

Escola de Pós-Graduação em Economia - EPGE

Fundação Getulio Vargas

Modelos Estatísticos de Segmentação da Estrutura a Termo: Testes Empíricos de Ajustes e Previsões

Dissertação submetida à Escola de Pós-Graduação em Economia

da Fundação Getulio Vargas como requisito de obtenção do

título de Mestre em Economia

Aluno: Axel André Simonsen

Orientador: Caio Almeida

Rio de Janeiro

Janeiro 2008

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Escola de Pós-Graduação em Economia - EPGE

Fundação Getulio Vargas

Modelos Estatísticos de Segmentação da Estrutura a Termo: Testes Empíricos de Ajustes e Previsões

Palavras Chave: Estrutura a Termo; Previsão; Finanças Empíricas;

Dissertação submetida à Escola de Pós-Graduação em Economia

da Fundação Getulio Vargas como requisito de obtenção do

título de Mestre em Economia

Aluno: Axel André Simonsen

Banca Examinadora:

Caio Ibsen Rodrigues de Almeida (EPGE/FGV)

Marco Antonio Bonomo (EPGE/FGV)

Marcelo Cunha Medeiros (PUC/RJ)

Rio de Janeiro

Janeiro 2008

Modelos Estatísticos de Segmentação da Estrutura a Termo: Testes Empíricos de Ajuste e Previsões.

Axel André Simonsen
EPGE/FGV
Mestrando

Caio Almeida
EPGE/FGV
Orientador

24 de abril de 2008

Resumo

Neste trabalho é proposta uma classe de modelos paramétricos para estrutura a termo de taxa de juros (ETTJ) em que diferentes segmentos possam ter características próprias, porém não independentes, o que é condizente com a teoria de preferências por *Habitat*. O modelo baseia-se em Bowsher & Meeks (2006) onde a curva é determinada por um spline cúbico nas yields latentes, mas difere no sentido de permitir diferentes funções de classe C^2 entre os segmentos, ao invés de polinômios cúbicos. Em particular usa-se a especificação de Nelson & Siegel, o que permite recuperar o modelo de Diebold & Li (2006) quando não há diferenciação entre os segmentos da curva. O modelo é testado na previsão da ETTJ americana, para diferentes maturidades da curva e horizontes de previsão, e os resultados fora da amostra são comparados aos modelos de referência nesta literatura. Adicionalmente é proposto um método para avaliar a robustez da capacidade preditiva do modelos. Ao considerar a métrica de erros quadráticos médios, os resultados são superiores à previsão dos modelos Random Walk e Diebold & Li, na maior parte das maturidades, para horizontes de 3, 6, 9 e 12 meses.

Sumário

1	Introdução	3
2	Revisao da Literatura	5
2.1	Diebold & Li (2006)	5
2.2	Bowsher & Meeks (2006)	5
2.3	Cochrane & Piazzesi (2005)	6
2.4	Kargin & Onatski(2007)	7
3	Modelo	7
3.1	Especificação do Modelo	8
3.2	Estimação no <i>Cross Section</i>	12
3.3	Previsão da curva	13
3.4	Método de Avaliação dos Modelos	13
3.4.1	1ª métrica - Erros Quadráticos Médios (EQM) Relativos	14
3.4.2	2ª métrica - Método de Robustez para Amostras Distintas no tamanho ou no período	14
4	Avaliação Empírica	15
4.1	Dados	15
4.2	<i>Factor Loadings</i>	16
4.3	Estimação	18
4.4	Previsão	21
4.4.1	Modelos de Previsão dos Parâmetros	22
4.5	Resultados	23
4.5.1	Escolha do parâmetro p nos modelos com descontinuidade nos <i>loadings</i>	23
4.5.2	Previsão Fora da Amostra	24
4.5.3	Comparação Direta entre Modelos	28
5	Conclusões	30
6	Apêndice	31

1 Introdução

A importância da modelagem da ETTJ é clara por ser um pilar na determinação de preço dos ativos, decisões de investimento ou na criação de instrumentos financeiros. Existem diversas classes de modelos, que podem ser mais adequadas de acordo com a necessidade dos agentes. Instituições financeiras, no papel de intermediadoras do mercado (como exemplo tesouraria de bancos comerciais), tomam posições (brutas) diárias que são muitas vezes bem superiores ao próprio patrimônio líquido. Surge assim a necessidade de hedge diário onde os modelos que ajustam os preços observados perfeitamente, são mais demandados. Para instituições responsáveis por gestão de recursos, sejam de terceiros ou mesmo próprios, as estratégias são geralmente de prazo mais longo, fazendo com que modelos com consistência intertemporal sejam mais adequados. Ainda sob ótica de gestão, modelos que consideram variáveis macroeconômicas possuem o apelo de permitir que os gestores incorporem suas expectativas sobre as futuras variáveis macro na determinação da curva futura.

Embora muito tenha sido feito no desenvolvimento destes modelos, a literatura diretamente voltada à previsão não é tão extensa. Alguns trabalhos importantes nesta área foram feitos por Fama & Bliss (1987), Cochrane & Piazzesi (2005), Diebold & Li (2006), Bowsher & Meeks (2006) e Kargin & Onatski (2007). Os dois primeiros olham para as taxas *forward* para prever comportamento das taxas spot (Fama & Bliss) e excesso de retorno dos títulos de renda fixa (Cochrane & Piazzesi). A previsão das taxas futuras diretamente é analisada pelos demais. Diebold & Li (2006), basearam-se na estrutura de componentes exponenciais proposta por Nelson & Siegel (1987), onde a dinâmica da curva é dirigida por um vetor de parâmetros tridimensional. Suas principais vantagens são a parcimônia do modelo, a interpretação destes fatores e os bons resultados fora da amostra, especialmente para horizonte de 12 meses, que supera uma ampla classe de modelos alternativos. Bowsher & Meeks propuseram um modelo *Functional Plus Noise* (FSN) em que a curva de juros é dirigida por um polinômio cúbico em taxas latentes, onde os resultados mostraram-se superiores aos apresentados por Diebold & Li para previsões um mês à frente. Kargin & Onatski utilizaram uma técnica de autoregressão funcional, em que decompõem a ETTJ em fatores preditivos e obtiveram uma melhora nos resultados em relação à modelos que extraem os fatores principais. Uma outra abordagem na literatura de previsão é a incorporação de variáveis macroeconômicas no modelo de ETTJ, com destaque ao trabalho de Ang & Piazzesi (2003) que foi um dos precursores desta abordagem. Diebold, Rudebusch & Aruoba (2005) incorporam aos fatores latentes gerados pela especificação de Nelson & Siegel (1987) indicadores de atividade real, inflação e instrumento de política monetária encontrando evidências do efeito da inclusão destas variáveis na explicação de movimentos futuros da ETTJ. Monch (2006) usa a taxa curta de juros e os componentes principais de um conjunto extenso de variáveis macroeconômicas como fatores que explicam a ETTJ, conseguindo bons resultados fora da amostra em especial para maturidades curtas.

Fama & Bliss sugerem que modelos com melhores características na *cross section* possuem piores implicações preditivas. Na mesma linha, Cochrane & Piazzesi mostraram que, embora 3 fatores representem os principais movimentos da ETTJ, um quarto fator com pouca influência nestes movimentos é relevante ao se considerar exercícios de previsão. O foco deste trabalho será desenvolver um modelo estatístico que possua boa propriedade no ajuste em *cross section* e ao mesmo tempo consiga gerar resultados satisfatórios na previsão de taxa para horizontes distintos (de 1 a 12 meses à frente). O modelo proposto possui como ingrediente principal a segmentação da ETTJ, de tal forma que partições da curva possam ter dinâmicas próprias, mas não independentes. Esta abordagem é motivada pela hipótese de segmentação de mer-

cado de Culbertson(1957), em que investidores heterogêneos possuem forte preferências por determinadas maturidades. Os exemplos clássicos para esta teoria são a demanda dos fundos de pensão por títulos com maturidades longas, enquanto que empresas de bens de consumo se importam apenas com maturidades curtas, dada a características de seu passivos. No trabalho de Vayanos & Vila (2007), uma evidência neste sentido é observada no período em que houve uma reforma regulatória nos planos de previdência da UK, que gerou implicações diretas sobre as características da ETTJ apenas nas maturidades longas.

A principal crítica desta teoria, apontada por Cox,Ingersoll & Ross(1985) é a implicação que títulos de maturidades semelhantes (e com mesmas características) possam ser negociados a preços bem diferentes. Uma teoria menos extrema é baseada na teoria de Preferência por Habitat. Vayanos & Vila (2007) formalizam um modelo em que a economia não possui apenas investidores heterogêneos (no sentido de demandarem taxas em maturidades diferentes) mas também há arbitradores avessos ao risco. Assim, por um lado existe uma a força atuando em diferentes maturidades caracterizada por demandas heterogêneas, mas ao mesmo tempo as taxas não são completamente independentes pois existem arbitradores que se aproveitam das distorções, garantindo que títulos com maturidades semelhantes sejam negociado à preços parecidos.

O segundo ingrediente do modelo será a forma de incorporar a segmentação na construção da curva. O modelo de Bowsheer & Meeks(2006) assume que a ETTJ é caracterizada por uma função latente construída através da metodologia de spline cúbico sobre as yields latentes. Repare que, ao usar esta formulação, cada segmento da curva é determinado por polinômios cúbicos com coeficientes distintos, conectados por restrições de continuidade e suavidade nos nós dos segmentos. Estas restrições são coerentes com a teoria de Preferências por Habitat no sentido de conectar os diferentes segmentos de forma suave, agindo na mesma direção dos agentes arbitradores que impedem que títulos com maturidades semelhantes difiram muito nos preços. Neste trabalho, a construção da curva será semelhante, com a diferença que permitirá diferentes funções de classe C^2 em cada segmento. Por fim, o terceiro ingrediente será escolher dentro da classe de funções C^2 , uma função que se adeque bem para a modelagem da ETTJ tanto no ajuste *cross section* quanto para previsão. Foi escolhida a forma paramétrica de Nelson & Siegel que possui a vantagem de recuperar o modelo de Diebold & Li no caso em que os segmentos forem homogêneos (mesmos coeficientes).

Os modelos que juntam estes três ingredientes possuirão 3 *loadings* e serão denominados modelos $SL3$ (*segmented loadings*), onde a segmentação será caracterizada ao usar diferentes funções entre os segmentos. Embora a especificação inicial possua $3k$ coeficientes a serem estimados, a imposição das restrições reduzirá a dimensionalidade à apenas 3 fatores, que caracterizarão a dinâmica da curva independente do número de segmentos k . Um segundo modelo proposto é um subproduto do $SL3$, em que as restrições de continuidade e suavidade no ajuste do *cross section* são substituídas por restrições de igualdade dos parâmetros o que gera uma segunda forma de segmentação, este modelo será denotado por $\overline{SL3}$.

Os modelos são testados para previsão da ETTJ americana para diferentes horizontes de previsão. A avaliação da qualidade das previsões usará a métrica adotada pelos trabalhos feitos nesta área (Diebold & Li, Bowsheer & Meeks e outros) que consiste em analisar o Erro Quadrático Médio das previsões fora da amostra. Adicionalmente, é proposto um método de avaliação de robustez destes modelos. Este método considera duas dimensões diferentes, a primeira dimensão consiste em variar o número de observações que serão usados para o exercício fora da amostra e a segunda dimensão avalia o comportamento dos resultados para diferentes janelas amostrais de mesmo tamanho. Em particular, o modelo $\overline{SL3}$ foi o que apresentou melhores resultados

para o período fora da amostra de 1994:01 a 2000:12, superando o modelo de Diebold & Li (2006) e Random Walk na maior parte das maturidades, para os horizontes de 3,6 9 e 12 meses, enquanto o *SL3* – 3 foi o que se mostrou mais robusto sob o método proposto.

O trabalho será organizado da seguinte maneira: a seção 2 descreve o modelo, as especificações dos modelos alternativos que serão comparados, o método de previsão e a métrica que será usada para avaliação. Na seção 3 será feita a parte empírica do trabalho e a seção 4 a conclusão.

2 Revisao da Literatura

2.1 Diebold & Li (2006)

Especificam um modelo para a ETTJ que é uma variação do modelo de Nelson & Siegel¹ dada por

$$y_t(\tau) = \beta_{1t} + \beta_{2t} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} \right) + \beta_{3t} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \right) \quad (1)$$

$$\hat{\beta}_{i,t+h/t} = \hat{c}_i + \hat{\gamma}_i \hat{\beta}_{it}, \quad i = 1, 2, 3 \quad (2)$$

onde τ representa a maturidade e t o período. Mostram que os coeficientes β_{it} podem ser interpretados como os fatores correspondentes ao nível (β_{1t}), inclinação (β_{2t}) e curvatura (β_{3t}) da ETTJ. O modelo é estimado em 2 estágios, primeiro os coeficientes são estimados na dimensão do *cross section* (equação (1)), obtendo assim um série temporal. Em seguida, estimam-se os coeficientes da dinâmica (equação (2)) assumindo um processo *AR*(1) para cada fator individualmente. Para o exercício empírico, assume-se que a série $\{\lambda_t\}$ é constante com valor dado exogenamente, o que torna o problema de estimação dos demais parâmetros linear, o que simplifica bastante o procedimento computacional.

