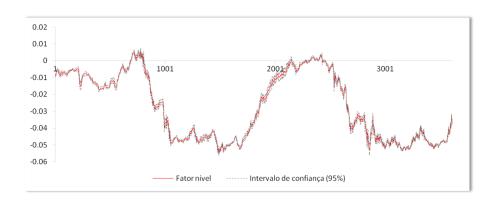


Figura 9 – SEM variáveis macroeconômicas: Fator inclinação – I_t

Nota: Dados de outubro 1997 a novembro 2011.

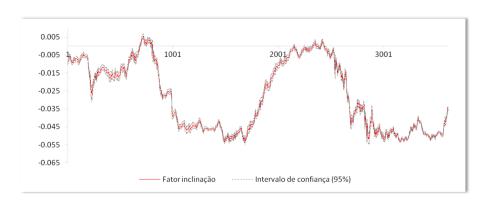


Figura 10 – COM variáveis macroeconômicas: Fator inclinação – I_t

0.04 0.03 0.02 0.01 0 -0.01 1 -0.02 -0.03 -0.04 -0.05 -0.06 -0.06 -0.07 Fator nível Intervalo de confiança (95%)

Figura 11 – SEM variáveis macroeconômicas: Fator curvatura – C_t

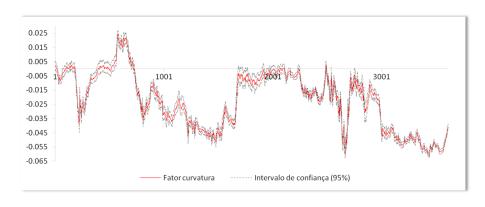


Figura 12 – COM variáveis macroeconômicas: Fator curvatura – C_t

Nota: Dados de outubro 1997 a novembro 2011.

Nas figuras seguintes, mostramos os fatores latentes estimados pelos modelos COM e SEM informação macroeconômica colocados no mesmo gráfico para facilitar a visualização das diferenças. Os três fatores apresentam uma diferença visível no início da série, sendo que o fator curvatura foi quem, aparentemente, apresentou uma maior diferença entre os dois modelos. Note que esta diferença é importante no ajuste do modelo, como será visto nas tabelas com o ajuste in-sample para cada método.

0.075 0.07 0.065 0.05 0.05 0.04 0.04 0.04 0.035 0.03 1 1001 2001 3001 — SEM VARIÁVEL MACROECONÔMICA — COM VARIÁVEL MACROECONÔMICA

Figura 13 – Fator Nível – N_t

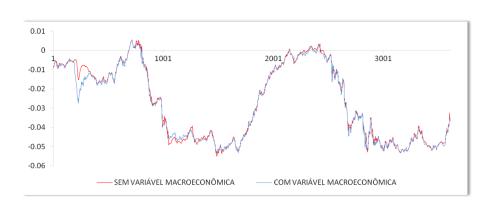


Figura 14 – Fator Inclinação — I_t

Nota: Dados de outubro 1997 a novembro 2011.

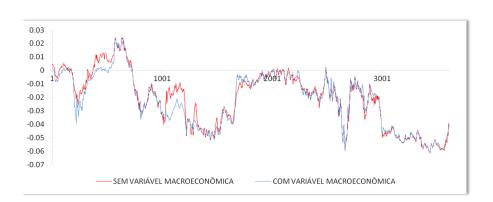


Figura 15 – Fator curvatura – C_t

As figuras 16, 17, 18, 19, 20, 21 e 22 a seguir contêm a evolução das volatilidades no tempo referentes à equação de medida, aos fatores latentes *nível*, *inclinação* e *curvatura* e às variáveis macroeconômicas utilização da capacidade industrial, taxa básica de juros americana e inflação respectivamente, estimadas para o modelo COM variáveis macroeconômicas usando MIDAS.

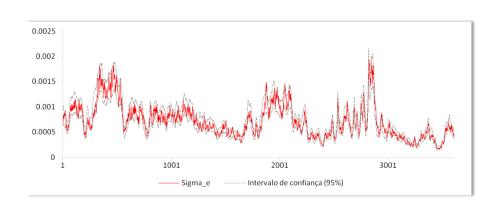


Figura 16 – Volatilidade estocástica: Erro de Medida -
 σ_e

Nota: Dados de outubro 1997 a novembro 2011.

