



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

碩士學位 請求論文

指導教授 黃 壽 星

장단기 기억 신경망을 활용한 빅데이터
기반의 수익률 곡선 예측

成均館大學校 一般大學院

퀀트應用經濟學科

宋 柄 憲

碩
士
學
位
請
求
論
文
예측 장단기 기억 신경망을 활용한 빅데이터 기반의 수익률 곡선
2
0
2
1
宋
柄
憲

碩士學位 請求論文

指導教授 黃 壽 星

장단기 기억 신경망을 활용한 빅데이터
기반의 수익률 곡선 예측

Forecasting Yield Curve with Long Short-Term
Memory Network in Data-rich Environment

成均館大學校 一般大學院

퀀트應用經濟學科

宋 柄 憲

碩士學位 請求論文

指導教授 黃 壽 星

장단기 기억 신경망을 활용한 빅데이터 기반의 수익률 곡선 예측

Forecasting Yield Curve with Long Short-Term
Memory Network in Data-rich Environment

이 論文을 經濟學 碩士學位請求論文으로 提出합니다.

2021 年 4 月 日

成均館大學校 一般大學院

퀀트應用經濟學科

宋 柄 憲

이 論文을 宋柄憲의 經濟學
碩士學位 論文으로 認定함.

2021 年 6 月 日

審査委員長

審査委員

審査委員

목차

제 1 장 서론.....	1
1. 연구 배경과 목적.....	1
2. 연구 동기 및 방법.....	2
3. 선행 연구.....	4
제 2 장 이론적 배경 및 연구 설계.....	6
1. Nelson-Siegel 모형군.....	6
1) Dynamic Nelson-Siegel 모형.....	6
2) Macro-Finance Dynamic Nelson-Siegel 모형.....	10
3) h-step Ahead Rolling Window Forecast.....	11
2. Deep Neural Network.....	12
1) Recurrent Neural Network (순환 신경망).....	12
2) Long Short-Term Memory Network (장단기 기억 신경망).....	13
제 3 장 연구 결과.....	16
1. 활용 데이터.....	16
2. 모형 설정 및 예측 결과.....	17
1) Macro-Finance Dynamic Nelson-Siegel 모형.....	17
2) Data-scarce Macro-Finance LSTM 모형.....	18
3) Data-rich Macro-Finance LSTM 모형.....	20
3. 모형 간 비교.....	20
1) 수익률 곡선 비교.....	20
2) RMSE (Root Mean Squared Error).....	25
3) Diebold-Mariano test.....	27
제 4 장 결론 및 향후 과제.....	29
참 고 문 헌.....	31
ABSTRACT.....	34

표목차

- <표 1> Long Short-Term Memory Network 기반 모형의 Cross Validation 결과
- <표 2> 비교 대상 모형별 예측오차(RMSE)
- <표 3> Diebold-Mariano 검정 결과

그림목차

- <그림 1> Dynamic Nelson-Siegel 모형의 파라미터별 만기에 따른 요인계수의 변화
- <그림 2> h-step ahead Rolling Window Forecasting 의 구조
- <그림 3> Recurrent Neural Network 의 구조
- <그림 4> Recurrent Neural Network 의 입출력 구조
- <그림 5> Long Short-Term Memory Network 의 입출력 구조
- <그림 6> MF-DNS 모형의 수익률 곡선 예측 결과
- <그림 7> sMF-LSTM 모형의 수익률 곡선 예측 결과
- <그림 8> rMF-LSTM 모형의 수익률 곡선 예측 결과
- <그림 9> 실제 미 국채 수익률(3 개월물)과 모형별 예측 결과 간 비교
- <그림 10> 실제 미 국채 수익률(6 개월물)과 모형별 예측 결과 간 비교
- <그림 11> 실제 미 국채 수익률(1 년물)과 모형별 예측 결과 간 비교
- <그림 12> 실제 미 국채 수익률(5 년물)과 모형별 예측 결과 간 비교
- <그림 13> 실제 미 국채 수익률(10 년물)과 모형별 예측 결과 간 비교
- <그림 14> 시간에 따른 모형별 예측오차(RMSE)의 변화

논문요약

장단기 기억 신경망을 활용한 빅데이터 기반의 수익률 곡선 예측

최근 기계학습 알고리즘 관련 연구가 진행됨에 따라 경제학 분야에서도 이를 접목하고자 하는 시도가 점차 증가하고 있다. 특히 심층 신경망 관련 연구를 통해 차원의 저주 문제를 극복하고 기존 연구 대비 더 많은 변수를 활용하는 것이 가능함을 보인 바 있다. 이에 본 연구에서는 Long Short-Term Memory Network 를 활용하여 미국 국채 수익률을 예측하는 데에 FRED 의 여러 거시경제변수를 사용하는 것을 시도하였다. 연구 결과 예측 대상 기간 및 사용한 설명변수의 수에 따라 일부 차이가 있으나, 수익률곡선의 표본 외 예측 면에서 LSTM 를 활용한 모형의 예측력이 Macro-Finance Dynamic Nelson-Siegel 모형을 활용한 rolling window forecast 대비 다소 우수한 성과를 보이는 것으로 나타났다.

주제어 : 예측, 수익률 곡선, Nelson-Siegel 모형,

Long Short-Term Memory Network, Macro-Finance 모형

제 1 장 서 론

1. 연구 배경과 목적

수익률 곡선(yield curve)은 채권의 신용 위험(default or credit risk), 유동성 위험(liquidity risk) 등 잔여 만기를 제외한 다른 요인이 모두 동일한 채권의 수익률을 잔존만기에 따라 표시하여 나타낸 것으로 이자율 기간구조(term structure of interest rate)라고도 부른다. 다른 조건이 일정할 때 잔존 만기가 길어질수록 채권 가격의 변동성은 커지고 유동성은 감소하는 경향이 있으므로 수익률 곡선은 일반적으로 우상향하는 형태를 보인다(Campbell, 1995).

이러한 이자율 기간구조는 채권시장 및 파생상품시장의 이해에 중요한 요소이므로 다양한 접근 방법을 통해 이를 이해, 추정, 예측하고자 하는 여러 노력이 이어지고 있다. Nelson-Siegel 모형(Nelson & Siegel, 1987)이나 Arbitrage-Free 모형(Ang & Piazzesi, 2003)부터 시작하여 Nelson-Siegel-Svensson 모형(Svensson, 1994), Macro-Finance Dynamic Nelson-Siegel 모형(Diebold, Rudebusch & Aruoba, 2006), Björk-Christensen 모형(Björk & Christensen, 1999) 등으로 이어지는 각종 경제 이론 기반, 통계적 기반의 연구뿐만 아니라, 최근에는 많은 연구를 통해 빠르게 발전하고 있는 기계학습(machine learning)과 심층신경망(deep neural network) 등의 알고리즘을 활용하는 실험적인 연구까지 확장되고 있다(Berg, 2018; Kožíšek, 2018).

본 연구에서는 이러한 기계학습 분야에서 많이 사용되고 있는 Recurrent neural network(이하 RNN)의 일종인 Long Short-Term Memory Network(이하 LSTM)를 활용하여 이자율 기간구조를 예측하고 그

성과를 분석한다. 이를 위하여 Diebold et al. (2006)에서 제시된 Macro-Finance Dynamic Nelson-Siegel(이하 MF-DNS) 모형을 벤치마크로 활용한다. 이 모형은 Dynamic Nelson-Siegel 모형에 사용되는 세가지 잠재적 요인(수준, 기울기, 곡도)에 다양한 거시경제변수를 더한 모형으로 기존 Dynamic Nelson-Siegel 모형보다 예측력이 더 좋은 것으로 알려져 있다(Diebold et al., 2006; 박원주, 임병인, 전용일, 2013).

구체적으로는 LSTM에 다양한 거시경제변수를 더한 모형을 Macro-Finance Long Short Term Memory Network(이하 MF-LSTM) 모형이라 할 때 이 기계학습 기반의 모형이 기존 MF-DNS 모형 대비 표본 외 예측력을 높이는지 분석을 실시하였다. 또한 기존 MF-DNS 모형과 동일하게 거시경제변수를 3개만 사용한 Data-scarce MF-LSTM 모형(이하 sMF-LSTM)과, 가용 설명변수를 모두 사용한 Data-rich MF-LSTM 모형(이하 rMF-LSTM)을 추정하여 예측력을 비교하였다.

2. 연구 동기 및 방법

다양한 거시변수들이 이자율 기간구조와 상호 영향을 주고 받는 상황에서 이를 추정하기 위한 모형에 어떤 변수를 포함시키거나 제거해야 할지 결정하는 것은 쉽지 않다. 이 때 선택할 수 있는 방법 중 하나는 연구자가 타당하다고 생각하는 이론이나 선행 연구에 근거하여 중요 변수를 선택적으로 결정하는 것이지만, 이자율 기간구조 및 수익률곡선에 대한 다양한 연구와 이론이 존재하는 만큼 앞서 진행된 여러 연구에서도 각각의 추정 모형에 포함되는 거시변수들이 서로 다르다(구병수, 2020). 이처럼 모형 불확실성이 존재하는 상황에서 특정 거시변수들만을 선택적으로 중요하다고 판단하는 것은 바람직하지 않을뿐더러, 실제로는 중요한 변수임에도 모형에는 이를 누락할 위험도 있다. 더욱이 채권시장과 연관된 거시경제변수들 간에는 상관관계가 높기 때문에 특정 변수가 모형에 포함되었는지 여부가 결과를 크게 달라지게 할 수 있다.

이러한 문제를 완화하고 적절한 거시경제변수를 선택하기 위하여 전통적으로는 BIC, AIC, R^2 , C_p , Bayes Factor 등과 같은 계량적인 기준을 토대로 전진 선택법(forward selection), 후진 제거법(backward elimination), 단계적 선택법(stepwise selection) 등을 사용하였다. 하지만 이 경우 검토해야할 설명변수의 수가 늘어나게 되어 최적 모형을 도출하는 것에 계산상 많은 노력이 요구된다(computationally expensive). 다른 방법으로는 특정 모형을 연구자가 자의적으로 결정하는 것이 아니라 잠재적으로 가능한 모든 모형을 조합하여 가중평균함으로써 모형 불확실성까지 추정에 반영하는 베이지안 접근법(Bayesian approach)이다(구병수, 2020). 베이지안은 모형 또한 확률변수라는 시각으로 접근하여 고안된 방법으로, 모형 불확실성까지 추정에 반영하므로 예측 면에서는 최적의 결과를 얻을 수 있다고 알려져 있지만(Madigan & Raftery, 1994), 계산이 복잡하고 많은 시간이 소요된다는 한계점이 있다.