Testam o modelo para a previsão da ETTJ americana para horizontes de 1,6 e 12 meses. Os principais resultados apresentados foram:

- Simplicidade da metodologia;
- boa aderência aos fatos estilizados como persistência no nível das taxas, flexibilidade de formatos da curva, maiores volatilidades em maturidades menores e baixa persistência nos *spreads*;
- excelentes resultados de previsão relativos à modelos concorrentes nesta literatura, em especial para horizonte de 12 meses.

2.2 Bowsher & Meeks (2006)

Definem a classe de modelos *Functional Plus Noise* (FSN) em que a ETTJ é dada por um funcional latente acrescido dos termos de erro. O funcional é unicamente determinado por um spline cúbico natural nas taxas latentes avaliadas nos nós dos splines. O mapeamento destas taxas latentes para as taxas latentes em outras maturidades é realizado através de uma transformação

¹Outras extensões da classe de modelo de Nelson & Siegel são estudadas por Pooter (2007) onde são comparados resultados de ajuste da amostra e previsão para diversas variações desta forma paramétrica. Também analisam diferentes métodos de previsão. Encontram resultados superiores à modelos concorrentes usando especificação de 4 fatores e estimação em 1 estágio usando Filtro de Kalman.

linear cuja matriz é determinística. O modelo então pode ser escrito na representação estado-espço linear, permitindo o uso do Filtro de Kalman para computar a Quasi-Verossimilhança Gaussiana e, assim, a estimação pode ser feita em um estágio apenas (o que difere de Diebold & Li (2006))². A curva observada é portanto uma transformação linear das taxas latentes acrescida do termo de erro (equação de observação) e as taxas latentes possuem a dinâmica dada pela equação de estado. Adicionalmente, como as taxas latentes são cointegradas, a equação de estado é escrita na forma de modelo de correção de erros onde os *spreads* das taxas latentes aparecem como regressores. O modelo denominado FSN-ECM para a série temporal de taxas observadas N-dimensional $\{y_t(\tau)\}_{t=1}^T$ é dado por

$$y_t(\tau) = S_\gamma(\tau) + \varepsilon_t = W(k, \tau)\gamma_t + \varepsilon_t \quad (3)$$

$$\Delta\gamma_t = \alpha(\beta'\gamma_t - \mu_s) + \Psi\Delta\gamma_t + \nu_t \quad (4)$$

onde $S_\gamma(\tau)$ é o funcional latente e $W(k, \tau)$ é a matriz determinística que depende apenas das maturidades τ e dos nós k . Os detalhes da construção desta matriz estão em Porier (1973, pp.517-518). O vetor de cointegração β é tal que $\beta'\gamma_t$ representa os *spreads* das taxas latentes. Variações desta especificação geral foram testadas, a saber, variações no número de nós (5 ou 6) e restrições sobre a matriz os coeficientes e sobre a estrutura dos erros.

Testam o modelo para a previsão da ETTJ americana para o horizontes de 1 mês. Os principais resultados apresentados foram:

- O FSN possibilita redução de dimensionalidade na estimação de tal forma que consegue estimar um número elevado de maturidades da curva sem aumentar o número de parâmetros a serem estimados (a dimensão destes dependerão apenas do número de nós fixados);
- excelentes resultados de previsão relativos à modelos concorrentes nesta literatura, entre eles o de Diebold & Li, no qual é superior significativamente para o horizonte de 1 mês (mas não comentam sobre previsões de prazos mais longos);
- concluem que os *spreads* das taxas de fato contribuem na previsão, corroborando algumas evidências neste sentido encontradas por outros autores.

2.3 Cochrane & Piazzesi (2005)

Estudam a variação no excesso de retorno esperado de títulos de renda fixa (*zero-coupon*) do governo americano. Encontram um único fator, dado por combinação linear de taxas *forward*, que consegue prever o retorno em excesso esperado para títulos com maturidades entre 1 a 5 anos com R^2 de até 44% (exercícios desta natureza feitos na literatura encontraram R^2 na ordem de 20%, ver Fama & Bliss (1987)) Este fator de previsão de retornos é não correlacionado com os movimentos associados ao nível, inclinação e curvatura descrito pela maior parte dos modelos de ETTJ. Fazem diferentes testes estatísticos como uma medida de avaliar a robustez nos resultados encontrados nos quais destacam-se:

- Rejeição da hipótese das expectativas (não previsibilidade) sobre diversos testes distintos, todos avaliados à 1% de significância;
- ao passo que o quarto fator explica apenas 2% da variação das taxas, consegue explicar 24% do retorno esperado dos títulos.

²Diebold, Rudebusch & Aruoba (2006) utilizam Filtro de Kalman.

2.4 Kargin & Onatski(2007)

Estudam a previsão de processos funcionais autoregressivos onde desenvolvem uma nova técnica chamada *Predictive Factor Decomposition* (PFD) para estimar o operador autoregressivo. Esta técnica consiste em encontrar uma aproximação de posto reduzido para este operador, com a métrica de minimizar o quadrado da norma dos erros de previsão. O método é ilustrado com um exemplo da estrutura a termo das taxas futuras de Eurodollar. O modelo é dado por:

$$f_{t+h} = \rho[f_t] + \varepsilon_{t+h}$$

Onde f_t um elemento do espaço de Hilbert H , ρ um operador linear limitado em H e ε_t é um H - ruído branco forte. A estimação consiste em encontrar o valor futuro de \hat{f}_{t+1} que resolve:

$$\min_{\hat{f}_{t+1}} E \left\| f_{t+1} - \hat{f}_{t+1}(f_1, f_2, \dots, f_n) \right\|^2$$

Uma aproximação razoável da solução é $\hat{\rho}[f_t]$, onde $\hat{\rho}$ é um estimador consistente de ρ . A principal dificuldade em estimar ρ é sua infinita-dimensionalidade. Como apenas um número finito de dados é observado, é necessária uma técnica de redução de dimensionalidade. A técnica de PCA usa os primeiros autovetores do operador de covariância amostral como base para a redução de dimensionalidade. Os autores argumentam que este método não é o mais adequado para a previsão pois o maior autovetor associado ao operador de covariância de f_t não será necessariamente o autovetor que possuirá melhor capacidade preditiva para f_{t+h} . Como exemplo citam a literatura que relaciona aproximadamente 95% dos movimentos da ETTJ aos três principais componentes ao passo que recentes pesquisas (Cochrane & Piazzesi (2005)) destacam que fator adicional à estes três com pouca relevância na explicação dos movimentos possui alta relevância na previsão. As principais contribuições obtidas foram:

- Desenvolvimento da nova técnica para estimar o operador autoregressivo;
- na seção empírica, para uma amostra relativa à um período de crescimento normal, o PFD supera o método de componentes principais e consegue resultados semelhantes à modelos desta literatura,

3 Modelo

Conforme exposto na seção anterior, o modelo de Bowsheer & Meeks (2006) assume que a curva pode ser descrita por um funcional latente acrescido dos termos de erro. O arcabouço do modelo *SL3* basea-se nesta estrutura flexibilizando este funcional para ser combinação linear de diferentes funções de classe C^2 entre os segmentos. Um ponto importante é que Bowsheer & Meeks usam o polinômio cúbico justificando que é a forma funcional que minimiza o *penalty* sobre a suavidade (dada pela integral do quadrado da segunda derivada da função) entre todas as funções de classe C^2 . Repare que a generalização não será no sentido de permitir outras funções de classe C^2 e sim permitir que segmentos distintos possuam funções distintas. Esta distinção é fundamental para gerar o efeito de segmentação nos parâmetros. A especificação geral do modelo *SL3* é dada por

$$Y_t(\tau) = S_\gamma(\tau) + \varepsilon_t = Z(k, \tau)\theta_t + \varepsilon_t$$

$$\theta_t = a + b\theta_{t-1} + \nu_t$$

A construção da matriz Z , as características dos vetores de fatores latentes θ_t e modelo de estimação são os objetos que serão tratados nesta seção. O modelo será estimado em dois estágios, o primeiro estágio envolve a estimação dos fatores no *cross section* e o segundo a estimação dos parâmetros autoregressivos que determinam a dinâmica destes fatores.

3.1 Especificação do Modelo

Seja x a maturidade e $[0, T_M]$ o intervalo de maturidades para qualquer período, onde T_M deve ser superior à qualquer maturidade da amostra. O primeiro passo consiste em especificar o conjunto de nós do modelo,

$$\Delta(k) = \{0 < x_1 < \dots < x_{k-1}\}, \quad x_{k-1} < x_k = T_M$$

onde $k-1$ nós particionam o intervalo em k segmentos. Este conjunto será o mesmo para todos os períodos e modelos que usam este arcabouço. A estrutura a termo para o período t (cujo índice será ocultado nesta seção), será dirigido por um funcional latente $f(\cdot)$:

$$f(x) = \sum_{i=1}^k f_i(x) \chi_{A_i} \quad (5)$$

onde χ_{A_i} representa a função indicadora do conjunto A_i . dado por

$$A_i \begin{cases} [x_0, x_1], & i = 1 \\ (x_{i-1}, x_i], & i = 2, \dots, k \end{cases} \quad (6)$$

e as funções $f_i(\cdot)$ são da seguinte forma:

$$\begin{aligned} f_i(x) &= a_i + b_i g_i(x) + c_i h_i(x), \quad a_i, b_i, c_i \in \mathbb{R}, \quad i = 1, \dots, k \\ \text{s.a. } g_i, h_i &\in C^2(\bar{A}_i) + \text{restrições, } \bar{A}_i \text{ é o fecho de } A_i \end{aligned} \quad (7)$$

O segundo passo será a imposição de restrições que garantam a continuidade e suavidade deste funcional. Como no interior dos segmentos as funções são de classe C^2 , basta impor restrições sobre as junções dos segmentos (conjunto de nós $\Delta(k)$), que serão dadas pelas equações (8), (9) e (10):

$$f_i(x_i) = f_{i+1}(x_i), \quad i = 1, \dots, k-1 \quad (8)$$

$$f'_i(x_i) = f'_{i+1}(x_i), \quad i = 1, \dots, k-1 \quad (9)$$

$$f''_i(x_i) = f''_{i+1}(x_i), \quad i = 1, \dots, k-1 \quad (10)$$

A flexibilização gerada pela escolha de funções distintas permite que cada segmento tenha uma dinâmica própria, embora não independente, uma vez que as restrições conectam estas dinâmicas, o que gera uma analogia à teoria de preferência por *habitat*. O sistema de restrições pode ser reescrito matricialmente por

$$R\beta = 0 \quad (11)$$

onde

$$\beta_{(3k \times 1)} = [a_1 \quad b_1 \quad c_1 \quad \dots \quad a_k \quad b_k \quad c_k]' \quad (12)$$

e, definindo os seguintes vetores linha para simplificar notação

$$X_1^i(x) = [1 \quad g_i(x) \quad h_i(x)] \quad (13)$$

$$X_2^i(x) = [0 \quad g'_i(x) \quad h'_i(x)] \quad (14)$$

$$X_3^i(x) = [0 \quad g''_i(x) \quad h''_i(x)] \quad (15)$$

temos que R é dado por:

$$R_{[3(k-1)] \times (3k)} = \begin{bmatrix} X_1^1(x_1) & -X_1^2(x_1) & 0_{3 \times 1} & \dots & 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} \\ 0_{3 \times 1} & X_1^2(x_2) & -X_1^3(x_2) & \dots & 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} & \dots & 0_{3 \times 1} & X_1^{k-1}(x_{k-1}) & -X_1^k(x_{k-1}) \\ X_2^1(x_1) & -X_2^2(x_1) & 0_{3 \times 1} & \dots & 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} & \dots & 0_{3 \times 1} & X_2^{k-1}(x_{k-1}) & -X_2^k(x_{k-1}) \\ X_3^1(x_1) & -X_3^2(x_1) & 0_{3 \times 1} & \dots & 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} & \dots & 0_{3 \times 1} & X_3^{k-1}(x_{k-1}) & -X_3^k(x_{k-1}) \end{bmatrix} \quad (16)$$

Note que embora o funcional latente esteja definido para todo o intervalo $[0, T_M]$, a estimação dos parâmetros que o definem somente terá um conjunto finito de taxas observáveis que serão dadas pelo vetores Y e τ (respectivas maturidades)

$$Y = [y_1 \quad \dots \quad y_M] \quad (17)$$

$$\tau = [t_1 \quad \dots \quad t_M] \quad (18)$$

A matriz que faz o mapeamento das maturidades observadas nos respectivos segmentos será denotada por $W_{(\tau)}$, uma forma de interpretar esta matriz é que são os *loadings factors* distintos para cada segmento caracterizados por $X_1^i(x), x \in A_i$. A construção desta matriz é detalhada a seguir:

Dado $(y_j, t_j) \in Y \times \tau$, então $\exists i$ tal que $t_j \in A_i$. A j -ésima linha³ de $W_{(\tau)}$ é definida por

$$W_{(\tau)}(j, \cdot) = [0_{3 \times 1} \quad \dots \quad 0_{3 \times 1} \quad X_1^i(\tau_j) \quad 0_{3 \times 1} \quad \dots \quad 0_{3 \times 1}] \quad (19)$$

onde é denotada por matriz dos *loadings* originais (para distinguir da matriz de *loadings* com restrições embutidas, que será visto mais a frente). Como exemplo, seja o conjunto de taxas $Y = \{y_1, y_2, y_3\}$ com as respectivas maturidades $\tau = \{t_1, t_2, t_3\}$ em que

$$\begin{aligned} 0 &< t_1 < x_1 \\ x_{i-1} &< t_2 < x_i \\ x_{k-1} &< t_3 < x_k \end{aligned}$$

³Supondo $M - 1 > i_{(t_j)} > 2$

então

$$W_{(\tau)} = \begin{bmatrix} X_1^1(t_1) & 0_{3 \times 1} & \cdots & 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} & \cdots & 0_{3 \times 1} \\ 0_{3 \times 1} & 0_{3 \times 1} & \cdots & 0_{3 \times 1} & X_1^i(t_2) & 0_{3 \times 1} & \cdots & 0_{3 \times 1} \\ 0_{3 \times 1} & 0_{3 \times 1} & \cdots & 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} & \cdots & X_1^k(t_3) \end{bmatrix}$$

O modelo possui a seguinte representação matricial⁴

$$Y = W\beta + \varepsilon \quad s.a. \quad R\beta = 0 \quad (20)$$

Como as restrições são em igualdade, é possível incorporá-las diretamente na especificação de Y com um pouco de álgebra matricial. Note que a dimensão de R é $3(k-1) \times 3k$. Como por construção, as $3(k-1)$ linhas de R são linearmente independentes, o posto de R é igual a $3(k-1)$, o que permite particionar esta matriz em uma submatriz quadrada de dimensão $3(k-1)$ com posto cheio e seu complemento com dimensão $3(k-1) \times 3$.

$$R = \begin{bmatrix} R_1 & R_2 \end{bmatrix}$$

então, usando o fato de que R_1 é inversível, pode-se reescrever a restrição (11.):

$$\begin{aligned} R_1\beta_1 + R_2\beta_2 &= 0 \\ \beta_1 &= -R_1^{-1}R_2\beta_2 \end{aligned} \quad (21)$$

Reorganizando (20), e substituindo (21):

$$\begin{aligned} Y &= W_1\beta_1 + W_2\beta_2 + \varepsilon \\ Y &= Z\beta_2 + \varepsilon \end{aligned} \quad (22)$$

onde $Z_{(\tau)}$ é denominada matriz de *loadings restritos*, dada por:

$$Z_{n \times 3} = (W_2 - W_1R_1^{-1}R_2) \quad (23)$$

A equação final (22) possui além dos 3 parâmetros (β_2), os parâmetros que definem as funções em cada segmento. Estes parâmetros adicionais tornam o problema de estimação não linear o que torna a estimação muito mais custosa e complexa em termos computacionais. Neste trabalho, estes parâmetros serão tomados exogenamente implicando que a matriz Z deve ser exógena, o que torna o problema linear em β_2 . A incorporação das restrições implica em uma redução da dimensionalidade (se $k > 1$), passando de $3k$ para apenas 3 parâmetros a serem estimados (considerando o problema linear). Essencialmente, isto é reflexo de o modelo possuir $3k$ parâmetros, $\theta = \{a_i, b_i, c_i\}_{i=1}^k$ e $3(k-1)$ restrições, o que resulta em 3 parâmetros livres, independente do número de nós escolhido. Note que esta implicação é desejável uma vez que existe um consenso na literatura que 3 fatores explicam a maior parte dos movimentos da curva, sendo interpretados como nível, inclinação e curvatura (Litterman & Scheinkman (1991)). Porém, como notado anteriormente um número maior de fatores pode ser importante para previsão (ver Cochrane & Piazzesi), por isso caracterizaremos modelos com mais fatores livres em trabalho subsequente.

Na seção empírica, serão considerados três casos particulares dos modelos *SL3*:

⁴O subscrito indicando qual conjunto de maturidades a matriz W é avaliada será ocultado pois neste trabalho este conjunto será fixo para todos períodos, vale ressaltar que a metodologia pode ser implementada com conjunto de maturidades variando no tempo diretamente.

- Modelo 1: Dielbold & Li

$$g_i(x) = \frac{e^{-\lambda x}}{\lambda x}, \quad h_i(x) = \frac{e^{-\lambda x}}{\lambda x} - e^{-\lambda x}, \quad \forall i \quad (24)$$

- Modelo 2: $SL3$

A formulação geral é baseada no modelo de Dielbold & Li, alterando-se os *factor loadings* de tal forma a permitir diferentes formas funcionais para cada segmento. Embora isto gere uma descontinuidade nas junções destes *loadings*, as restrições impostas garantem que a curva continuará contínua e suave. Os *loadings restritos* neste caso serão contínuos. Neste modelo, a segmentação pode ser observada pelo valor dos parâmetros dos $3k$ parâmetros que serão diferentes entre si, embora todos sejam combinações lineares dos mesmos 3 fatores livres (denote $\gamma_{3 \times 1}$). Como exemplo, se $a_1(\gamma) \neq a_2(\gamma)$, então a interpretação é que o fator associado ao nível da curva no primeiro segmento é diferente ao fator associado ao nível no segundo.

O modelo é dado por

$$g_i(x) = \frac{e^{-\lambda \phi_i(x,p)}}{\lambda \phi_i(x,p)}, \quad i = 1, \dots, k \quad (25)$$

$$h_i(x) = \frac{e^{-\lambda \phi_i(x,p)}}{\lambda \phi_i(x,p)} - e^{-\lambda \phi_i(x,p)}, \quad i = 1, \dots, k \quad (26)$$

onde $p \in \mathbb{R}$ e

$$\frac{\partial^2 \phi_i(x,p)}{\partial x \partial p} = 0, \quad \forall x \in [0, T_M] \quad (27)$$

Uma forma natural seria assumir a formulação original e permitir que o parâmetro λ , variasse com o segmento i , que pode ser obtida com a especificação trivial⁵. A vantagem da especificação $\phi_i(x,p)$ é permitir uma diferenciação para todos os segmentos através de um único parâmetro p e permitir formatos um pouco mais flexíveis (é possível tirar o peso de maturidades longas sobre o expoente). Neste trabalho adota-se a seguinte forma funcional para ϕ :

$$\phi_i(x,p) = x - x_{i-1}(1-p), \quad x \in A_i \quad (28)$$

As implicações da escolha deste formato de $\phi_i(\cdot)$ sobre os *loadings* são discutidas na seção empírica. Como exemplo, fixe um i , a restrição (8) fica:

$$\begin{aligned} & a_i + b_i \frac{e^{-\lambda[x_i - x_{i-1}(1-p)]}}{\lambda[x_i - x_{i-1}(1-p)]} + c_i \left[\frac{e^{-\lambda[x_i - x_{i-1}(1-p)]}}{\lambda[x_i - x_{i-1}(1-p)]} - e^{-\lambda[x_i - x_{i-1}(1-p)]} \right] \\ &= a_{i+1} + b_{i+1} \frac{e^{-\lambda x_i p}}{\lambda x_i p} + c_{i+1} \left[\frac{e^{-\lambda x_i p}}{\lambda x_i p} - e^{-\lambda x_i p} \right] \end{aligned}$$

- Modelo 3: $\overline{SL3}$

As funções escolhidas para os segmentos são as mesmas do modelo $SL3$. A diferença é no tipo de restrição pois impõe-se a igualdade dos coeficientes de segmentos distintos, isto é,

⁵ $\phi_i(x,p) = \frac{\lambda_i}{\lambda}$

$a_i = a_j, b_i = b_j$ e $c_i = c_j \forall i, j \in \{1, \dots, k\}$. A segmentação neste caso será diretamente na curva, ou seja, os coeficientes serão iguais, mas os *loadings restritos* terão descontinuidades geradas pela segmentação, implicando em descontinuidade na curva. Embora seja uma característica indesejável em termos de ajuste do *cross section*, este modelo é suavizado na etapa de previsão e gera resultados bem interessantes fora da amostra. A matriz \bar{R} possui a mesma estrutura da matriz R (ver matriz (16)), só que é avaliada em outros vetores linhas $\bar{X}_j^i(x) = \bar{X}_j$ (independentes do segmento e o valor não depende mais de x) dados por:

$$\bar{X}_1 = [1 \ 0 \ 0] \quad (29)$$

$$\bar{X}_2 = [0 \ 1 \ 0] \quad (30)$$

$$\bar{X}_3 = [0 \ 0 \ 1] \quad (31)$$

e portanto, \bar{R} é dado por

$$\bar{R}_{[3(k-1)] \times (3k)} = \begin{bmatrix} \bar{X}_1 & -\bar{X}_1 & 0_{3 \times 1} & \dots & 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} \\ 0_{3 \times 1} & \bar{X}_1 & -\bar{X}_1 & \dots & 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} & \dots & 0_{3 \times 1} & \bar{X}_1 & -\bar{X}_1 \\ \bar{X}_2 & -\bar{X}_2 & 0_{3 \times 1} & \dots & 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} & \dots & 0_{3 \times 1} & \bar{X}_2 & -\bar{X}_2 \\ \bar{X}_3 & -\bar{X}_3 & 0_{3 \times 1} & \dots & 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0_{3 \times 1} & 0_{3 \times 1} & 0_{3 \times 1} & \dots & 0_{3 \times 1} & \bar{X}_3 & -\bar{X}_3 \end{bmatrix} \quad (32)$$

3.2 Estimação no *Cross Section*

Os parâmetros β, λ e p podem ser estimados através do procedimento de mínimos quadrados não lineares. Diebold & Li propõe, entretanto, um procedimento mais simples, que consiste em fixar o valor do parâmetro λ e estimar os coeficientes no *cross section* por mínimos quadrados ordinários. Seguindo esta metodologia, λ será fixado *ex ante* assim como o parâmetro p para os modelos SLs, implicando que o processo de estimação se reduz ao vetor de parâmetros β . então, para λ e p fixos (equivalentemente a dizer que a matriz Z da equação.(22) é exógena), o problema é dado por

$$\min_{\beta_2} (Y - Z\beta_2)'(Y - Z\beta_2)$$

cuja solução será um vetor com 3 parâmetros:

$$\hat{\beta}_2 = (Z'Z)^{-1}Z'Y$$

O modelo $\overline{SL3}$ é calculado de maneira análoga exceto que a matriz Z da equação (22) é denotada por \bar{Z} ao substituí as as restrições por \bar{R} , com o estimador deste modelo denotado por $\hat{\beta}_2$. O vetor de parâmetros do modelo de Diebold & Li, que será usado para comparação dos modelos será denotado por $\tilde{\beta}$, que é obtido de forma análoga dada pelos *loadings* da equação (1).

3.3 Previsão da curva

Agora será adicionada à notação dos parâmetros o índice t referente à data da observação Y_t . A extração dos coeficientes descrita na seção anterior será realizada para cada data t , gerando assim uma série temporal para os vetores de coeficientes, que será denotada por $\{\theta_t^i\}_{t=1}^J$, onde $i \in \{1, 2, 3\}$ representa o modelo e $\theta_t^1 = \tilde{\beta}_t$, $\theta_t^2 = \hat{\beta}_{2t}$ e $\theta_t^3 = \widehat{\beta}_{2t}$. O escalar J é o tamanho da amostra que será usada na estimação.