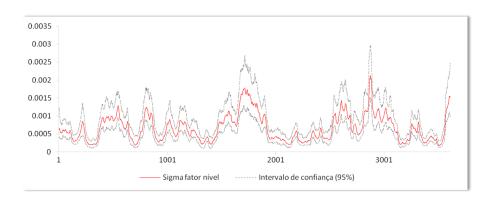


Figura 17 – Volatilidade estocástica: Nível - $\sigma_{N,t}$

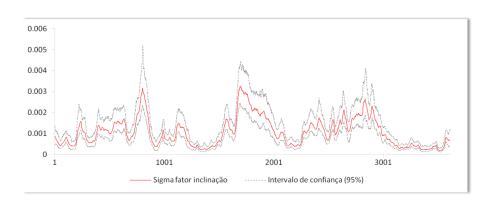


Figura 18 — Volatilidade estocástica: Inclinação - $\sigma_{I,t}$

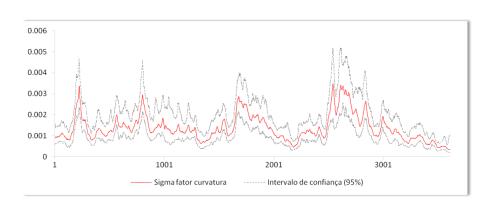


Figura 19 — Volatilidade estocástica: Curvatura - $\sigma_{C,t}$

Nota: Dados de outubro 1997 a novembro 2011.

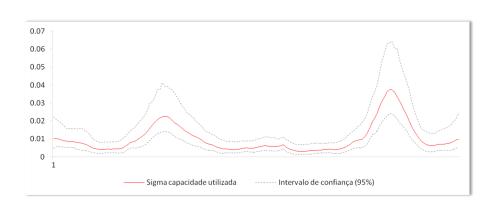


Figura 20 – Volatilidade estocástica: Capacidade Utilizada - $\sigma_{UCI,t}$

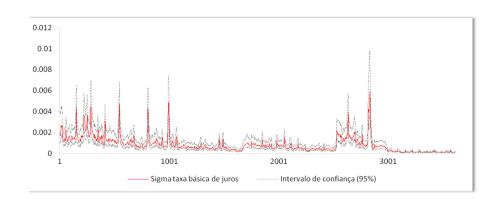


Figura 21 – Volatilidade estocástica: Forward Funds - Taxa Básica de Juros - $\sigma_{FFR,t}$

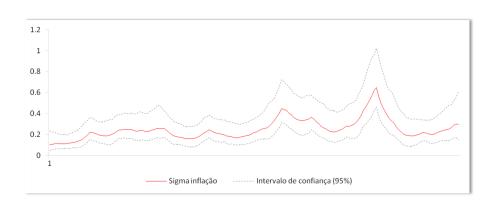


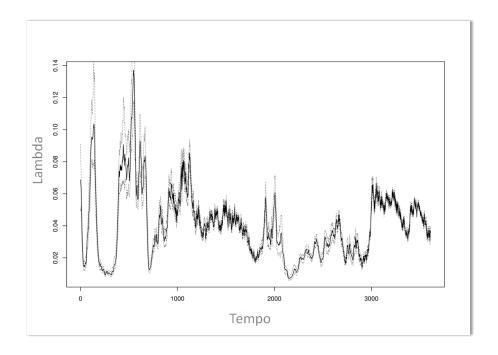
Figura 22 — Volatilidade estocástica: Inflação - $\sigma_{Infl,t}$

Nota: Dados de outubro 1997 a novembro 2011.

Pelas figuras de 16-22, podemos perceber que as volatilidades se alteram bastante ao longo do tempo, reforçando a necessidade já mencionada de considerarmos uma estrutura de volatilidade estocástica multivariada no nosso modelo, capturando, desta forma, os fatos estilizados das séries financeiras. Note que a volatilidade condicional não é somente importante para os fatores *nível*, *inclinação* e *curvatura* que são observados em frequências diárias, mas também é possível notar mudanças relevantes na volatilidade condicional da utilização da capacidade industrial e inflação, refletindo as mudanças observadas na Economia Americana neste período.

Por fim, temos a figura 23 abaixo que nos mostra a evolução no tempo do parâmetro de decaimento λ_t . Note que o modelo captura uma grande variação no parâmetro de decaimento no tempo, consistente com as mudanças nos padrões e formatos da curva de juros neste período. Os resultados mostram claramente a diferença nos *spread*'s entre as taxas dos períodos de 1997 a 2001 e de 2006 a 2008, que apresentaram baixos *spread*'s

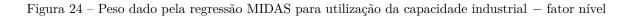
entre as diversas maturidades, e os períodos de 2002 a 2006 e de 2008 a 2011 com uma grande dispersão entre *yields* de diversas maturidades.



