

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

LUCAS OLIVEIRA CALDELLAS DE QUEIROZ

ANÁLISE E ESTIMAÇÃO DA ESTRUTURA A
TERMO DA TAXA DE JUROS COM ABORDAGEM
BAYESIANA

Porto Alegre

2016

LUCAS OLIVEIRA CALDELLAS DE QUEIROZ

**ANÁLISE E ESTIMAÇÃO DA ESTRUTURA A
TERMO DA TAXA DE JUROS COM
ABORDAGEM BAYESIANA**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Mestre em Economia, com ênfase em Economia Aplicada.

Orientador: Prof. Dr. João Frois Caldeira

Porto Alegre

2016

CIP - Catalogação na Publicação

Queiroz, Lucas Oliveira Caldellas

Análise e Estimação da Estrutura a Termo da Taxa de Juros com abordagem bayesiana / Lucas Oliveira Caldellas Queiroz. -- 2016.

51 f.

Orientador: João Fróis Caldeira.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Ciências Econômicas, Programa de Pós-Graduação em Economia, Porto Alegre, BR-RS, 2016.

1. Estrutura a Termo da Taxa de Juros. 2. Hipótese das Expectativas. 3. Econometria. 4. Modelo de Fatores. 5. Modelo de Nelson-Siegel. I. Caldeira, João Fróis, orient. II. Título.

LUCAS OLIVEIRA CALDELLAS DE QUEIROZ

ANÁLISE E ESTIMAÇÃO DA ESTRUTURA A TERMO DA TAXA DE JUROS COM ABORDAGEM BAYESIANA

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Mestre em Economia, com ênfase em Economia Aplicada.

Aprovada em Porto Alegre, 10 de Janeiro de 2017.

BANCA EXAMINADORA:

Prof. Dr. João Frois Caldeira- Orientador
UFRGS

Prof. Dr. Flávio Augusto Ziegelmann
UFRGS

Prof. Dr. Márcio Poletti Laurini
USP

Prof. Dr. Tiago Wickstrom Alves
UNISINOS

AGRADECIMENTOS

Agradeço ao meu orientador, Prof. João Caldeira, por me introduzir ao tema, ao CNPQ pelo apoio financeiro e a todos aqueles que de alguma forma me ajudaram neste período.

*“A tarefa não é tanto ver aquilo que ninguém viu,
mas pensar o que ninguém ainda pensou
sobre aquilo que todo mundo vê.”
(Arthur Schopenhauer)*

RESUMO

Este trabalho analisa e modela a Estrutura a Termo das Taxas de Juros objetivando ao teste da Hipótese das Expectativas(HE) na ponta curta da curva de juros e a uma aplicação da teoria de Markowitz (1952) no mercado de renda fixa utilizando a estrutura proposta por Caldeira, Moura e Santos (2015). Para estes fins foram utilizados dados dos contratos futuros de 1 dia dos depósito interbancários (DI1) negociados na BMF interpolados em maturidades fixas, sendo utilizados em base semanal quando do teste da HE e em base diária para a construção dos portfólios de mínima variância. Os resultados encontrados para o teste da HE sugerem a invalidade da teoria, uma vez que o prêmio de risco é se mostra ajustável a um modelo GARCH-M e, portanto, variante no tempo. Os portfólios de mínima variância ajustados nas versões irrestrita e restrita (*duration* máxima de 1 ano) se mostraram consistentes, tendo superado quase a totalidade dos fundos analisados. O portfólio de mínima variância irrestrito obteve o maior Índice de Sharpe no período analisado.

Palavras-chaves: Estrutura a Termo da Taxa de Juros. Hipótese das Expectativas. GARCH-M. Modelo de Fatores. Nelson-Siegel. Estado de Espaço. Modelo de Mínima Variância.

ABSTRACT

This work analyzes and model the Term Structure of Interest Rates seeking testing Expectation Hypothesis in the short end of the Yield Curve and to apply the portfolio theory to the fixed income context using the framework proposed by Caldeira, Moura e Santos (2015). We used a database of constant maturities interbank deposits's future contracts. The results suggest Expectation Hypothesis doesn't hold and risk premium could be modeled by a GARCH-M framework, being time variant. The bond portfolio optimized were, in general, consistent with high sharpe ratio relative to other funds and beaten the chosen benchmark during the period analyzed.

Keywords: Yield Curve Term Structure. Expectation Hypothesis. GARCH-M. Factor Model. Nelson-Siegel. State Space. Minimum Variance Model.

LISTA DE ABREVIATURAS E SIGLAS

BM&F	Bolsa de Mercadoria e Futuros
CDI	Cédulo de Depósito Interbancário
DI1	Contrato Futuro de Depósito Interbancário de 1 dia
DNS	Modelo Dinâmico de Nelson-Siegel
ECM	Modelo de Correção de Erros
ETTJ	Estrutura a Termo da Taxa de Juros
GED	Distribuição de Erros Generalizados
HE	Hipótese das Expectativas
HEP	Hipótese das Expectativas Puras
HRP	Retorno pelo período do investimento
IRF-M	Índice de Renda Fixa de Mercado
IS	Índice de Sharpe
LTN	Letras do Tesouro Nacional
MCMC	Monte Carlo por Cadeias de Markov
VAR	Modelo de Vetores Autorregressivos

SUMÁRIO

1	INTRODUÇÃO	10
2	A ESTRUTURA A TERMO DAS TAXAS DE JUROS	13
2.1	Conceitos Básicos	13
2.2	A Hipótese das Expectativas	15
2.3	Fatos Estilizados da ETTJ	17
3	MODELOS DA CLASSE DE NELSON-SIEGEL	18
3.1	O Modelo Básico	18
3.2	Representação em Espaço de Estado	20
3.3	O filtro de Kalman	22
3.4	A abordagem Bayesiana	24
3.4.1	<i>O Suavizador</i>	25
3.4.2	<i>A estimação do VAR</i>	26
3.4.3	<i>O estimador MCMC</i>	28
3.4.3.1	<i>A Amostragem de Gibbs</i>	29
3.4.4	<i>Escolha das priors</i>	30
4	A ALOCAÇÃO DE RECURSOS	32
4.1	Estratégia de portfólio no mercado de renda fixa	33
4.1.1	<i>A distribuição dos log-retornos</i>	33
5	RESULTADOS	35
5.1	Dados	35
5.2	Teste da Hipótese das Expectativas	37
5.3	Retorno das Carteiras Otimizadas	40
6	CONSIDERAÇÕES FINAIS	45
	REFERÊNCIAS	47
	APÊNDICE - FUNDOS DE INVESTIMENTOS COMPARADOS	50

1 INTRODUÇÃO

A estrutura a termo das taxas de juros (ETTJ), ou simplesmente curva de juros, é um dos diversos pontos de ligação entre a teoria de finanças e a teoria econômica. Uma ramificação da teoria econômica, a economia monetária busca investigar a relação entre variáveis econômicas reais em níveis agregados (produto real, taxas reais de juros, etc) e as variáveis nominais (taxa de inflação, taxas nominais de juros, taxa de câmbio nominal e a oferta de moeda). Em Walsh (2010) é enfatizado a importância da distinção entre taxas reais e nominais de juros para que se compreenda as questões de economia monetária. Contudo, outra distinção aconselhada é a em relação às taxas de curto e longo prazo. Essa relação se deve ao fato de que mudanças nas taxas de juros de curto prazo, usada pela autoridade monetária como instrumento de política, afetarão o gasto agregado a depender se as taxas reais de juros de longo prazo forem afetadas. Assim, ainda segundo Walsh (2010), as políticas baseadas em juros reduzem o papel da demanda por moeda e elevam a importância da curva de juros futuros na transmissão da ação de política monetária para a economia real. No mercado financeiro a modelagem da curva de juros é essencial para a precificação de ativos e construção de operações de *hedge*. Desta forma, identificar os fatores que levam a alterações no formato da curva de juros torna-se crucial para formuladores política econômica e participantes do mercado financeiro em geral.

Passados mais de um século após sua formulação por Fisher (1896), a Hipótese das Expectativas (HE) é a teoria mais conhecida e estudada sobre a ETTJ. Segundo esta tese em sua versão pura (HEP), a taxa de juros à vista de longo prazo é a média de longo prazo das expectativas das futuras taxas à vista de curto prazo. A implicação mais direta desta suposição é que, em um mundo com investidores neutros ao risco, títulos com maturidades diferentes seriam substitutos perfeitos, mantido o mesmo risco de crédito. Assim, o retorno esperado ao se comprar títulos com vencimento em 1 mês e levá-los até o vencimento, consecutivamente durante 12 meses, é igual ao de se comprar, no início do mesmo período, um título com maturidade em 12 meses e levá-lo até seu vencimento. De forma semelhante, sob a HEP, o retorno ao se comprar um título com vencimento em 1 mês e mantê-lo até seu vencimento será igual ao de se comprar o título com vencimento em 12 meses e vendê-lo após 1 mês.

Se a HEP vale, a tarefa de compreender a relação entre as taxas correntes e as expectativas das taxas futuras, incumbida a economistas, formadores de política econômica e profissionais dos mercados financeiros, está facilitada. A inclinação da curva de juros à vista é ocasionada pela expectativa do mercado quanto aos valores futuros das taxas à vista e mudanças na inclinação serão totalmente explicadas por alterações nessas expectativas. Se, por outro lado, a HE não é válida em nenhuma de suas versões, o prêmio de risco

é variável, tornando-se difícil para o agente discernir entre mudanças das taxas à vista causadas por alterações nas expectativas das taxas futuras ou mudanças das taxas à vista causadas por variações no prêmio de risco (BRITO; DUARTE; GUILLÉN, 2003).

Embora quase a totalidade da literatura que objetiva verificar a validade da HE se baseie em testar o poder preditivo das taxas forward em relação às taxas à vista futuras, há autores que buscam fazer tal teste em uma abordagem mais aplicada aos operadores do mercado financeiro. A idéia, neste caso, é que se vale a HEP, estratégias que envolvam compra de títulos de longo prazo para venda antecipada não deverão gerar excessos de retornos se comparado às estratégias simples de *buy and hold* com títulos curtos, com maturidade igual ao horizonte de investimento. Evidências de persistência de excesso de retorno indicariam a existência de prêmio de risco, invalidando a hipótese. Do ponto de vista prático, caso a ETTJ possua inclinação positiva e estável no tempo, a estratégia com títulos de longo prazo deverá obter resultados superiores, evidenciando a existência do prêmio de risco.

Ainda que seja foco de diversos estudos, como exposto acima, a modelagem da ETTJ pode ser ainda mais explorada em alguns campos de pesquisa. Em finanças o modelo de Markowitz (1952) segue ainda como peça fundamental para a montagem de portfólios de renda variável. Contudo, raros são os trabalhos que buscam aplicá-lo ao contexto da renda fixa. Dentre as causas para este fato está a dificuldade em se modelar os retornos e a matriz de covariância desses ativos. Neste trabalho utilizamos as derivações propostas em Caldeira, Moura e Santos (2015) para aplicação de um modelo de variância mínima em um conjunto de títulos pré-fixados, sendo estimado em sua versão comum e também com restrição na *duration* do portfólio.

Para a realização desta dissertação foram utilizados contratos de DI futuro de 1 dia negociados na BMF, interpolados em maturidades constantes de 1, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 42, 48 e 60 meses, tidos como proxy para a curva de juros zero cupom brasileira. Com a globalização e integração dos mercados financeiros mundiais, nem sempre as trajetórias dimensionadas pelos bancos centrais quanto às taxas de juro em suas respectivas economias podem ser sustentadas. Dessa forma, investimentos do setor público e privado e os financiamentos existentes estão sujeitos à volatilidade das taxas de juro. Os contratos futuros desta variável financeira vêm, justamente, criar uma alternativa para o gerenciamento deste tipo de risco. Eles surgiram nos anos 1970, nos EUA, quando o fim da conversibilidade ao dólar e das paridades fixas entre moedas abriu espaço às flutuações das taxas de câmbio e às consequentes oscilações das taxas de juro dos diferentes países.

Além desta introdução, o trabalho é composto de mais quatro seções. O primeiro capítulo aborda os conceitos básicos sobre a estrutura a termo das taxas de juros, explicando ainda a teoria da Hipótese das Expectativas e alguns fatos estilizados. O segundo capítulo apresenta o modelo de Nelson-Siegel e sua representação em Espaço de Estado, além do

processo de estimação bayesiana. O terceiro capítulo aborda a otimização de portfólio de renda fixa e as derivações para o vetor de retornos e a matriz de covariância. Nos quarto capítulo são apresentados os resultados observados, sendo apresentadas em seguida as considerações finais.

2 A ESTRUTURA A TERMO DAS TAXAS DE JUROS

De acordo com Fabozzi (2007), qualquer conjunto de títulos de dívida de mesma qualidade creditícia pode ser descrito com relação às taxas de retorno esperadas e seus correspondentes prazos de vencimento (maturidade). A essa forma de descrição convencionou-se chamar Estrutura a Termo das Taxas de Juros (ETTJ), Curva de Rendimentos ou Curva de Juros.

Na literatura sobre a estrutura a termo das taxas de juros é comum verificar o uso de títulos zero cupom. Essa denominação é dada àqueles títulos em que há em seu fluxo de caixa somente o valor pago no período inicial e o valor recebido no vencimento. Os títulos públicos pré-fixados zero-cupom tem um valor de resgate fixo (valor de face) e a sua taxa de juros é dada unicamente pela diferença entre o valor de face e o valor pago pelo título, sendo por isso chamado por Rossi (1996) de títulos de desconto puro.

Como notado em Svensson (1994), a necessidade de se usar títulos zero cupom (ou transformar em zero cupom os títulos que pagam cupom) surge em função de que títulos com a mesma maturidade porém com diferentes taxas de cupom apresentam diferentes taxas de rendimento até o vencimento (*yield to maturity*). Por definição, fica então claro que uma curva de juros construída com tais taxas não refletirá apropriadamente a estrutura a termo das taxas de juros (que tem como objetivo justamente sumarizar a relação entre os retornos até o vencimento e as maturidades).

2.1 Conceitos Básicos

Podemos definir a taxa interna de retorno, ou simplesmente o *yield to maturity* do título, como a taxa que desconta o seu fluxo de pagamentos, obtendo então seu preço. Para um título no período corrente t com maturidade em τ e preço P , temos:

$$P_t(\tau) = \frac{1}{(1 + Y_t(\tau))^\tau}, \quad (2.1)$$

em que $Y_t(\tau)$ é o *yield* em sua forma simples¹. No entanto, é geralmente mais conveniente trabalhar com a taxa em capitalização contínua, em logarítmo. Assim, reescrevemos a equação para o preço.

$$P_t(\tau) = e^{-\tau y(\tau)}, \quad (2.2)$$

onde $y(\tau)$ representa o *yield* na forma de capitalização contínua (a que chamaremos no decorrer do texto somente de *yield* ou taxa) e $e^{-\tau y(\tau)}$ é o valor presente de R\$1 recebível τ

¹ Forma simples pode ser entendida como capitalização em tempo discreto. Em sua versão em tempo contínuo, o *yield* é apresentado como $y(\tau) = \log(1 + Y_t(\tau))$.

períodos à frente, descontado à taxa $y(\tau)$, dos pagamentos restantes ao detentor do título. Tomando o logaritmo, temos a seguinte equação para o log-preço:

$$p_t(\tau) = -\tau y(\tau). \quad (2.3)$$

Da equação acima depende-se que τ representa a variação proporcional no preço deste título a cada variação percentual em $y(\tau)$. Neste caso, um título com maturidade em 1 ano perderia 1% de seu valor para cada elevação de 1% em sua taxa, enquanto um título de 5 anos teria seu preço reduzido em 5% para igual variação em sua taxa. Essa medida de risco de taxa de juros é conhecida como *duration*. Assim, a variação no log-preço é composta da variação na maturidade e da variação na taxa negociada². Um título negociado antes de seu vencimento obtém retorno pelo período do investimento, e não conhecido é antecipadamente. No que se segue, chamaremos tal retorno de HPR.