A partir destas séries, pode-se inferir sobre a dinâmica e previsão dos coeficientes e, conseqüentemente, da ETTJ. Defina, para uma data fixada j , a amostra de tamanho J por

$$C_j = \{\theta_{j-J+1}, \dots, \theta_{j-1}, \theta_j\}$$

Gera-se uma previsão h passos a frente através do operador $F_h^i(\cdot)$

$$F_h^i(C_j) = \hat{\theta}_{j+h}^i \quad (33)$$

onde h representa o horizonte da previsão. A previsão dos coeficientes, dada pelo operador $F_h^i(\cdot)$ pode ser por métodos variados, sejam lineares ou não. Os modelos que serão adotados neste trabalho serão descritos na seção empírica. Uma vez obtido a previsão dos parâmetros, o mapeamento para a previsão das taxas será dado por:

$$\hat{Y}_{j+h}^i = Z^i \hat{\theta}_{j+h}^i \quad (34)$$

Repare que, independente do modelo, o mapeamento das taxas nos coeficientes estimados é linear o que facilita a implementação. Como exemplo, será usado como *benchmark* o modelo "random walk", indexado por $i = 0$:

$$\hat{\theta}_{j+h}^i = F_h^0(C_j) = Y_j ; Z_{(\tau)}^0 = I_M \implies \hat{Y}_{j+h}^0 = Z_{(\tau)}^0 \hat{\theta}_{j+h}^0 = I_M Y_j = Y_j$$

3.4 Método de Avaliação dos Modelos

O exercício que será feito é a previsão das taxas para um conjunto de maturidades da curva, para diferentes horizontes (h). As avaliações serão feitas sob três métricas: a magnitude dos erros de previsão, a comparação relativa com a previsão "random walk" e a consistência destes resultados sob diferentes períodos amostrais. O método de avaliação é descrito abaixo:

Seja $A = \{Y_0, Y_1, \dots, Y_T\}$ a amostra completa, onde $Y_i \in \mathbb{R}^M$ é o vetor de taxas observadas para M maturidades. Defina o tamanho J da subamostra de estimação deste conjunto, e denote por A_j , onde $j \leq T - h$ é o índice do período final da subamostra, por:

$$A_j = \{Y_{j-J+1}, \dots, Y_{j-1}, Y_j\}$$

Para esta subamostra vamos estimar um conjunto de coeficientes C_j e conseqüentemente uma previsão um passo a frente \hat{Y}_{j+h}^i (ver seção 3.3). Define-se o vetor de erros de previsão do modelo i , um passo a frente por

$$e_{j+h}^i = Y_{j+h} - \hat{Y}_{j+h}^i$$

e N como o número de observações que serão usadas para fazer a avaliação da previsão fora da amostra. Assim, gera-se uma série de tamanho N

$$\{e_{w+h}^i\}_{w=j}^{j+N-1} \quad \text{onde} \quad j + N - 1 \leq T$$

na qual serão baseados os critérios de avaliação.

3.4.1 1ª métrica - Erros Quadráticos Médios (EQM) Relativos

Repare que $e_{j+h}^i \in \mathbb{R}^M$, isto é, para cada j fixado, têm-se os erros de previsão para M maturidades da curva. Denote $E_{m,h}^i \in \mathbb{R}^N$ como o vetor de erros da m -ésima maturidade referente ao modelo i , para previsão h períodos a frente, então

$$EQM_{m,h}^i = \frac{E_{m,h}^{i'} E_{m,h}^i}{N} \quad ; \quad m = 1, \dots, M \quad ; \quad i = 0, 1, \dots$$

O segundo passo é avaliar o EQM do modelo i em relação à algum *benchmark*, que será indexado por b . Valores negativos representam melhoras percentuais em relação ao *benchmark*.

$$V_{m,h}^{i,b} = \frac{EQM_{m,h}^i - EQM_{m,h}^b}{EQM_{m,h}^b}; \quad m = 1, \dots, M \quad ; \quad i = 0, 1, \dots \quad (35)$$

É comum na literatura usar a previsão por "*Random Walk*" como este benchmark. Este modelo será indexado por $b = 0$.

3.4.2 2ª métrica - Método de Robustez para Amostras Distintas no tamanho ou no período

A métrica anterior avalia apenas um período específico com tamanho N . O método proposto consiste em avaliar os resultados para tamanhos e períodos amostrais distintos, ou seja, se o experimento for repetido para amostras diferentes, será que os modelos preservam suas qualidades de previsão? Esta é uma característica desejável pois constitui uma forma de verificar a robustez destes modelos.

Para fazer a avaliação de um modelo i , deve-se fixar os parâmetros⁶ N, J, j, b, i, h . Para simplificação de notação, faça $\Upsilon = \{N, J, b, i, j, h\}$. Assim, o valor (35) será denotado por V_m^Υ . Considere primeiro alterar a amostra variando K_1 janelas amostrais consecutivas, ou seja, variar o indexador de período do conjunto Υ . Denotaremos Υ_t o conjunto de índices fixados alterando-se apenas o índice $j = t$. Neste exercício, há duas dimensões que devem ser avaliadas, a primeira sendo o período e a segunda a maturidade da curva. Gera-se então uma sequência de valores

$$\{V_m^{\Upsilon_t}\}_{t=j-K_1}^j, \quad j \geq K_1, \quad m = 1, \dots, M \quad (36)$$

que fornecerá uma superfície nestas duas dimensões. Esta superfície será relativa ao *benchmark* e portanto, pontos abaixo de zero indicam melhores previsões relativas. Para sintetizar a informação em cada instante de tempo em apenas um escalar, propoe-se olhar também a média entre as maturidades.

$$\{VM^{\Upsilon_t}\}_{t=j-K_1}^j, \quad j \geq K_1; \quad (37)$$

$$\text{onde } VM^{\Upsilon_t} = \frac{1}{M} \sum_{m=1}^M V_m^{\Upsilon_t}$$

⁶Os índices representam:

N : tamanho da amostra de previsão
 J : tamanho da subamostra de estimação
 j : período final da subamostra
 b : modelo benchmark
 i : modelo
 h : horizonte de previsão

Considere agora o caso de variar o tamanho da amostra de previsão. Denotaremos Υ_w o conjunto de índices fixados alterando-se apenas o índice $N = w$. Novamente teremos uma superfície representada pelas diferentes maturidades e pelos diferentes tamanhos de amostras de previsão.

$$\{V_m^{\Upsilon_w}\}_{w=N-K_2}^N, \quad K_2 < N, \quad m = 1, \dots, M$$

e analogamente

$$\{\overline{VM}^{\Upsilon_w}\}_{w=N-K_2}^N, \quad K_2 < N \quad (38)$$

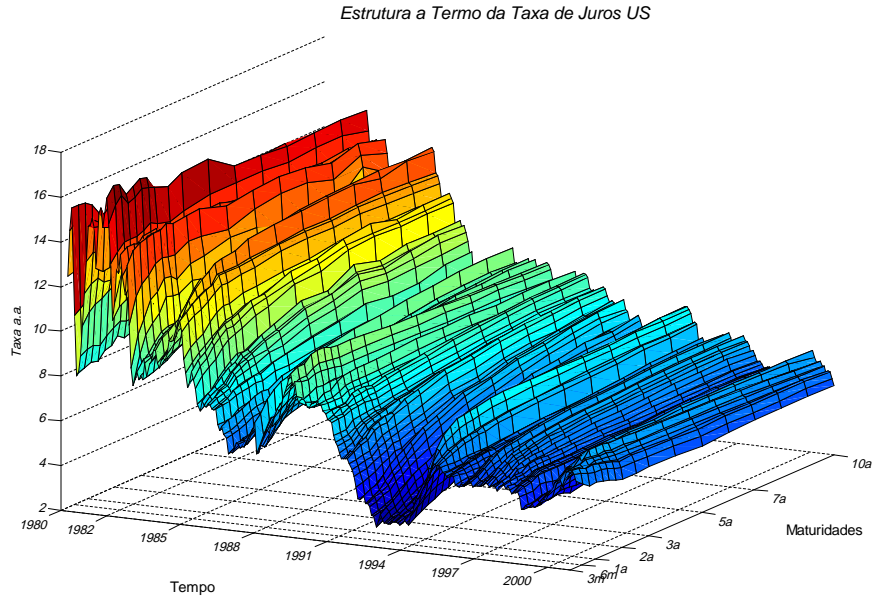
$$\text{onde } VM^{\Upsilon_w} = \frac{1}{M} \sum_{m=1}^M V_m^{\Upsilon_w}$$

4 Avaliação Empírica

Esta seção será dividida em 5 partes, a primeira será a descrição dos dados, a segunda uma análise dos *loadings*, a terceira e quarta parte serão a estimação dos coeficientes e previsão das curvas e por fim, na quinta parte será a avaliação da previsão dos modelos.

4.1 Dados

A base de dados utilizada foi a mesma usada no trabalho de Dielbold Li, já tratada conforme descrito em seu paper (2006), referente à taxa de final de mês (média de *bid-ask*) dos títulos do governo americano de cupom zero (*zero cupom US Treasuries*). O período da amostra vai de Janeiro de 1980 à Dezembro de 2000. As maturidades coletadas também seguem critério adotado por estes autores, fixadas por toda a amostra em 3,6,9,12,15,18,21,24,30,36,48,60,72,84,96,108 e 120 meses. A figura 1 apresenta estes dados.



4.2 Factor Loadings

Nesta seção será examinada a propriedade dos modelos SL3 (*segmented loadings 3*). Os *loadings* propostos pelo modelo de Nelson-Siegel são polinômios com decaimento exponencial que permitem uma interpretação aplicada à modelagem de curva de juros. Conforme destacado por Diebold & Li, o primeiro fator, por ser constante, pode ser visto com fator de longo prazo (não é afetado pela maturidade). O segundo começa em 1 e cai exponencialmente para zero na medida em que a maturidade aumenta, sendo interpretado como fator de curto prazo. O terceiro fator, começa em zero, é crescente inicialmente e, a partir de determinada maturidade, começa a cair monotonicamente a zero, refletindo um componente de médio prazo. Os fatores associado à estes *loadings* são associados ao nível, inclinação e curvatura da estrutura a termo, conforme interpretação usual da literatura relacionada. No primeiro gráfico da figura 2, são mostrados o 2º e 3º *loadings* deste modelo.

A hipótese de segmentação da curva será feita através da flexibilização nos loadings Considere o segundo *loading factor* do modelo de Nelson-Siegel $g_i(x) = \frac{e^{-\lambda_i x}}{\lambda_i x}$, $\lambda_i = \bar{\lambda} \forall i$, uma primeira maneira de usar a estrutura de segmentação é usar um parâmetro λ_i distinto para cada partição A_i . Os *loadings* originais são representados pela matriz W (ver matriz (19)), com dimensão $n \times 3k$, onde $g_i(\cdot)$ e $h_i(\cdot)$ são dados pela equação (24). Note que, para uma maturidade arbitrária, a respectiva linha da matriz possui três valores diferentes de zero, referentes às colunas associadas ao segmento que pertence. Para facilitar a visualização gráfica, agrupamos esta matriz em um matriz \tilde{W} com apenas 3 colunas. No segundo gráfico da Figura 2 mostra-se um exemplo para o conjunto de nós $\Delta_{(k)} = \{1, 2, 4, 18, 24, 84\}$ e $\lambda = \{0.10, 0.085, 0.077, 0.053, 0.042, 0.030\}$, ou seja, λ_i decrescente entre as partições.

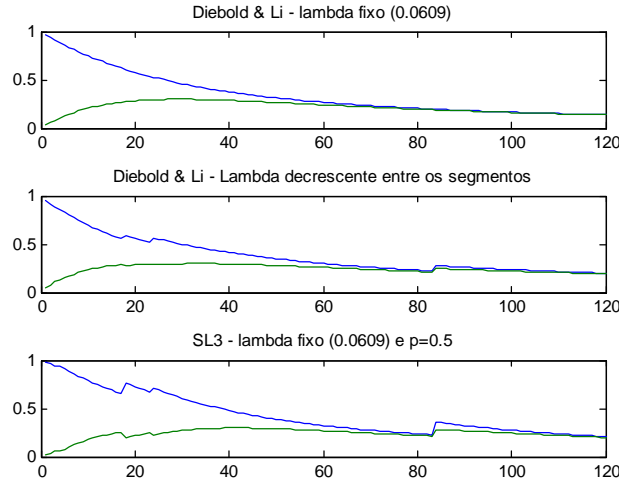


Figura 2: *Factors Loadings* (por maturidade)

Embora a especificação de λ_i distintos permita a descontinuidade, note que não gera muita flexibilidade pois o "salto" no *loading* nos nós, para valores de x_i elevados, é muito pequeno.

A forma alternativa é usar a classe de modelos SL, em que a descontinuidade não atua sobre o parâmetro λ e sim sobre o termo da maturidade x . Recordando a especificação do SL (seção 3.1), o 2º *loading* é dado por $g_i(x) = \frac{e^{-\lambda\phi_i(x,p)}}{\lambda\phi_i(x,p)}$ com $\phi_i(x,p) = x - x_{i-1}(1-p)$, ou seja, no limite com $p = 0$, o *loading* volta a valer 1, caracterizando um *reset* em cada nó, enquanto que para $p = 1$, a especificação original de Nelson-Siegel é recuperada (não há descontinuidade nos nós). Repare que, embora também seja possível escolher $\lambda = 0$, isto implicará em uma função constante para todo $x \in A_i$, que não ocorre no caso de $p = 0$. Além disso optou-se por esta formulação por depender de apenas um parâmetro. Os *loadings* do modelo *SL3*, para o caso de segmentação em ambos, estão no terceiro grafico da figura 2. Uma potencial vantagem em permitir a segmentação é diminuir a multicolinearidade existente entre os *loadings*, em especial para maturidades longas, o que afeta a precisão dos estimadores.