 ${
m Figura} \ 23$ — Parâmetro de decaimento: λ_t

Nota: Dados de outubro 1997 a novembro 2011.

Para o modelo que considera as variáveis macroeconômicas, seguem os gráficos com cada um dos pesos atribuídos pela regressão MIDAS.



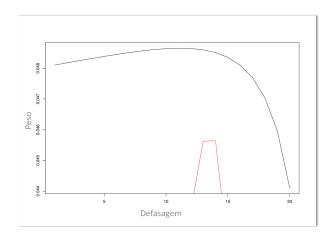


Figura 25 – Peso dado pela regressão MIDAS para utilização da capacidade industrial – fator inclinação

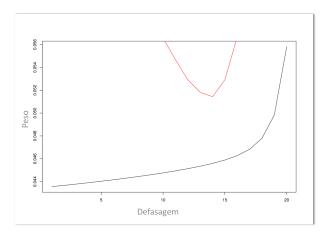
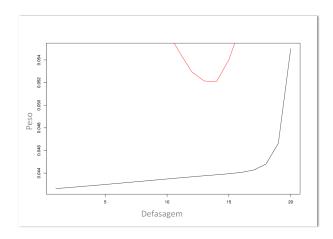


Figura 26 – Peso dado pela regressão MIDAS para utilização da capacidade industrial – fator curvatura



Nota: Dados de outubro 1997 a novembro 2011.

As figuras 24, 25 e 26 são referentes à variável utilização da capacidade industrial. Podemos ver que os pesos atribuídos para o fator nível são os únicos que começam crescendo e depois decrescem rapidamente conforme aumentamos o fator defasagem. Já para os fatores inclinação e curvatura, os pesos são crescentes, começando com um crescimento lento até a defasagem em torno de 17, aumentando rapidamente o crescimento a partir de então, atribuindo-se um peso maior para as maiores defasagens.

900 900 900 15 20 Defasagem

Figura 27 – Peso dado pela regressão MIDAS para Inflação – fator nível

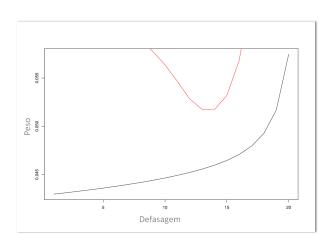


Figura 28 – Peso dado pela regressão MIDAS para Inflação – fator inclinação

Figura 29 – Peso dado pela regressão MIDAS para Inflação – fator curvatura

Nas figuras 27, 28 e 29, temos os pesos atribuídos à variável inflação, e podemos observar que, para os três fatores da curva yield, os pesos são sempre crescentes, aumentando a velocidade de crescimento em torno da defasagem 17. Tanto para a variável utilização da capacidade industrial, quanto para inflação, aos coeficientes ϕ_{45} e ϕ_{65} (ver equações 3.18 e 3.19 respectivamente), que são referentes à variável taxa de juros básica americana, foram atribuídos o valor zero, pois, assim, não temos problema com a identificação dos parâmetros do modelo.

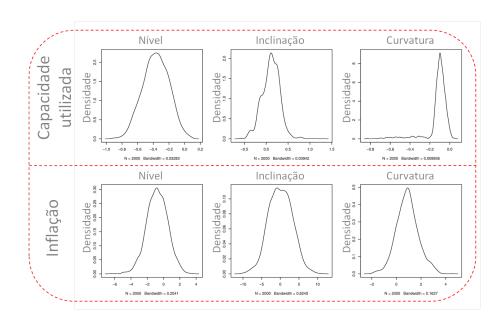


Figura 30 – Peso dado pela regressão MIDAS x coeficientes VAR(1)

A figura 30 nos mostra os pesos obtidos pela regressão MIDAS multiplicados por seus respectivos coeficientes da estrutura VAR(1), equações 3.18 e 3.19 para utilização da capacidade industrial e inflação respectivamente.