Considerando o caso onde um título de longo prazo é mantido por um ano e depois negociado, temos o seguinte log HPR:

$$r_{t+1}(\tau) = p_{t+1}(\tau - 1) - p_t(\tau) \quad (2.4)$$

$$= y_t(\tau) - (\tau - 1)(y_{t+1}(\tau - 1) - y_t(\tau)). \quad (2.5)$$

Agora consideremos o caso em que o título é mantido por dois anos.

$$r_{t+2}(\tau) = p_{t+2}(\tau - 2) - p_t(\tau) \quad (2.6)$$

$$= y_t(\tau) - (\tau - 2)(y_{t+2}(\tau - 2) - y_t(\tau)). \quad (2.7)$$

Contudo, sabemos que:

$$r_{t+2}(\tau) = r_{t+1,t+2}(1) + r_{t+1}(\tau), \quad (2.8)$$

em que $r_{t+1,t+2}(1)$ é o log-retorno do título entre o primeiro e o segundo ano. Fazendo iterativamente, chegamos às seguintes equações:

$$r_{t+3}(\tau) = r_{t+2,t+3}(1) + r_{t+1,t+2}(1) + r_{t+1}(\tau) \quad (2.9)$$

⋮

$$r_{t+n}(\tau) = r_{t+(n-1),t+n}(1) + \dots + r_{t+1,t+2}(1) + r_{t+1}(\tau). \quad (2.10)$$

A equação (2.10) nos mostra que o *yield* é uma média aritmética dos log-retornos de 1 ano do título durante seu período de maturação. Entretanto, como os títulos são negociados diariamente, é possível demonstrar, da mesma maneira, que o *yield* é uma média aritmética dos log-retornos de 1 dia do título durante seu período de maturação.

² No vencimento $\tau=0$, levando o preço obrigatoriamente ao valor de face

É também comum a análise dos yield em termos relativos a alguma maturidade. A diferença entre o yield de longo prazo e o yield de curto prazo é chamada de *spread*. A diferença entre o log HPR do título de longo prazo e o do de curto prazo é chamada excesso de retorno. Tomando como base o título de 1 ano, o excesso de retorno em 1 ano é igual ao *spread* inicial menos o produto da variação na maturidade do título pela variação na taxa, como pode ser notado pela equação abaixo:

$$r_{t+1}(\tau) - y_t(1) = y_t(\tau) - y_t(1) - (\tau - 1)(y_{t+1}(\tau - 1) - y_t(\tau)). \quad (2.11)$$

2.2 A Hipótese das Expectativas

Uma mais tradicional teoria sobre o comportamento dos *spreads* e, por consequência, do formato da curva de juros, é a Hipótese das Expectativas. Em essência, ela foi assim descrita, em sua forma pura (HEP), por Fisher (1896):

*that this interest is truly an average is attested both by the comparative stability of the rate of interest realized on long time bonds as compared with the fluctuations of the rate of interest in the short time money market[...]The investor who holds a bond a long time realizes an interest which is an "average" of the oscillating rates of those who speculate during the interim.*³

Nesta versão, agentes neutros ao risco possuem informação completa, de modo que todos os investidores possuam as mesmas expectativas para as taxas de juros futuras. A ausência de assimetria de informação faz com que os agentes escolham somente as maturidades que possuem as maiores expectativas de retorno no horizonte de investimento. Como consequência da tentativa de arbitragem entre as maturidades o prêmio de risco é levado a zero, implicando com que os títulos de diferentes maturidades sejam substitutos perfeitos. Deste modo, o valor esperado de estratégias que envolvam sequências de aplicações em títulos de curto prazo ou investimentos em títulos de longo prazo será o mesmo.

Assim exposto, sempre que o *yield* de um título longo excede o de um título curto, o primeiro deve subir de forma a gerar expectativa de perda de capital e eliminar a vantagem de um *yield* mais elevado. Pela equação (2.11) a expectativa de variação no *yield* do título longo será de $\frac{1}{\tau-1}$ vezes o *spread* corrente. De modo semelhante, é esperado que a média dos *yields* de curto prazo subam nos $\tau-1$ períodos restantes para igualar a taxa do título longo. Novamente de acordo com a equação (2.11), a expectativa de elevação das taxas curtas será $\frac{\tau}{\tau-1}$ vezes o *spread* corrente.

Para que a teoria seja testável é necessário que se faça suposições sobre como as expectativas são formadas. Sob expectativas racionais os agentes não cometem erros

³ [...]Que este rendimento é verdadeiramente uma média é atestada tanto pela estabilidade comparativa da taxa de juro realizada em obrigações de longo prazo em comparação com as flutuações da taxa de juro no mercado monetário de curto prazo [...] O investidor que detém um título por longo tempo obtém um rendimento que é uma "média" das taxas daqueles que especulam durante este período.

sistemáticos, de modo que o excesso de retorno é imprevisível, tendo média zero para o caso da HEP.

Reescrevendo a equação (2.11), temos a seguinte relação:

$$rx_{t+1}(\tau) = S_t(\tau, 1) - \beta(f_t(\tau - 1) - y_t(\tau)) = \theta^\tau + \varepsilon_{t+1} \quad (2.12)$$

em que $rx_{t+1}(\tau)$ é o excesso de retorno em relação ao título de maturidade de 1 período, $S_t(\tau, 1)$ é o *spread* observado no período t , β é igual a $(\tau - 1)$, $f_t(\tau - 1)$ é a taxa *forward* para a maturidade $\tau - 1$, observada em t , θ^τ é o prêmio de risco pela maturidade e ε_{t+1} é um ruído observado em $t + 1$ e não correlacionado com os ruídos de períodos anteriores.

Lima e Issler (2003), Marçal e Pereira (2007), Guillen e Tabak (2009) e Tabak (2009) são alguns dos trabalhos que testam a HE para o caso brasileiro. As hipóteses em geral usadas são de que as taxas *forward* seriam estimadores não viesados das taxas futuras à vista e os *spreads* entre taxas de longo prazo e de curto prazo seriam estimadores não viesados de futuras variações de curto prazo nas taxas longas. Não há consenso na literatura para o caso brasileiro para a rejeição ou aceitação completa da HE. Contudo, observa-se que os autores têm encontrado evidências para aceitação da HE em maturidades curtas e rejeição em maturidades longas.

Neste trabalho objetivamos ao teste da HE para a ponta curva da curva de juros, empregando uma metodologia alternativa. Não será imposta relação linear entre *spreads*, excessos de retornos e taxas *forward*, como de costume na literatura. Observaremos o excesso de retorno semanal obtido no mercado e faremos a análise e modelagem da série temporal gerada, levando em conta a heterocedasticidade condicional observada.

Engle, Lilien e Robins (1987) estenderam a análise básica de modelos com heterocedasticidade condicionada ao permitirem que a média da série seja dependente de sua própria variância condicional. Esta estrutura tem forte apelo em aplicações no mercado financeiro, visto que agentes avessos ao risco cobram um retorno maior para manterem títulos de maior risco. Dado que o risco de um título pode ser medido pela sua variância, o prêmio de risco nesse caso seria uma função crescente da variância condicional dos retornos. Utilizando a notação da equação (2.12), apresentamos abaixo o modelo:

$$rx_t(\tau) = \theta^\tau + \varepsilon_t \quad (2.13)$$

$$\theta^\tau = \mu + \delta h_t \quad \delta > 0 \quad (2.14)$$

$$h_t = \varphi_0 + \sum_{i=1}^q \varphi_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \gamma_i h_{t-i} \quad (2.15)$$

$$\varepsilon_t = \nu_t \sqrt{h_t}, \quad (2.16)$$

em que ν_t é um ruído branco e, como pode ser notado, caso o modelo seja homocedástico, h_t será uma constante e o modelo volta a ser o modelo comum com prêmio de risco constante.

Para a modelagem foi utilizada a Distribuição de Erros Generalizada (GED). Esta distribuição foi inicialmente utilizada no contexto de modelos ARCH por Nelson (1991). Ele nota que a GED tem a Normal como um caso especial, apresentando alternativas com maior ou menor grau de curtose, sendo muito atraente ao permitir excessos de curtose positivos e negativos com apenas 1 parâmetro.

2.3 Fatos Estilizados da ETTJ

A curva de juros se desenvolve de forma dinâmica no tempo, apresentando padrões já bem conhecidos pela literatura. Na análise das séries temporais geradas emergem alguns fatos estilizados. Primeiramente, o *yield* médio aumenta junto com a maturidade, formando uma curva média positivamente inclinada. Esse padrão sugere a existência de um prêmio pela maturidade, podendo ser causado por aversão a risco, preferência por liquidez ou por segmentação de mercado. O segundo fato surge da análise das volatilidades, decrescendo conforme o aumento da maturidade, possivelmente pelo fato de taxas longas serem médias de taxas mais curtas. Por fim, os *yields* apresentam geralmente elevada persistência, sendo captada pelos elevados coeficientes de autocorrelação. Em contraste, os *spreads* são menos voláteis e menos persistentes. Por isso, várias estratégias de gestão ativa de títulos são baseadas na propriedade de reversão à média dos *spreads* (DIEBOLD; RUDEBUSCH, 2013).

3 MODELOS DA CLASSE DE NELSON-SIEGEL

Segundo Diebold e Rudebusch (2013), diversos são os motivos que levaram o modelo proposto por Nelson e Siegel (1987) a se tornar muito popular entre analistas do mercado financeiro e banqueiros centrais, como notado por Svensson (1995), Gürkaynak, Sack e Wright (2007), Nyholm (2008) e Caldeira (2011). Sua versão dinâmica, proposta em Diebold e Li (2006), modelo DNS, também demonstrou bons resultados em termos de previsão e rapidamente viu seu modelo inicial ganhar extensões. Segundo Caldeira, Laurini e Portugal (2011), a principal contribuição dos autores foi a de mostrar que as previsões geradas por esta classe de modelos superam muitos outros modelos utilizados para este fim anteriormente, como os modelos de vetores autoregressivos (VAR) e de correção de erros (ECM).

Christensen, Diebold e Rudebusch (2009), Christensen, Diebold e Rudebusch (2011) e Pooter (2007) são alguns dos trabalhos que comparam o modelo DNS com e sem restrições de não-arbitragem. Golinski e Zaffaroni (2015) estendem o modelo para aplicações com modelagem de memória longa. Diebold, Rudebusch e Aruoba (2006) estimam o modelo utilizando a estrutura de espaço de estado e o filtro de Kalman, enquanto Caldeira, Moura e Portugal (2010) estimam o modelo utilizando o Filtro de Kalman e mostram que os resultados obtidos, em termos de previsão, foram melhores do que os obtidos com o método de dois passos, proposto inicialmente, para todas as maturidades quando são considerados horizontes de previsão de um mês, três meses e seis meses.

Dentre as diversas extensões possíveis ao modelo DNS, o uso da estatística bayesiana é uma das que possuem grande potencial, dados os resultados encontrados nesta literatura em termos de previsão. Embora modelos de fatores sejam sempre lembrados quando se tratam de dados de curva de juros, geralmente em dimensões elevadas, Bańbura, Giannone e Reichlin (2010) encontram evidências de que uma estimação bayesiana fazendo uso de prioris de Minnesota consegue previsões superiores a tais modelos. Carriero, Kapetanios e Marcellino (2012) também performam uma análise bayesiana em um modelo VAR para um conjunto de yields de grande dimensão, usando também prioris de Minnesota para redução de dimensionalidade, e conseguem resultados positivos em termos preditivos em relação ao passeio aleatório.

3.1 O Modelo Básico

Nelson e Siegel (1987) foram pioneiros no desenvolvimento de uma classe de modelos que busca representar a ETTJ através de um modelo de fatores que utiliza uma classe de funções, aquelas associadas à soluções de equações diferenciais ou à diferença.

Inicialmente eles supõe que as taxas *spot* seguem um processo gerado por uma equação diferencial de segunda ordem. Assim, como as taxas *forward* são previsões, elas seriam as soluções da equação. Sabendo que as taxas *forward* instantâneas são taxas *forward* com prazo mínimo entre a liquidação e a maturidade (ex.: 1 dia) o método mais parcimonioso encontrado pelos autores é obtido com a suposição de que as taxas *spot* seguem equações diferenciais ordinárias de 2ª ordem com raízes reais e idênticas. Logo, as taxas *forward* instantâneas para maturidades m são dadas por:

$$r(m) = \beta_0 + \beta_1(e^{-\frac{m}{\tau}}) + \beta_2(\frac{m}{\tau}e^{-\frac{m}{\tau}}) \quad (3.1)$$

Como visto anteriormente o *yield to maturity* de um título é uma média das taxas *forward* para um título zero-cupom. Assim, podemos calculá-lo da seguinte forma:

$$R(m) = \frac{\int_0^m r(x) dx}{m} = \beta_0 + \frac{(\beta_1 + \beta_2)(1 - e^{-\frac{m}{\tau}})}{\frac{m}{\tau}} - \beta_2 e^{-\frac{m}{\tau}} \quad (3.2)$$

Simplificando a notação fazemos $\frac{m}{\tau} = \lambda\tau$, sendo τ agora a referência para maturidade. Assim chegamos à representação do modelo dinâmico de Nelson e Siegel(DNS) proposta por Diebold e Li (2006), em que os parâmetros são variantes no tempo:

$$R(\tau) = \beta_{0t} + \beta_{1t}(\frac{1 - e^{\lambda\tau}}{\lambda\tau}) + \beta_{2t}(\frac{1 - e^{\lambda\tau}}{\lambda\tau} - e^{\lambda\tau}) \quad (3.3)$$

Analisando a equação acima, é possível observar que:

$$\lim_{t \rightarrow \infty} R(t) = \beta_0 \quad (3.4)$$

Por este motivo, em Nelson e Siegel (1987) o fator β_0 foi interpretado como o componente de longo prazo, sendo reinterpretado em Diebold e Li (2006) como o nível da curva de juros, por ter sua parcela de influência constante em todas as maturidades.

Na ponta curta da curva podemos observar que:

$$\lim_{t \rightarrow 0} R(t) = \beta_0 + \beta_1 \quad (3.5)$$

O fator β_1 foi então interpretado por Nelson e Siegel como o componente de curto prazo, uma vez que seu componente, $(\frac{1 - e^{\lambda t}}{\lambda t})$, começa em 1 e decai à 0 com taxa de decaimento λ . Ou seja, quanto maior o valor de λ mais curto será o período de influência de β_1 .

O componente de β_2 , $(\frac{1 - e^{\lambda t}}{\lambda t} - e^{\lambda t})$, tende a 0 quando $t \rightarrow 0$ e decai a 0 quando $t \rightarrow \infty$. A convexidade gerada é indicativa de que a principal influencia de β_3 se dá próxima ao meio do período em análise. Por este motivo tal fator pode ser entendido como o componente de médio prazo da curva ou, pela interpretação de Diebold e Li (2006), a curvatura.