이와 같이 변수 선택이라는 어려움 없이 주어진 모든 변수를 활용하고자 하는 또다른 접근 방법으로 기계학습이 있다. 특히 Barron(1993)의 연구에서는 Deep Neural Network 를 활용할 경우 많은 수의 설명변수를 사용할 때 발생하는 차원의 저주(Curse of Dimensionality)을 극복할 수 있다는 것을 보여주었다.¹ 기존 연구에서 널리 사용되는 선형회귀분석과 같은 비교적 단순한 분석 방법을 활용하는 경우에는 차원의 저주로 인하여 모형의 적합도가 크게 감소하므로, 여러 선행 연구들과 같이 일부 중요한 변수들만 선택하여야 하는 한계가 존재하는 반면 심층 신경망 모형에서는 이러한 문제를 극복할 수 있다는 것이다.

본 연구에서는 각종 선행연구에서 주로 활용되었던 거시경제변수인 기준금리(Federal funds rate), 제조업가동률(manufacturing capacity utilization), 물가상승률(inflation rate) 뿐만 아니라 FRED-MD 에서 제공하고 있는 각종 거시경제변수를 최대한 활용하여 수익률곡선의 예측력을 높일 수 있는지

¹차원의 저주란 데이터 학습을 위해 차원이 증가하게 되면서 전체 학습 데이터 수가 차원의 수보다 절대적 또는 상대적으로 적어지거나, 개별 차원 내에서 학습할 데이터 수가 적어지게(become sparse) 되어 모형의 성능이 저하되는 현상을 지칭

확인하였다. 기존의 MF-DNS 모형에서 사용하는 위 세 가지 거시경제변수만을 활용한 모형의 예측 결과와 여러가지 거시경제변수를 활용하여 수익률곡선을 예측한 rMF-LSTM 모형의 예측 결과를 비교하여, 기계학습을 이자율 기간구조 예측에 적용할 경우 보다 우수한 성과를 얻을 수 있는지 분석하였다.

3. 선행 연구

이자율 기간구조에 대한 연구는 방법론에 따라 상당히 다른 양상을 보이고 있다. 첫째는 거시경제변수를 명시적으로 고려하지 않고 직접 관측할 수 없는 잠재적 요인(latent factor)만을 활용하는 접근법이 있고, 둘째는 거시경제변수를 채권 수익률의 결정 변수로 활용하여 인플레이션 등 실물경제가 채권수익률에 미치는 영향력에 초점을 두고 접근한 방법이 있다.

Ang & Piazzesi(2003)는 선형 기간구조(affine term structure) 모형에 거시경제변수를 고려한 거시-금융 기간구조 모형(macro-finance model)을 최초로 제안하였다. 인플레이션, 실업률, 산업생산지수 등을 활용할 경우 잠재적 요인만을 고려했을 때보다 예측 성과가 더욱 뛰어남을 보였으나, 거시경제변수가 수익률곡선에 영향을 받지 않고 독립적으로 결정된다고 가정하고 분석을 진행하였다는 한계점이 있다.

Diebold et al. (2006)은 이러한 Ang & Piazzesi(2003)의 한계를 극복하기 위해 Dynamic Nelson-Siegel model 을 제시하였다. 거시경제변수와 수익률곡선의 양방향 관계를 가정하고 분석을 진행하였으며, 그 결과 거시경제변수가 잠재적 요인에 미치는 영향이 더 크다는 결과를 제시하였다. 이들의 연구에서는 이자율 기간구조 추정 시 Nelson-Siegel model 에 동태적 접근법을 접목하였으며 거시경제변수로 제조업가동률(manufacturing capacity utilization), 물가상승률(inflation rate), 기준금리(Federal funds rate)를 활용하였다. 해당 연구에서 제시된 Dynamic Nelson-Siegel model 은

모형의 추정 및 확장이 쉽고, 표본 내 적합도와 표본 외 예측력이 모두 우수하여 이후 수익률 곡선 추정에 많이 활용되고 있다. Rudebusch & Wu (2004)는 Affine Arbitrage-free model 을 기반으로 이자율 기간구조를 추정하는 방안을 제시하고, 거시경제변수를 활용하여 분석을 실시하였다. Ang, Piazzesi & Wei (2006)는 앞서 Ang & Piazzesi (2003)의 한계점을 극복할 수 있는 선형 기간구조 모형을 제시하여 수익률곡선의 거시경제 예측력에 관한 연구를 진행하였다. 선형 기간구조 모형은 무재정차익(arbitrage-free) 조건을 도입하는데 유리하다는 장점이 있으나 Duffie, Garleanu & Pedersen (2002)의 연구에서 지적된 바와 같이 전역 최적화 계수를 찾기 어렵고 표본 외 예측력이 낮다는 한계가 있다.

최근 국내 채권시장이 성장함에 따라 채권 관련 호가, 수익률 등의 거래 데이터가 축적되기 시작하면서 국내 채권시장을 대상으로 한 연구도 다수 진행되고 있다. 윤재호 (2011)는 Dynamic Nelson-Siegel model 에 경기동행지수, 소비자물가지수, 콜금리, 원달러환율을 고려한 Structural VAR(SVAR)을 활용하여 수익률 곡선을 추정한 바 있다. 또한 강규호, 오형석 (2015)은 소규모 개방경제인 한국의 여건을 고려하여 해외 요인을 다수 반영하고자 하였다. 이에 미국 장기금리가 포함된 선형 기간구조 모형을 활용하여 한국 수익률곡선에 구조적 변화가 발생하였음을 보였다. 정현석, 강규호 (2017)는 미국 장단기 금리를 모두 고려한 Dynamic Nelson-Siegel model 으로 한국의 수익률곡선의 동태성을 저변동성 상태와 고변동성 상태로 구분하여 접근하였으며 연구 결과 고변동성 상태일 때 한국의 장기금리는 미국의 장기금리에 크게 영향을 받고 있음을 보였다.

한편 기계학습을 활용한 국내 채권시장의 이자율 기간구조에 대한 분석은 그리 많지 않다. 외국 사례에서는 Berg (2018)가 Gaussian Restricted Boltzmann Machine 을 활용하여 기계학습 알고리즘을 스웨덴 국채 이자율의 기간구조를 분석하기 위한 연구에 사용하였다. 국채 수익률의 잠재적 요인(latent factor)만을 활용했다는 점, 하이퍼 파라미터의 미세조정(hyper-parameter fine tuning)을 위한 validation 과정에서의 데이터 누수(data leakage)가 있었다는 한계가 있으나 수익률 곡선 추정 및

예측에 있어 기계학습을 접목하는 방법의 가능성을 보여주었다는 점에서 의의가 있다. 또한 Kožíšek (2018)은 미국 국채 선물 시장의 일별 수익률 자료를 활용하여 AR, VAR, LSTM으로 각각 분석을 실시하고 그 결과를 비교하였다. 이 연구는 LSTM 알고리즘을 이자율 기간구조 분석에 활용했다는 점에서 의의가 있으나 잠재적 요인(latent factor) 외의 거시경제변수를 활용하지 않았다는 점에서 한계가 있다.

제 2 장 이론적 배경 및 연구 설계

1. Nelson-Siegel 모형군

1) Dynamic Nelson-Siegel 모형

Nelson & Siegel(1987)이 제시한 수익률곡선 함수는 만기가 τ 인 선도이자율(forward rate)이 중근해(equal roots)를 갖는 이차차분방정식의 해라고 정의한 것에서 도출되며, 선도이자율을 식 (1)과 같이 정의하면 잔존만기가 τ 인 수익률은 식 (2)와 같이 정의할 수 있다. 그 결과 이 식으로부터 수익률곡선 함수인 식 (3)을 도출할 수 있다.

$$f(t) = \beta_1 + \beta_2 e^{-\lambda\tau} + \beta_3 \lambda\tau e^{-\lambda\tau} \quad (1)$$

$$y(\tau) = \frac{1}{\tau} \int_0^\tau f(u) du \quad (2)$$

$$y(\tau) = \beta_1 + \beta_2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_3 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \quad (3)$$

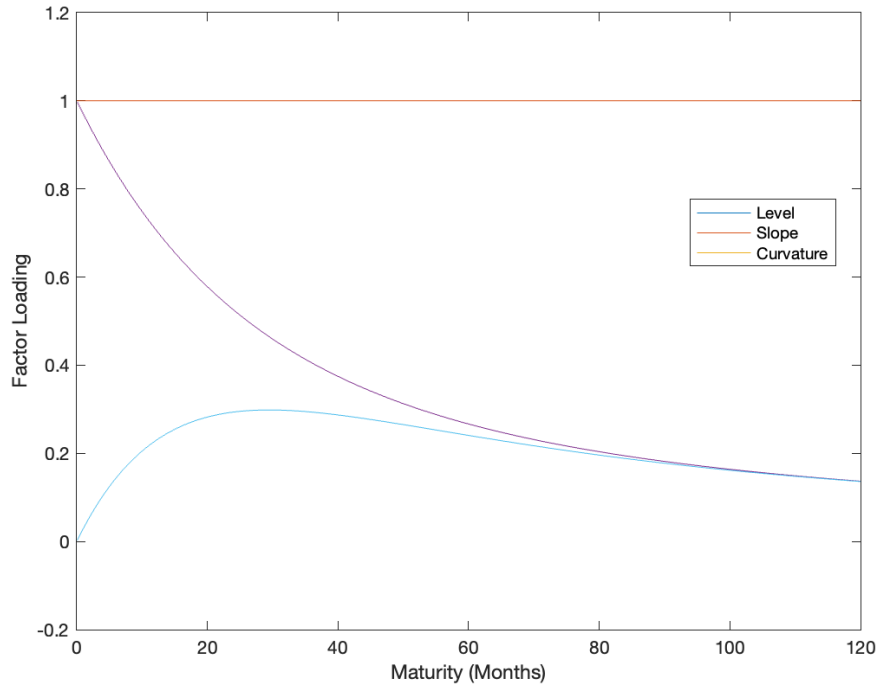
Diebold & Li (2006)는 식 (3)을 동태적으로 해석하여 식 (4)와 같은 수익률곡선의 함수를 유도하였다.

$$y_t(\tau) = \beta_{1t} + \beta_{2t} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_{3t} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \quad (4)$$

식 (4)의 Dynamic Nelson-Siegel model 에서 λ 값은 요인계수(factor loading) 값이 만기가 중간인 경우 최대값을 갖게되며, 이 때 λ 는 할인이 지수적으로 감소하는 정도를 나타내는 파라미터(parameter for exponential decay rate)로 정의된다. λ 값이 클수록 할인이 크고, λ 값이 작을수록 할인이 작다. 따라서 λ 값이 작을수록 식 (4)의 모형은 상대적으로 장기 이자율에 대해 더 높은 적합도를 갖게 된다. 식 (4)를 보면 β_{1t} 의 요인계수 값은 1로 만기에 따라 변화하지 않으므로 장기적 요소(long-term factor)로 고려할 수 있다. 즉 β_{1t} 의 변화는 잔존 만기에 관계 없이 모든 채권 수익률에 동일한 영향을 미치므로 수익률곡선의 수준(level) 요인으로 볼 수 있다. 또한 β_{2t} 의 요인계수 값인 $\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right)$ 는 값이 1부터 시작하여 만기가 증가함에 따라 0에 수렴하는 모습을 보이므로, β_{2t} 의 변화는 단기 수익률에 더 큰 영향을 준다고 볼 수 있다. 이러한 특성을 감안하였을 때 β_{2t} 는 수익률곡선의 기울기(slope) 요인으로 볼 수 있다. 마지막으로 β_{3t} 의 요인계수 값인 $\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right)$ 는 값이 0에서 시작, 만기가 증가함에 따라 점증하다가 일정 시점 이후 값이 다시 0으로 수렴하는 모습을 보인다. 즉 β_{3t} 의 변화는 단기 수익률이나 장기 수익률보다는 중기 수익률에 더 큰 영향을 미치므로 수익률곡선의 곡도(curvature) 요인으로 볼 수 있다.