Medidas de Ajuste

Uma medida fundamental do sucesso da metodologia proposta é o desempenho no ajuste da curva de juros para o período observado. Para isso, inicialmente comparamos o desempenho de um modelo SEM a inclusão de variáveis macroeconômicas, isto é, baseado somente nos yields da curva de juros, comparado o modelo completo COM variáveis macroeconômicas e a estrutura de ponderação ótima usando pesos obtidos com o procedimento MIDAS. Nas tabelas 3 e 4 abaixo, encontram-se algumas medidas sobre o ajuste dos modelos SEM e COM variáveis macro respectivamente. Em todas as tabelas, as medidas de desempenho estão usando como unidade de medida basis points.

SEM variáveis macroeconômicas

Tabela 3 – Modelo DNS SEM variáveis macroeconômicas

Maturidade (meses)	Erro Médio	Desvio-Padrão	R-EQM ¹	EAM^2
3	-2.90	6.95	7.54	5.44
12	2.70	7.61	8.07	6.15
24	3.43	5.34	6.35	4.95
36	-0.28	4.21	4.22	3.23
60	-1.86	5.39	5.70	4.27
84	-0.05	5.14	5.14	4.07
120	-7.37	7.02	10.18	8.30
240	7.20	8.40	11.07	8.01

COM variáveis macroeconômicas

Tabela 4 – Modelo DNS COM variáveis macroeconômicas

Maturidade (meses)	Erro Médio	Desvio-Padrão	R-EQM ¹	EAM^2
3	-3.24	6.58	7.34	4.99
12	2.33	7.322	7.68	5.79
24	3.29	5.82	6.68	5.08
36	-0.043	4.72	4.72	3.49
60	-0.99	4.908	5.00	3.74
84	0.99	4.99	5.09	3.970
120	-6.75	7.309	9.95	7.91
240	4.45	6.46	7.85	5.59

Nota: Dados de outubro 1997 a novembro 2011.

Das 8 maturidades analisadas, quase todas apresentaram um melhor ajuste pelo modelo COM informação macroeconômica, com exceção dos títulos com vencimentos até a maturidade de 24 e 36 meses, que possuem um erro quadrático médio menor para o modelo SEM as variáveis utilização da capacidade industrial, taxa básica de juros e inflação. Porém, o erro médio, em módulo, para estas duas maturidades é menor no modelo em que consideram-se as informações macroeconômicas.

É interessante observarmos que as taxas de juros de longo prazo são as que mais dependem das expectativas macroeconômicas. A partir disto, através das duas tabelas apresentadas acima, podemos constatar que, para as maturidades de períodos mais longos, o modelo estimado com as variáveis macroeconômicas e que utilizou uma agregação diferenciada através das regressões *MIDAS* apresentou um ajuste efetivamente melhor que o modelo em que não foram consideradas tais variáveis. Reforçando, assim, a necessidade de mantermos essas informações no modelo.

A seguir, temos a tabela 5, na qual constam informações do modelo DNS ajustado com variáveis macroeconômicas através da Inferência Bayesiana, considerando uma estrutura de volatilidade estocástica para os erros de medida, porém o método de agregação das variáveis com maior frequência utilizado foi média simples, considerando, assim, a ponderação de pesos iguais para cada observação. Ao compararmos as tabelas 4 e 5, é fácil notarmos que, para todas as maturidades, o modelo que considera pesos diferentes ao agregar as variáveis de maior frequência apresentou tanto o erro quadrático médio, quanto o erro absoluto médio menor. Desta forma, conseguimos mostrar que o uso da técnica MIDAS é de suma importância para um aperfeiçoamento do modelo, já que uma

¹Raiz quadrada do erro quadrático médio

²Erro Absoluto Médio

estrutura sub-ótima de determinação de pesos pode afetar severamente a estimação de parâmetros e o ajuste do modelo.

Tabela 5 – Modelo DNS COM variáveis macroeconômicas - Inferência Bayesiana - Ponderação MIDAS Pesos Iguais

Maturidade (meses)	Erro Médio	Desvio-Padrão	R - EQM^1	EAM^2
3	-2.9	20.63	20.83	5.12
12	3.16	21.44	21.67	6.77
24	4.16	20.87	21.28	5.57
36	0.56	20.49	20.5	3.59
60	-1.1	20.6	20.63	4.58
84	0.43	20.6	20.6	4.51
120	-7.39	21.34	22.58	9.03
240	6.13	21.12	21.99	6.9

Nota: Dados de outubro 1997 a novembro 2011.