3.2 Representação em Espaço de Estado

O trabalho de Diebold e Li (2006) reinterpretou o modelo seminal proposto em Nelson e Siegel (1987) introduzindo a dinâmica. Contudo, a estimação até então se dava em dois passos, fixando o parâmetro λ e obtendo os fatores pelo método dos mínimos quadrados. Em Diebold, Rudebusch e Aruoba (2006) o modelo passou a ser colocado já no formato de estado de espaço e os fatores estimados pelo Filtro de Kalman, sendo o λ e os demais hiperparâmetros estimados pelo método da Máxima Verossimilhança¹.

A equação de mensuração do modelo é gerada ao introduzirmos termos de erros ao modelo DNS determinístico. Os erros neste caso são interpretados como idiossincráticos, sendo específicos de cada maturidade. Nesse formato os *yields* ficam relacionados estocasticamente aos três fatores não observáveis (nível, inclinação e curvatura), aqui representados respectivamente por L_t , S_t e C_t ,

$$y_t = \Lambda f_t + \varepsilon_t, \quad (3.6)$$

sendo representada na forma matricial, conforme em Diebold, Rudebusch e Aruoba (2006), da seguinte maneira:

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \left(\frac{1-e^{\lambda\tau_1}}{\lambda\tau_1}\right) & \left(\frac{1-e^{\lambda\tau_1}}{\lambda\tau_1} - e^{\lambda\tau_1}\right) \\ 1 & \left(\frac{1-e^{\lambda\tau_2}}{\lambda\tau_2}\right) & \left(\frac{1-e^{\lambda\tau_2}}{\lambda\tau_2} - e^{\lambda\tau_2}\right) \\ \vdots & \vdots & \vdots \\ 1 & \left(\frac{1-e^{\lambda\tau_N}}{\lambda\tau_N}\right) & \left(\frac{1-e^{\lambda\tau_N}}{\lambda\tau_N} - e^{\lambda\tau_N}\right) \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} + \begin{pmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_N) \end{pmatrix}. \quad (3.7)$$

Para a especificação da dinâmica dos fatores latentes, a equação de transição toma a forma de um modelo autoregressivo vetorial de primeira ordem,

$$(f_t - \mu) = A(f_{t-1} - \mu) + \eta_t, \quad (3.8)$$

também apresentado a seguir na forma matricial.

$$\begin{pmatrix} L_t - \mu_L \\ S_t - \mu_S \\ C_t - \mu_C \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} L_{t-1} - \mu_L \\ S_{t-1} - \mu_S \\ C_{t-1} - \mu_C \end{pmatrix} + \begin{pmatrix} \eta_t(L) \\ \eta_t(S) \\ \eta_t(C) \end{pmatrix} \quad (3.9)$$

Para fins de estimação é necessário notar que o modelo descrito acima não pode ser diretamente implementado, visto que o estado estimado na equação de transição não é exatamente igual ao estado utilizado como variável independente na equação de

¹ Outros métodos de estimação dos hiperparâmetros são também possíveis. Como exemplo, Caldeira, Laurini e Portugal (2011) estimam os hiperparâmetros utilizando métodos bayesianos

mensuração. Para formularmos um modelo diretamente implementável procedemos com uma manipulação algébrica. Neste caso, chamaremos $(f_t - \mu)$ de x_t . Subtraímos então $\Lambda\mu$ dos dois lados da equação de transição:

$$y_t - \Lambda\mu = \Lambda f_t - \Lambda\mu + \varepsilon_t \quad (3.10)$$

$$= \Lambda(f_t - \mu) + \varepsilon_t \quad (3.11)$$

$$y_t - \Lambda\mu = \Lambda x_t + \varepsilon_t \quad (3.12)$$

$$y_t^* = \Lambda x_t + \varepsilon_t \quad (3.13)$$

Com esta manipulação, o modelo será estimado da seguinte forma,

$$y_t^* = \Lambda x_t + \varepsilon_t \quad (3.14)$$

$$x_t = Ax_{t-1} + \eta_t. \quad (3.15)$$

O VAR é um modelo genérico, sendo possível obter diversas especificações dentro do mesmo. Como padrão em outros trabalhos, o pressuposto que os fatores são estacionários no tempo, assim possível regredir essas variáveis em termos de seus desvios em relação às suas respectivas médias:

$$(\mathbf{f}_t - \boldsymbol{\mu}) = \mathbf{A}(\mathbf{f}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\eta}_t$$

Onde $\boldsymbol{\mu}$ é o vetor de médias dos fatores, $\boldsymbol{\eta}$ um vetor de erros aleatórios e \mathbf{A} é uma matriz 3×3 de parâmetros responsáveis pela relação linear entre os regressores e as variáveis dependentes.

Quando utilizados desvios em relação à média, a expressão acima torna-se desvantajosa por não possuir a estrutura necessária em Estado-Espaço, e a implementação no filtro não é direta. Para contornar tal problema, é possível reescrever o sistema da seguinte forma:

$$\mathbf{f}_t = \mathbf{C} + \mathbf{A}\mathbf{f}_{t-1} + \boldsymbol{\eta}_t$$

Em que $\mathbf{C} = (\mathbf{I} - \mathbf{A})\boldsymbol{\mu}$ é um vetor 3×1 de constantes.

Algumas suposições adicionais sobre os erros aleatórios devem ser feitas em torno do modelo para que o filtro de Kalman possa ser executado. A primeira delas refere-se aos vetores $\boldsymbol{\epsilon}_t$. Neste trabalho o pressuposto que os três fatores são suficientes para explicar quaisquer variáveis conjuntas que ocorrem às taxas dos títulos de diferentes maturidades, restando então a variação intrínseca a cada um deles, ou de outra forma, a idiossincrática.

A segunda suposição afeta o último grupo de erros aleatórios $\boldsymbol{\eta}$. Neste trabalho é suposto que os fatores, além da média, mantém a mesma variância ao longo do tempo, de

forma que $\boldsymbol{\eta}_t \sim N(\mathbf{0}_{3 \times 1}, \mathbf{Q})$. Ao contrário do que ocorre no outro grupo de erros, agora a matriz \mathbf{Q} é completa, ou seja, há covariâncias entre os fatores.

3.3 O filtro de Kalman

O filtro de Kalman é um método feito em cima da forma estado-espço de modo recursivo, sempre atualizando a variável omitida, ou de estado, de acordo com uma ponderação entre o que foi medido e o que foi predito. O filtro de Kalman então prova um meio de estimar $y_{t|t-1}$ baseando-se em informações até $t - 1$ e, só então, atualizar o componente permanente conforme a chegada de novas informações.

De modo geral, os modelos em estado-espço trabalham com dois tipos principais de equações, a equação de mensuração, e a de transição. Considerando o modelo deste trabalho:

$$\mathbf{y}_t = \mathbf{A}\mathbf{f}_t + \boldsymbol{\epsilon}_t \quad \text{equação mensuração} \quad (3.16)$$

$$\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{H}_t)$$

$$E(\mathbf{f}_t \boldsymbol{\epsilon}_t) = 0$$

$$\mathbf{f}_t = \mathbf{C} + \mathbf{A}\mathbf{f}_{t-1} + \boldsymbol{\eta}_t \quad \text{equação de transição} \quad (3.17)$$

$$\boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q})$$

$$E(\mathbf{f}_{t-1} \boldsymbol{\eta}_t) = 0$$

Onde é necessário assumir que os erros são não correlacionados, ou seja, $E(\boldsymbol{\epsilon}_t \boldsymbol{\eta}_t) = 0$.

As equações (3.16) e (3.17) de mensuração e de transição juntas formam o modelo estado-espço, as matrizes \mathbf{A} , \mathbf{H}_t e \mathbf{Q} são conhecidas em t , também chamadas de hiperparâmetros. Nesse sistema, já escrito na forma adequada, os fatores \mathbf{f}_t são estocásticos e $\mathbf{A}\mathbf{f}_t$ é conhecido como componente permanente do modelo. Para que o filtro funcione, é necessário ter em mãos um valor inicial para \mathbf{f}_0 e um para a sua matriz covariância \mathbf{P}_0 , os quais são determinados de acordo com a informação que o agente possui em $t = 0$. Este vetor e matriz também são conhecidos como *priori* do modelo. Com a *priori*, é possível prever qual será o valor de β_t quando $t = 1$.

O filtro então funciona recursivamente a cada chegada de nova informação, esse processo basicamente pode ser dividido em dois passos principais:

a) predição: antes do tempo t é necessário prever otimamente o valor que \mathbf{y}_t assumirá.

Essa predição é calculada com base em toda informação disponível até $t - 1$. De modo compacto o valor predito é escrito como $\mathbf{y}_{t|t-1}$ e é feito com base no vetor de fatores preditos para t , escrito compactamente por $\mathbf{f}_{t|t-1}$;

- b) **atualização**: essa etapa ocorre da chegada de uma nova informação, ou seja, novos valores das taxas de juros, dadas por \mathbf{y}_t . O filtro então calcula o erro de predição dessas taxas como $\boldsymbol{\epsilon}_{t|t-1} = \mathbf{y}_t - \mathbf{y}_{t-1}$. Com o erro de predição, os fatores podem ser atualizados, escritos por $\mathbf{f}_{t|t}$, e então voltar à etapa de predição, agora para \mathbf{y}_{t+1} , com base na informação disponível até t , chamado de $\mathbf{y}_{t+1|t}$.

O algoritmo pode ser sumarizado da seguinte forma:

- a) inicialmente são determinados valores para $\mathbf{f}_{0|0}$ e $\mathbf{P}_{0|0}$ com base em conhecimento prévio do pesquisador ou mesmo na estimação dos parâmetros com base numa pré-amostra;
- b) o segundo passo então consiste em calcular os valores dos parâmetros preditos, ela é feita através da equação de transição:

$$\mathbf{f}_{1|0} = \mathbf{C} + \mathbf{A}\mathbf{f}_{0|0} \qquad \mathbf{P}_{1|0} = \mathbf{A}\mathbf{P}_{0|0}\mathbf{A}' + \mathbf{Q}$$

- c) com a chegada de uma nova informação em $t = 1$, definida pelo vetor \mathbf{y}_1 , então o erro de predição e sua variância devem ser calculados da seguinte forma:

$$\begin{aligned} \boldsymbol{\epsilon}_{1|0} &= \mathbf{y}_1 - \mathbf{y}_{1|0} = \mathbf{y}_1 - \mathbf{\Lambda}\mathbf{f}_{1|0} \\ \boldsymbol{\phi}_{1|0} &= \mathbf{\Lambda}\mathbf{P}_{1|0}\mathbf{\Lambda}' + \mathbf{H}_{1|0} \end{aligned}$$

- d) com os resultados do item anterior, o fator $\mathbf{f}_{1|0}$ deve ser corrigido com base no erro de predição $\boldsymbol{\epsilon}_{1|0}$:

$$\begin{aligned} \mathbf{f}_{1|1} &= \mathbf{f}_{1|0} + \mathbf{P}_{1|0}\mathbf{\Lambda}\boldsymbol{\phi}_{1|0}^{-1}\boldsymbol{\epsilon}_{1|0} = \mathbf{B}_{1|0} + \mathbf{K}_1\boldsymbol{\epsilon}_{1|0} \\ \mathbf{P}_{1|1} &= \mathbf{P}_{1|0} - \mathbf{P}_{1|0}\mathbf{\Lambda}'\boldsymbol{\phi}_{1|0}^{-1}\mathbf{\Lambda}\mathbf{P}_{1|0} \end{aligned}$$

Onde \mathbf{K}_1 é conhecido como ganho de Kalman.

- e) com $\mathbf{f}_{1|1}$ e $\mathbf{P}_{1|1}$, o primeiro item deve ser retomado, prevendo $t = 2$ e assim por diante, recursivamente até o final da amostra em $t = T$.

Vale ressaltar que o ganho de Kalman $\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{\Lambda}'\boldsymbol{\phi}_{t|t-1}^{-1}$ é o responsável por determinar o peso dado à nova informação que chegou, definida em $\boldsymbol{\epsilon}_{t|t-1}$.

3.4 A abordagem Bayesiana

Na abordagem clássica, a estimação e consequente inferência sobre o vetor $\tilde{\mathbf{f}}_T = [\mathbf{f}_1 \ \mathbf{f}_2 \ \dots \ \mathbf{f}_T]'$ é feita com base nos hiperparâmetros do modelo \mathbf{A} , \mathbf{H}_t e \mathbf{Q} . Em contraste, a abordagem Bayesiana considera que todos os hiperparâmetros, e não apenas os fatores, são aleatórios. Outro ponto onde ambas abordagens divergem refere-se ao fato que a inferência sobre $\tilde{\mathbf{f}}_T$ é feita conjuntamente com os hiperparâmetros, e não condicionalmente.

Essa característica dessa abordagem torna a estimação via *Markov Chain Monte Carlo* (MCMC) vantajosa frente à alternativas clássicas, cujo procedimento desta última compreende rodar o filtro diversas vezes sob diferentes hiperparâmetros afim de que se encontre aqueles que maximizam a função de verossimilhança do modelo. Como poderá ser visto adiante na seção do estimador MCMC, a abordagem Bayesiana também implica em sucessivas aplicações do filtro de Kalman, todavia o mérito agora reside na natureza do estimador, o qual é rodado uma única vez com diversas iterações, sendo capaz de gerar distribuições chamadas de *posteriors* para todos os parâmetros do modelo.

A ideia básica por trás da abordagem Bayesiana é que ambas variáveis de estado e hiperparâmetros são tratados como dados omitidos, eles devem ser gerados via simulação por métodos MCMC e, a partir de então, a inferência poderá feita.

Ao tratar os parâmetros do modelo como aleatórios, deve-se atribuir distribuições de probabilidade para eles. As distribuições inicialmente atribuídas aos hiperparâmetros e variáveis de estado, também conhecidas como *prioris*, tem de refletir o conhecimento prévio do pesquisador a respeito do comportamento dessas variáveis antes da observação dos dados empíricos. Caso não haja muita informação a priori sobre a distribuição dos parâmetros, a distribuição *priori* é chamada de difusa, uma vez que é pouco informativa, e existem métodos para incorporar essa ausência de informação à estimação.

Ao retirar uma amostra da população, a *priori* deve ser atualizada de forma que a nova informação possa ser incorporada, procedimento qual é feito com base no teorema de Bayes. Como exemplo tome o modelo vigente, neste caso o conjunto de hiperparâmetros e variáveis de estado a serem estimados podem ser representados em θ . Desta maneira pelo teorema de Bayes:

$$\begin{aligned}
 f(Y|\theta) &= \frac{h(Y, \theta)}{g(\theta)}, && \text{rearranjando os termos:} \\
 h(Y, \theta) &= f(Y|\theta)g(\theta) \\
 &= p(\theta|Y)m(Y) \\
 p(\theta|Y) &= \frac{f(Y|\theta)g(\theta)}{m(Y)} && (3.18)
 \end{aligned}$$

Ainda, a distribuição da amostra dada pela função de probabilidade $m(Y)$ pode ser ignorada, uma vez que ela não depende de θ , portanto ela pode ser considerada uma constante e descartada sem perda relevante de informação sobre o conjunto de hiperparâmetros e variáveis de estado:

$$p(\theta|Y) \propto f(Y|\theta)g(\theta) \quad (3.19)$$

A probabilidade de obter as variáveis θ dado que a amostra Y foi retirada equivale à multiplicação da probabilidade da amostra Y ser retirada, dado que os parâmetros são θ e a distribuição do conjunto paramétrico θ .