Contudo, os *loadings* mostrados na figura 2, em geral não são os *loadings* que serão usados para a estimação. Conforme visto na seção 3.1 a incorporação das restrições altera estes *loadings*, assim, deve-se analisar a matriz Z , para entender quais são os *loadings* efetivamente atuantes:

$$Y = W\beta + \varepsilon \quad s.a. \quad R\beta = 0 \implies Y = Z\beta + \varepsilon$$

Nesta parte que os modelos *SL3* e $\overline{SL3}$ diferem. A transformação da matriz W para a matriz Z implica em uma redução da dimensão de $n \times 3k$ para $n \times 3$, para os dois modelos. Entretanto no primeiro, as restrições de continuidade e suavidade impostas sobre o funcional latente implicam que o *loading* restrito também será contínuo e suave, enquanto para o segundo, as restrições (igualdade nos parâmetros) apenas replicam os *loadings*. Na figura 3a mostra-se esta transformação. Os valores de p são aqueles que serão usados no exercício de previsão fora da amostra e o critério de escolha é descrito na seção empírica⁷.

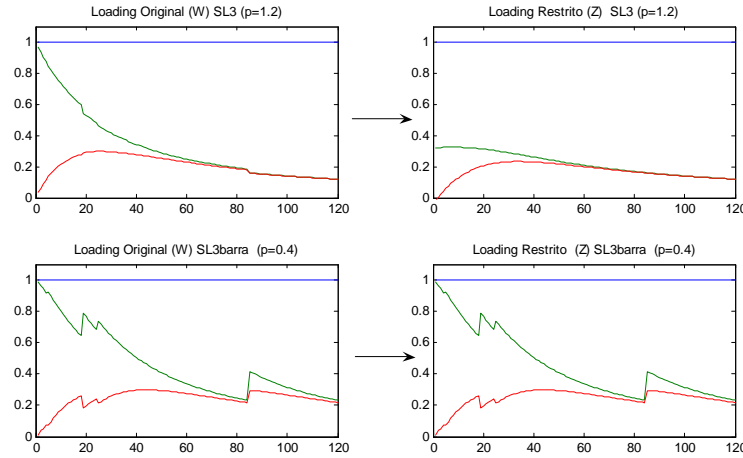


Figura 3a: *Loading Factor* Originais (matriz W) e Restritos (matriz Z).

Nota-se que, dependendo do tipo de segmentação imposta sobre os *loadings* originais, a multicolinearidade nos *loadings* restritos pode piorar (*SL3* com $p = 1.2$). Uma alternativa natural ao

⁷Vale lembrar que, para mesmos valores de p , os *loadings* originais são os mesmos nos dois modelos (a matriz W é a mesma o que muda são as restrições $R \neq \bar{R}$).

$SL3$, é a segmentação em um único *loading*. Neste caso, o efeito sobre a redução da multicolinearidade dos *loadings* tende a ser mais relevante, pois o salto em um *loading* não é acompanhado do salto do outro *loading*. Na figura 3b apresentamos os modelos $SL3 - 2$ e $SL3 - 3$ que representam segmentação imposta sobre o 2º e o 3º *loadings* originais respectivamente. Para efeito ilustrativo, utilizamos $p = 0.5$ e também foi feita uma normalização nos *Loadings Restritos* de tal forma que o valor máximo deles é igual a 1.

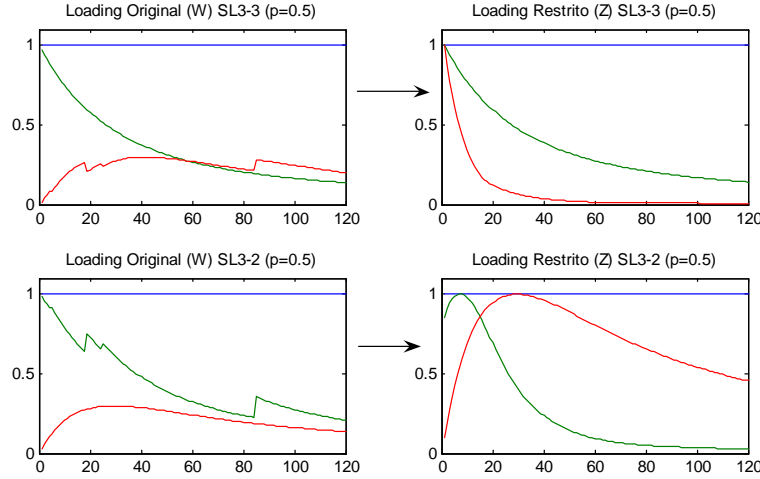


Figura 3b: *Loading Factor* Originais (matriz W) e Restritos (matriz Z), para os modelos $SL3$ com segmentação em apenas um *loading* original.

4.3 Estimação

Serão avaliados 5 modelos:

- Diebold & Li (DL)
- $SL3$, $SL3 - 2$ e $SL3 - 3$
- $\overline{SL3}$

Repare que todos os modelos são baseados na forma paramétrica proposta por Diebold & Li, nos quais o parâmetro λ determina a taxa de decaimento exponencial. Será adotado um valor fixo para λ de 0.0609, baseado no paper de Diebold & Li(2006). Adicionalmente, para os modelos segmentados, os valores fixados para p serão os valores escolhidos no exercício dentro da amostra que será detalhado na subseção de resultados. Usou-se $p = 1.2$ e $p = 0.4$ para os modelos $SL3$ e $\overline{SL3}$ respectivamente. Vale notar que, idealmente, λ e p deveriam também ser estimados, o que gera um custo computacional elevado, pois torna o problema não linear.

Para o modelo $\overline{SL3}$, devido às restrições de igualdade, os coeficientes serão iguais nos diferentes segmentos. A interpretação da segmentação é que mesmos coeficientes atuam sobre a curva por segmentos de *loadings* distintos. No modelo $SL3$ a segmentação possui uma interpretação diferente onde, embora o sistema seja reduzido a três fatores, os $3k$ fatores originais podem ser recuperados usando a equação (19). Para entender tal efeito, considere a curva de juros do dia

31/08/1998, para o conjunto de nós $\Delta_{(k)} = \{1, 2, 4, 18, 24, 84\}$. Considerando o modelo $SL3$ com $p = 1.2$, há 6 segmentos e portanto serão 18 coeficientes originais, conforme mostrado abaixo:

Fatores Estimados (Restritos)			Fatores Originais			
1 fator	2 fator	3 fator	Segmento	1 fator	2 fator	3 fator
5.08	0.09	-0.54	1	5.45	-0.52	-0.86
			2	5.31	-0.38	-0.70
			3	5.20	-0.27	-0.56
			4	5.10	-0.20	-0.31
			5	5.07	-0.19	-0.24
			6	5.08	0.09	-0.54

Aplicando-se a metodologia descrita na seção (3.2), foram geradas as séries de coeficientes de cada modelo. Na Fig4a, são mostrados os gráficos da média das curvas geradas pelos modelos e média das curvas observadas.

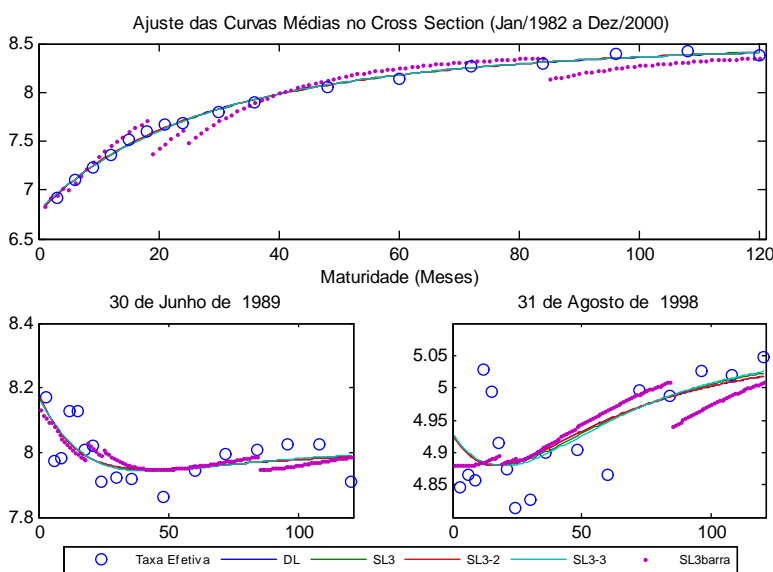


Figura 4a: Ajuste dos Modelos na dimensão cross section, em relação à média observada no período e em relação à dias específicos.

Nota-se que o modelo $\overline{SL3}$ não funciona bem para o ajuste da curva no *cross section*, uma vez que erra muito e possui características indesejáveis como as descontinuidades nos nós. Estes erros são diretamente relacionados à restrição de igualdade dos fatores entre os segmentos, ou seja, o mesmo fator que atua sobre a parte curta da curva atua sobre a parte longa, mas o *loading* é diferente, o que gera esse efeito. Entretanto, como será mostrado adiante, possui boa capacidade preditiva. O modelo de Diebold & Li apresenta um ótimo ajuste enquanto que os demais são muito semelhantes a este. É evidente que tal semelhança decorre do fato que o parâmetro p escolhido para o $SL3$ é muito próximo de 1, o que torna seus coeficientes muito próximos. Os modelos $SL3$ neste caso, são apenas "perturbações" no modelo de Diebold & Li, em que as diferenças médias não passam de 2 pontos base, conforme mostrado na figura 4b, primeiro gráfico. Interessante que o comportamento das diferenças aparenta ter uma simetria em relação à distancia do parâmetro p de 1 (se $p = 1$, a diferença nos três modelos é zero).

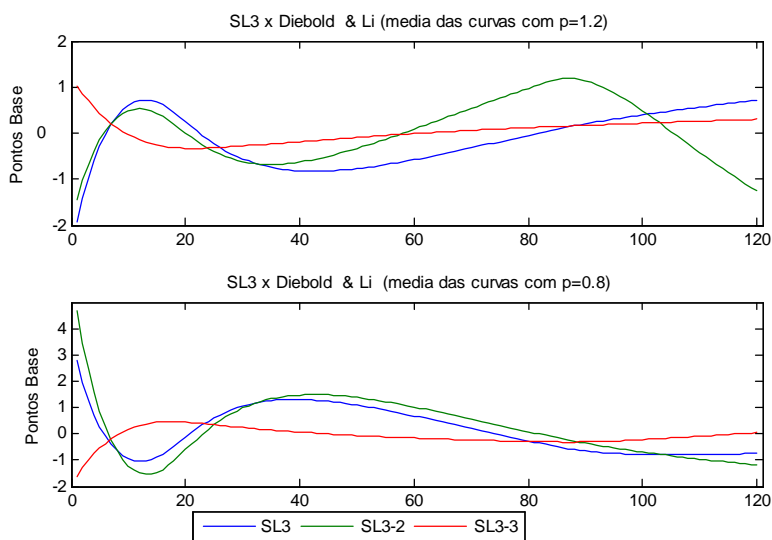


Figure 4b: Diferença das curvas estimadas médias no período de Jan/1982 a Dez/2000 entre os modelos SL3 e o Diebold & Li, para duas escolhas do parâmetro p .

Os comportamento dos erros do modelo de Diebold & Li são mostrados na figura 4c, note que a magnitude dos erros é muito baixa, em torno de 5 pontos base, refletindo o ótimo ajuste destes

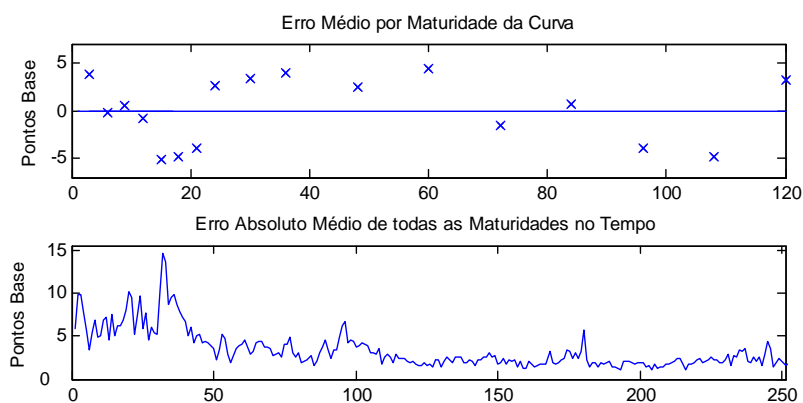


Figura 4c: Erros médios por maturidade e evolução dos erros absolutos médios para o modelo de Diebold & Li

A seguir são mostradas as séries dos fatores estimados e a correlação destes para os modelos Diebold & Li, $SL3$ e $SL3$. A interpretação de cada fator é feita diretamente associando à nível, inclinação e curvatura respectivamente (Litterman & Scheinkman(1991)). Um dos fatos estilizados na literatura é que as taxas apresentam forte persistência, que é capturada pelo fator referente ao nível da curva enquanto que os *spreads* apresentam menor persistência, representados pelo fator relacionado à inclinação. Na figura 5, vemos que todos modelos capturam esta dinâmica, mas a diferenciação dos modelos segmentados é que o fator relacionado à inclinação

(2º fator) possui maior variabilidade. Em particular, o modelo $\overline{SL3}$ apresenta menor correlação entre os fatores, o que talvez possa explicar a melhor qualidade na previsão.