Para verificar a influência da metodologia de estimação, também realizamos a estimação do modelo por máxima verossimilhança usando a decomposição do erro de previsão usando filtro de Kalman, metodologia utilizada em Diebold et al. (2006). Note que o método padrão usando o filtro de Kalman utilizada em Diebold et al. (2006) limita a estimação a uma formulação em espaço de estado padrão, o que restringe o modelo a um decaimento constante no tempo (mas estimado pelo modelo) e não permite a utilização da estrutura de volatilidade estocástica para erros de medida e para os fatores latentes. Assim esta comparação permite analisar o impacto tanto do método de estimação quanto destas duas restrições no ajuste para a curva de juros.

As Tabelas 6 e 7 mostram os resultados de ajuste para o modelo utilizando somente yields e o modelo com a inclusão das variáveis macroeconômicas estimados pelo método baseado no filtro de Kalman e máxima verossimilhança. É possível notar que o desempenho preditivo é bastante inferior ao obtido pelo método Bayesiano e a estrutura MIDAS. Embora a inclusão de variáveis macroeconômicas melhore de forma geral o ajuste obtido em relação ao modelo que utiliza somente *yields*, o desempenho geral da estimação baseada no filtro de Kalman é dominada pela estimação Bayesiana proposta nesse trabalho para o período em análise.

¹Raiz quadrada do erro quadrático médio

²Erro Absoluto Médio

Tabela 6 – Modelo DNS SEM variáveis macroeconômicas - Filtro de Kalman

Maturidade (meses)	Erro Médio	Desvio-Padrão	R-EQM ¹	EAM^2
3	-15.28	46.04	48.51	41.11
12	-2.04	36.14	36.20	30.88
24	0.02	29.81	29.81	24.91
36	-5.94	29.29	29.88	23.77
60	-12.04	32.97	35.09	28.57
84	-11.74	34.68	36.61	30.11
120	-18.03	39.84	43.72	35.56
240	2.26	40.54	40.59	35.70

Tabela 7 – Modelo DNS COM variáveis macroeconômicas - Filtro de Kalman

Maturidade (meses)	Erro Médio	Desvio-Padrão	R-EQM ¹	EAM^2
3	-3.24	22.26	22.49	17.01
12	-0.52	13.96	13.97	8.18
24	0.39	12.05	12.06	5.32
36	-1.63	13.28	13.38	7.05
60	0.23	16.69	16.69	10.76
84	3.94	16.70	17.16	10.76
120	-2.74	15.00	15.24	9.04
240	10.61	22.90	25.24	18.73

 $^{^1\}mathrm{Raiz}$ quadrada do erro quadrático médio

²Erro Absoluto Médio

 $^{^1\}mathrm{Raiz}$ quadrada do erro quadrático médio

²Erro Absoluto Médio

6 Conclusão

Este trabalho contribui para a literatura de modelos dinâmicos para a estrutura a termo de taxas de juros introduzindo uma estrutura de amostragem de frequências mistas nesta classe de modelos. A metodologia proposta é especialmente importante porque amplia as aplicações práticas desta classe de modelos, permitindo utilizar a informação presente em variáveis macroeconômicas observadas mensalmente para a análise de curvas diárias de juros. Embora exista uma ampla classe de modelos de macro-finanças combinando informações da curva de juros com variáveis macroeconômicas, a estimação destes modelos é sempre baseada na frequência de amostragem mais baixa das variáveis macroeconômicas, usando dados mensais ou mesmo trimestrais. Entretanto, apesar destes modelos serem extremamente úteis na análise de questões econômicas, como o impacto de mudanças na curva de juros sobre inflação e atividade econômica, sua aplicação para atividades como gestão de risco e apreçamento de ativos é limitada já que estas questões precisam ser implementadas na frequência mais elevada possível.

Nós implementamos uma versão do modelo dinâmico de Nelson e Siegel (1987) usando Mixed Data Sampling (MIDAS), generalizando os trabalhos de Diebold e Li (2006) e, especialmente, o fundamental modelo de macro-finanças de Diebold et al. (2006). Nesta implementação, nós também capturamos diversas características importantes das curvas de juros e informações macroeconômicas nas duas frequências de amostragem analisadas. Como os dados de curvas de juros são observados em frequência diária, é fundamental considerar as características de volatilidade condicional autoregressiva existentes neste processo. A estrutura de volatilidade estocástica multivariada introduzida neste modelo é importante tanto em termos dos procedimentos de inferência, que caracterizam corretamente os intervalos de credibilidade para parâmetros e fatores latentes na presença de heteroscedasticidade condicional, quanto para o próprio ajuste do modelo aos dados observados. A captura de processos de mudança na variância condicional também é importante para a aplicação do modelo na gestão financeira, já que captura o fato que o risco existente na curva de juros está variando no tempo.