A distribuição $p(\theta|Y)$ é chamada de *posteriori*, equivalentemente, ela representa a distribuição de θ atualizada em relação à Y , e $g(\theta)$ é a distribuição a *priori* dos parâmetros. Nem sempre é possível obter uma forma analítica para a f.d.p. *posteriori*, quando sim, é dito que a função de verossilhança e a *priori* são conjugadas, caso contrário, métodos de simulação são necessários para que a maximização da *posteriori* em relação à θ seja feita. O estimador *Markov Chain Monte Carlo* é um desses métodos e será abordado em maior detalhamento adiante.

3.4.1 O Suavizador

Neste trabalho é abordado o método sugerido em Carter e Kohn (1994) o qual é capaz de gerar todo vetor $\tilde{\mathbf{f}}_T$ a partir da distribuição com base em toda a amostra $p(\tilde{\mathbf{f}}_T|\tilde{\mathbf{y}}_T)$. Esse método consiste em obter a distribuição $p(\tilde{\mathbf{f}}_T|\tilde{\mathbf{y}}_T)$ construída pela seguinte cadeia:

$$\begin{aligned} p(\tilde{\mathbf{f}}_T|\tilde{\mathbf{y}}_T) &= p(\mathbf{f}_T|\tilde{\mathbf{y}}_T)p(\mathbf{f}_{T-1}|\mathbf{f}_T, \tilde{\mathbf{y}}_T)p(\mathbf{f}_{T-2}|\mathbf{f}_{T-1}, \mathbf{f}_T, \tilde{\mathbf{y}}_T) \\ &\quad \times p(\tilde{\mathbf{f}}_{T-3}|\mathbf{f}_{T-2}, \mathbf{f}_{T-1}, \mathbf{f}_T, \tilde{\mathbf{y}}_T) \end{aligned}$$

Levando-se em consideração que \mathbf{f}_T , condicionado a \mathbf{f}_{T-1} e $\tilde{\mathbf{y}}_t$ não possui mais informação sobre \mathbf{f}_{T-2} além da contida em \mathbf{f}_{T-1} , então os termos condicionantes podem ser omitidos de forma a simplificar a expressão. Ainda, se considerarmos que $\tilde{\mathbf{y}}_T$ não contém informação alguma para gerar \mathbf{f}_{t-i} além daquela disponível em \mathbf{f}_{t-i+1} e $\tilde{\mathbf{y}}_{t-i}$, a expressão pode ser reduzida a:

$$p(\tilde{\mathbf{f}}_T|\tilde{\mathbf{y}}_T) = p(\mathbf{f}_T|\tilde{\mathbf{y}}_T) \prod_{t=1}^{T-1} p(\mathbf{f}_t|\mathbf{f}_{t+1}, \tilde{\mathbf{y}}_t)$$

Em termos mais práticos, deve-se gerar então todo o vetor $\tilde{\mathbf{f}}_T|\tilde{\mathbf{y}}_T$ a partir de \mathbf{f}_T , o qual é obtido pelo tradicional método do filtro de Kalman, somente então $\mathbf{f}_t|\tilde{\mathbf{y}}_T$ é gerado de modo retroativo a partir da distribuição condicionada ao vetor gerado na iteração anterior $p(\mathbf{f}_t|\mathbf{f}_{t+1}, \tilde{\mathbf{y}}_t)$ para $t = T-1, T-1, \dots, 1$, feita através de um suavizador. Esse método é amplamente utilizado na abordagem Bayesiana quanto ao cômputo da distribuição *posteriori* da marginal dos fatores \mathbf{f}_t .

A diferença entre o filtro e o suavizador basicamente refere-se aos dados considerados para a estimação dos estados \mathbf{f}_t , ao passo que o filtro se utiliza dos dados passados e do atual, dados por $\tilde{\mathbf{y}}_{t-1}$ e \mathbf{y}_t respectivamente, o suavizador se utiliza dos dados futuros. Após a obtenção da *posteriori* dos estado através do filtro, \mathbf{f}_T , cujo processo é progressivo em $t = 1, \dots, T$, então o suavizador é capaz de gerar todos os estados condicionados a toda a amostra de forma retroativa, em $t = T-1, \dots, 1$.

A derivação de $\mathbf{f}_{t|t, f_{t+1}} = E(\mathbf{f}_t|\mathbf{f}_{t|t}, \mathbf{f}_{t+1})$ equivale a atualizar a estimação de \mathbf{f}_t combinando $\mathbf{f}_{t|t}$ com as informações disponíveis no vetor gerado \mathbf{f}_{t+1} na recursão anterior. Desta forma, \mathbf{f}_{t+1} pode ser considerado um vetor adicional de dados observados que contém informação além daquela contida em $\mathbf{f}_{t|t}$. As equações de atualização podem ser escritas da seguinte forma:

$$\begin{aligned}\mathbf{f}_{t|t, f_{t+1}} &= E(\mathbf{f}_t|\tilde{\mathbf{y}}_t, \mathbf{f}_{t+1}) \\ &= E(\mathbf{f}_t|\mathbf{f}_{t|t}, \mathbf{f}_{t+1}) \\ &= \mathbf{f}_{t|t} + \mathbf{P}_{t|t} \mathbf{A}' \mathbf{P}_{t+1|t}^{-1} (\mathbf{f}_{t+1|T} - \mathbf{C} - \mathbf{A} \mathbf{f}_{t|t})\end{aligned}\quad (3.20)$$

$$\begin{aligned}\mathbf{P}_{t|t, P_{t+1}} &= \text{Cov}(\mathbf{f}_t|\tilde{\mathbf{y}}_t, \mathbf{f}_{t+1}) \\ &= \text{Cov}(\mathbf{f}_t|\mathbf{f}_{t|t}, \mathbf{f}_{t+1}) \\ &= \mathbf{P}_{t|t} + \mathbf{P}_{t|t} \mathbf{A}' \mathbf{P}_{t+1|t}^{-1} (\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}) \mathbf{P}_{t+1|t}^{-1} \mathbf{A} \mathbf{P}_{t|t}\end{aligned}\quad (3.21)$$

De modo geral, a derivação do suavizador é a mesma do que a dada pelo filtro de Kalman, considerando-se que agora os dados utilizados são os futuros.

3.4.2 A estimação do VAR

O filtro de Kalman e suavizador são capazes de gerar todo o vetor \mathbf{f}_t para $t = 1, \dots, T$, pressupondo que os hiperparâmetros \mathbf{A}, \mathbf{Q} e \mathbf{H}_t são conhecidos a priori. Todavia, da abordagem clássica, geralmente esses parâmetros também devem ser estimados, cujo procedimento é feito de modo exógeno ao filtro de Kalman e suavizador. Na prática, convém rodar o filtro sob diversos valores distintos dessas variáveis e então verificar a escolha de parâmetros que melhor de ajusta aos dados.

A abordagem Bayesiana, por outro lado, trata a estimação dessas variáveis de modo conjunto ao filtro de Kalman. A cada iteração, o filtro é rodado e uma sequência de

vetores, chamados aqui de $\mathbf{f}_t^{(j)}$, é gerada, para $j = 1, \dots, J$. Com esses dados em mãos, é possível estimar um modelo VAR Bayesiano a cada nova iteração, logo, J valores da matriz de parâmetros do VAR e matriz de covariância dos erros serão obtidas, definidas respectivamente por \mathbf{A} e \mathbf{Q} . Tal como nos demais casos, o período de *burn-in* deve ser igualmente respeitado por questão de convergência das distribuições. Com os dados restantes em mãos é possível construir as *posteriors* e a inferência poderá ser feita.

A estimação Bayesiana necessita da incorporação do conhecimento prévio do pesquisador a respeito do comportamento dos parâmetros, de outro modo, a partir desta distribuição o estimador se ajustará aos dados, então é de suma importância que se escolha bem a distribuição *priori*. Para facilitar a explanação, tome novamente o modelo VAR, onde o índice j foi omitido por facilidade de notação:

$$\mathbf{f}_t = \mathbf{C} + \mathbf{A}\mathbf{f}_{t-1} + \boldsymbol{\eta}_t$$

Alternativamente, uma forma muito comum de reescrever este mesmo modelo é:

$$\mathbf{f}'_t = \begin{bmatrix} 1 & \mathbf{f}'_{t-1} \end{bmatrix} \times \begin{bmatrix} \mathbf{C}' \\ \mathbf{A} \end{bmatrix} + \boldsymbol{\eta}'_t \quad (3.22)$$

$$= \mathbf{x}'_t \mathbf{A}^* + \boldsymbol{\eta}'_t \quad (3.23)$$

Definindo \mathbf{y} como o vetor $MT \times 1$ que empilha \mathbf{f}_t , \mathbf{X} a matriz $T \times (M + 1)$ que empilha x_t e $\boldsymbol{\epsilon}$ o vetor $MT \times 1$ de erros aleatórios, o sistema VAR pode ainda ser reescrito como:

$$\mathbf{y} = (\mathbf{I}_M \otimes \mathbf{X})\boldsymbol{\alpha} + \boldsymbol{\epsilon}$$

Onde $\boldsymbol{\alpha} = \text{vec}(\mathbf{A})$ é o vetor de parâmetros empilhados, ou vetorizados.

Se os erros desse modelo são gaussianos, é comum atribuir uma distribuição *priori* normal para o vetor $\boldsymbol{\alpha}$. Quando for o caso da normalidade nos dados, então o inverso da matriz de covariâncias segue uma distribuição Wishart, ou de forma similar, a matriz de covariâncias tem uma distribuição Wishart inversa. De forma compacta essa relação é descrita como:

$$\boldsymbol{\alpha} | \mathbf{Q} \sim N(\boldsymbol{\alpha}, \boldsymbol{\Sigma}) \quad (3.24)$$

$$\mathbf{Q} \sim iW(\mathbf{Q}, \underline{\nu}) \quad (3.25)$$

Com as distribuições *priori* em mãos, a *posteriori* conjunta toma a seguinte forma:

$$p(\mathbf{A}, \mathbf{Q} | \tilde{\mathbf{y}}_T) \propto |\mathbf{Q}|^{-(T+M+1)/2} \exp \left[-\frac{1}{2} (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' (\mathbf{Q}^{-1} \otimes \mathbf{X}' \mathbf{X}) (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}) \right] \quad (3.26)$$

$$\times \exp \left[-\frac{1}{2} (\boldsymbol{\alpha} - \underline{\boldsymbol{\alpha}})' \underline{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\alpha} - \underline{\boldsymbol{\alpha}}) \right] \quad (3.27)$$

Onde o primeiro grupo exponencial representa a função de verossimilhança do modelo. Deste modo, é possível obter as *posteriors* condicionais como:

$$\boldsymbol{\alpha} | \mathbf{Q}, \tilde{\mathbf{y}}_T \sim N(\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\Sigma}}) \quad (3.28)$$

$$\bar{\boldsymbol{\Sigma}} = (\underline{\boldsymbol{\Sigma}} + \mathbf{Q}^{-1} \otimes \mathbf{X}' \mathbf{X})^{-1} \quad (3.29)$$

$$\bar{\boldsymbol{\alpha}} = \bar{\boldsymbol{\Sigma}} [\underline{\boldsymbol{\Sigma}}^{-1} \underline{\boldsymbol{\alpha}} + (\mathbf{Q}^{-1} \otimes \mathbf{X}' \mathbf{X}) \hat{\boldsymbol{\alpha}}] = \bar{\boldsymbol{\Sigma}} [\underline{\boldsymbol{\Sigma}}^{-1} \underline{\boldsymbol{\alpha}} + \text{vec}(\mathbf{X}' \mathbf{Y} \mathbf{Q}^{-1})] \quad (3.30)$$

E a *posteriori* condicional para \mathbf{Q} segue como:

$$\mathbf{Q} | \mathbf{A}, \tilde{\mathbf{y}}_T \sim iW(\bar{\mathbf{Q}}, \bar{\nu}) \quad (3.31)$$

$$\bar{\mathbf{Q}} = \underline{\mathbf{S}} + (\mathbf{Y} - \mathbf{X} \mathbf{A}^*)' (\mathbf{Y} - \mathbf{X} \mathbf{A}^*) \quad (3.32)$$

$$\bar{\nu} = T + \underline{\nu} \quad (3.33)$$

3.4.3 O estimador MCMC

Quando as distribuições *posteriori* tem forma funcional conhecida, a inferência estatística é direta, todavia, nem sempre elas estão dispostas com uma forma analítica tratável. O método *Markov Chain Monte Carlo*, também conhecido como MCMC, é uma forma de estimar a distribuição *posteriori* via simulação, driblando o desconhecimento do formato das *posteriors*.

O MCMC gera valores aleatórios a partir da distribuição condicional $f(\boldsymbol{\theta}^{(j+1)} | \boldsymbol{\theta}^{(j)})$. Quando $f(\cdot)$ for estacionária, então o vetor inicial de parâmetros $\boldsymbol{\theta}^{(0)}$ pode ser gerado a partir da real *posteriori*, conseqüentemente todos os demais $\boldsymbol{\theta}^{(j)}$, para $j = 1, \dots, J$ serão obtidos a partir desta mesma distribuição.

Quando não dispomos da forma analítica da *posteriori*, o vetor inicial de parâmetros $\boldsymbol{\theta}^{(0)}$ parte de um ponto arbitrário do espaço paramétrico, nasce então o problema da convergência. Ao passo que os valores $\boldsymbol{\theta}^{(j)}$ vão sendo gerados, estes convergem à verdadeira distribuição dos parâmetros, porém o número de iterações necessárias a convergência é o real problema.

Sendo possível gerar infinitos valores $\boldsymbol{\theta}^{(j)}$ tal que $j \rightarrow \infty$, a convergência é garantida. Como obter infinitos valores de θ é impossível, a aproximação deve ser feita de modo a gerar J parâmetros diferentes, com J suficientemente grande a fim de que se

obtenha um nível convergência desejado, caso contrário, estimações imprecisas podem ocorrer. Dentre os J valores gerados, comumente os B primeiros devem ser descartados, uma vez que podem encontrar-se em pontos pouco verossímeis do espaço paramétrico. Esse período de descarte também é conhecido como *burn-in*.

De modo geral não há uma regra clara para definir o tamanho do período *burn-in*, caso o pesquisador tenha uma boa experiência, ou informação a priori sobre o provável valor assumido pelos parâmetros, este assumirá $\boldsymbol{\theta}^{(0)}$ de modo que o mesmo esteja próximo da média da distribuição *posteriori*. Um segundo ponto deve ser levado em consideração ao definir o tamanho do *burn-in*, a questão convergência relaciona-se ao quão auto-correlacionados estão esses valores gerados. Caso essa auto-correlação seja alta, a velocidade de convergência será mais lenta e mais iterações serão necessárias.

3.4.3.1 A Amostragem de Gibbs

Esse tipo de amostragem é uma simulação tipo Monte Carlo feita em Cadeias de Markov. Sua motivação reside no fato de que nem sempre é possível transformar a *posteriori* numa forma analítica tratável do ponto de vista matemático.