다음 그림 1은 각 파라미터 별로 만기 변화에 따른 요인계수 값의 변화를 나타낸 것으로, 이와 같은 요인계수 값의 특징에 따라 Nelson-Siegel model 을 3 요인 모형(3-factor model)으로 해석할 수 있다.

<그림 1> Dynamic Nelson-Siegel 모형의 파라미터별 만기에 따른 요인계수의 변화



그러나 Nelson-Siegel model 에서 수익률곡선을 설명하는 수준, 기울기, 곡도 세 가지의 잠재적 요인들은 실제로는 관측이 불가능한 변수이다. 따라서 모형을 추정하는 시점에서는 상태공간모형(state space model)으로 표현하여 Kalman filter 를 활용하여 추정할 수 있다. Diebold et al. (2006)의 표기 방식을 따라 위 식 (4)를 아래 식 (5)와 같이 표현하여, Nelson-Siegel model 의 각 파라미터를 이자율 기간구조의 수준, 기울기, 곡도 요인으로 해석할 수 있다.

$$y_t(\tau) = L_t + S_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + C_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \quad (5)$$

식 (5)를 수준, 기울기, 곡도 세 가지 요인의 상태공간모형으로 표현하면, 우선 1 차 벡터자기회귀모형(VAR, Vector Autoregressive model)을 따르는 상태방정식(state equation)으로 이들을 식 (6)과 같이 표현할 수 있으며, 측정방정식(measurement equation)은 식 (7)과 같이 표현할 수 있다.

$$\begin{pmatrix} L_t - \mu_L \\ S_t - \mu_S \\ C_t - \mu_C \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} L_{t-1} - \mu_L \\ S_{t-1} - \mu_S \\ C_{t-1} - \mu_C \end{pmatrix} + \begin{pmatrix} \eta_t^L \\ \eta_t^S \\ \eta_t^C \end{pmatrix} \quad (6)$$

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1 - e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1 - e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1 - e^{-\lambda\tau_2}}{\lambda\tau_1} & \frac{1 - e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1 - e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1 - e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} + \begin{pmatrix} \epsilon_t(\tau_1) \\ \epsilon_t(\tau_2) \\ \vdots \\ \epsilon_t(\tau_N) \end{pmatrix} \quad (7)$$

행렬 형태로 표현된 식 (6)과 식 (7)을 간단히 정리하여 Nelson-Siegel 모형을 식 (8)과 같이 상태공간모형으로 표현할 수 있다.

$$\begin{aligned} f_t &= Af_{t-1} + (I - A)\mu + \eta_t \\ y_t &= \Lambda f_t + \epsilon_t \\ \begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix} &\sim WN \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 \\ 0 & H \end{pmatrix} \right] \end{aligned} \quad (8)$$

$$\text{where } A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad f_t = \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix},$$

$$\text{and } \Lambda = \begin{pmatrix} 1 & \frac{1 - e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1 - e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1 - e^{-\lambda\tau_2}}{\lambda\tau_1} & \frac{1 - e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1 - e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1 - e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix}$$

식 (8)에서의 H 행렬은 만기별 채권 수익률의 오류간에 서로 상관관계가 없음을 가정하여 대각 행렬(diagonal matrix)로 가정한다. 반면 Q 행렬은 수익률곡선의 세 잠재요인들 간에 서로 상관관계가 존재할 수 있으므로 대각 행렬이 아닌 것으로 가정한다.

본 연구에서 식 (8)을 추정하기 위하여 Kalman filter 를 사용하였다. 해당 모형에서 추정하여야 할 파라미터는 상태방정식(state equation)인 A 행렬의 원소, 상태변수의 평균, 관측방정식 λ 행렬에 포함되어 있는, Q 행렬과 H 행렬의 원소들이다. 당기의 수익률과 그 전기의 정보가 주어진 상태에서 상태변수에 대한 반복적인 예측은 식 (9)와 같이 이루어지며, 이 때 Mean Square Error 행렬(MSE matrix)은 식 (10)과 같이 계산할 수 있다.

$$f_{t|t-1} = E_{t-1}[f_t] = Af_{t-1} + (I - A) \quad (9)$$

$$\Sigma_{t|t-1} = E_{t-1}[(f_t - f_{t|t-1})(f_t - f_{t|t-1})'] = A\Sigma_{t-1}A' + Q \quad (10)$$

관측방정식을 이용하면 조건부 예측오차와 공분산은 식 (11), 식 (12)와 같으며 이를 이용한 교정방정식은 식 (13)과 식 (14)와 같다. 이 때 앞선 내용과 같이 H 행렬은 대각 행렬로 가정하며, A 행렬과 Q 행렬은 AR(1)을 따른다는 가정 하에 추정을 진행한다.

$$v_t = y_t - \Lambda f_{t|t-1} \quad (11)$$

$$Var(v_t) = F_t = \Lambda \Sigma_{t|t-1} \Lambda' + H \quad (12)$$

$$f_{t|t} = E_t[f_t] = f_{t|t-1} - f_{t|t-1} \Lambda' F_t^{-1} \Lambda \Sigma_{t|t-1} \quad (13)$$

$$\Sigma_t = E_t[(f_t - f_{t|t})(f_t - f_{t|t})'] = \Sigma_{t|t-1} - \Sigma_{t|t-1} \Lambda' F_t^{-1} \Lambda \Sigma_{t|t-1} \quad (14)$$

2) Macro-Finance Dynamic Nelson-Siegel 모형

Diebold, Rudebusch & Aruoba (2006)의 연구에서는 Nelson-Siegel 모형을 확장하여 거시경제변수와 수준, 기울기, 곡도 간의 양방향적인(bi-directional) 관계를 고려하여 모형을 추정하였다. 그러나 이들의 연구 결과 수준, 기울기, 곡도와 같은 잠재적 요인이 거시경제변수에 미치는 영향은 통계적으로 유의하지 않다. 또한 박원주 등(2013)에 의하면 잠재적인 요인이 거시경제변수의 영향을 받아 변화하지만, 이들 요인이 거시경제에 미치는 영향력은 미미함을 보였다. 이에 따라 본 연구에서는 단방향적인(uni-directional) 관계를 가정하고 추정하였으며, 이 경우 거시경제변수를

상태방정식(state equation)에 외생변수로서 추가하여 모형을 식 (15)와 같이 단순하게 표현할 수 있다.

즉, 식 (8)에서 상태방정식만을 일부 변형시킬 뿐 관측방정식(observation equation)은 동일하다.

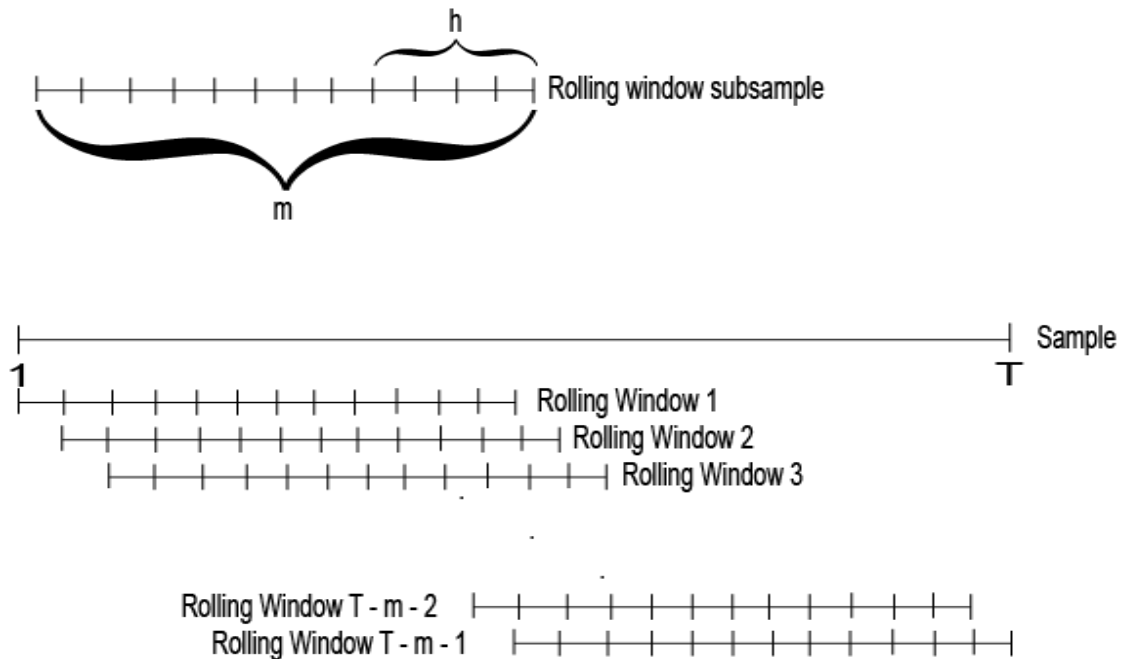
$$\begin{pmatrix} L_t - \mu_L \\ S_t - \mu_S \\ C_t - \mu_C \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} L_{t-1} - \mu_L \\ S_{t-1} - \mu_S \\ C_{t-1} - \mu_C \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} \Delta FFR_{t-1} \\ \Delta CPI_{t-1} \\ \Delta MCU_{t-1} \end{pmatrix} + \begin{pmatrix} \eta_t^L \\ \eta_t^S \\ \eta_t^C \end{pmatrix} \quad (15)$$

거시경제변수로 활용한 자료는 Federal Fund Rate(ΔFFR_{t-1}), Consumer Price Index(ΔCPI_{t-1}), Manufacturing Capital Utilization(ΔMCU_{t-1})이다. 식 (15)에 표기한 것과 같이 각각의 자료는 그 전기 대비 증가율을 활용하였다.