Outro ponto importante é que, na nossa parametrização, o parâmetro de decaimento λ é tratado como um fator latente adicional. Isso permite que o modelo se adapte às diversas mudanças de padrão e formato observados nas curvas de juros em períodos longos de tempo. Isso é especialmente importante já que com o uso de informações macroeconômicas o intervalo de tempo analisado é usualmente maior que nas análises com somente dados diários. Os ganhos obtidos no ajuste do modelo com este parâmetro de de-

caimento variante no tempo indicam que esta característica é fundamental para a correta modelagem das curvas de juros em períodos longos.

Também é importante notar que o trabalho introduz algumas contribuições metodológicas na análise econométrica de modelos com *Mixed Data Sampling*. Nosso modelo utiliza uma estrutura complexa de fatores latentes, tratando como processos não-observados tanto os fatores de *nível*, *inclinação* e *curvatura*, quanto o processo de volatilidade estocástica multivariado. Nós implementamos uma análise Bayesiana de processos MIDAS em um modelo de dimensão elevada, enquanto que as aplicações usuais de modelos MIDAS são baseadas na estimação frequentista sem variáveis latentes.

A análise realizada neste trabalho focou nos aspectos financeiros de modelos de estrutura a termo de taxas de juros, mas é importante notar que um aspecto fundamental é a importância macroeconômica dos resultados, isto é, olhando a transmissão de informações financeiras para a macroeconomia. Neste aspecto a contribuição da metodologia MIDAS é importante já que mostra o padrão de incorporação das informações diárias de curvas de juros nas variáveis macroeconômicas. Acreditamos que este aspecto é um dos desenvolvimentos possíveis para metodologia proposta neste trabalho que pode ser importante na análise macroeconômica, e ainda precisa ser explorado de forma adequada.

ANDERSEN, T. G. et al. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, v. 96, n. 453, p. 42–55, 2001. Disponível em: http://amstat.tandfonline.com/doi/abs/10.1198/016214501750332965. Citado na página 38.

ANDREOU, E.; GHYSELS, E.; KOURTELLOS, A. Regression models with mixed sampling frequencies. *Journal of Econometrics*, v. 158, n. 2, p. 246 – 261, 2010. ISSN 0304-4076. <ce:title>Specification Analysis in Honor of Phoebus J. Dhrymes</ce:title>. Disponível em: http://www.sciencedirect.com/science/article-/pii/S0304407610000072>. Citado 2 vezes nas páginas 33 e 50.

ANG, A.; PIAZZESI, M. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics*, v. 50, n. 4, p. 745 – 787, 2003. ISSN 0304-3932. Disponível em: http://www-sciencedirect.com/science/article/pii/S0304393203000321. Citado na página 27.

BAROSSI-FILHO, M.; ACHCAR, J. A.; SOUZA, R. M. Modelos de volatilidade estocastica em series financeiras: uma aplicacao para o ibovespa. *Economia Aplicada (Impresso)*, v. 14, p. 25–40, 2010. Citado na página 30.

BLACK, F.; SCHOLES, M. The pricing of options and corporate liabilities. *Journal of Political Economy*, The University of Chicago Press, v. 81, n. 3, p. pp. 637–654, 1973. ISSN 00223808. Disponível em: http://www.jstor.org/stable/1831029. Citado na página 20.

BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, v. 32, p. 307–327, 1986. Citado na página 30.

BROTO, C.; RUIZ, E. Estimation methods for stochastic volatility models: a survey. *Journal of Economic Surveys*, Blackwell Publishing Ltd/Inc., v. 18, n. 5, p. 613–649, 2004. ISSN 1467-6419. Disponível em: http://dx.doi.org/10.1111/j.1467-6419.2004.00232.x. Citado 2 vezes nas páginas 30 e 38.

CALDEIRA, J. F.; LAURINI, M. P.; PORTUGAL, M. S. Bayesian inference applied to dynamic Nelson-Siegel model with stochastic volatility. *Brazilian Review of Econometrics*, v. 30, n. 1, p. 123–161, 2010. Citado na página 29.

CARNERO, M. A.; PENA, D.; RUIZ, E. Persistence and kurtosis in garch and stochastic volatility models. *Journal of Financial Econometrics*, v. 2, n. 2, p. 319–342, 2004. Disponível em: http://jfec.oxfordjournals.org/content/2/2/319.abstract. Citado na página 30.