Para construir o modelo Bayesiano dinâmico de Nelson-Siegel, é necessário selecionar:

- os valores iniciais para todas as matrizes de covariâncias dos erros do modelo em $t = 1, \dots, T$, alocados em $\tilde{\mathbf{H}}_T^{(0)}$;
- um valor inicial para a matriz $\mathbf{A}^{(0)}$ e outro para $\mathbf{Q}^{(0)}$;
- os valores iniciais \mathbf{f}_0 e \mathbf{P}_0 para iniciar o filtro de Kalman para a estimação dos fatores;
- uma *priori* para a matriz \mathbf{Q} , dada por $\overline{\mathbf{Q}}$, para a sua estimação via distribuição Wishart inversa;
- os valores iniciais das variáveis $h_{i,0}$ e $\text{Var}(h_{i,0})$ para iniciar o filtro de kalman.
- uma *priori* para a matriz $\boldsymbol{\sigma}_\omega$, igualmente diagonal, onde a estimação da *posteriori* é feita individualmente em cada elemento, dada pela distribuição Gama inversa;

Então, para $j = B + R$ valores dos parâmetros gerados, quando os B valores iniciais são descartados no *burn-in*, deve-se:

1. inicialmente gerar os parâmetros da regressão $\tilde{\mathbf{f}}_T$ pelo filtro de Kalman tradicional com base na distribuição *posteriori*

$$\mathbf{f}_t | \tilde{\mathbf{Y}}_t, \tilde{\mathbf{H}}_t^{(j-1)}, \mathbf{A}^{(j-1)}, \mathbf{Q}^{(j-1)} \sim N(\overline{\mathbf{f}}_t, \overline{\mathbf{H}}_t)$$

- e então rodado o suavizador no modo convencional;
2. próximo passo consiste em estimar a matriz \mathbf{Q} com base na distribuição *posteriori* $\mathbf{Q}|\mathbf{f}_t^{(j)}, \mathbf{A}^{(j-1)} \sim iW(\bar{\mathbf{Q}}, \bar{v})$.
 3. então a matriz de transição do sistema \mathbf{A} pode ser obtida através da estimação do VAR e um valor é gerado para essa matriz pela *posteriori* $\mathbf{A}|\tilde{\mathbf{f}}_T^{(j)}, \mathbf{Q}^{(j)} \sim N(\bar{\mathbf{A}}, \mathbf{Q})$;
 4. o próximo passo é gerar as log-volatilidades $\mathbf{h}_T^{(j)}$ a partir da distribuição *posteriori* condicional $\mathbf{h}_T|\tilde{\mathbf{f}}_T, \mathbf{V}_e t a^{j-1}, \tilde{\boldsymbol{\delta}}_T^{(j-1)} \sim N(\mathbf{s}_{T|T}, \mathbf{V}_{T|T})$, cujos parâmetros são obtidos do filtro de Kalman. Então para $t = T - 1, \dots, 1$ os $\mathbf{h}_t^{(j)}$ podem ser obtidos do suavizador, cuja *posteriori* condicional é dada pela distribuição de $\mathbf{h}_T|\tilde{\mathbf{f}}_t$;
 5. deve-se então gerar a matriz $\boldsymbol{\sigma}_\omega^{(j)}$ da *posteriori* condicional $\boldsymbol{\sigma}_\omega|\tilde{\mathbf{h}}_T^{(j)} \sim iW(\bar{\boldsymbol{\sigma}}_\omega, \bar{v}_\omega)$;
 6. deve-se então repetir o procedimento $j = B + R$ vezes, afim de que sejam descartados os B valores iniciais obtidos dos parâmetros por questão de convergência;
 7. quando $j > B$ então entende-se que os valores gerados de todos os parâmetros estimados advém de distribuições muito próximas de suas respectivas reais *posteriors*. Então seus valores são guardados para o cômputo de suas médias e demais momentos amostrais necessários à inferência estatística e construção da função impulso-resposta.

3.4.4 Escolha das *prioris*

A escolha das *prioris* na estimação bayesiana toma um papel crucial se levarmos em consideração que ela afeta a performance do estimador quanto a velocidade de convergência à real distribuição *posteriori*, caso seja mal especificada, poderá causar grande dano ao trabalho, levando a estimativas imprecisas.

Uma boa opção, e já amplamente utilizada, é a *priori* de Minnesota. Ela foi elaborada por pesquisadores da *University of Minnesota* e do FED de Minneapolis e pode ser, sobretudo, atribuída aos trabalhos de Litterman (1979), Litterman (1985) e Litterman (1986). A proposta inicial para esta distribuição era simplificar a modelagem de dados macroeconômicos via incorporação de restrições, as quais eram feitas para facilitar a computação dos modelos da época. Essa *priori* foi concebida com foco nos fatos estilizados apresentados pelos dados macroeconômicos e ganhou popularidade devido à sua simplicidade e resultados satisfatórios.

Muito da proposta original de Litterman modificou-se com o tempo, surgindo deste modo diversas especificações para as distribuições, portanto a *priori* de Minnesota pode ser entendida como um grupo de distribuições que visam incorporar fatos estilizados dos dados em questão, representados pela crença do pesquisador antes de rodar o modelo. Não obstante, dados em finanças apresentam algumas características que justificam o uso de uma *priori* de Minnesota.

Tome novamente o sistema VAR com M variáveis dependentes, esse sistema tem sua matriz de parâmetros representada por \mathbf{A}_M , cujos elementos são definidos por α_{ij} , com $i, j = 1, \dots, M$. Dentre os diversos valores esperados para α_{ij} , deve ser levado em consideração que o tratamento do modelo ocorre em torno das médias dos fatores, ou seja, o VAR é tido como estacionário. Quando tal, uma opção para a média desses parâmetros é colocá-los como $\underline{\alpha}_{ij} = 0$. Alternativamente, é possível se se escolha algum valor no intervalo $(0, 1)$ de acordo com a persistência esperada dos fatores para retornarem às suas médias, valores maiores que 0 geralmente são indicados para aqueles parâmetros correspondentes às variáveis auto-defasadas.

Ademais, a *priori* deve atribuir valores para a matriz de covariâncias dos parâmetros. Chamemos essa matriz de $\underline{\mathbf{P}}$, cuja ordem é M^2 . Uma configuração diagonal pode ser especificada, indicando que não são esperadas correlações entre diferentes parâmetros α_{ij} . Se pegarmos cada equação do VAR individualmente, indexadas por i , chamamos a matriz de covariâncias dos coeficientes relativos à i -ésima equação como $\underline{\mathbf{P}}_i$. Ainda, os elementos da diagonal de $\underline{\mathbf{P}}_i$ são chamados de $\underline{\mathbf{P}}_{i,jj}$, deste modo a *priori* de Minnesota atribui os seguintes valores à esses elementos:

$$\underline{\mathbf{P}}_{i,jj} = \begin{cases} \underline{a}_1 & \text{para os coeficientes das variáveis defasadas nelas mesmas;} \\ \underline{a}_2 \frac{\sigma_{ii}}{\sigma_{jj}} & \text{para os coeficientes das demais variáveis defasadas.} \end{cases}$$

A grande vantagem dessa *priori* de Minnesota é evitar complicações na escolha correta da especificação da *priori* no caso Bayesiano, reduzindo o problema à escolha dos três escalares \underline{a}_1 , e \underline{a}_2 .

Outro aspecto importante desta *priori* ocorre da especificação dos valores \underline{a}_i , comumente as próprias variáveis endógenas defasadas tem um poder explicativo para elas mesmas em t maior do que as demais variáveis endógenas defasadas. Isso pode ser expresso então colocando $\underline{a}_1 > \underline{a}_2$. Em outras palavras, o pesquisador tem uma crença mais forte de que os parâmetros em $\boldsymbol{\alpha}$ são iguais 0 para aqueles referentes às defasagens das outras variáveis endógenas do que os parâmetros em $\boldsymbol{\alpha}$ referentes à elas mesmas.

Por fim, vale ressaltar que a *priori* de Minnesota não é específica completamente todos os parâmetros necessários para a estimação e inferência Bayesiana. Essa *priori* não diz nada sobre o comportamento da matriz de covariâncias dos erros do modelo, definidos em \mathbf{H}_t .

4 A ALOCAÇÃO DE RECURSOS

A teoria de portfólio desenvolvida por Markowitz (1952) ainda é uma das bases da teoria moderna de finanças, sendo amplamente utilizada por gestores de recursos. Nela, Markowitz propôs um modelo aonde os investidores se importam apenas com o risco, medido pela matriz de covariância dos retornos dos ativos, e o retorno esperado da carteira, de modo que para duas carteiras com o mesmo retorno esperado, a de menor risco será a preferida. Na abordagem conhecida por média-variância o problema do investidor é escolher a ponderação que dará a cada ativo, de modo que maximize o retorno esperado a um dado nível de risco escolhido. Matematicamente podemos definir o problema de otimização do investidor da seguinte forma:

$$\begin{aligned} & \max_{w_t} && w_t' \mu_{r_t|t-1} \\ & \text{sujeito a} && \\ & && w_t \Sigma_{r_t|t-1} w_t' \leq b_i \\ & && w_t' l = 1 \end{aligned}$$

onde w_t é o vetor de pesos a ser otimizado, $r_{t|t-1}$ é o vetor de previsões um passo à frente dos retornos e $\Sigma_{r_t|t-1}$ é a matriz, predita um passo à frente, de covariância condicional dos retornos.

Uma abordagem alternativa à carteira de média-variância é a carteira de mínima variância. Neste contexto, o problema de otimização do investidor é escolher a ponderação de cada ativo, de modo a obter o portfólio de menor risco.

$$\begin{aligned} & \min_{w_t} && w_t \Sigma_{r_t|t-1} w_t' \\ & \text{sujeito a} && \\ & && w_t' l = 1. \end{aligned}$$

Para gestão de riscos em portfólios de renda fixa é padrão o cálculo da *duration* como medida de mitigação do risco de taxa de juros. Assim como em Caldeira, Moura e Santos (2015), apresentamos uma extensão ao modelo de mínima variância para o caso de renda fixa, na qual é imposta uma restrição à *duration* do portfólio.

$$\begin{aligned} & \min_{w_t} && w_t \Sigma_{r_t|t-1} w_t' \\ & \text{sujeito a} && \\ & && \tau' w_t \leq D \\ & && w_t' l = 1, \end{aligned}$$

em que D é a *duration* a ser indexada o portfólio.

4.1 Estratégia de portfólio no mercado de renda fixa

Embora muito utilizadas no contexto de portfólio de ações, a otimização baseada em média-variância ainda não é tão popular em carteiras de renda fixa. Caldeira, Moura e Santos (2015) apresentam dois motivos que explicam este fato. A primeira razão seria a baixa volatilidade histórica e estabilidade dessa classe de ativos, fazendo com que não houvessem grandes incentivos para que se usassem metodologias sofisticadas de gerenciamento de risco neste mercado. Contudo, os autores notam que a turbulência nos mercados globais têm levado a uma elevação na volatilidade dos preços dos títulos, passando então a abordagem por otimização de portfólio a ter importância, inclusive, no processo de diversificação do risco através de distintas maturidades. O segundo motivo, como defendido também por Korn e Koziol (2006) e Puhle (2008), é que a não popularidade da abordagem de Markowitz ao mercado de renda fixa se deve a dificuldades na modelagem de retornos e da matriz de covariância dos títulos.

Diante desse fato, Caldeira, Moura e Santos (2015) propuseram uma nova abordagem para a obtenção do vetor de retornos esperados e de sua matriz de covariância, tendo como foco os modelos de fatores. O uso destes modelos é uma evolução ao trabalho de Puhle (2008), visto que os modelos de fatores, em especial os modelos da classe de Nelson Siegel, se provaram muito eficientes em prever a curva de juros, conforme Diebold e Li (2006).

4.1.1 A distribuição dos log-retornos

O sistema formado pelas equações do modelo DNS implica a seguinte distribuição para a previsão um passo à frente dos *yields* dos títulos zero cupom continuamente compostos e de seus dois momentos:

$$y_{t|t-1} \sim N(\mu_{y_{t|t-1}}, \Sigma_{y_{t|t-1}}) \quad (4.1)$$

$$\mu_{y_{t|t-1}} = \Gamma + \Lambda \hat{f}_{t|t-1} \quad (4.2)$$

$$\Sigma_{y_{t|t-1}} = \Lambda Q_{t|t-1} \Lambda' + \Sigma \quad (4.3)$$

aonde $\hat{f}_{t|t-1}$ é o valor esperado dos fatores no período t , condicionado à informação disponível em $t-1$. De outra forma, podemos notar que $\hat{f}_{t|t-1} = E_{t-1}[f_t] = \mu + \Upsilon \hat{f}_{t-1|t-1}$ são previsões dos estados baseadas nas estimativas filtradas do modelo, $\hat{f}_{t-1|t-1}$. Note que a matriz de covariância do estado verdadeiro, mas não observado (f_t), seria Ω . Contudo, ao utilizar a estimativa filtrada, utilizamos a matriz de covariância dos estados filtrados, $Q_{t|t-1} = Var_{t-1}[f_t] = \Omega + \Upsilon Q_{t-1|t-1} \Upsilon'$, e não a matriz de covariância dos fatores não observados, Ω . Assim, levamos em consideração o fato de não conhecermos os verdadeiros estados não observados e a consequente incerteza gerada nas estimativas do filtro de Kalman.

Para calcularmos o retorno realizado, $r_{\tau_i,t}$, ao mantermos o título durante o período entre $t-1$ e t , considerando o decréscimo na maturidade de τ_i para τ_{i-1} , usamos a seguinte expressão para o log-retorno:

$$r_{\tau_i,t} = \log \left(\frac{P_{\tau_{i-1},t}}{P_{\tau_i,t-1}} \right) = \log P_{\tau_{i-1},t} - \log P_{\tau_i,t-1} = -\tau_{i-1} \cdot y_{i-1,t} + \tau_i \cdot y_{i,t-1} \quad (4.4)$$

Assim, com as equações acima é possível notar que as previsões um passo á frente dos log-retornos dos títulos zero cupom, $r_{\tau_i,t|t-1}$ são normalmente distribuídas, com a média dada por:

$$\mu_{r_{\tau_i,t|t-1}} = -\tau_{i-1} \cdot \mu_{r_{\tau_{i-1},t|t-1}} + \tau_i \cdot y_{i,t-1}, \quad (4.5)$$

e uma matriz de covariância $\Sigma_{r_{t|t-1}}$ com os elementos da diagonal dados por:

$$\sigma_{r_{\tau_i,t|t-1}}^2 = \tau_{i-1}^2 \cdot \left[\underbrace{\Lambda_{i-1} Q_{t|t-1} \Lambda'_{i-1} + \sigma_{i-1}^2}_{\sigma_{y_{i-1},t|t-1}^2} \right] \quad (4.6)$$

aonde Λ_{i-1} é a linha $(i-1)$ da matriz Λ , σ_{i-1}^2 é o $(i-1)$ -ésimo elemento da diagonal de Σ , $\sigma_{y_{i-1},t|t-1}$ é o $(i-1)$ -ésimo elemento da diagonal de $\Sigma_{y_{t|t-1}}$, e $i = 2, \dots, N$. Os elementos fora da diagonal principal, ou seja, a covariância entre os retornos esperados $r_{\tau_i,t|t-1}$ e $r_{\tau_j,t|t-1}$ são dados por:

$$\sigma_{r_{\tau_i}, r_{\tau_j}} = \tau_{i-1} \tau_{j-1} \cdot \left[\underbrace{\Lambda_{i-1} Q_{t|t-1} \Lambda'_{j-1} + \sigma_{i-1,j-1}}_{\sigma_{y_{i-1},y_{j-1}}} \right], \quad (4.7)$$

aonde $\sigma_{i-1,j-1}$ é o elemento $(i-1, j-1)$ de Σ , e $\sigma_{y_{i-1},y_{j-1}}$ é o elemento $(i-1, j-1)$ da matriz de covariância dos *yields* esperados, $\Sigma_{y_{t|t-1}}$.

5 RESULTADOS

Nesta seção descreveremos os dados utilizados e os resultados obtidos neste trabalho.