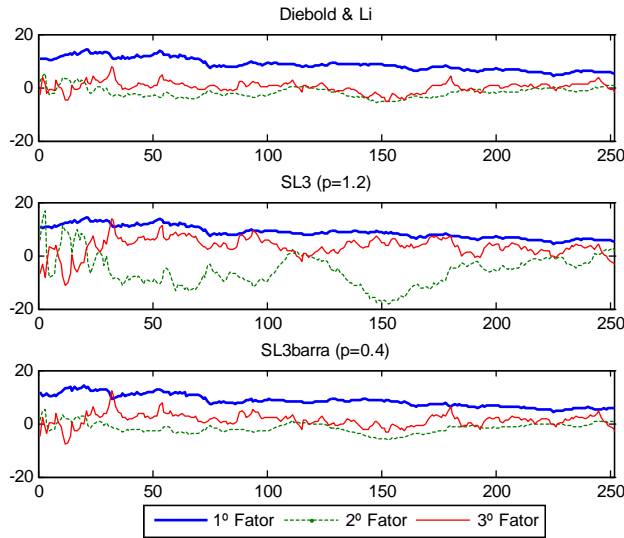


Figura 5: Série temporal dos fatores

Correlação dos Fatores

Diebold & Li	SL3 (p=1.2)			SL3barra (p=0.4)		
1.00				1.00		
-0.02	1.00			0.02	1.00	
0.22	0.22	1.00		0.07	-0.01	1.00
			1.00			
			0.17	-0.68	1.00	
						1.00

Repare que a extração destes fatores envolveu 252 estimações distintas por Mínimos Quadrados Ordinários. Uma forma de diagnosticar a qualidade destas estimações é olhar para a estatística t destes coeficientes. A série destas estatísticas foi significativa a 5% em 100% dos casos (252 regressões) para o fator associado ao nível (intercepto), em todos os modelos. Contudo, para os outros dois fatores, há períodos em que estes não são estatisticamente significantes. A frequência de estatísticas t significantes para os 5 modelos analisados é reportada na tabela abaixo.

Modelo	Significantes à 10%			Significantes à 5%		
	1 fator	2 fator	3 fator	1 fator	2 fator	3 fator
Diebold Li	100%	96%	86%	100%	95%	77%
SL3	100%	85%	84%	100%	81%	80%
SL3-2	100%	96%	80%	100%	95%	73%
SL3-3	100%	77%	84%	100%	73%	77%
SL3barra	100%	96%	72%	100%	95%	65%

4.4 Previsão

Seguindo a metodologia apresentada na seção (3.2) o primeiro passo é definir o tamanho da amostra que será utilizada para a estimação. Como período inicial da amostra foi usado Jan/1985 e como período inicial de previsão Jan/1994 (defina data como $j + h$). Assim, o tamanho da janela amostral de estimação será a quantidade de períodos entre estas datas menos o horizonte

de previsão, ou seja, $J = 108 - h$. O período de previsão será de Jan/1994 a Dez/2000 (84 observações). Para cada modelo i e horizonte h deve-se gerar as séries de previsão dos parâmetros

$$\{\hat{\theta}_k^i\}_{k=j+h}^{84+j+h}, \quad h \in \{1, 3, 6, 9, 12\}, \quad i \in \{0, 1, \dots, 5\} \quad (39)$$

4.4.1 Modelos de Previsão dos Parâmetros

Seja $l \in \{1, 2, 3\}$ o índice de cada fator e $i \in \{0, 1, \dots, 5\}$ o índice da cada modelo, sendo que o random walk é representado pelo índice 0. Serão avaliados 2 especificações lineares autoregressivas de previsão dos coeficientes dos modelos⁸:

$$\theta_{t+h}^{i,l} = a_{i,l} + b_{i,l}\theta_t^{i,l} + \varepsilon_{t+h} \implies \hat{\theta}_{t+h}^{i,l} = \hat{a}_{i,l} + \hat{b}_{i,l}\theta_t^{i,l}, \quad \forall l \quad (40)$$

$$\theta_{t+1}^{i,l} = a_{i,l} + b_{i,l}\theta_t^{i,l} + v_{t+1}, \implies \hat{\theta}_{t+h}^{i,l} = \hat{a}_{i,l} \sum_{k=0}^{h-1} (\hat{b}_{i,l})^k + (\hat{b}_{i,l})^h \theta_t^{i,l}, \quad \forall l \quad (41)$$

$$\text{onde } h \in \{1, 3, 6, 9, 12\}, i \in \{1, \dots, 5\}$$

As duas especificações caracterizam um processo $AR(1)$ mas são diferentes para horizontes de previsão superiores à um mês ($h \neq 1$). Na primeira, os fatores associados à $t + h$ são regredidos diretamente nos fatores defasados no horizonte, seguindo critério adotado por Diebold & Li (2006) em que basea-se no método de mínimos quadrados avaliados na função perda de previsão h -passos à frente. A segunda especificação é um $AR(1)$ tradicional onde a previsão de h -passos à frente é feita indutivamente. O modelo VAR também foi testado para a previsão destes fatores mas não apresentou resultados melhores. Diebold & Li justificam que embora o VAR capture a relação conjunta das variáveis, pode gerar um *overfitting* dentro da amostra o que pode prejudicar a capacidade preditiva. Um segundo argumento a favor da especificação $AR(1)$ é a parcimônia do modelo no sentido de implicar na estimação de um número bem menor de parâmetros.

Para a previsão serão estimadas 84 regressões distintas, com mesmo tamanho de janela amostral. Uma forma de diagnosticar a qualidade destes modelos é olhar para a autocorrelação dos resíduos. Adotou-se olhar para 3 regressões (1, 42, 84) que acreditamos ser representativas das demais uma vez que as amostras diferem pouco (existe sobreposição uma vez que é uma janela móvel). Na figura 6, mostra-se as funções de autocorrelação para a regressão da primeira janela amostral (que mostrou-se semelhante às regressões 42 e 84), de cada um dos parâmetros do modelo SL3 usando o critério (41). Para $h = 1$, os resíduos são bem comportados, o que não ocorre para $h = 12$. Apesar disto, esta primeira especificação é a que será usada para seguir a metodologia proposta por Diebold & Li⁹.

⁸Não se aplica ao modelo random walk.

⁹Temos uma nota comparando os resultados ao usar os modelos de previsão (40) e (41).

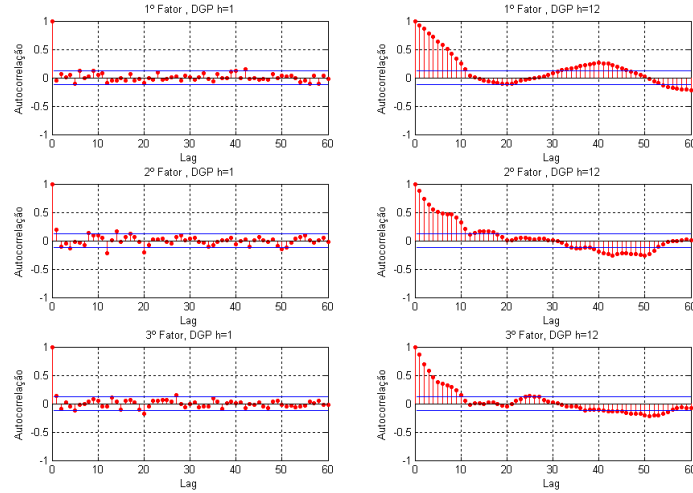


Fig5a: Funções de Autocorrelação para o modelo SL3 com h=1 e h=12

Uma vez obtidas as séries de previsão dos coeficientes, usa-se a equação (34) para obter as respectivas séries de previsões das M maturidades da curva $\{F_h^i(C_{k-h})\}_{k=j+h}^{84+j+h}$, $F_h^i(C_j) \in \mathbb{R}^M$. Um tratamento deve ser feito na previsão do modelo $\overline{SL3}$, que como comentado antes, gera uma curva com descontinuidade nos nós. Assim, para este modelo, existe uma etapa adicional que é impor às taxa previstas um formato de curva contínua e suave. Adotou-se um spline cúbico, que será denotado por um operador S (descrito no apêndice) sobre as maturidades previstas:

$$\hat{Y}_{j+h}^5 = S(Z_{(\tau)}^5 F_h^5(C_j)) = S(Z_{(\tau)}^5 \hat{\theta}_{j+h}^5) \quad (42)$$

4.5 Resultados

4.5.1 Escolha do parâmetro p nos modelos com descontinuidade nos *loadings*

Os modelos $SL3$ e $\overline{SL3}$ requerem a definição do parâmetro p . Definiu-se o critério para escolha de p aquele que, para cada modelo, minimiza o erro quadrático médio de previsão em uma subamostra de treino, dentro da amostra. O período de previsão dentro da amostra será de Jan/1988 a Dez/1993, e a janela de estimação começa com o período de Jan/1982 a Dez/1987 (ajustado para o horizonte de previsão, diminuído a janela amostral). O erro quadrático médio foi calculado para previsões com horizontes de 1, 3, 6 e 12 meses. O método de previsão dentro da amostra usado para escolha dos parâmetros foi a equação(41) por apresentar maior robustez no sentido de o parâmetro melhor não variar muito com o horizonte de previsão.

Modelo	Parâmetro "p" ótimo dentro da amostra (período 1988:01 a 1993:12)			
	h=3	h=6	h=9	h=12
SL3	1.21	1.21	1.21	1.21
SL3-2	1.16	1.21	0.10	0.10
SL3-3	1.16	1.21	1.21	1.16
S13 barra	0.39	0.34	0.29	0.20

4.5.2 Previsão Fora da Amostra

Para comparação dos modelos propostos, serão mostradas 3 dimensões relevantes. A primeira é olhar o comportamento dos erros de previsão para as maturidades da curva, fixado um horizonte de previsão e um período. A segunda e terceira serão justamente variar os horizontes de previsão e os períodos de avaliação.

Nas tabelas 2a , 2b e 2c são mostrados os resultados em termos de raiz do erros quadráticos médios das previsões no período analisado (Jan/1994 a Dez/2000), para horizontes de 1, 6 e 12 meses, valores em azul e vermelho representam os melhores e piores resultados para determinada maturidade. O primeiro ponto que deve ser destacado é que os resultados obtidos para previsão 1 mês a frente não são bons pois nenhum modelo consegue superar de maneira significativa os erros obtidos pelo Random Walk e há uma dispersão entre os modelos no sentido que não há como inferir qual tem melhor performance relativa. Para horizontes superiores a um mês, os modelos em sua maioria, erram em média menos do que a previsão por Random Walk, indicando que a curva possui alguma informação em seu comportamento passado que ajuda a prever o comportamento futuro, característica já destacada pela literatura como Cochrane/Piazessi, Fama, Diebold & Li, Bowsher & Meeks entre outros. O segundo ponto é que os modelos *SL3* geram resultados semelhantes (em magnitude e formato) por possuírem, essencialmente, a mesma estrutura paramétrica a menos de pequenas modificações.