CASELLA, G.; GEORGE, E. I. Explaining the gibbs sampler. *The American Statistician*, v. 46, n. 3, p. 167–174, 1992. Disponível em: http://www.tandfonline.com/doi/abs/10-.1080/00031305.1992.10475878. Citado na página 47.

CHIB, S.; ERGASHEV, B. Analysis of multifactor affine yield curve models. *Journal of the American Statistical Association*, v. 104, n. 488, p. 1324–1337, 2009. Disponível em: http://amstat.tandfonline.com/doi/abs/10.1198/jasa.2009.ap08029. Citado na página 30.

- DAI, Q.; SINGLETON, K. J. Expectation puzzles, time-varying risk premia, and affine models of the term structure. *Journal of Financial Economics*, v. 63, n. 3, p. 415 441, 2002. ISSN 0304-405X. Disponível em: http://www.sciencedirect.com/science/article-/pii/S0304405X02000673. Citado na página 20.
- DIEBOLD, F. X.; LI, C. Forecasting the term structure of government bond yields. Journal of Econometrics, v. 130, n. 2, p. 337 – 364, 2006. ISSN 0304-4076. Disponível em: http://www.sciencedirect.com/science/article/pii/S0304407605000795. Citado 12 vezes nas páginas 15, 17, 20, 24, 25, 26, 28, 29, 31, 35, 36 e 81.
- DIEBOLD, F. X.; RUDEBUSCH, G. D. Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach. [S.l.]: Princeton University Press, 2013. Citado 2 vezes nas páginas 20 e 27.
- DIEBOLD, F. X.; RUDEBUSCH, G. D.; ARUOBA, S. B. The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, v. 131, n. 1–2, p. 309 338, 2006. ISSN 0304-4076. Disponível em: http://www.sciencedirect.com/science/article/pii/S030440760500014X. Citado 12 vezes nas páginas 20, 21, 22, 26, 28, 29, 32, 36, 37, 62, 78 e 81.
- DUFFEE, G. Term premia and interest rate forecasts in affine models. *Journal of Finance*, v. 57, p. 405–443, 2002. Citado na página 22.
- DUFFIE, D.; KAN, R. A yield-factor model of interest rates. *Mathematical Finance*, Blackwell Publishing Ltd, v. 6, n. 4, p. 379–406, 1996. ISSN 1467-9965. Disponível em: http://dx.doi.org/10.1111/j.1467-9965.1996.tb00123.x. Citado na página 20.
- ENGLE, R. Autoregressive conditional heteroskedasticity with estimates of the variance of uk inflation. *Econometrica*, v. 50, p. 987–1008, 1982. Citado na página 30.
- FELLER, W. An Introduction to Probability Theory and Its Applications. third. New York: Wiley, 1968. Citado na página 39.
- FRANKEL, J. A.; LOWN, C. S. An indicator of future inflation extracted from the steepness of the interest rate yield curve along its entire length. *The Quarterly Journal of Economics*, v. 109, n. 2, p. 517–530, 1994. Disponível em: http://qje.oxfordjournals-org/content/109/2/517.abstract. Citado na página 24.
- GHYSELS, E.; SINKO, A.; VALKANOV, R. Midas regressions: Further results and new directions. *Econometric Reviews*, v. 26, n. 1, p. 53–90, 2007. Disponível em: http://www.tandfonline.com/doi/abs/10.1080/07474930600972467>. Citado 3 vezes nas páginas 33, 50 e 51.
- HAMILTON, J. D. *Time Series Analysis*. [S.l.]: Princeton University Press, 1994. Citado na página 36.

HAUTSCH, N.; YANG, F. Bayesian inference in a stochastic volatility nelson-siegel model. *Computational Statistics & Data Analysis*, v. 56, n. 11, p. 3774–3792, 2012. Disponível em: http://EconPapers.repec.org/RePEc:eee:csdana:v:56:y:2012:i:11:p:3774-3792. Citado na página 29.