5.1 Dados

Os dados utilizados neste trabalho serão as taxas dos contratos futuros de depósito interbancário (DI1), preço de fechamento, negociados na BM&FBovespa. Estes são contratos futuros em que o ativo subjacente é a média da taxa de juros apurada diariamente sobre os depósitos interbancários (CDI), capitalizada durante o período de maturidade do título, τ . Tais contratos são padronizados e têm seu valor no vencimento de (R\$ 100.000,00). Assim, o valor corrente pode ser encontrado descontando-se tal valor pela taxa negociada, base 252 dias úteis. Caldeira, Laurini e Portugal (2011) apresentam a fórmula de cálculo do resultado financeiro de um investimento em um contrato de DI futuro por um preço DI^* e mantido até o vencimento, na maturidade τ :

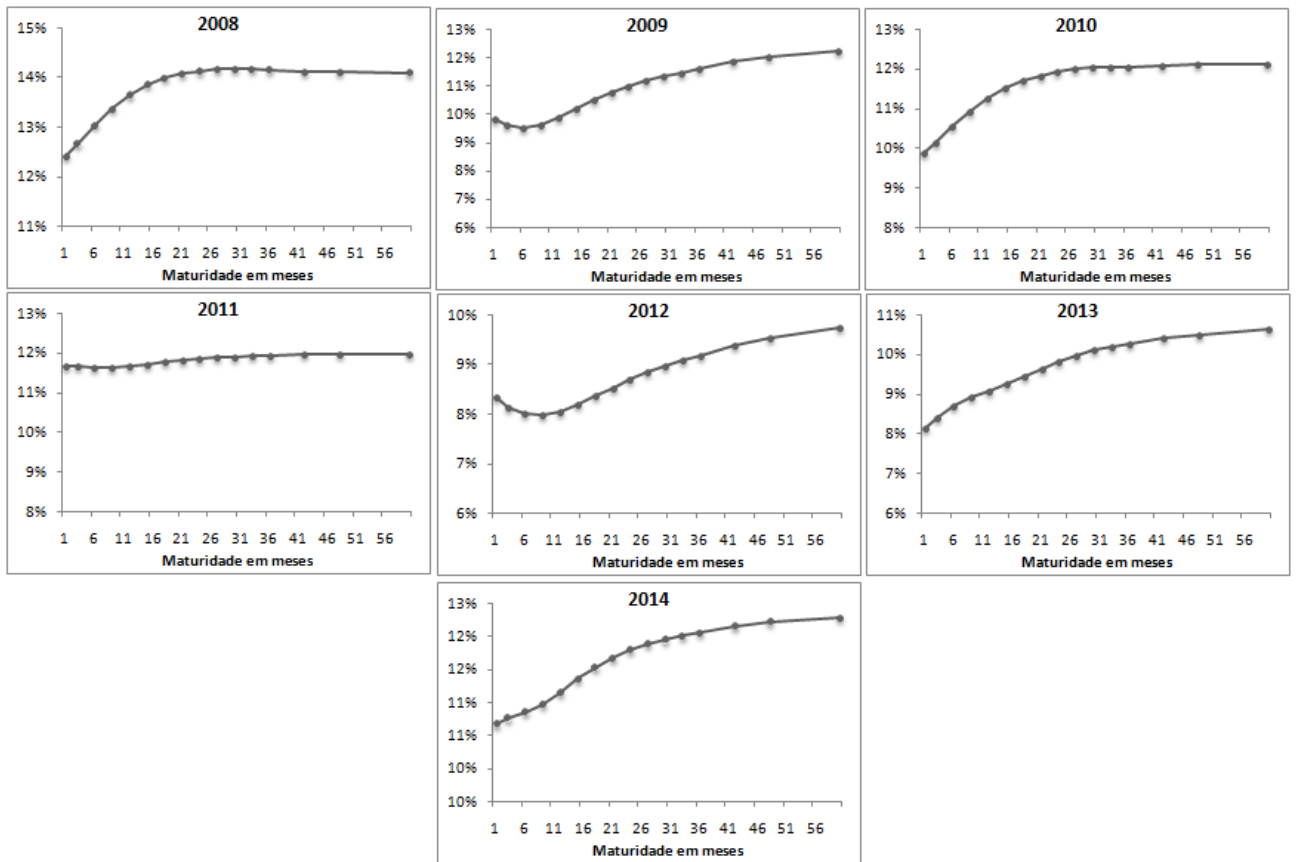
$$100.000 \left(\frac{\prod_{i=1}^{\zeta(t,\tau)} (1 + y_i)^{\frac{1}{252}}}{(1 + DI)^{\frac{\zeta(t,\tau)}{252}}} - 1 \right), \quad (5.1)$$

onde y_i representa a taxa DI ($i - 1$) dias após o dia de negociação e ζ é uma função representativa do número de dias entre t e τ .

O contrato futuro de DI é, portanto, similar a um título zero cupom, com a exceção de que existem pagamentos de ajustes marginais: todos os dias, o fluxo de caixa é igual à diferença entre o preço de ajuste do dia atual e o preço de ajuste do dia anterior corrigido pelo CDI do dia anterior (JÚNIOR, 2013).

As curvas de juros médias anuais são apresentadas abaixo. É possível verificar no período, diversas das formas as quais ela pode tomar, sendo positivamente inclinada em média.

Figura 1 – Curva de Juros média ano



Fonte: elaboração própria (2016)

Os dados utilizados neste trabalho se comportam como o previsto pelos fatos estilizados. Como pode ser visto pela tabela abaixo, a curva de juros média é positivamente inclinada, as maturidades mais longas são menos voláteis e os *spreads* se mostram menos persistentes que os *yields*.

Tabela 1 – Estatísticas Descritivas - Dados Mensais

Yields									
	M1	M3	M6	M9	M12	M24	M36	M48	M60
$\bar{y}(\%)$	10,63	10,64	10,73	10,85	11,00	11,54	11,80	11,94	12,02
$\hat{\sigma}_y(\%)$	2,01	1,98	1,97	1,97	1,97	1,85	1,72	1,64	1,59
$\hat{\rho}_y(1)$	0,98	0,98	0,98	0,97	0,97	0,95	0,94	0,93	0,93
$\hat{\rho}_y(6)$	0,61	0,60	0,57	0,55	0,53	0,49	0,46	0,45	0,44

Spread									
	M3	M6	M9	M12	M24	M36	M48	M60	
$\bar{S}(\%)$	0,01	0,09	0,22	0,36	0,91	1,17	1,31	1,39	
$\hat{\sigma}_S(\%)$	0,27	0,56	0,77	0,92	1,20	1,30	1,35	1,39	
$\hat{\rho}_S(1)$	0,87	0,89	0,88	0,89	0,89	0,90	0,90	0,90	
$\hat{\rho}_S(6)$	0,14	0,15	0,18	0,20	0,28	0,33	0,38	0,40	

Fonte: elaboração própria (2016)

Para a estratégia de portfólios de renda fixa otimizados foram utilizados como *benchmark* os índices IRF-M e IRF-M1. O IRF-M mede a evolução, a preços de mercado, de uma carteira de Letras do Tesouro Nacional (LTN). Por isso, ele é muito utilizado como indexador de carteiras de renda fixa pré-fixadas. O IRF-M1 utiliza metodologia semelhante, porém restringe os títulos a possuírem maturidade máxima de até 1 ano.

A fim de analisarmos a *performance* do portfólio otimizado também em relação aos pares de mercado, serão utilizados os retornos de 13 fundos, retirados da base da Economática. Os critérios de seleção para os fundos foram: (i) a data de início (o fundo deveria possuir retornos em todos o período de tempo em que o portfólio foi testado), (ii) o *benchmark* deveria ser IRF-M ou IRF-M1 e, (iii) os fundos deveriam possuir prazo de resgate D+0. A descrição dos fundos pode ser encontrada no apêndice deste trabalho.

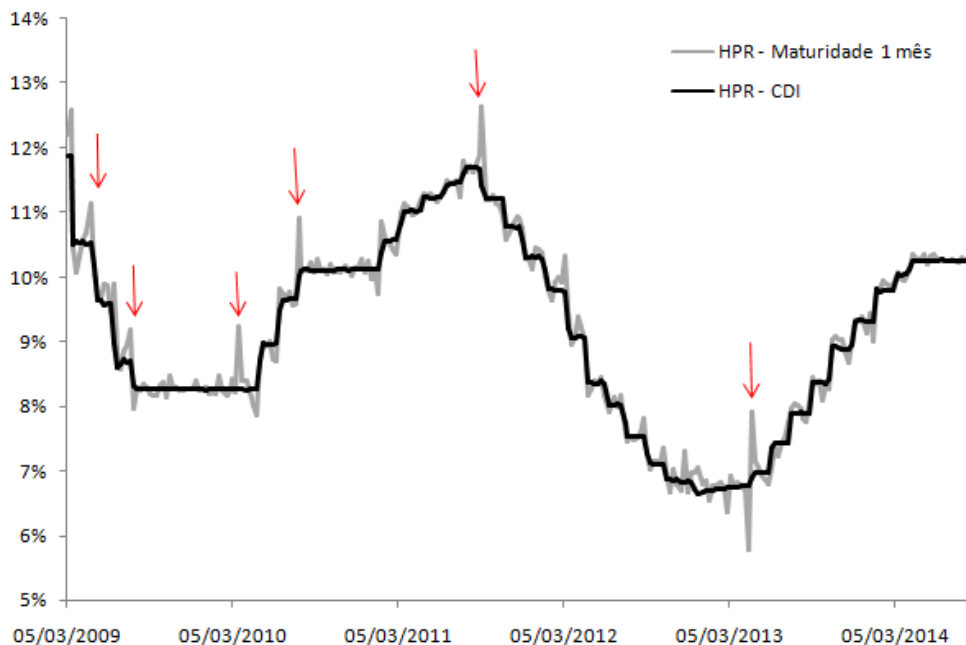
5.2 Teste da Hipótese das Expectativas

Como apresentado anteriormente, a Hipótese das Expectativas, sob expectativas racionais, prevê que o excesso de retorno observado no mercado ao longo do tempo entre um título de longo prazo e um título de curto prazo deve se assemelhar a um ruído branco. A análise consiste então em observar e modelar os dados, verificando suas propriedades estatísticas e confrontando com as previstas pela teoria.

Optou-se por analisar séries de excesso de retorno em base semanal, utilizando o CDI para cálculo do excesso de retorno. Os dados compreendem o período entre 05/03/09 e 04/09/2014, totalizando 275 observações. Para fins de análise preliminar, apresentamos o

gráfico de rentabilidade diária do título com maturidade de 1 mês e a rentabilidade diária do CDI, ambos com valor anualizado.

Figura 2 – Retornos diários (%a.a.)



Fonte: elaboração própria (2016)

Pelo gráfico acima podemos notar uma característica presente no mercado de juros futuros. Como destacado pelas setas, os excessos de retorno dos títulos de 1 mês em relação ao CDI são bastante significativos em momentos que precedem alterações no ciclo de taxa de juros. Este fato nos reforça a idéia de que agentes demandam prêmios de risco para títulos de maturidade superior. Em seguida iniciamos o teste na série gerada.

O primeiro teste consistiu em verificar a autocorrelação das séries e do quadrado das séries utilizando o teste de Ljung-Box.

Tabela 2 – Teste de Ljung-Box (p-valor)

lag	M1	M1 ²
1	0,93	0,03
2	0,96	0,04
3	0,62	0,14
4	0,66	0,24

Fonte: elaboração própria (2016)

Podemos notar que não há evidência de autocorrelação na média, porém há sinais de heterocedasticidade condicionada. Assim, fazemos uso dos modelos da família

ARCH/GARCH. Para selecionar o que melhor se ajusta nos balizaremos pelos critérios de informação. Abaixo segue a estimação para a série de excesso de retornos da maturidade de 1 mês.

Tabela 3 – Critérios de Informação - Maturidade 1 mês

	(0,1)	(0,2)	(1,0)	(1,1)	EGARCH (1,1)
Akaike	-9,94	-9,93	-8,01	-9,94	-9,99
Bayes	-9,88	-9,88	-7,96	-9,90	-9,90
Shibata	-9,44	-9,93	-8,01	-9,95	-9,99
Hannan-Quinn	-9,92	-9,91	-7,99	-9,93	-9,95

Fonte: elaboração própria (2016)

Após selecionado o modelo, testamos a estabilidade dos parâmetros e do modelo por meio do teste de Nyblom-Hansen. Os dados abaixo sugerem a estabilidade do modelo e dos parâmetros individualmente.

Tabela 4 – Teste de Estabilidade de Nyblom-Hansen

Estatística Conjunta:	1,3232		
Estatísticas Individuais:			
μ	0,10		
δ	0,12		
φ	0,44		
γ	0,13		
λ	0,15		
Valores Críticos Assintóticos	(10%	5%	1%)
Estatística Conjunta:	1,49	1,68	2,12
Estatísticas Individuais:	0,35	0,47	0,75

Fonte: elaboração própria (2016)

O modelo estimado obteve os seguintes parâmetros:

$$rx_t(\tau) = \theta^\tau + \varepsilon_t \quad (5.2)$$

$$\theta^\tau = -0,0002 + 0,1589h_t \quad (5.3)$$

$$\ln(h_t) = -0,039 \frac{\varepsilon_{t-1}^2}{h_{t-1}^{0,5}} + 0,121 \left| \frac{\varepsilon_{t-1}^2}{h_{t-1}^{0,5}} \right| + 0,895 \ln(h_{t-1})$$

(5.5)

$$\varepsilon_t = \nu_t \sqrt{h_t}. \quad (5.6)$$

	Estimativa	Erro Padrão	valor t	Pr(> t)
μ	-0.0002	0.000021	-9.5259	0.0000
δ	0.1589	0.009495	16.7378	0.0000
φ	-0.0393	0.007885	-4.9846	0.0000
γ	0.8952	0.000425	2353.8603	0.0000
λ	0.1212	0.014459	8.3864	0.0000

Fonte: elaboração própria (2016)

Dado que há evidências de que os parâmetros são constantes no tempo e estatisticamente significantes, o último passo consiste em testar a série dos resíduos padronizados, ν , que deve ter as propriedades de um ruído branco. Como notado na tabela abaixo, a série não possui evidências de autocorrelação a 5% de significância.

Tabela 5 – Teste de Ljung-Box (p-valor)

lag	ν	ν^2
1	0,92	0,08
2	0,95	0,08
3	0,62	0,14
4	0,66	0,24

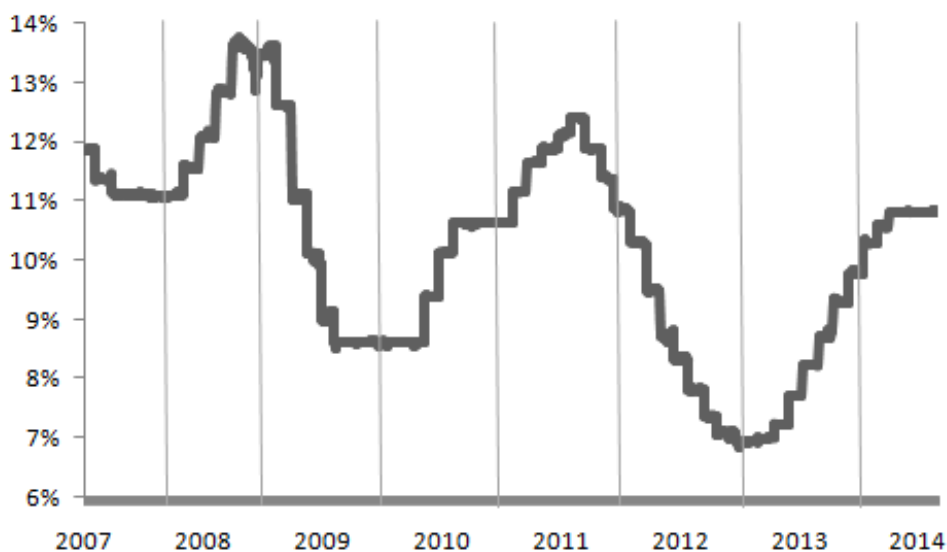
Fonte: elaboração própria (2016)

5.3 Retorno das Carteiras Otimizadas

Para o *backtest* da estratégia, o modelo de variância mínima foi utilizado em sua variante irrestrita e em uma versão aonde a *duration* é restrita ao máximo de 252 dias. O cálculo da matriz de covariância e a estimação do modelo foram realizados utilizando-se janelas móveis de 252 dias. O primeiro resultado da estratégia ocorre em 15/06/2007 e o último retorno é contabilizado em 08/09/2014, gerando 1.806 observações.