3) h-step Ahead Rolling Window Forecast

다음 <그림 2>는 h-step ahead rolling window forecast 기법을 시각화한 것이다. 총 T 기간의 표본 자료(sample) 중 m 기간의 부표본(subsample)을 활용하여 총 $(T - m - 1)$ 회에 걸쳐 다음 h 기간 동안의 값을 예측한다. 이러한 방법을 통해 지나치게 먼 과거의 자료가 과도한 영향을 미치지 않도록 하는 동시에 새로운 정보를 적절히 반영하여 가까운 미래 시점을 예측할 수 있다. 본 연구에서는 여러 기간에 걸친 모형 간 예측력을 비교하기 위하여 MF-DNS 모형에 one-step ahead rolling window forecasting 기법을 접목한 예측을 실시하였다.

<그림 2> h-step ahead Rolling Window Forecasting 의 구조



출처: MathWorks Help Center

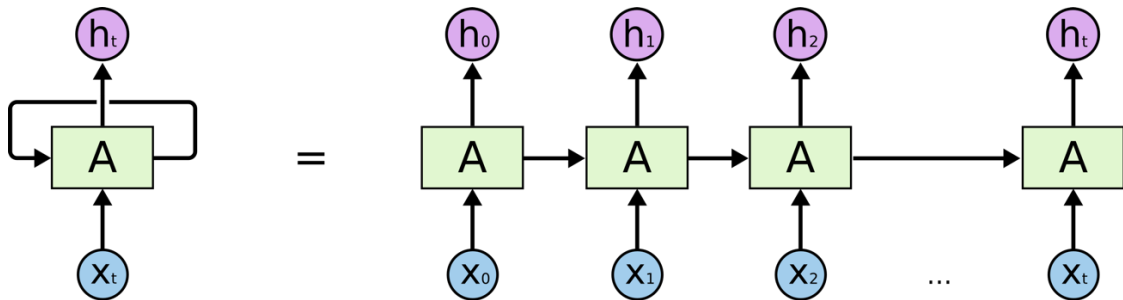
2. Deep Neural Network

1) Recurrent Neural Network (순환 신경망)

분석에 활용한 물가상승률 등의 거시경제변수는 과거의 상태가 현재의 상태에 지속적인 영향을 미치는 순차적 데이터(sequential data)이다. 일반적인 회귀분석 기반의 기계학습 알고리즘 – 선형 회귀분석, 서포트 벡터 머신(SVM, support vector machine), 합성곱 신경망(CNN, convolutional neural network) 등 –은 순차적 데이터를 분석하는 데에 일정한 한계가 있어 주로 횡단면 자료(cross-sectional data)를 활용하여 분석을 실시하게 된다. 하지만 거시경제와 관련된 많은 변수들은 과거의 정보를 충분히 활용하여야만 함의를 적절히 분석하고 해당 변수에 대한 예측력을 높일 수 있다. 따라서

이러한 한계를 극복하기 위하여 출력값을 다시 입력으로 피드백(feedback)할 수 있는 재귀적(recursive) 구조를 갖는 신경망을 구성하여 활용한다.

<그림 3> Recurrent Neural Network의 구조



출처: Colah (2015)

RNN 은 내부에 순환하는 구조가 존재하는 신경망(neural network)으로 정보가 지속되는 것을 돕는다(Hochreiter & Schmidhuber, 1997). <그림 2>는 일반적인 순환 신경망 구조를 시각화한 것으로, 각 네트워크는 다음 네트워크에 정보를 전달하며 사슬과 같은 구조를 형성한다.

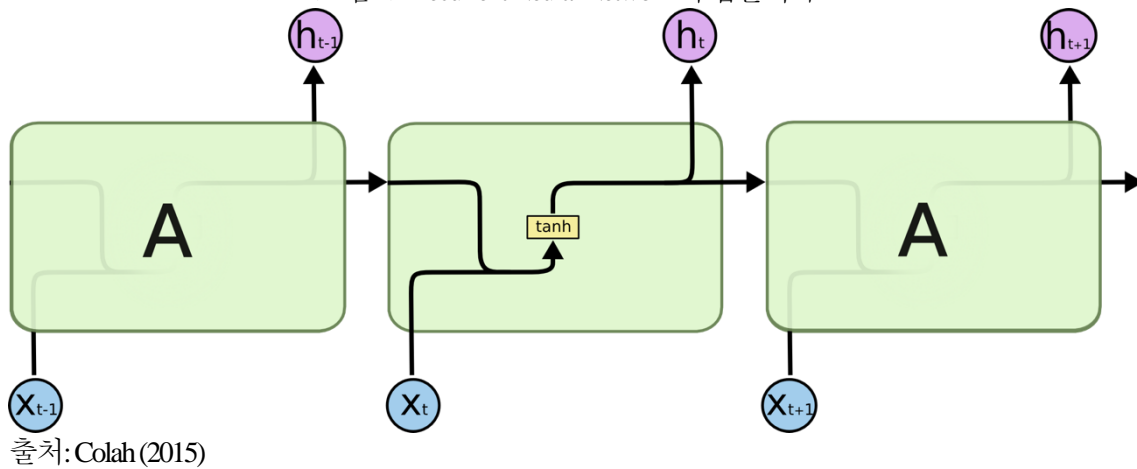
2) Long Short-Term Memory Network (장단기 기억 신경망)

장기 시계열(time series-) 또는 순차형 자료(sequential data)를 분석하는 상황과 같이 정보의 흐름이 확대되어야 하는 경우 일반적인 순환 신경망으로는 정확한 예측이 어렵다는 한계가 있다. 향후 예측을 위해서는 과거 시계열 또는 순차형 자료를 신경망에 유지하기 위한 '기억'이 네트워크 내에서 오래도록 지속될 필요가 있으나, 기억이 네트워크 내에서 지속되기 어려운 구조가 순환 신경망의 주요 한계점으로 지적되었다.

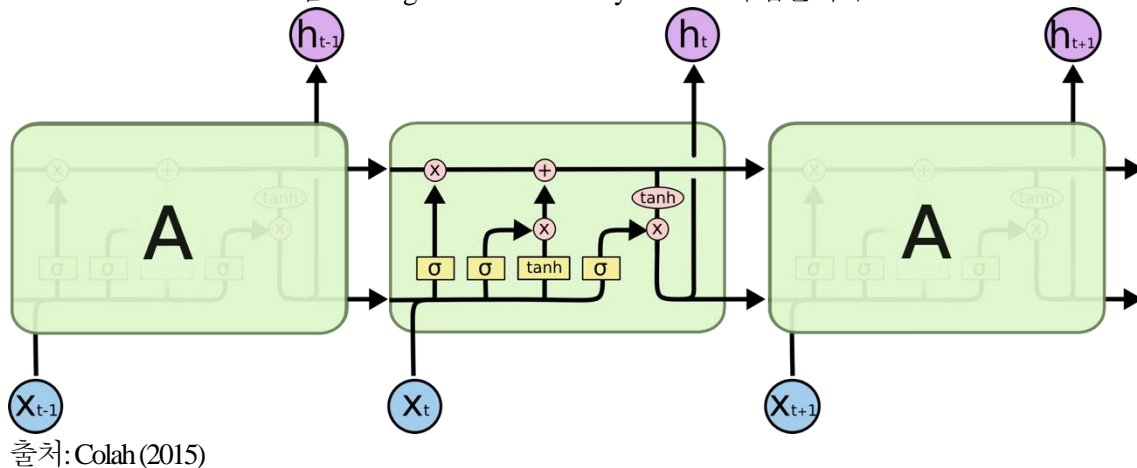
이러한 문제점은 vanishing gradient problem 이라고 하며, 각 뉴런 간 네트워크의 가중치(weight)가 지속적으로 갱신(update)되는 과정에서 각 가중치에 대한 갱신 비율(gradient)에 1 보다 작은 값이 지속적으로 곱해지면서 결국 그 값이 0으로 수렴하게 되는 현상이다. 이와 같은 문제가 발생하는 경우 먼 과거의 정보 및 상태는 현재 시점의 학습에 영향을 미치지 못하게 되어 해당 정보가 매우

중요한 정보일지라도 네트워크 내에서의 영향력이 소실된다. 과거의 정보 및 상태가 비교적 오랫동안 기억되어야 하는 장기 의존성이 필요한 경우 전통적인 RNN 모형으로는 적절한 분석 및 예측이 어려운 한계가 있다. 이러한 기억의 '장기 의존성'을 가능케 한 신경망 모형이 LSTM이다. LSTM 은 일반적인 RNN 보다 내부구조가 복잡하여 비교적 오랫동안 정보를 '기억'할 수 있다(Hochreiter & Schmidhuber, 1997; Karpathy, Johnson & Fei-Fei, 2015).

<그림 4> Recurrent Neural Network의 입출력 구조



<그림 5> Long Short Term Memory Network의 입출력 구조



<그림 5>는 LSTM의 구조를 시각화한 것이다(Colah, 2015). 학습이 장기에 걸쳐 오래 계속되는 경우 비교적 초기에 학습한 내용에 대하여 기억이 유실되는 *vanishing gradient problem*을 개선하고자 하였다.<그림 4>에서 볼 수 있는 RNN의 input, forget, output 구조에 <그림 5>과 같이 gate cell을 더하여 그 결과값을 조정함으로써 RNN의 한계를 극복하고자 하였다. Input gate는 새로이 입력되는 정보의 저장 여부를, forget gate는 이전 상태 정보의 저장 여부를, output gate는 갱신된 cell의 출력값을 제어하는 역할을 한다.

LSTM을 포함한 대부분의 신경망 모형 분석에서는 모형의 크기를 결정하는 cell size, 데이터 학습을 하는 횟수인 epoch, 모형의 과적합(overfitting)을 막기 위한 dropout rate, 한 번에 처리할 데이터 인스턴스(data instance)의 크기를 정하는 batch size 등 각종 초매개변수(hyper-parameter)를 적절히 설정할 필요가 있다. 이러한 hyper-parameter는 연구자 임의로 설정하는 것 또한 가능하나, 설정에 따라 모형의 적합성이나 예측력 등의 결과가 크게 달라지게 되므로 주의를 기울여야 한다. 본 연구에서는 예측오차가 가장 작은 모형을 찾을 수 있도록 grid search 방식으로 cross validation을 수행하여 최적 hyper-parameter를 구한 다음², 학습 데이터(training set)로 신경망을 학습시켜 모형을 추정하였다. 본 연구에서는 LSTM 모형의 예측력을 높이기 위하여 one-step ahead recursive forecasting (또는 iterated forecasting)을 수행하였다.