- HEATH, D.; JARROW, R.; MORTON, A. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, The Econometric Society, v. 60, n. 1, p. pp. 77–105, 1992. ISSN 00129682. Disponível em: http://www.jstor.org/stable/2951677. Citado na página 20.
- HILLS, S. E.; SMITH, A. F. Parameterization issues in bayesian inference. *Bayesian statistics*, Oxford University Press, New York, v. 4, p. 227–246, 1992. Citado na página 48.
- HILLS, S. E.; SMITH, A. F. Diagnostic plots for improved parameterization in bayesian inference. *Biometrika*, Biometrika Trust, v. 80, n. 1, p. 61–74, 1993. Citado na página 48.
- HORDAHL, P.; TRISTANI, O.; VESTIN, D. A joint econometric model of macroeconomic and term-structure dynamics. *Journal of Econometrics*, v. 131, p. 405 444, 2006. ISSN 0304-4076. Disponível em: http://www.sciencedirect.com/science-/article/pii/S0304407605000175. Citado na página 27.
- KASS, R. E. et al. Markov chain monte carlo in practice: A roundtable discussion. *The American Statistician*, v. 52, n. 2, p. 93–100, 1998. Disponível em: http://amstat.tandfonline.com/doi/abs/10.1080/00031305.1998.10480547. Citado na página 32.
- KOOP, G. Bayesian Econometrics. [S.l.]: John Wiley & Sons Ltd, 2003. Citado 2 vezes nas páginas 39 e 40.
- KOOPMAN, S. J.; MALLEE, M. I.; WEL, M. van der. Analyzing the term structure of interest rates using the dynamic Nelson-Siegel model with time-varying parameters. *Journal of Business and Economic Statistics*, v. 28, n. 3, p. 329–451, 2010. Citado na página 29.
- LAURINI, M. P.; CALDEIRA, J. F. A macro-finance term structure model with multivariate stochastic volatility. *Texto para Discussao Serie Economia FEA-RP USP*, 2013. Citado 3 vezes nas páginas 54, 57 e 58.
- LAURINI, M. P.; HOTTA, L. K. Bayesian extensions to diebold-li term structure model. *International Review of Financial Analysis*, v. 19, n. 5, p. 342 350, 2010. ISSN 1057-5219. Disponível em: http://www.sciencedirect.com/science/article/pii-/S1057521910000578. Citado 3 vezes nas páginas 31, 55 e 57.
- LITTERMAN, R. B.; SCHEINKMAN, J. Common factors affecting bond returns. *The Journal of Fixed Income*, v. 1, n. 1, p. 54–61, 1991. Citado 3 vezes nas páginas 20, 25 e 26.
- MULLER, P. A generic approach to posterior integration and gibbs sampling. *Statistics and Computing (to appear)*, 1991. Citado na página 49.

NELSON, C. R.; SIEGEL, A. F. Parsimonious modeling of yield curves. *The Journal of Business*, The University of Chicago Press, v. 60, n. 4, p. pp. 473–489, 1987. ISSN 00219398. Disponível em: http://www.jstor.org/stable/2352957. Citado 9 vezes nas páginas 15, 17, 20, 21, 23, 25, 31, 36 e 81.

- ROBERT, C. P.; CASELLA, G. Monte Carlo Statistical Methods (Springer Texts in Statistics). Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005. ISBN 0387212396. Citado 5 vezes nas páginas 45, 46, 47, 48 e 49.
- ROMER, D. Advanced Macroeconomics. [S.l.]: McGraw-Hill, 2012. Citado na página 20.
- RUDEBUSCH, G. D.; WU, T. A macro-finance model of the term structure, monetary policy and the economy*. *The Economic Journal*, Blackwell Publishing Ltd, v. 118, n. 530, p. 906–926, 2008. ISSN 1468-0297. Disponível em: http://dx.doi.org/10.1111/j-.1468-0297.2008.02155.x. Citado na página 27.
- SIEGEL, A. F.; NELSON, C. R. Long-term behavior of yield curves. *Journal of Financial and Quantitative Analysis*, v. 23, p. 105–110, 3 1988. ISSN 1756-6916. Disponível em: http://journals.cambridge.org/article_S0022109000012953. Citado na página 36.
- TAYLOR, S. J. Modelling Financial Time Series. [S.l.]: World Scientific Publishing Co. Pte. Ltd, 1986. Citado 3 vezes nas páginas 30, 37 e 57.
- TIERNEY, L. Markov chains for exploring posterior distributions. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 22, n. 4, p. pp. 1701–1728, 1994. ISSN 00905364. Disponível em: http://www.jstor.org/stable/2242477. Citado na página 49.
- WASSERMAN, L. All of Statistics: A Concise Course in Statistical Inference. [S.l.]: Springer Publishing Company, Incorporated, 2010. ISBN 1441923225, 9781441923226. Citado na página 44.