O período analisado se mostrou fundamental para a avaliação do modelo, visto que acontecimentos com impactos distintos ocasionaram ciclos de alta, baixa e de manutenção da taxa básica de juros, nos permitindo avaliar o desempenho em diferentes contextos da política monetária.

Figura 3 – CDI (%a.a.)

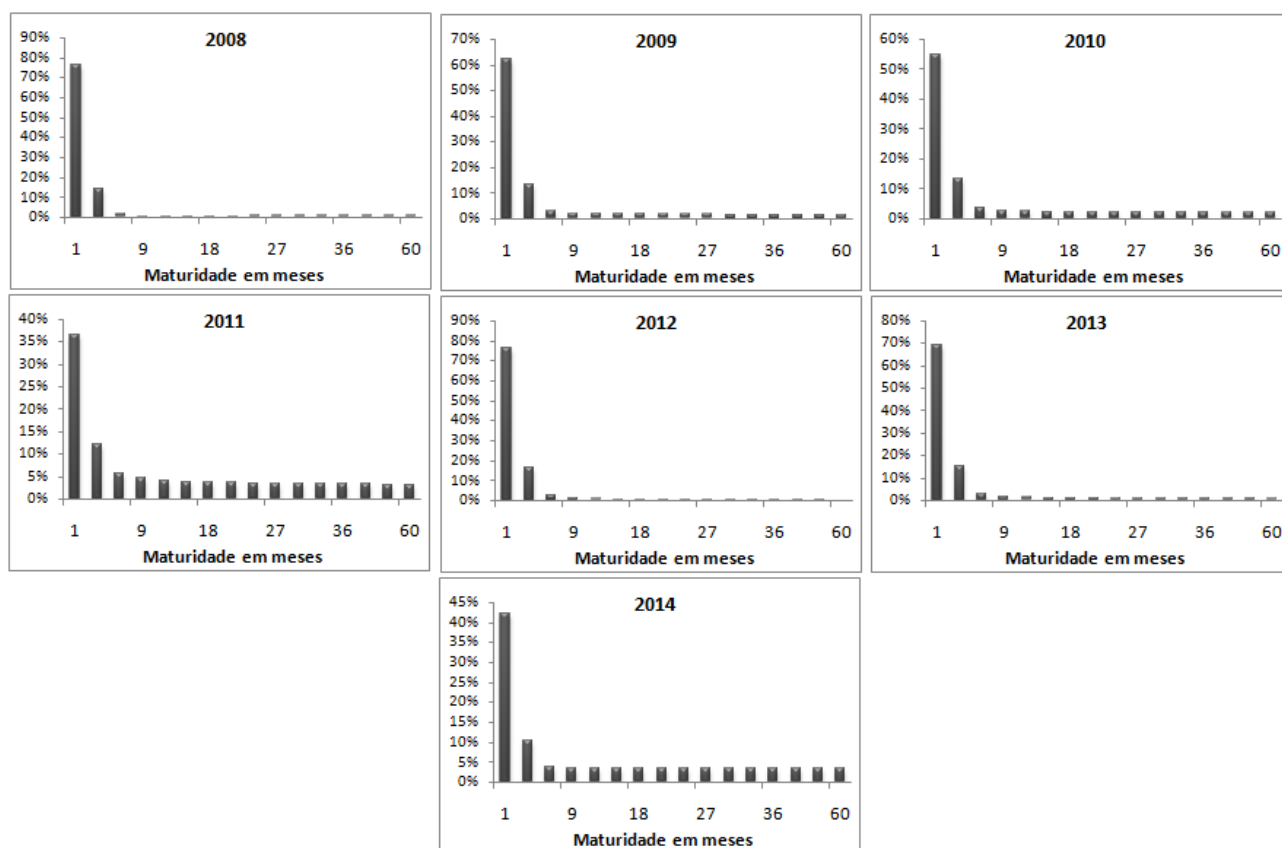


Fonte: elaboração própria (2016)

O ano de 2008 foi marcado por um forte crescimento interno, liderado pela demanda, pelo forte influxo de capitais e a apreciação dos termos de troca. No mesmo período, o país obteve elevação em seu *rating*, concedido pela agências de classificação de risco, e passou a figurar na lista dos países com "grau de investimento". Neste contexto, o Banco Central optou por uma política monetária restritiva e consequente aumento nas taxas de juros durante todo o ano, gerando desvalorização na taxa dos títulos pré-fixados e consequentemente perda de capital nestes portfólios. Observa-se que, em média, os fundos analisados gerenciaram o risco e conseguiram obter rendimento médio de 12,63% contra o CDI de 12,32%. O portfólio otimizado teve alocação média acima de 70% em títulos de 1 mês, minimizando o risco de taxa de juros e obtendo 12,98% em rendimento.

Como pode ser notado pelos gráficos abaixo, entre os anos de 2009 e 2011 o portfólio de mínima variância passou a reduzir sua alocação em títulos curtos, suavizando suas posições em todas as maturidades disponíveis e aumentando a *duration* da carteira. A estratégia de tomada de risco, neste caso, foi compensada com um prêmio sobre o CDI, visto que este obteve retorno de 34,48% no período enquanto o portfólio de mínima variância obteve 40,51%. Os gestores dos fundos pré-fixados analisados também se expuseram a títulos mais longos, como pode ser notado pelo rendimento de 39,08% no período. Abaixo apresentamos as alocações médias anuais da estratégia, em cada vértice:

Figura 4 – Alocações médias anuais

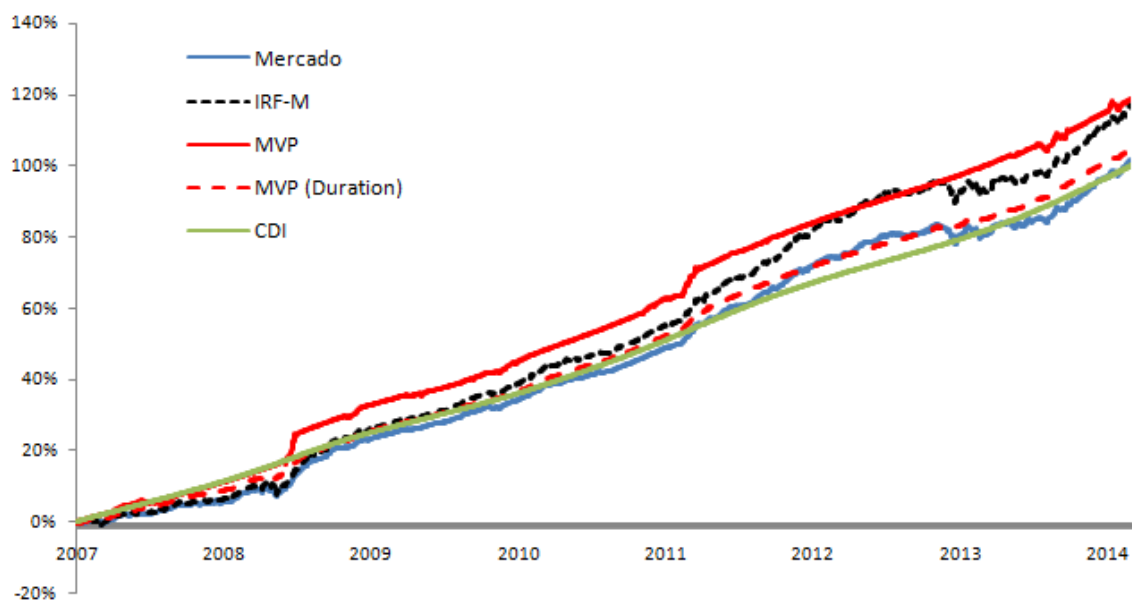


Fonte: elaboração própria (2016)

O ano de 2012 foi marcado pela tentativa do governo em estimular a economia. Instituições financeiras públicas, como o Banco do Brasil e Caixa Econômica Federal, iniciaram processo de redução de juros cobrados aos clientes. Ainda, o Banco Central continuou de forma acelerada, o ciclo de queda das taxas de juros iniciado no ano anterior, permitindo que títulos pré-fixados marcados a mercado obtivessem constantes elevações de preço e gerando retornos acima das taxas previamente contratadas. O modelo otimizado, no entanto, identifica excesso de volatilidade no período e reduz a *duration* do portfólio. Esse fato faz com que a estratégia tenha rendimento muito semelhante ao CDI no período (8,35% contra 8,37% do CDI), embora não acompanhe o rendimento dos fundos pré-fixados, que foi de 12,23% no ano de 2012.

No ano de 2013 a carteira de mínima variância manteve a mesma estratégia do ano anterior. É possível perceber que os fundos pré-fixados pouco conseguiram reverter de suas posições em 2012, incorrendo em grande perda de capital e rendimento médio de 2,17%. O portfólio otimizado manteve posições em vértices curtos e obteve rendimento mais próximo ao CDI (7,47% contra 8,02% do CDI).

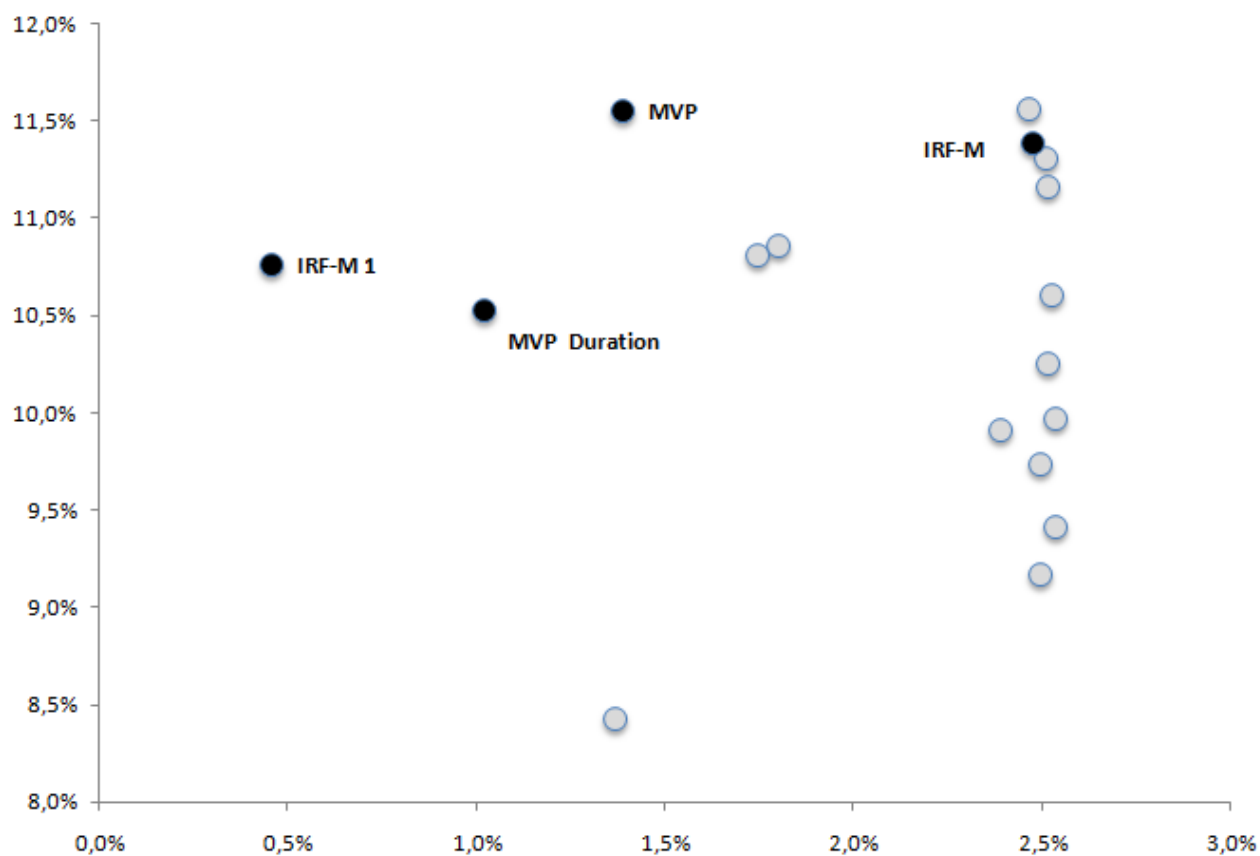
Figura 5 – Comparativo de Retornos Acumulados



Fonte: elaboração própria (2016)

A estratégia utilizada, que se baseou em rebalanceamentos mensais (21 dias), obteve um retorno de 118,94% no período analisado contra 116,7% do IRF-M e 101,94% de rendimento do CDI no mesmo período. Dada a baixa variância das carteiras otimizadas, a estratégia obteve Índice de Sharpe (IS) igual a 0,95, enquanto o *benchmark* obteve IS igual a 0,47. No mesmo período o IRF-M1 atingiu retorno de 108,04%, superando o modelo restrito na *duration* que obteve retorno de 104,99%. Abaixo apresentamos gráfico com dispersão de retornos. É possível notar que os fundos mantêm gestão de risco ativa de modo que o risco anualizado seja muito semelhantes entre si e em relação ao IRF-M, embora haja bastante dispersão entre os retornos. Ainda, pela tabela a seguir também pode ser verificada que a estratégia de mínima variância adotada apresentou perda máxima diária constantemente abaixo das apresentadas pelos fundos de mercado.

Figura 6 – Risco x Retorno (%a.a.)



Fonte: elaboração própria (2016)

Tabela 6 – Perda Máxima Diária

	Mercado	MVP	MVP(Duration)
2007	-0,89%	-0,47%	-0,32%
2008	-1,49%	-0,47%	-0,80%
2009	-0,33%	-0,41%	-0,16%
2010	-0,19%	-0,17%	-0,09%
2011	-0,27%	-0,30%	-0,11%
2012	-0,32%	-0,09%	-0,02%
2013	-0,58%	-0,07%	-0,20%
2014	-0,30%	-0,44%	-0,11%

Fonte: elaboração própria (2016)

6 CONSIDERAÇÕES FINAIS

Este trabalho objetivou o estudo sobre a estrutura a termo da curva de juros. Inicialmente foi apresentada a Hipótese das Expectativas, teoria padrão entre as que buscam explicar o comportamento da curva de juros. Sob a HE, o excesso de retorno entre títulos de longo e curto prazo deveria se comportar como um ruído branco, sendo portanto imprevisível. Diversos trabalhos que objetivam ao teste de tal teoria se propõem à análise de um modelo linear de regressão em que as taxas *forward* são preditoras de variações nas futuras taxas de curto prazo, expurgando os possíveis efeitos de heterocedasticidade com o uso de estimadores robustos. Não há consenso na literatura para a validade de tal teoria no mercado brasileiro. Contudo, alguns autores têm encontrado evidências para aceitação da HE em maturidades curtas.