해당 예측은 다음과 같이 진행한다. 학습 데이터를 통해 추정한 모형에 t 기의 설명변수를 사용하여 $(t + 1)$ 기를 예측하고, 그 예측 결과와 평가 데이터(test set)의 차이를 학습하여 모형의 신경망을 갱신(update)한다. 그 뒤 갱신된 모형에 $(t + 1)$ 기의 설명변수를 사용하여 $(t + 2)$ 기를 예측하고, 실제 데이터와의 차이를 학습하여 신경망을 재차 갱신한다. 위 과정을 반복하여 전체 평가 데이터 기간에 대한 예측값을 구하는 것을 one-step ahead recursive forecasting이라고 한다.

² 해당 cross-validation의 결과는 <표 1>을 참조

제 3 장 연구 결과

1. 활용 데이터

본 연구는 Federal Reserve Economic Data(FRED)의 자료를 활용하여 분석을 진행하였다. McCracken & Ng (2016)은 134 종의 거시경제변수에 대한 정보를 1959 년 1 월부터 최근 시점까지의 월(분기) 단위의 시계열 자료로 정리하였고, 이를 FRED Monthly(Quarterly) Database(이하 FRED-MD, FRED-QD))라고 명명하여 온라인에 공개 및 지속적으로 업데이트하여 많은 연구자들과 다양한 데이터 수요자들에게 도움을 주고 있다. 본 연구에서는 해당 자료에 수록된 최대한 많은 거시경제변수를 활용하고자 하였으나 몇몇 변수들의 경우 시계열이 불완전하거나 중도에 새로이 추가된 것이 있어 해당 자료는 제외하고 분석을 실시하였다. 또한 분석 대상으로 활용한 시계열 변수들의 정상성(stationarity)을 확보할 수 있도록 McCracken 등(2016)이 제시한 데이터 변환 방법(Data Transformation Code)을 기반으로 차분, 로그 변환 등 필요한 변환을 복합적으로 실시하였다. 이와 같은 변환 과정을 거쳐 1959 년 3 월부터 2019 년 12 월까지 730 개월, 113 개의 거시경제변수 및 5 개의 국채 수익률 자료를 활용하여 분석 및 예측을 진행하였다.³

³ 데이터에 대한 상세한 설명과 그 변환 과정은 McCracken & Ng(2016)의 부록을 참조

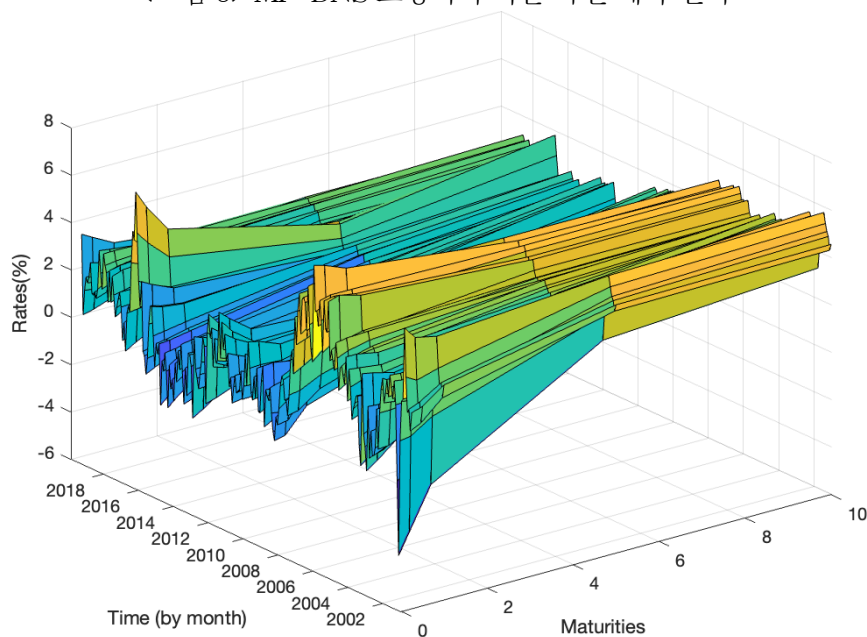
2. 모형 설정 및 예측 결과

1) Macro-Finance Dynamic Nelson-Siegel 모형

앞서 2 장에서 언급한 총 3 가지의 거시경제변수(Federal Fund Rate, Manufacturing Capital Utilization, Consumer Price Index)를 이용하여 MF-DNS 모형을 추정하였다. 그 다음 one-step ahead rolling window forecasting 방법을 활용하여 1 개월 후의 만기별 미국 국채 수익률에 대한 예측을 실시하였고 그 결과는 아래 그림 5 와 같다.

본 연구에서는 730 기(개월)의 표본 자료 중 502 기간의 부표본을 활용하여 다음 1 기간의 수익률 곡선을 예측하는 것을 총 228 회 반복하여($h = 1, m = 502, T = 730$), 2001 년 1 월부터 2019 년 12 월까지의 자료를 예측하였다.

<그림 6> MF-DNS 모형의 수익률 곡선 예측 결과

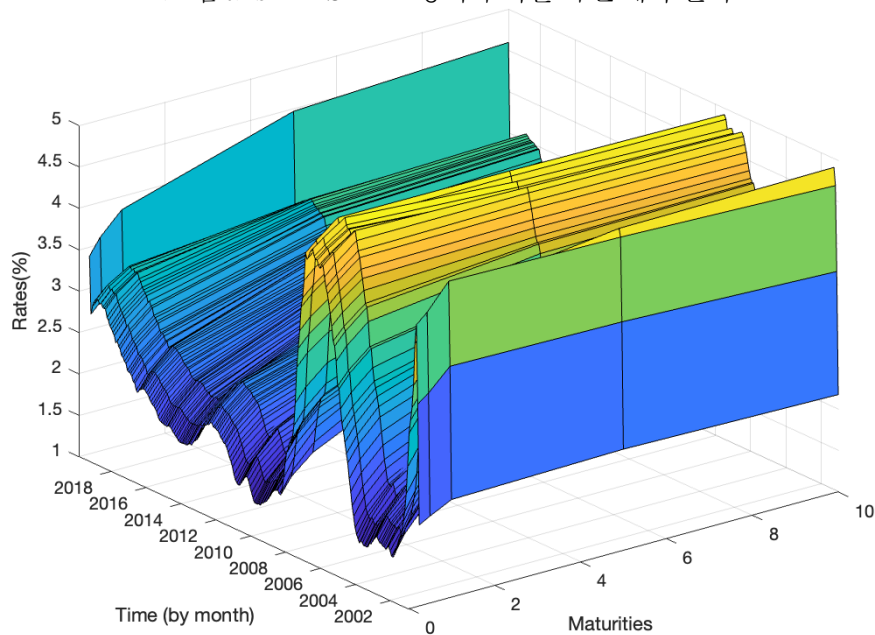


2) Data-scarce Macro-Finance LSTM 모형

MF-DNS 모형과 같이 3 개의 거시경제변수와 과거의 국채 수익률을 이용하여 sMF-LSTM 모형을 설정하고 예측을 실시하였다. 총 730개월의 데이터를 학습(training-), 검정(validation-), 평가 데이터(test set)로 나누어 cross validation 을 통해 모형을 탐색하였으며, 이 때 정보 누수(data leakage)를 막기 위하여 시간순서에 따라 분리하여 학습 및 검정을 진행하였다⁴. 다음 <표 1>에 정리한 cross validation 결과에 따라 각 모형별로 예측오차인 RMSE 를 가장 작게 하는 hyper-parameter 를 설정하여 신경망 학습 및 예측을 실시하였다.

sMF-LSTM 모형은 <표 1>에 제시된 cross validation 결과에 따라 hyper-parameter 를 각각 epoch 2,000, batch size 20, drop-out rate 0.5 로 설정하고, one-step ahead recursive forecasting 방법으로 예측을 진행하였다. 예측 결과는 다음 <그림 7>과 같다.

<그림 7> sMF-LSTM 모형의 수익률 곡선 예측 결과



⁴ 총 730개월의 데이터를 학습 데이터(1959년 3월~1990년 12월, 382개월), 검정 데이터(1991년 1월~2000년 12월, 120개월), 평가 데이터(2001년 1월~2019년 12월, 228개월)로 나누어 분석을 실시함

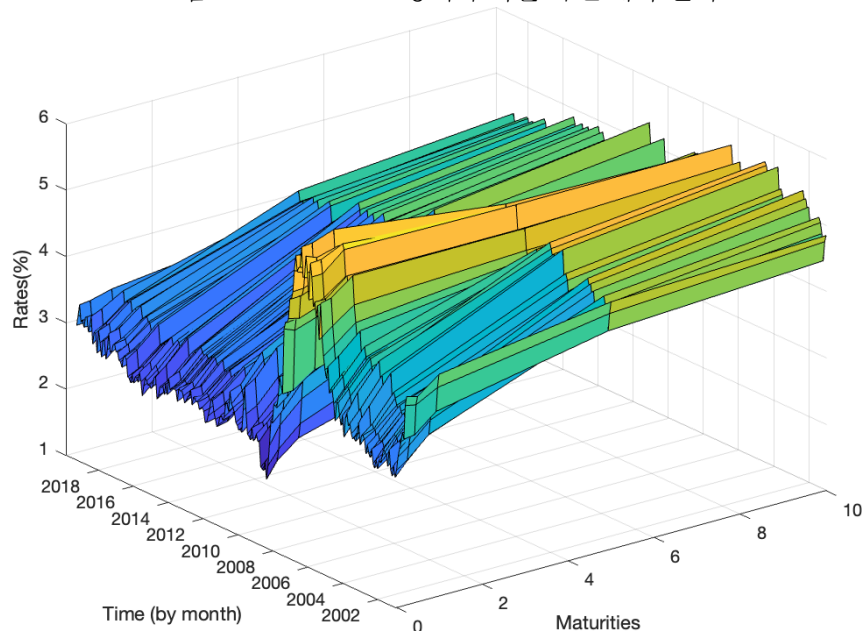
<표 1> Long Short Term Memory Network 기반 모형의 Cross Validation 결과