Horizonte de Previsão: 1 mês	Maturidades (Meses)					
	3	12	36	60	84	120
Random Walk	0.18	0.24	0.28	0.27	0.26	0.25
(vies)	0.03	0.02	0.01	(0.00)	(0.01)	(0.01)
Diebold & Li	0.17	0.23	0.27	0.28	0.27	0.25
(vies)	(0.01)	0.05	(0.02)	(0.06)	(0.00)	(0.03)
SL3	0.17	0.23	0.28	0.29	0.27	0.25
(vies)	(0.02)	0.03	(0.06)	(0.08)	(0.02)	(0.05)
SL3-2	0.17	0.23	0.28	0.29	0.27	0.25
(vies)	(0.01)	0.03	(0.05)	(0.07)	(0.01)	(0.05)
SL3-3	0.17	0.23	0.27	0.28	0.27	0.25
(vies)	(0.04)	0.04	(0.03)	(0.07)	(0.01)	(0.03)
SL3 barra	0.18	0.22	0.28	0.29	0.27	0.41
(vies)	(0.06)	(0.01)	(0.01)	(0.10)	(0.05)	0.14

Tabela 2a: Raiz do Erros Quadráticos Médios de Previsão e vieses do erro (fonte menor) no Período de Jan/1994 a Dez/2000 para o horizonte de 6 meses

Horizonte de Previsão: 6 meses	Maturidades (Meses)					
	3	12	36	60	84	120
Random Walk	0.60	0.78	0.87	0.86	0.81	0.75
(vies)	0.22	0.18	0.10	0.05	0.02	(0.02)
Diebold & Li	0.54	0.67	0.75	0.77	0.74	0.69
(vies)	0.09	0.16	0.00	(0.11)	(0.11)	(0.18)
SL3	0.53	0.66	0.79	0.82	0.78	0.73
(vies)	0.07	0.05	(0.15)	(0.23)	(0.20)	(0.25)
SL3-2	0.57	0.70	0.83	0.86	0.83	0.76
(vies)	0.02	0.01	(0.18)	(0.28)	(0.26)	(0.24)
SL3-3	0.49	0.66	0.76	0.78	0.75	0.70
(vies)	(0.03)	0.06	(0.05)	(0.14)	(0.13)	(0.20)
SL3 barra	0.51	0.66	0.72	0.72	0.71	0.66
(vies)	0.12	0.17	0.09	(0.07)	(0.05)	0.12

Tabela 2b: Raiz do Erros Quadráticos Médios de Previsão e vieses do erro (fonte menor) no Período de Jan/1994 a Dez/2000 para o horizonte de 6 meses

Horizonte de Previsão: 12 meses	Maturidades (Meses)					
	3	12	36	60	84	120
Random Walk	1.01	1.19	1.23	1.18	1.10	1.05
(vies)	0.42	0.39	0.24	0.13	0.06	(0.03)
Diebold & Li	0.74	0.75	0.81	0.89	0.91	0.92
(vies)	0.13	0.21	(0.05)	(0.26)	(0.33)	(0.48)
SL3	0.71	0.70	0.80	0.91	0.92	0.94
(vies)	0.12	0.09	(0.21)	(0.40)	(0.44)	(0.55)
SL3-2	0.87	0.96	1.02	1.04	1.00	0.97
(vies)	(0.28)	(0.39)	0.26	0.31	0.34	0.25
SL3-3	0.66	0.71	0.81	0.90	0.91	0.92
(vies)	0.02	0.12	(0.09)	(0.29)	(0.35)	(0.49)
SL3 barra	0.79	0.79	0.78	0.82	0.79	0.70
(vies)	0.32	0.23	0.01	(0.25)	(0.25)	0.05

Tabela 2c: Raiz do Erros Quadráticos Médios de Previsão e vieses do erro (fonte menor) no Período de Jan/1994 a Dez/2000 para o horizonte de 12 meses

Os modelos SL3, que impõem a segmentação da curva, apresentaram bons resultados, onde foram superiores na maior parte das maturidades da curva, para ambos horizontes analisados. Embora a diferença não seja muito significativa, há uma evidência a favor desta modelagem. Para facilitar visualização dos resultado apresentados nas tabela 2b e 2c, as figuras 6a e 6b representam os mesmos resultados adicionando também os horizontes de 3 e 9 meses, só que em termos relativos aos erros de previsão do modelo benchmark (Random Walk). Valores negativos representam ganhos percentuais em relação ao erro médio do modelo benchmark, sendo esta a métrica descrita pela equação (35). Nota-se que o formato dos ganhos relativos é bem semelhante entre os modelos, caracterizados por menores ganhos nas maturidades mais longas da curva, resultado encontrado por Diebold & Li e replicado pelos modelos SLs dado que possuem mesma forma paramétrica. Um ponto importante é notar que o modelo $\overline{SL3}$ difere dos demais por manter o comportamento de erro mais estável entre as maturidades, o que implica em melhores previsões para a ponta longa da curva em relação aos demais.

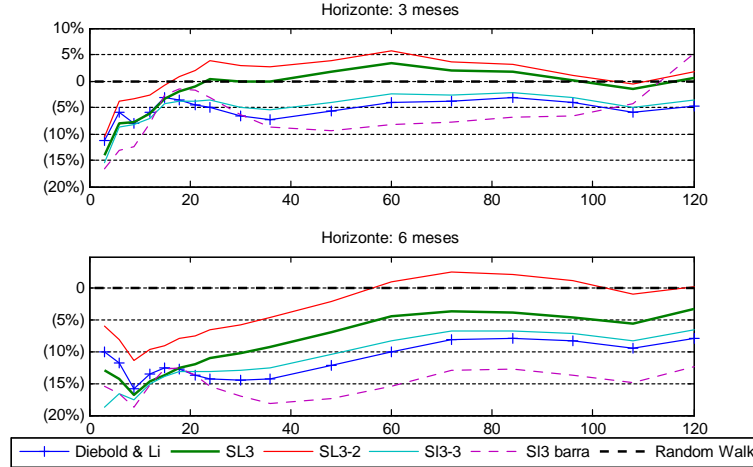


Figura 6a: Percentual de Erros Quadráticos Médios de Previsão dos modelos relativos ao modelo Random Walk para os horizontes de 3 e 6 meses

Uma outra maneira de comparar os modelos é olhar na dimensão dos horizontes de previsão. Para este período fixado, os modelos preservaram a ordenação para horizontes distintos, com o $\overline{SL3}$ obtendo os melhores resultados. Mas vale ressaltar que os parâmetros escolhidos para cada horizonte foram independentes, embora para alguns modelos foram bem parecidos. O

modelo que apresentou o pior resultado foi $SL3 - 2$ (descontinuidade no loading de inclinação) e, inclusive, foi o mais instável na escolha do parâmetro dentro da amostra.

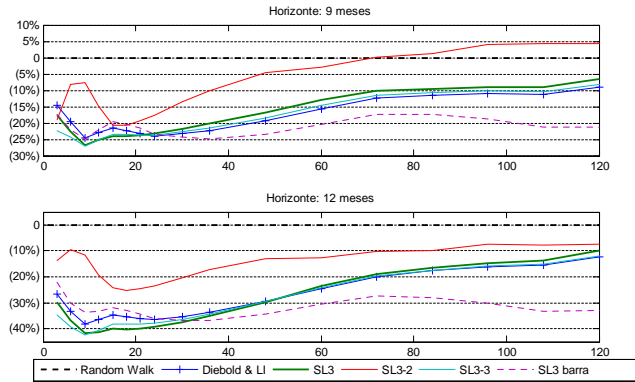


Figura 6b: Percentual de Erros Quadráticos Médios de Previsão dos modelos relativos ao modelo Random Walk para os horizontes de 9 e 12 meses

Outra dimensão que deve ser olhada como medida de robustez dos modelos é o comportamento dos erros de previsão destes modelos para diferentes tamanhos e janelas amostrais. Os resultados obtidos anteriormente eram referentes à um período fora da amostra (Jan/1994 até Dez/2000), totalizando 84 observações. Será usada a métrica descrita na seção 3.4.2 que sintetiza os ganhos na dimensão das maturidades em um escalar definido pela média. A evolução deste escalar será computada sob duas outras dimensões, a primeira variando o tamanho amostral (equação (38)) e a segunda variando a janela de previsão (equação (37)). O primeiro exercício mantém a data inicial do período fora da amostra em Jan/1994 e reduz gradativamente o final da amostra de Dez/2000 a Set/1997, ou seja, a medida de erro relativo será avaliada para amostras com 44 a 84 observações, variando unitariamente. Os resultados para os horizontes de previsão de 3,6,9 e 12 meses são mostrados na Figura 7.

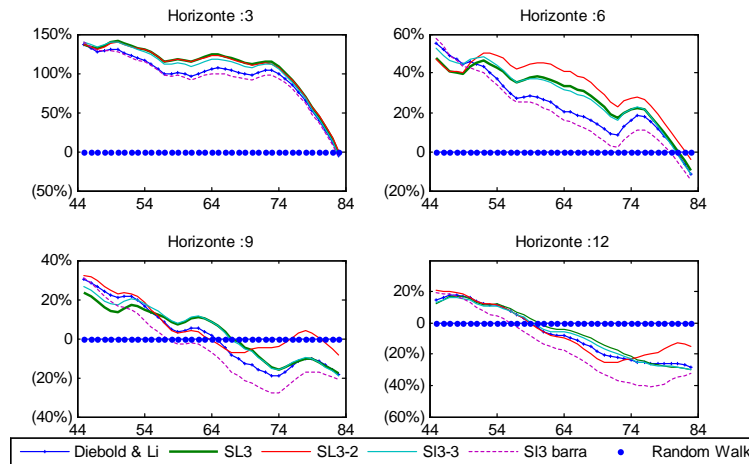


Figura 7: Média de Erros Relativos ao Modelo Random Walk para diferentes tamanhos de amostra. A média dos erros é entre as maturidades. No eixo das abcissas está o tamanho da

amostra. $N = 40$ representa o período de previsão de 1994:01 a 2000:12 e $N = 44$ 1994:01 a 1997:09

Em primeiro lugar, o ponto mais importante é que os modelos são melhores que a previsão Random Walk (nível zero) apenas para os tamanhos de amostra maiores, ou seja, os resultados de boa capacidade preditiva descritos pela literatura com período fora da amostra a partir de Jan/1994 podem não ser válidos ao se usar número de observações menor. O segundo ponto é que, conclusões acerca de comparações entre modelos olhando apenas um tamanho de janela fixo também podem ser inválidas. Basta ver que as curvas se cruzam para alguns modelos, indicando que a comparação relativa é sensível ao tamanho amostral. Assim, incluir uma análise de robustez desta natureza parece ser um procedimento mais adequado. Ao olhar sob esta ótica, o modelo $\overline{SL3}$ aparenta ser o modelo com melhores resultados enquanto o modelo $SL3 - 2$ é o que apresenta mais instabilidade.

O segundo exercício consiste em manter fixo o número de observações em 84 e variar o período da amostra. Para tanto, é necessário que seja feita uma nova escolha de parâmetros para os modelos SLs, para que não haja sobreposição do período fora da amostra que começa em Jan/1994 com o período em que os parâmetros foram escolhidos (Jan/1988 a Dez/1993). Assim, a janela de estimação na escolha dentro da amostra que era de Jan/1982 a Dez/1987 foi deslocada 18 meses¹⁰ (Jul/1980 a Jun/1986) e o período de previsão dentro da amostra também (Jul/1986 a Jun/1992). Os parâmetros encontrados foram:

Modelo	Parâmetro "p" ótimo dentro da amostra (período 1986:07 a 1992:06)			
	h=3	h=6	h=9	h=12
SL3	1.13	1.13	0.54	0.54
SL3-2	1.13	1.42	0.47	0.40
SL3-3	1.16	1.21	1.22	1.16
SI3 barra	0.61	0.68	0.68	0.61

Dados estes parâmetros, o exercício de previsão foi realizado para o a janela com início em (Jul/1992 - Jun/1999), descolando unitariamente (inclui mês novo e retira o último) até (Jan/1994 - Dez/2000), que recupera o período em que foi feito o exercício inicial. Os resultados obtidos são mostrados na Figura 8.

O modelo $\overline{SL3}$, sob esta ótica foi levemente superior ao modelo de Diebold Li. O modelo que teve um destaque foi o $SL3 - 3$. Vale a pena destacar novamente que, sob esta ótica, fica evidente que os resultados de boa capacidade preditiva destes modelos são bem sensíveis ao período em que o exercício é feito. Se todos os modelos fossem testado 2 anos antes, nenhum conseguiria ser melhor em média que a previsão por Random Walk. Evidente que esta conclusão olha a média dos ganhos relativos entre todas as maturidades, o ideal é olhar para a dimensão das maturidades também, ou seja, será que o ganho do modelo $SL3 - 3$ é superior em todas as maturidades? Ou é muito superior em uma maturidade de tal forma que mais do que compensa a perda nas demais? Na seção final será explorada esta dimensão.

¹⁰ Este exercício também foi repetido para 12 meses, onde os resultados principais foram preservados. Pra 24 meses, a escolha dos parâmetros p ficou instável, e o resultado dos modelos melhores sobre a otica da janela de 18 e 12 meses ficaram pra ao Diebold & Li.

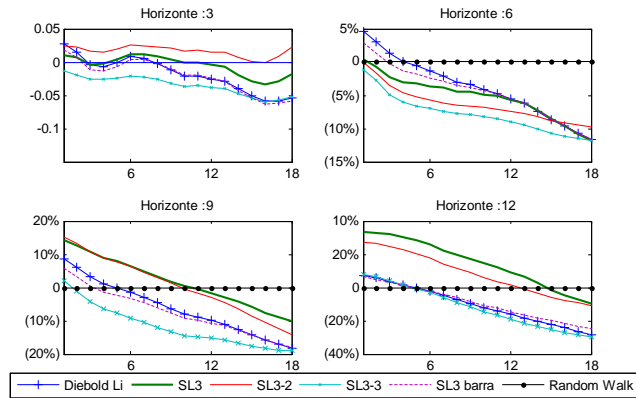


Figura 8: Média de erros Relativos ao modelo Random Walk para amostras em diferentes períodos consecutivos no tempo (janela móvel do período fora da amostra).

