GPIF Working Paper

An examination of the predictive power of the term spread for the equity risk premium

March 30, 2026

Amane Saito

Satoshi Takase

EXECUTIVE SUMMARY

- The term spread (TMS) has been widely studied, primarily in the United States, as one of the predictors of the equity risk premium (ERP). Research that involves forecasting has the potential to contribute to risk management and return enhancement in asset management. Therefore, examining the stability and replicability of its predictive power is of practical importance. Accordingly, this paper examines the predictive power of the TMS from the perspectives of reproducibility and replicability, which are commonly employed as criteria for assessing whether an analysis is scientific.
- In the empirical analysis, we use the same data as in prior research on the United States to extract the cyclical components of the TMS and to estimate its predictive power, thereby examining reproducibility. We also extend the analysis to Europe and Japan, and examine regional differences and time variation in predictive power across the United States, Europe, and Japan, which corresponds to replicability. The results confirm the reproducibility of prior findings, while showing that the predictive power varies across regions and over time, and is not uniform.

(Note) This working paper is a compilation of the research results of GPIF staff, and the contents and opinions expressed in the text do not represent the official views of GPIF. In the preparation of this working paper, we received highly valuable comments from Mr. Kensuke Fukunaga, a staff member of GPIF, for which we express our sincere gratitude.

1. Introduction

The Government Pension Investment Fund, hereafter referred to as GPIF, has undertaken the reconstruction of its equity active portfolio based on a scientific approach, as described in its annual report (GPIF 2025). In addition, GPIF (2024) highlights the usefulness of scientifically identifying causal relationships in measuring the effects of stewardship activities and ESG investment. Such a scientific approach, which differs from discretionary and qualitative judgment that depends on individual decision makers, is conducive to the stability and replicability of empirical findings. Thus, its implementation across a broad range of asset management related analytical activities is likely beneficial. In particular, with respect to economic and market analyses that involve forecasting, quantitatively examining the characteristics of business cycles and asset price fluctuations may contribute to risk management and return enhancement. Therefore, examining whether empirical findings are stable and replicable constitutes an issue of practical importance. Accordingly, this paper takes as an example economic and market analyses that involve forecasting, which have been conducted in the literature, and examines whether such analyses contribute to the scientific approach pursued by GPIF.

To assess whether an analysis contributes to a scientific approach, we refer to an evaluation framework proposed for economic research, which classifies assessment criteria into two broad categories[1], reproducibility and replicability (Dreber and Johannesson 2025). Reproducibility refers to the ability to obtain results similar to those of the original study when the same data are used. Replicability refers to the ability to obtain results similar to those of the original study when different data are used. In economics, prior studies have reexamined published findings (Dewald et al. 1986; McCullough 2006; Chang and Li 2022). However, much of this literature has focused primarily on reproducibility, and from the perspective of assessing whether an analysis contributes to a scientific approach, it is desirable to examine replicability as well.

As a subject of examination, among economic and market analyses that involve forecasting, it is appropriate to focus on those for which alternative datasets are relatively accessible. Accordingly, taking into account practical usefulness, this paper focuses on analyses of the term structure of interest rates, for which data availability is relatively high across regions and time periods. Short term interest rates reflect current economic fundamentals and are primarily determined by central bank monetary policy. By contrast, long term interest rates incorporate a premium over short term rates that compensates investors for risks associated with holding bonds over an extended horizon, including price fluctuation risk and liquidity risk. The difference between long term and short term interest rates, namely the term spread (TMS), is regarded as a variable that reflects bond market participants' expectations regarding future economic and financial conditions. From a practical perspective, while the release of macroeconomic data related to the real economy often involves reporting lags, the TMS is available on a timely basis and constitutes a dataset that can be effectively utilized in empirical analysis. Previous studies have used the TMS, as a measure of the term structure, and have examined its predictive power for business cycles, personal consumption growth, and stock returns (Campbell 1987; Harvey 1988; Fama and French 1989; Estrella and Hardouvelis 1991).

Among the prior studies, we place particular emphasis on Faria and Verona (2020), which extracts short term, medium term, and long term cyclical components from the TMS using statistical techniques when forecasting the equity risk premium (ERP)[2] in the United States. This study facilitates a clear understanding of the characteristics of the predictive power of the TMS, and we consider that it provides a suitable foundation for our analysis. The ERP refers to the additional rate of return that investors require for holding equities, which are considered to involve greater risk than safe assets such as government bonds. In general, when the ERP increases, this implies that investors' risk aversion intensifies and that a higher return is demanded for holding equities. In such periods, investors adopt a more cautious stance toward equity investment, and since stock prices tend to be relatively low, expected future returns are likely to be higher.