Epoch	Batch Size	Drop-out Rate	sMF-LSTM		rMF-LSTM	
			Validation RMSE	Validation Loss	Validation RMSE	Validation Loss
500	5	0.2	6.4913	3.6031	6.2483	19.5208
500	5	0.4	5.4067	3.2884	6.5972	21.7618
500	5	0.5	4.7625	3.0863	3.0027	4.5082
500	10	0.2	4.7542	3.0836	3.6513	6.6660
500	10	0.4	6.4369	3.5880	6.6811	22.3188
500	10	0.5	4.2455	2.9139	3.6602	6.6985
500	20	0.2	6.4958	3.6044	8.8591	39.2423
500	20	0.4	7.7545	3.9381	3.6023	6.4882
500	20	0.5	7.4389	3.8572	2.5981*	3.3752*
1000	5	0.2	5.7091	3.3791	4.4200	9.7684
1000	5	0.4	6.3074	3.5517	2.8756	4.1344
1000	5	0.5	8.5929	4.1456	4.7219	11.1482
1000	10	0.2	5.3516	3.2716	5.3863	14.5062
1000	10	0.4	6.4248	3.5846	3.2127	5.1606
1000	10	0.5	8.5641	4.1386	3.3763	5.6996
1000	20	0.2	4.8188	3.1044	6.6716	22.2548
1000	20	0.4	8.1646	4.0409	3.8111	7.2622
1000	20	0.5	6.9185	3.7198	5.5072	15.1645
2000	5	0.2	4.6281	3.0424	3.3877	5.7382
2000	5	0.4	8.7691	4.1879	3.8109	7.2615
2000	5	0.5	8.0231	4.0058	3.6018	6.4865
2000	10	0.2	6.1667	3.5119	2.9491	4.3486
2000	10	0.4	5.5689	3.3373	4.1108	8.4494
2000	10	0.5	7.7802	3.9447	3.0685	4.7078
2000	20	0.2	4.8659	3.1196	3.0709	4.7154
2000	20	0.4	5.0874	3.1898	2.6939	3.6286
2000	20	0.5	3.8051*	2.7587*	4.2753	9.1389

3) Data-rich Macro-Finance LSTM 모형

sMF-LSTM 모형에서 활용한 3 개의 거시경제변수와 과거 국채 수익률뿐만 아니라 활용 가능한 113 개의 거시경제변수를 모두 사용하여 rMF-LSTM 모형을 설정하고 예측을 실시하였다. <표 1>의 결과에 따라 cross validation 을 통해 hyper-parameter 를 각각 epoch 500, batch size 20, drop-out rate 0.5 로 설정하고 분석을 진행하였다. 별도로 언급하지 않은 부분은 sMF-LSTM 과 동일한 조건 및 방법으로 예측을 실시하였다. 예측 결과는 아래 <그림 8>과 같다.

<그림 8> rMF-LSTM 모형의 수익률 곡선 예측 결과



3. 모형 간 비교

1) 수익률 곡선 비교

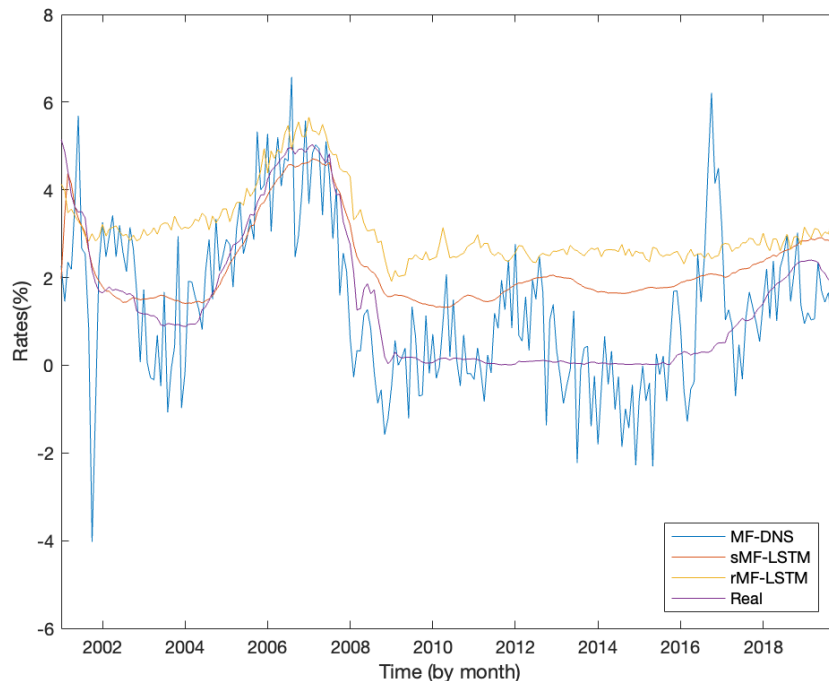
모형 간 예측력 비교를 위하여 다음 <그림 9>, <그림 10>, <그림 11>에 각각 3 개월 만기, 6 개월, 1 년 만기의 국채의 실제 수익률과 예측 수익률을 도식하였다. 상대적으로 국채 만기가 단기인 경우 MF-

DNS 모형은 시기와 관계없이 실제 금리의 수준(level)과는 가장 유사한 예측 결과를 보이고 있으나, 실제 수익률 대비 예측값의 변동 폭이 지속적으로 크게 계산되어 그 안정성이 떨어진다는 한계가 있다.

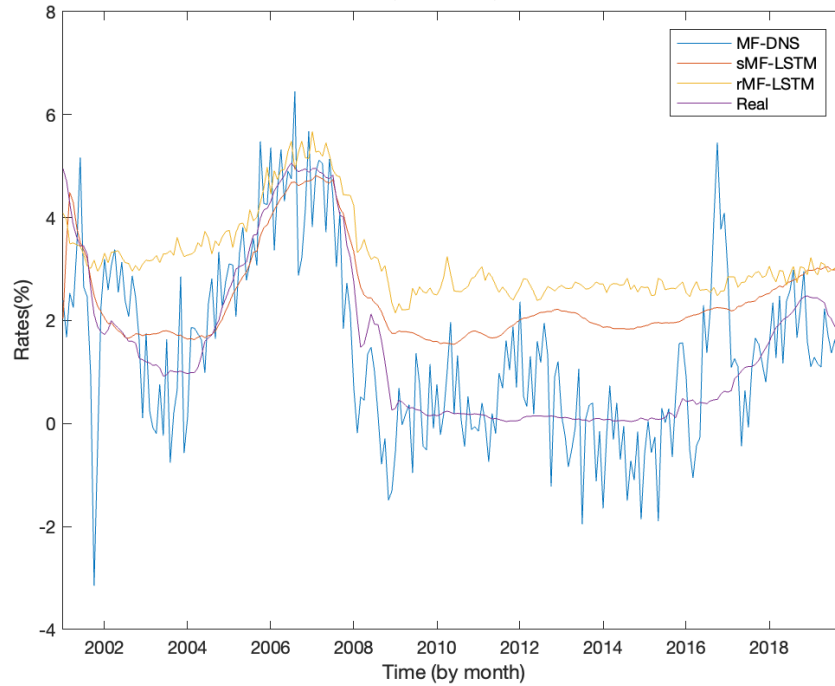
sMF-LSTM 모형은 2008 년 금융위기 이전까지 실제 수익률의 변화 추이(dynamics)와 가장 유사한 예측 결과를 보이고 있다. 2008 년 이후에도 실제 수익률의 변화 정도는 유사하게 예측하였으나, 다만 미 연준의 양적완화 실시 이후로 0%에 가깝게 유지되는 금리의 수준(level)과는 지속적으로 1~2%p 가량의 차이를 보였다.

rMF-LSTM 모형은 실제 수익률 대비 지속적으로 높은 수준(level)을 갖는 수익률 예측 결과를 보이고 있으며, MF-DNS 모형, sMF-LSTM 모형과 비교하였을 때 실제 수익률의 움직임을 잘 반영한다고 보기 어렵다.

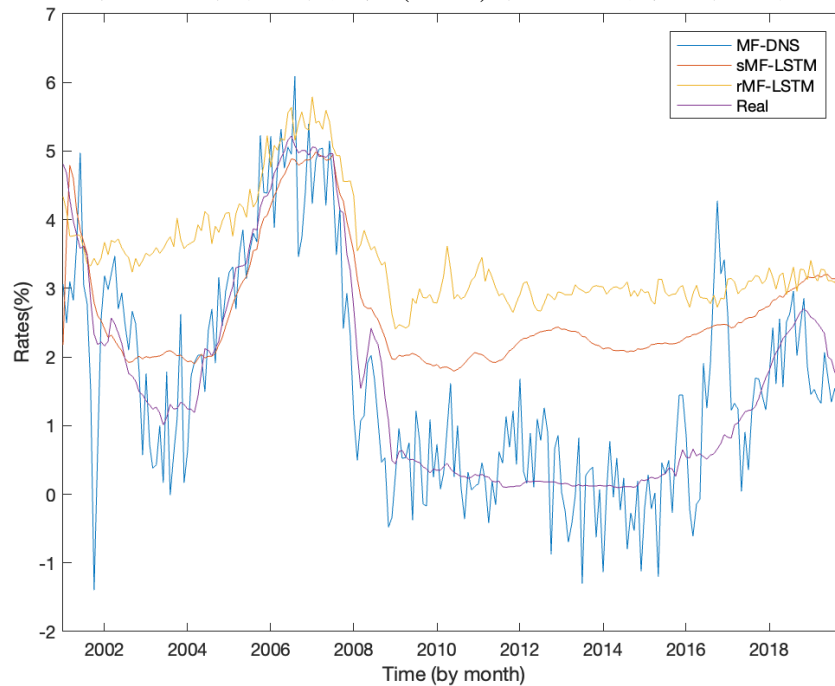
<그림 9> 실제 미 국채 수익률(3개월물)과 모형별 예측 결과 간 비교



<그림 10> 실제 미 국채 수익률(6개월물)과 모형별 예측 결과 간 비교



<그림 11> 실제 미 국채 수익률(1년물)과 모형별 예측 결과 간 비교

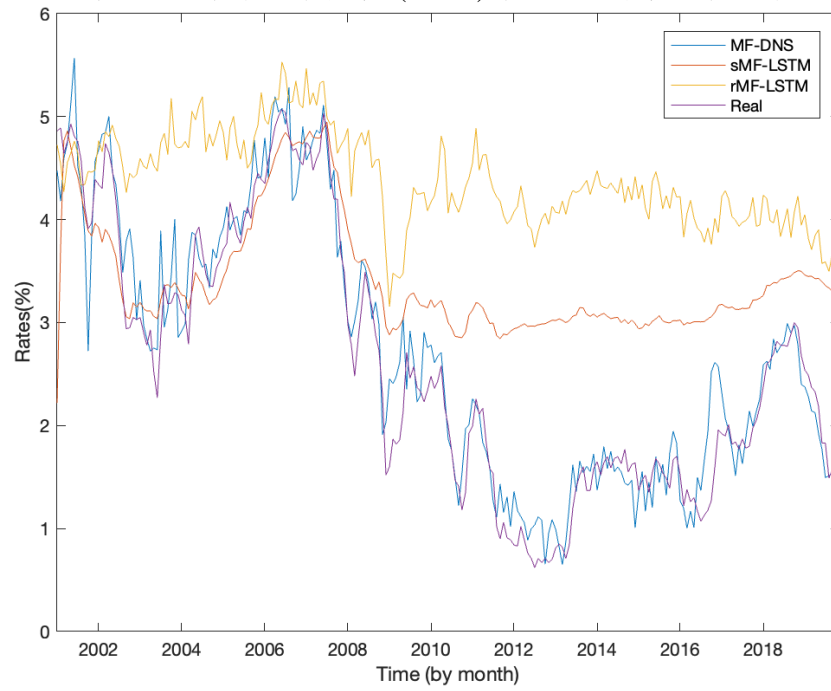


다음 <그림 12>, <그림 13>은 각각 5년, 10년 만기의 국채 수익률과 예측값을 비교하여 도시한 것이다. 앞서 단기채권의 수익률을 예측한 것과 달리 MF-DNS 모형의 예측이 실제 수익률과 상당히