Dados estes ingredientes (as 3 dimensões de avaliação) pode-se chegar à algumas conclusões sobre os modelos propostos. Os modelos $SL3-3$ e $\overline{SL3}$ superam de forma consistente os demais modelos, apresentando robustez na dimensão da janela e tamanho amostrais, sendo também robustos na dimensão do horizonte de previsão. Estes resultados sugerem que a segmentação da ETTJ de fato gera uma capacidade preditiva superior aos modelos concorrentes. Entretanto, um cuidado deve ser tomado nesta conclusão pois, para períodos além de 2000, houve aparentemente uma mudança na dinâmica dos erros de previsão. Vale ressaltar que foi proposta uma estrutura geral que gera características desejáveis para uma ETTJ e os testes foram feitos para uma forma funcional particular específica, dada pela equação (28). É possível que haja outras formas funcionais que gerem resultados melhores de previsão, que fica para estudos futuros.

4.5.3 Comparação Direta entre Modelos

Foi visto que uma dimensão de análise envolve as maturidades e a outra a janela das observações fora da amostra, uma forma de olhar estas duas dimensões conjuntamente é gerar uma superfície dos erros relativos (ver equação (36)) à algum benchmark, onde novamente, os valores abaixo de zero representam uma melhoria, com a única diferença que agora o modelo benchmark não é uma reta constante em zero e sim um plano. Um exercício interessante é comparar o modelo de Diebold & Li com $\overline{SL3}$ e $SL3-3$ apresentados na figura 8, expandindo na dimensão das maturidades. Na figura 9 estes três modelos são comparados com a previsão pelo modelo Random Walk, para o horizonte de 6 meses. Nota-se que todos possuem um comportamento semelhante quanto ao erro de previsão, justificado por serem provenientes da mesma forma paramétrica. Superam o random walk significativamente mais no final da amostra. Os modelos $SL3$ conseguem resultados melhores que os do random walk um pouco mais cedo do que o modelo de Diebold & Li.

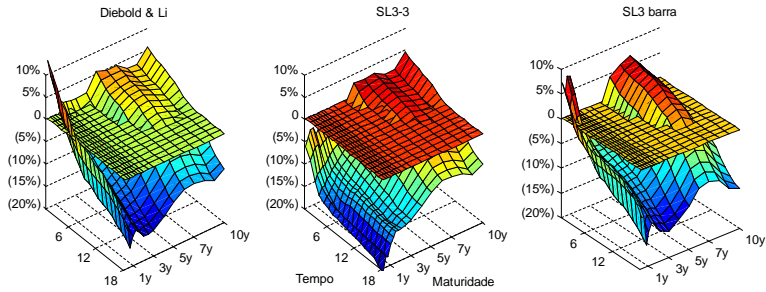


Figura 9: Erros Quadráticos Médios relativos ao modelo Random Walk por maturidade da curva e por período amostral, para o horizonte de 6 meses.

Uma forma mais direta de comparar modelos é tomar um modelo como benchmark. Nas Figura 10a e 10b o modelo de Diebold & Li é usado como benchmark. No horizonte de previsão de 6 meses nota-se que o modelo $SL3 - 3$ é superior em praticamente todas as maturidades ao modelo Diebold & Li (exceto apenas no final do período), com destaque para as maturidades curtas, enquanto que o modelo $SL3$ é superior nas maturidades longas.

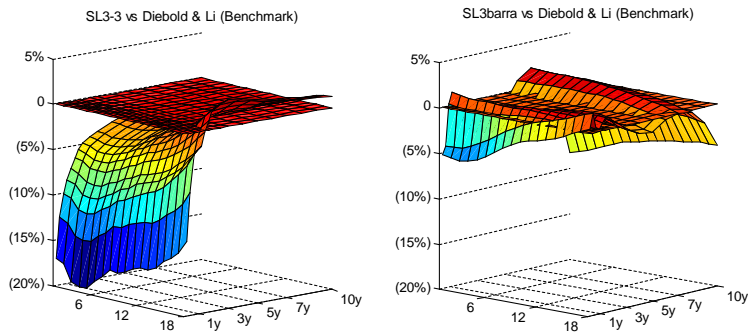


Figura 10a: Erros Quadráticos Médios de Previsão para o horizonte de 6 meses do modelo $SL3-3$ e $SL3$ barra relativos ao modelo de Diebold & Li.

Ao considerarmos o horizonte de 12 meses, o padrão de erro relativo é parcialmente preservado mas os ganhos são menores em razão da previsão pelo modelo de Diebold & Li possuir seus melhores resultados neste horizonte.

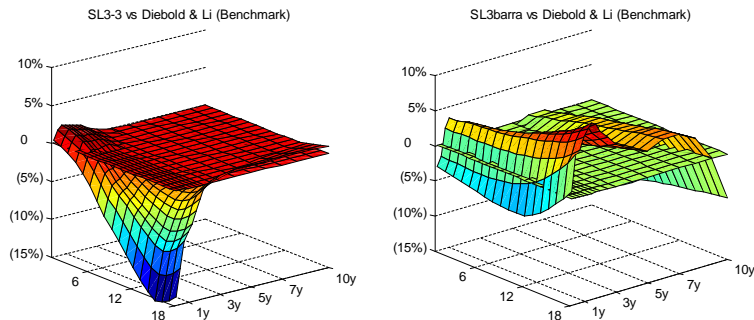


Figura 10b: Erros Quadráticos Médios de Previsão para o horizonte de 12 meses do modelo $SL3-3$ e $SL3$ barra relativos ao modelo de Diebold & Li.

5 Conclusões

Neste trabalho foi proposta uma nova classe de modelos estatísticos para a ETTJ com a vantagem de poder incorporar uma dinâmica segmentada para a curva. Esta segmentação é condizente com a teoria de preferência por Habitat, onde diferentes maturidades possuem demandas próprias, sendo reguladas em parte por arbitradores que negociam principalmente maturidades próximas. Esta estrutura é flexível no sentido de poder gerar o modelo de Diebold & Li como caso particular.

Estes modelos foram usados para prever a ETTJ americanas e foram comparados aos modelos *benchmark* da literatura, Diebold & Li e random walk. Adicionalmente foi proposto um método de avaliação da qualidade de previsões que gera mais robustez nos resultados encontrados e aponta para o possível problema de não se observar determinadas dimensões, uma vez que os resultados são sensíveis às amostras.

Duas variações do modelo SL3 mostraram-se superiores em diversas dimensões, a saber, horizonte de previsão (exceto para 1 mês a frente), maturidades da curva e robustez para amostras distintas. Com estes resultados há evidências que a modelagem da ETTJ com uma estrutura segmentada melhora a capacidade preditiva, dado os ganhos relativos ao modelo random walk. Os modelos $SL3 - 3$ (segmentação no 3º-fator), possuem uma estrutura mais coerente que o $SL3$, mas seus resultados são piores na parte longa da curva. Ressalta-se que os resultados deste trabalho foram obtidos com uma especificação funcional particular, podendo ser alterada de forma a melhorar tanto a qualidade da estimação dos coeficientes como também a robustez na escolha do parâmetro exógeno que determina as características dos *loadings* originais (p).

Vale considerar que algumas variáveis do modelo foram tomadas *ad hoc*, o que dá espaço para pesquisas futuras que podem gerar resultados melhores. Como exemplo as escolhas dos nós que determinam as partições dos segmentos (usou-se o mesmo conjunto de Bowsher&Meeks), o valor do parâmetro λ (usou-se o mesmo escolhido no trabalho do Diebold & Li). Com relação aos métodos de previsão, é possível melhorar a qualidade da estimação dos parâmetros se esta for feita em apenas 1 estágio evitando o acúmulo dos resíduos, usando por exemplo dados em Painel ou Filtro de Kalman.

Outras extensões possíveis são os modelos com 4 *loadings* baseados na estrutura de Nelson-Siegel Extendido, que geram implicações diferentes pois o número de parâmetros passa de $3k$ para $4k$, onde o número de fatores livres passa a depender do número de nós k . Por fim, dado que parte relevante da literatura de previsão de ETTJ considera modelos com variáveis macroeconômicas, é possível também estudar a relação da dinâmica destes segmentos com estas variáveis.

6 Apêndice

Suavização da Previsão do Modelo $\overline{SL3}$

Conforme discutido na seção 4.4.1, a curva prevista pelo modelo $\overline{SL3}$ possui descontinuidade nos nós. A idéia de suavizar a curva basea-se em estimar o funcional latente $f_\gamma(\tau)$ descrito por Bowsher & Meeks (2006), usando o conjunto de taxas previstas pelo modelo $\overline{SL3}$. O funcional latente possui características desejáveis para a ETTJ por ser contínuo e suave.

Seja $X \in \mathbb{R}^M$ um conjunto de taxas, $\gamma \in \mathbb{R}^{k+2}$ um conjunto de taxas latentes que separam o funcional linear f_γ em k segmentos, $W_\tau \in \mathbb{R}^{M \times k+2}$ a matriz determinística¹¹ que mapeia as respectivas maturidades de X (conjunto τ) nas $k+2$ maturidades de γ (junções dos segmentos + maturidade inicial e final) e $\varepsilon \in \mathbb{R}^M$.

$$X = f_\gamma + \varepsilon = W(k, \tau)\gamma + \varepsilon$$

Como este funcional é linear no conjunto de taxas latentes podemos estimar $\hat{\gamma}$ por Mínimos Quadrados Ordinários¹² e, portanto, a estimativa do funcional latente \hat{f}_γ , avaliado em um conjunto de maturidades $\tilde{\tau}$ arbitrário é dada por:

$$\hat{f}_\gamma = W_{\tilde{\tau}}\hat{\gamma} = W_{\tilde{\tau}}(W'_{\tilde{\tau}}W_{\tilde{\tau}})^{-1}W'_{\tilde{\tau}}X = P_{W_{\tilde{\tau}}}X$$

Fazendo o vetor X ser o vetor de previsões dado por $Z^5_{(\tau)}\hat{\theta}^5_{j+h}$ (ver equação (42)), com respectivo conjunto de maturidades τ , obtemos a previsão suavizada do modelo $\overline{SL3}$:

$$\hat{Y}^5_{j+h} = S(Z^5_{(\tau)}F^5_h(C_j)) = P_{W_\tau}Z^5_{(\tau)}F^5_h(C_j)$$

A figura 11 mostra um exemplo desta suavização:

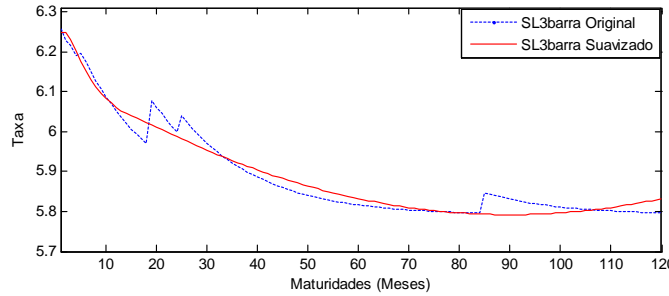


Figura 11: Exemplo de Suavização da Curva gerada pelo modelo $\overline{SL3}$.

¹¹Os detalhes da construção de $W(k, \tau)$ estão em Porier (1973, pp.517-518).

¹²Supondo que a distribuição de ε seja multivariada normal com matriz de covariância diagonal.

Referências

- [1] Ang, A. , Piazzesi, M. 2003. A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomics and Latent Variables. *Journal of Monetary Economics*, 50(4), 745-787.
- [2] Bowsher, C.G. , Meeks, R. , 2006. High Dimensional Yield Curves: Model and Forecasting. Working Paper.
- [3] Cochrane, J.H., Piazzesi, M., 2005. Bond Risk Premia. *The American Economic Review*, 138-160.
- [4] Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of the term structure of interest rates. *Econometrica* 53, 385-407.
- [5] Culbertson, J.,1957. The Term Structure of Interest Rates. *Quartely Journal of Economics*, 71, 485-517.
- [6] Diebold F. X., Li, C., 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics* 130, 337-364.
- [7] Diebold F. X., G.D. Rudebusch, Aruoba S.B., 2006. The macroeconomy and the yield curve. *Journal of Econometrics* 131, 309-338.
- [8] Fama,E. , Bliss, R. 1987. The Information in Long-Maturity Forward Rates. *American Economic Review*, 77(4), pp.680-92.
- [9] Kargin, V. , Onatski, A., 2007. Curve Forecasting by Functional Autoregression. Working Paper.
- [10] Litterman, R., Scheinkman, J., 1991. Common Factors affecting bond returns. *Journal of Fixed Income* 1, 54-61.
- [11] McCulloch, J.H.,1971. Measuring the Term Structure of interest rates. *Journal of Business*, 19-31.
- [12] Monch, E., 2006. Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-Augmented VAR Aproach. Working Paper.
- [13] Nelson, C. R. , A.F. Siegel (1987). Parsimonius Modeling Of Yield Curves. *Journal of Business* 60, 473-489.
- [14] Poirier, D.J., 1973. Piecewise regression using cubic spline. *Journal of the American Statistical Association* 68, 515-524.
- [15] Pooter, M. , 2007. Examining the Nelson-Siegel Class of term Structure Models. Working Paper.
- [16] Vayanos, D., Vila, J. , 2007. A Preference-Habitat Model of Term Structure of Interest Rates. Working Paper.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)