Nos propusemos a testar a hipótese das expectativas através da análise da série de excessos de retorno observada durante o período de março de 2009 à setembro de 2014, modelando explicitamente a heterocedasticidade encontrada. Os resultados encontrados sugerem a rejeição da hipótese de que o prêmio de risco seja constante. Como apresentado, foi possível ajustar um processo GARCH-M, na linha proposta por Engle, Lilien e Robins (1987), no qual a média da série é afetada pela volatilidade. Entretanto, deve-se atentar para o fato de que o resultado pode ser sensível ao período analisado e à frequência dos dados utilizados.

Em um segundo momento foi utilizado o modelo de variância mínima para otimizar um portfólio de títulos de renda fixa. Embora tal modelagem seja de amplo conhecimento no mercado financeiro e na academia, não há ainda extensa literatura sobre o tema quando do uso em mercados de renda fixa. Utilizando as derivações propostas por Caldeira, Moura e Santos (2015), estimamos a estratégia e comparamos com um *benchmark*, IRF-M, e os fundos indexados a ele, apresentados no capítulo anterior. A estratégia foi capaz de apresentar retorno elevado e risco substancialmente inferior aos pares de mercado. Dentre estes, somente o fundo da gestora Western Asset conseguiu obter retorno acima do *benchmark* no período analisado.

E a estratégia adotada conseguiu, de forma mecânica, fazer movimentos de elevação e redução da *duration* da carteira muito semelhantes aos dos gestores analisados, para o período entre 2009 e 2011, tendo como base os retornos financeiros apresentados, apresentando ainda volatilidade e perda máxima diária substancialmente inferiores aos dos fundos de mercado. Por sua característica conservadora de minimizar a variância, o modelo optou por manter forte alocação em títulos de curto prazo nos anos de 2012 e 2013. Essa opção gerou resultados inferiores aos fundos pré-fixados no ano de 2012, tendo retorno compensado no ano seguinte e mantendo um retorno acumulado nestes dois

anos também acima do rendimento dos fundos pré-fixados, se mostrando como estratégia apropriada em um período mais longo. O modelo com restrição de *duration*, entretanto, não conseguiu rentabilidade acima do IRF-M1. Como esperado, a estratégia apresentou volatilidade reduzida no período e rendimento acumulado acima do CDI.

REFERÊNCIAS

- BAÑBURA, M.; GIANNONE, D.; REICHLIN, L. Large bayesian vector auto regressions. *Journal of Applied Econometrics*, Wiley Online Library, Ontario, v. 25, n. 1, p. 71–92, 2010.
- BRITO, R.; DUARTE, A. J.; GUILLÉN, O. T. O prêmio pela maturidade na estrutura a termo das taxas de juros brasileiras. *Banco Central do Brasil Working Paper Series*, Brasília, v. 72, 2003.
- CALDEIRA, J. F. Estimação da estrutura a termo da curva de juros no brasil através de modelos paramétricos e não paramétricos. *Análise Econômica*, Porto Alegre, v. 29, n. 55, 2011.
- CALDEIRA, J. F.; LAURINI, M. P.; PORTUGAL, M. S. Inferência bayesiana aplicada ao modelo dinâmico de nelson-siegel com volatilidade estocástica nos fatores. *Revista Brasileira de Econometria*, Rio de Janeiro, 2011.
- CALDEIRA, J. F.; MOURA, G. V.; PORTUGAL, M. S. Efficient yield curve estimation and forecasting in brazil. *Revista EconomiA*, [S.I.], 2010.
- CALDEIRA, J. F.; MOURA, G. V.; SANTOS, A. A. Bond portfolio optimization using dynamic factor models. [S.I.], 2015. Disponível em: <<http://ssrn.com/abstract=2079318>>. Acesso em: 20 dez. 2015.
- CARRIERO, A.; KAPETANIOS, G.; MARCELLINO, M. Forecasting government bond yields with large bayesian vector autoregressions. *Journal of Banking & Finance*, Elsevier, Londres, v. 36, n. 7, p. 2026–2047, 2012.
- CARTER, C. K.; KOHN, R. On gibbs sampling for state space models. *Biometrika*, London, v. 81, n. 3, p. 541–553, 1994.
- CHRISTENSEN, J. H.; DIEBOLD, F. X.; RUDEBUSCH, G. D. An arbitrage-free generalized nelson–siegel term structure model. *The Econometrics Journal*, Wiley Online Library, [S.I.], v. 12, n. 3, p. C33–C64, 2009.
- CHRISTENSEN, J. H.; DIEBOLD, F. X.; RUDEBUSCH, G. D. The affine arbitrage-free class of nelson–siegel term structure models. *Journal of Econometrics*, Elsevier, North-Holland, v. 164, n. 1, p. 4–20, 2011.
- DIEBOLD, F. X.; LI, C. Forecasting the term structure of government bond yields. *Journal of Econometrics*, Elsevier, North-Holland, v. 130, n. 2, p. 337–364, 2006.
- DIEBOLD, F. X.; RUDEBUSCH, G. D. *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach*. [S.I.]: Princeton University Press, 2013.
- DIEBOLD, F. X.; RUDEBUSCH, G. D.; ARUOBA, S. B. The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of econometrics*, Elsevier, North-Holland, v. 131, n. 1, p. 309–338, 2006.

- ENGLE, R. F.; LILIEN, D. M.; ROBINS, R. P. Estimating time varying risk premia in the term structure: The arch-m model. *Econometrica: Journal of the Econometric Society*, JSTOR, [S.I.], p. 391–407, 1987.
- FABOZZI, F. J. *Fixed income analysis*. [S.I.]: John Wiley & Sons, 2007. v. 6.
- FISHER, I. Appreciation and interest. *Publications of the American Economic Association*, JSTOR, [S.I.], v. 11, n. 4, p. 1–98, 1896.
- GOLINSKI, A.; ZAFFARONI, P. Long memory affine term structure models. [S.I.], 2015. Disponível em: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1787037>. Acesso em: 18 nov. 2015.
- GUILLEN, O. T. D. C.; TABAK, B. M. Characterising the brazilian term structure of interest rates. *International Journal of Monetary Economics and Finance*, Inderscience, [S.I.], v. 2, n. 2, p. 103–114, 2009.
- GÜRKAYNAK, R. S.; SACK, B.; WRIGHT, J. H. The u.s. treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, Elsevier, [S.I.], v. 54, n. 8, p. 2291–2304, 2007.
- JÚNIOR, C. de L. C. Estratégia de trading utilizando o modelo dinâmico de nelson-siegel. 2013.
- KORN, O.; KOZIOL, C. Bond portfolio optimization: A risk-return approach. *The Journal of Fixed Income*, Institutional Investor Journals, [S.I.], v. 15, n. 4, p. 48–60, 2006.
- LIMA, A. M. C.; ISSLER, J. V. A hipótese das expectativas na estrutura a termo de juros no brasil: uma aplicação de modelos de valor presente. *Revista brasileira de economia*, SciELO Brasil, [S.I.], v. 57, n. 4, p. 873–898, 2003.
- LITTERMAN, R. B. *Techniques of forecasting using vector autoregressions*. [S.I.], 1979. Disponível em: <<https://www.minneapolisfed.org/research/wp/wp115.pdf>>. Acesso em: 20 dez. 2015.
- LITTERMAN, R. B. *A Bayesian Procedure for Forecasting with Vector Autoregressions and Forecasting with Bayesian Vector Autoregressions—four Years of Experience*. [S.I.]: Federal Reserve Bank of Minneapolis, 1985.
- LITTERMAN, R. B. Forecasting with bayesian vector autoregressions - five years of experience. *Journal of Business & Economic Statistics*, Washington, v. 4, n. 1, p. 25–38, 1986.
- MARÇAL, E. F.; PEREIRA, P. V. A estrutura a termo das taxas de juros no brasil: testando a hipóteses de expectativas. *Pesquisa e Planejamento Econômico*, [S.I.], v. 37, n. 1, p. 113–148, 2007.
- MARKOWITZ, H. Portfolio selection. *The journal of finance*, Wiley Online Library, Nova York, v. 7, n. 1, p. 77–91, 1952.
- NELSON, C. R.; SIEGEL, A. F. Parsimonious modeling of yield curves. *Journal of business*, JSTOR, [S.I.], p. 473–489, 1987.

- NELSON, D. B. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, JSTOR, [S.I.], p. 347–370, 1991.
- NYHOLM, K. *Strategic Asset Allocation in Fixed Income Markets: A Matlab Based User's Guide*. [S.I.]: John Wiley & Sons, 2008.
- POOTER, M. D. Examining the nelson-siegel class of term structure models: In-sample fit versus out-of-sample forecasting performance. [S.I.], 2007. Disponível em: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=992748>. Acesso em: 01 nov. 2014.
- PUHLE, M. *Bond Portfolio Optimization*. [S.I.]: Springer Science & Business Media, 2008. v. 605.
- ROSSI, J. W. A estrutura a termo da taxa de juros: uma síntese. *Pesquisa e Planejamento Econômico*, Brasília, v. 26, n. 3, 1996.
- SVENSSON, L. E. *Estimating and interpreting forward interest rates: Sweden 1992-1994*. [S.I.], 1994.
- SVENSSON, L. E. Estimating forward interest rates with the extended nelson & siegel method. *Sveriges Riksbank Quarterly Review*, [S.I.], v. 3, n. 1, p. 13–26, 1995.
- TABAK, B. Testing the expectations hypothesis in the brazilian term structure of interest rates: a cointegration analysis. *Applied Economics*, Taylor & Francis, [S.I.], v. 41, n. 21, p. 2681–2689, 2009.
- WALSH, C. E. *Monetary Theory and Policy*. [S.I.]: MIT press, 2010.

APÊNDICE - FUNDOS DE INVESTIMENTOS COMPARADOS

Tabela 7 – Fundos de Investimento IRF-M

Nome	CNPJ	Classificação Anbima	Gestora	Benchmark	Data de início
BB Prev RF	7111384000169	RF Indexados	BB DTVM	IRF-M	08/12/2004
BB RF Mkt. Plus	4128704000103	RF Indexados	BB DTVM	IRF-M	21/03/2005
BB RF Pre LP	8080680000102	RF Indexados	BB DTVM	IRF-M	29/06/2006
BB Top RF	5949051000188	RF Indexados	BB DTVM	IRF-M	24/03/2005
HSBC FI RF	7364756000169	RF Duração Livre	HSBC	IRF-M	31/08/2005
Itau LT Capital	2935788000153	RF Indexados	Itau Unibanco	IRF-M	14/03/2005
Itau LT Maxi	2935821000145	RF Indexados	Itau Unibanco	IRF-M	14/03/2005
Itau RF Pre	7104761000132	RF Indexados	Itau Unibanco	IRF-M	23/06/2005
Itau RF Pre LP FI	7096486000152	RF Indexados	Itau Unibanco	IRF-M	06/04/2005
Itau Year Fix	5381727000180	RF Indexados	Itau Unibanco	IRF-M 1	14/03/2005
Unibanco RF LP FI	5756479000104	RF Indexados	Itau Unibanco	IRF-M	11/03/2005
Uniclass RF LP FIC	5756488000103	RF Indexados	Itau Unibanco	IRF-M	11/03/2005
Western Asset Prev	7894727000109	RF Indexados	Western Asset	IRF-M	25/05/2007

Fonte: Economática

Tabela 8 – Risco x Retorno (Jul/07 - Set/14)

	BB Prev	BB RF Mkt.	BB RF Pre	BB Top RF	HSBC FI RF	Itau LT Capital	Itau LT Maxi	Itau RF Pre	Itau RF Pre LP	Itau Fix	Unibanco RF LP	Uniclass RF LP	Western Asset Prev	IRF-M	MVP	IRF-M 1	MVP	MVP Duration
Risco (%a.a)	1,74%	2,52%	2,51%	2,51%	1,80%	2,53%	2,53%	2,39%	2,51%	1,37%	2,49%	2,49%	2,46%	2,47%	1,38%	2,47%	1,38%	1,02%
Retorno (% a.a)	10,81%	10,62%	10,26%	11,16%	10,86%	9,42%	9,98%	9,92%	11,31%	8,43%	9,74%	9,18%	11,57%	11,40%	11,55%	10,76%	10,53%	10,53%
Risco (%a.a)	0,10%	4,24%	4,21%	4,21%	2,32%	4,04%	4,04%	3,89%	3,93%	3,05%	3,98%	3,98%	3,67%	3,80%	2,22%	3,80%	2,22%	1,84%
Retorno 2008	11,85%	13,80%	13,34%	14,29%	13,87%	10,81%	11,38%	12,35%	13,93%	10,59%	12,09%	11,50%	14,39%	13,88%	18,98%	13,93%	18,98%	13,53%
Risco (%a.a)	0,11%	1,92%	1,94%	1,94%	1,10%	2,02%	2,02%	1,79%	2,02%	1,40%	1,77%	1,77%	2,01%	2,01%	1,14%	2,01%	1,14%	0,81%
Retorno 2009	9,47%	10,99%	11,12%	12,03%	11,55%	10,58%	11,14%	10,77%	12,26%	9,81%	11,54%	10,99%	12,13%	12,47%	10,12%	11,56%	10,12%	10,48%
Risco (%a.a)	1,23%	1,25%	1,25%	1,25%	0,66%	1,35%	1,35%	1,22%	1,35%	0,67%	0,61%	0,57%	1,32%	1,35%	0,87%	1,35%	0,87%	0,63%
Retorno 2010	11,10%	10,76%	10,40%	11,30%	10,45%	10,08%	10,64%	10,41%	11,76%	8,75%	9,64%	9,08%	11,66%	11,87%	11,16%	10,37%	11,16%	10,61%
Risco (%a.a)	1,88%	1,82%	1,82%	1,82%	1,00%	1,86%	1,86%	1,60%	1,86%	0,45%	1,61%	1,61%	1,83%	1,86%	1,53%	1,86%	1,53%	0,97%
Retorno 2011	14,20%	13,66%	13,30%	14,21%	13,74%	12,65%	13,23%	12,74%	14,38%	9,95%	13,61%	13,03%	14,49%	14,45%	14,79%	12,61%	14,79%	13,37%
Risco (%a.a)	1,98%	1,94%	1,93%	1,93%	1,23%	1,95%	1,95%	1,58%	1,95%	0,35%	1,30%	1,31%	2,04%	1,95%	0,33%	1,95%	0,33%	0,15%
Retorno 2012	13,94%	13,26%	12,90%	13,80%	11,40%	12,10%	12,66%	11,87%	13,79%	7,18%	11,15%	10,61%	14,33%	13,97%	8,35%	9,44%	8,35%	8,56%
Risco (%a.a)	3,10%	3,13%	3,10%	3,10%	2,88%	3,09%	3,09%	3,03%	3,09%	0,52%	3,09%	3,09%	3,05%	3,06%	0,28%	3,06%	0,28%	0,92%
Retorno 2013	2,12%	1,83%	1,32%	2,13%	3,29%	0,98%	1,48%	1,69%	2,50%	5,12%	1,60%	1,09%	3,12%	2,56%	7,47%	2,56%	7,47%	6,37%
Risco (%a.a)	2,16%	2,17%	2,16%	2,16%	2,48%	2,19%	2,19%	2,19%	2,20%	0,30%	2,20%	2,19%	2,27%	2,21%	1,99%	2,21%	1,99%	0,69%
Retorno 2014	9,38%	8,96%	8,72%	9,32%	9,41%	8,45%	8,82%	8,83%	9,57%	5,43%	8,80%	8,42%	10,05%	9,53%	6,56%	9,53%	6,56%	7,70%