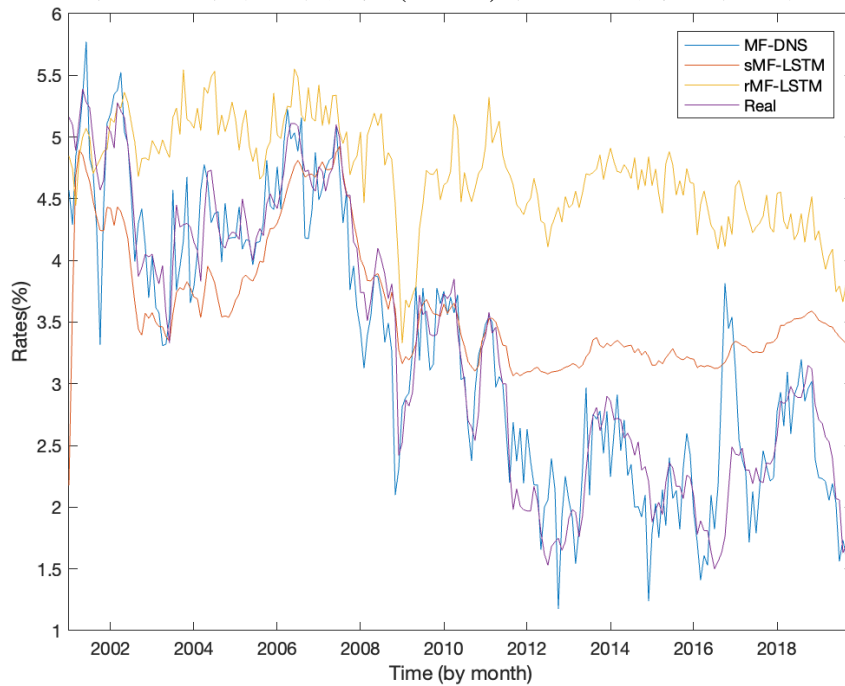
유사한 모습을 보인다. 채권은 만기가 길어짐에 따라 수익률의 변동성이 커지는 경향이 있는데(Campbell, 1995), 이에 따라 상대적으로 변동성이 큰 MF-DNS 모형의 예측값과 실제 수익률 간의 격차가 더욱 줄어들게 되었다.

sMF-LSTM 모형은 단기채권 수익률을 예측한 것과 같이 2008 년 이전의 실제 수익률 변화를 비교적 잘 예측하고 있다. 다만 만기 증가에 따라 국채 수익률의 변동성이 커지는 것에 비하여 sMF-LSTM 모형의 장기채권 수익률 예측 결과의 변동 폭이 크지 않아 실제 수익률과의 격차가 더 크게 발생하였다. rMF-LSTM 모형은 단기채권 수익률을 예측했을 때와 동일하게 실제 수익률과 모형 예측값이 작지 않은 차이를 보이므로 sMF-LSTM 모형 대비 실제 수익률의 변화를 잘 예측한다고 보기 어렵다.

<그림 12> 실제 미 국채 수익률(5년물)과 모형별 예측 결과 간 비교



<그림 13> 실제 미 국채 수익률(10년물)과 모형별 예측 결과 간 비교



이와 같이 예측 대상 기간에 따라 모형의 예측력 차이가 발생하는 원인은 학습 가능한 데이터의 존재 여부의 차이로 보인다. 글로벌 금융위기로 인하여 2009 년 3 월부터 실시된 미 연준의 양적 완화(quantitative easing, QE), 제로금리 정책(Zero Interest Rate Policy, ZIRP) 채택 등 전례 없는 새로운 통화정책 충격(monetary policy shock)은 모형 설계 및 학습 시점에 주어진 과거 데이터만으로는 학습이 불가능하므로 LSTM 모형의 예측 결과는 실제 수익률곡선과 차이가 발생하게 된다. 새롭게 주어진 거시경제변수와 실제 국채 수익률 정보를 활용하여 학습을 진행하더라도 해당 통화정책 충격을 반영하는 정보의 수는 정책 충격 이전의 정보 대비 그 수가 적어 예측에 미치는 영향이 상대적으로 작다. 이로 인해 통화정책 충격 이후 MF-LSTM 모형의 예측력은 그 이전과 비교하였을 때 더 낮아지게 된다. 이는 다음 <그림 14>와 <표 2>에서 다시 확인할 수 있다.

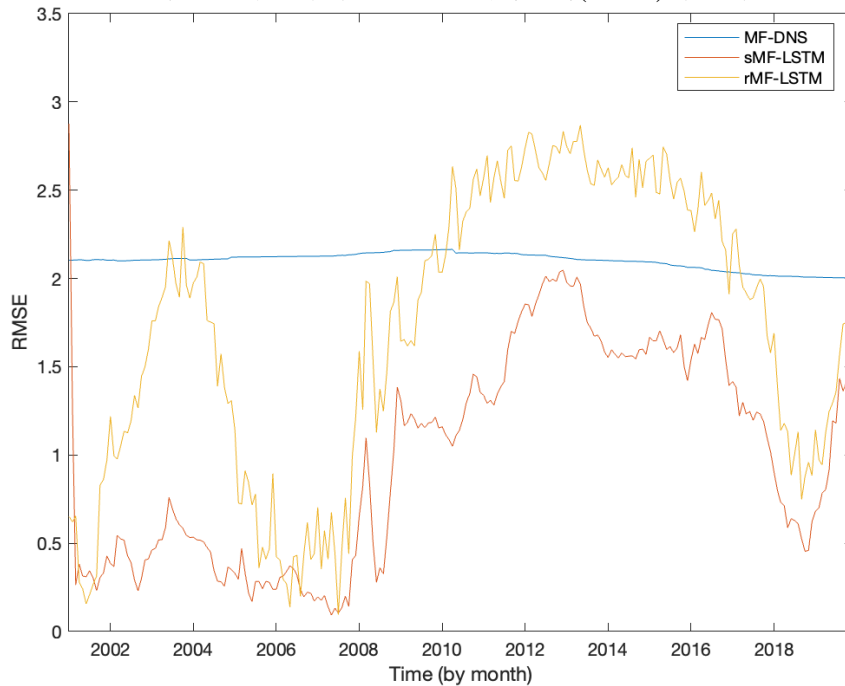
2) RMSE (Root Mean Squared Error)

표본 외 예측에서 각 모형들 간의 예측 성과를 비교하기 위하여 RMSE (Root Mean Squared Error)를 비교하였다. 각 만기별 수익률에 대한 h 기 후 시점 예측값의 RMSE 는 다음 식 (16)과 같이 계산한다.

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=T_R+1}^{T_p} (y_{t+h}(\tau) - \widehat{y_{t+h}}(\tau))^2} \quad (16)$$

이 때, T_p 는 표본 외 예측기간 중 h 기 이후 실제 수익률과 예측 수익률의 차이를 구하는데 사용한 기간으로, 본 연구에서는 2001 년 1 월부터 2019 년 12 월까지의 1 개월 후 시점($h = 1$)의 수익률에 대한 예측을 실시하였다. 각 모형별 예측오차는 아래 <표 2>와 같이 sMF-LSTM, rMF-LSTM, MF-DNS 순으로 커지며 전체 시계열에 대한 RMSE 의 추이는 아래 <그림 14>와 같다.

<그림 14> 시간에 따른 모형별 예측오차(RMSE)의 변화



<표 2> 비교 대상 모형별 예측오차(RMSE)

모형별 RMSE	MF-DNS	sMF-LSTM	rMF-LSTM
전체 기간	2.101	1.005	1.734
2001~2007	2.115	0.393	1.010
2008~2014	2.134	1.422	2.350
2015~2019	2.034	1.278	1.884

<그림 14>와 <표 2>를 보면 시기에 따라 LSTM 기반 모형의 예측력이 크게 달라짐을 확인할 수 있다. 특히 sMF-LSTM 모형은 2008년 이전의 예측력이 매우 뛰어나지만, 미 연준의 양적완화 정책 기간 중에는 그 예측력이 상대적으로 저하되는 것을 볼 수 있다. 2014년 10월말 양적완화 정책 종료 후 미국 내 경제 회복세와 함께 2016년부터 시작된 금리 인상 기간을 거치며 모형의 예측력이 다시 높아지는 모습을 보이기도 하였다. rMF-LSTM 모형은 RMSE만을 단순 비교하는 경우에는 MF-DNS

대비 예측력이 뛰어난 것으로 볼 수 있으나, 정확한 비교를 위하여 그 차이가 통계적으로 유의한지 Diebold-Mariano test 를 통해 이를 살펴볼 필요가 있다.

3) Diebold-Mariano test

각 모형의 예측오차 또한 추정치이므로, 비교 대상 모형 간 예측오차의 크기를 단순히 비교하는 것이 아니라 가설검증 과정을 거쳐 그 우월성을 판단하여야 한다. 이를 위하여 Diebold & Mariano (2002)가 제시한 Diebold-Mariano test(이하 DM 검정)를 실시하여 기준 모형과 비교 모형 간 예측오차 차이가 통계적으로 유의한지 검정한다. 본 연구에서의 표본 외 예측 구간이 228 기(개월)이므로, Harvey, Leybourne & Newbold (1997)가 소표본 문제 극복을 위해 제시한 Modified DM 검정이 아닌 일반적인 DM 검정을 실시하였다. DM 검정에 사용한 귀무가설 및 대립가설은 아래와 같다.

$$\begin{cases} H_0: \bar{d} = 0 \\ H_1: \bar{d} > 0 \end{cases}$$

$$where \bar{d} = \frac{1}{T_p} \sum_{t=1}^T (e_{t+h|t}^1)^2 - (e_{t+h|t}^2)^2$$

비교 대상인 두 모형의 예측오차 차이의 평균의 분산의 추정량인 $\widehat{V}(\bar{d})$ 은 식 (17)과 같이 나타낼 수 있으며, 여기서 추정된 DM 검정통계량은 식 (18)과 같이 나타낸다(Mariano, 2002). 여기서 $\widehat{V}(\bar{d})$ 는 \bar{d} 의 분산을 나타내고, \bar{d} 는 비교 대상인 두 모형의 MSE 차이의 평균, $\widehat{\gamma}_k$ 는 k 기 시차 간의 자기공분산을 나타낸다.

$$\widehat{V}(\bar{d}) = \frac{1}{T_p} \left[\widehat{\gamma}(0) + 2 \sum_{k=1}^{h-1} \widehat{\gamma}_k \right] \quad (17)$$

$$DM = \left(\frac{1}{\widehat{V}(\bar{d})} \right)^{\frac{1}{2}} \times \bar{d} \quad (18)$$

$$\widehat{Y}_k = \frac{1}{T_p} \sum_{t=T_R+1}^{T_p} (d_t - \bar{d})(d_{t-k} - \bar{d}) \quad (19)$$

본 연구에서는 총 세 모형을 비교하므로 총 3 회의 DM 검정을 실시하였다. 검정 결과는 아래 <표 3>과 같다. sMF-LSTM 모형은 전 기간에 걸쳐 다른 두 모형 대비 뛰어난 예측력을 보이고 있다. rMF-LSTM 모형은 2008 년~2014 년중 기간에는 MF-DNS 모형 대비 오히려 나쁜 예측력을 보였으며, 2015 년~2019 년중 기간의 두 모형 간 예측력 차이는 통계적으로 유의하지 못하다는 것을 확인할 수 있다⁵.

<표 3> Diebold-Mariano 검정 결과

DM 검정통계량 (p-value)	(1) MF-DNS (2) sMF-LSTM	(1) MF-DNS (2) rMF-LSTM	(1) rMF-LSTM (2) sMF-LSTM
전체 기간	33.465*** ($< 2.2\text{e-}16$)	4.544*** ($4.602\text{e-}06$)	19.167*** ($1.428\text{e-}49$)
2001~2007	42.208*** ($3.655\text{e-}58$)	18.834*** ($5.209\text{e-}32$)	6.158*** ($1.259\text{e-}08$)
2008~2014	18.495*** ($1.765\text{e-}11$)	-5.515 (1.000)	28.333*** ($9.739\text{e-}45$)
2015~2019	16.13*** ($1.324\text{e-}16$)	1.083 (0.014)	10.725*** ($1.088\text{e-}15$)

⁵ 두 모형간 예측력 차이 검정(two-sided DM test) 결과는 DM=1.083, p-value=0.2832으로, two-sided DM test의 귀무가설인 $\bar{d} = 0$ 을 기각할 수 없음

제 4 장 결론 및 향후 과제

본 연구는 FRED-MD 에 수록되어 있는 1959 년 1 월부터 2019 년 12 월까지의 113 종의 거시경제변수를 기반으로 Long Short-Term Memory Network 를 활용하는 모형을 설정하는 경우 기존 Diebold et al. (2006)에서 제시한 Macro-Finance Dynamic Nelson-Siegel(MF-DNS) 모형과 비교하여 더욱 우수한 이자율 기간구조 예측 모형을 설정할 수 있는지 조사하였다. 비교 기준이 되는 MF-DNS 모형을 추정한 뒤 이와 동일한 거시경제변수(제조업가동률, 물가상승률, 기준금리)만을 사용한 Data-scarce Macro-Finance LSTM(sMF-LSTM) 모형, FRED-MD 에 수록된 거시경제변수를 최대한 활용한 Data-rich Macro-Finance LSTM(rMF-LSTM) 모형을 각각 설정하여 예측을 실시하고 이를 비교하였다.

본 연구의 의의는 다음과 같다. 첫째, LSTM 기반 모형을 사용하는 경우 MF-DNS 모형보다 우수한 예측력을 보이는 것을 확인하였다. 특히 rMF-LSTM 모형의 경우 전 예측기간에 걸쳐 가장 우수함을 보이는 것을 통해 동일한 수의 변수를 사용하는 경우 LSTM 을 사용하는 것이 기존 MF-DNS 모형 대비 미국 국채 수익률 곡선 예측에 더욱 유용함을 보였다. 둘째, 단순히 심층 신경망 모형을 사용하는 것만으로는 차원의 저주, 즉 과적합으로 인한 성능 저하 문제를 피하기 어렵다는 것을 확인하였다. 세 가지의 거시경제변수만을 활용한 sMF-LSTM 모형이 전 예측 기간에 걸쳐 rMF-LSTM 모형 대비 더 낮은 예측오차를 보였으며, rMF-LSTM 모형과 MF-DNS 모형을 비교하였을 때에도 rMF-LSTM 모형이 더욱 우수하다고 할 수 없다. 이는 Barron (1993)의 연구 결과로부터 추측한 것과는 반대되는 결과로, LSTM 의 적용만으로는 차원의 저주를 피하기 어려움을 확인할 수 있다.

본 연구는 추후 다음 세 가지 측면에 대하여 보완하는 것을 통해 더욱 우수한 예측 모델을 추정할 수 있을 것으로 예상된다. 첫째, 김도완 (2011)의 연구와 같이 3, 6, 12 개월 등 본 연구에서 확인한 1 개월 후 시점에 대한 예측보다 더 긴 시계의 국채 수익률 예측에 있어서도 LSTM 모형의 성과가 MF-DNS 모형 대비 우수한지 추가적인 확인이 필요하다. 둘째, Kuo & Mallick approach (1998), Gibbs Variable Selection (Dellaportas, Karlis & Xekalaki, 1997), Stochastic Search Variable Selection (George & McCulloch, 1993)과 같은 베이지안 변수 선택 방법과의 교차 적용을 통해 본 연구에서 제시한 sMF-LSTM 모형 대비 더욱 뛰어난 예측력을 가진 Data-rich LSTM 모형을 설정할 수 있는지 추가적인 조사가 필요하다. 셋째, 과거 자료를 통해 학습할 수 없었던 새로운 충격(shock)이 발생하는 경우 이를 보정하거나 빠르게 학습할 수 있는 방법에 대한 추가적인 연구가 필요하다. 추후 연구에서 이러한 한계점을 보완한다면 더욱 우수한 예측력의 수익률 곡선 예측 기법을 구현할 수 있을 것으로 기대된다.

참 고 문 헌

- 강규호, 오형석. (2015). 미국 장기시장금리 변동이 우리나라 금리기간구조에 미치는 영향 분석 및 정책적 시사점. **BOK 경제연구**.
- 구병수. (2020). 베이지안 변수선택을 이용한 한국 수익률곡선 추정. **BOK 경제분석**, 26(1).
- 김도완. (2011). 2 국가 Nelson-Siegel 모형을 이용한 해외요인이 한국 채권시장에 미치는 영향분석 및 예측. **한국은행 경제분석**, 17(3), 89-130.
- 박원주, 임병인, 전용일. (2013). Macro-Finance 모형을 이용한 국고채 수익률 추정 및 예측. **한국경제연구**, 31(2), 187-210.
- 윤재호. (2011). SVAR 를 이용한 거시. 금융 기간구조 모형 분석. **금융경제연구**, 466.
- 정현석, 강규호. (2017). 미국 시장금리가 우리나라 수익률곡선에 미치는 비대칭적 영향 분석. **경제학연구**, 65(4), 159-202.
- Ang, A., & Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary economics*, 50(4), 745-787.
- Ang, A., Piazzesi, M., & Wei, M. (2006). What does the yield curve tell us about GDP growth? *Journal of Econometrics*, 131(1-2), 359-403.
- Barron, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information theory*, 39(3), 930-945.
- Berg, M. (2018). Modeling the Term Structure of Interest Rates with Restricted Boltzmann Machines.
- Björk, T., & Christensen, B. J. (1999). Interest rate dynamics and consistent forward rate curves. *Mathematical Finance*, 9(4), 323-348.

- Campbell, J. Y. (1995). Some lessons from the yield curve. *Journal of economic perspectives*, 9(3), 129–152.
- Colah. (2015). Understanding LSTM Networks. Retrieved 2021-05-28 from <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Dellaportas, P., Karlis, D., & Xekalaki, E. (1997). Bayesian analysis of finite poisson mixtures. *Athens University of Economics and Business, Greece*.
- Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2), 337–364.
- Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1), 134–144.
- Diebold, F. X., Rudebusch, G. D., & Boragan Aruoba, S. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, 131(1–2), 309–338. <https://doi.org/10.1016/j.jeconom.2005.01.011>
- Duffie, D., Garleanu, N., & Pedersen, L. H. (2002). Securities lending, shorting, and pricing. *Journal of Financial Economics*, 66(2–3), 307–339.
- George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423), 881–889.
- Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting*, 13(2), 281–291.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735–1780.
- Karpathy, A., Johnson, J., & Fei-Fei, L. (2015). Visualizing and understanding recurrent networks. *arXiv preprint arXiv:1506.02078*.
- Kožíšek, J. (2018). Forecasting Term Structure of Government Bonds Using High Frequency Data.

- Kuo, L., & Mallick, B. (1998). Variable selection for regression models. *Sankhyā: The Indian Journal of Statistics, Series B*, 65–81.
- Madigan, D., & Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association*, 89(428), 1535–1546.
- Mariano, R. S. (2002). Testing forecast accuracy. *A companion to economic forecasting*, 2, 284–298.
- MathWorks. *Rolling-Window Analysis of Time-Series Models*. Retrieved 2021-05-28 from <https://kr.mathworks.com/help/econ/rolling-window-estimation-of-state-space-models.html>
- McCracken, M. W., & Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & economic statistics*, 34(4), 574–589.
- Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of business*, 473–489.
- Rudebusch, G. D., & Wu, T. (2004). The recent shift in term structure behavior from a no-arbitrage macro-finance perspective. *FRB of San Francisco Working Paper*(2004–25).
- Svensson, L. E. (1994). *Estimating and interpreting forward interest rates: Sweden 1992–1994*.

ABSTRACT

Forecasting Yield Curve with Long Short Term Memory Network in Data-rich Environment

Byeongheon Song

Department of Quantitative Applied Economics

Sungkyunkwan University

The recent progress on machine learning algorithms incurred various researches trying to graft machine learning onto the field of economics. Some of the prior studies on deep learning algorithms proved that applying deep neural network algorithms may evade the curse of dimensionality, which enabled utilizing many independent variables into one model. This study tried to apply Long Short Term Memory Network to forecast the yield curve of US Treasury Bonds and Bills, utilizing various macro-economic variables from FRED as explanatory variables in the model. The result of the study showed that out-of-sample forecasting power – of which varied from the forecasting period and the number of explanatory variables – of the LSTM model is superior to the Macro-Finance Dynamic Nelson-Siegel model with one-step ahead rolling window forecasting.