[1] Dreber and Johannesson (2025) classifies reproducibility into three categories, computational reproducibility, recreate reproducibility, and robustness reproducibility. Replicability is classified into two categories, conceptual replicability, which refers to verification using new data and analytical methods that differ from those in the original study, and direct replicability, which refers to verification using new data while employing the same analytical methods as in the original study. In this paper, we focus on direct replicability.

[2] Four concepts of the ERP can be distinguished, namely the historical risk premium, which is based on realized past data, the expected risk premium, which is derived from market participants' expectations, the required risk premium, which represents the additional return that investors demand in order to hold the market portfolio, and the implied risk premium, which is calculated from a pricing model under the assumption that observed market prices are correct (Fernandez 2023). In this paper, we use the historical risk premium.

Conversely, when the ERP declines, investors' risk tolerance strengthens, equities are more readily chosen even with lower excess returns, and stock prices tend to be relatively high.

When conducting forward looking analysis, it is necessary to take into account structural changes in the economy and financial markets. In particular, when the term structure of interest rates is included in the analysis, it is desirable to consider the effects of unconventional monetary policy implemented by central banks. While traditional monetary policy has primarily targeted short term interest rates, unconventional monetary policy also targets long term interest rates and the quantity of liquidity supplied to the market. Engstrom and Sharpe (2018) points out that the TMS incorporates market expectations regarding future monetary policy. In recent years, quantitative easing (QE), which targets the quantity of liquidity supplied to the market, has been implemented intermittently. Large scale purchases of long term government bonds under quantitative easing reduce the risk premium on long term bonds irrespective of underlying economic fundamentals, thereby compressing the TMS. Accordingly, it is plausible to consider that QE affects the TMS through its influence on market expectations as well as through actual asset purchases. By contrast, in many cases equity markets are not directly subject to asset purchases, and therefore quantitative easing does not generate a direct compression of the ERP[3]. As a result, the TMS and the ERP may exhibit divergent movements, and it is plausible to consider that the predictive power of the TMS for the ERP may weaken. However, if the portfolio rebalancing effect of quantitative easing (Selgrad 2023), induces capital flows from bond markets into equity markets, or if the increase in cross asset correlations associated with quantitative easing (Williams 2014), also arises in the relationship between long term interest rates and equities, the predictive power of the TMS may not necessarily decline. Since the effects of quantitative easing on financial markets remain only partially understood, it is desirable to take QE into account when analyzing the predictive power of the TMS.

This paper analyzes the predictive power of the TMS for the ERP based on prior studies, and conducts an examination from the perspectives of reproducibility and replicability as proposed by Dreber and Johannesson (2025). In order to investigate the variables in greater detail, we focus on the short term, medium term, and long term cyclical components extracted from the TMS by means of wavelet decomposition, following Faria and Verona (2020). In the empirical analysis, we first collect the same data as in Faria and Verona (2020) and examine reproducibility. We then extend the scope of the analysis across regions and time periods in order to examine replicability. Since Faria and Verona (2020) focuses on the United States, it remains an open question whether the documented predictive power reflects a phenomenon specific to the United States or whether it also holds when the analysis is extended to financial markets in other regions, which corresponds to regional replicability. In addition, we examine predictive power while taking into account time variation in economic and financial conditions across regions, including QE, which corresponds to regional and temporal replicability.

The contribution of this paper is to provide a quantitative analysis of the predictive power of the TMS for the ERP in Europe and Japan, extending a literature that has primarily focused on the United States, and to clarify whether such predictive power exhibits regional replicability. In addition, by analyzing the time variation

[3] As part of monetary policy, asset purchases that indirectly include equities may also be conducted.

in predictive power within each region, including periods characterized by QE, this paper demonstrates whether replicability holds across regions and over time.

The remainder of this paper is organized as follows. Section 2 reviews the related literature. Section 3 examines the reproducibility of the predictive relationship between the TMS and the ERP, and extends the analysis across regions in order to assess regional replicability. Section 4 analyzes time variation in the predictive power of the TMS within each region, and examines regional and temporal replicability. Section 5 concludes and discusses directions for future research.

2. Related literature

2.1 Reproducibility in economic research

Examination of empirical methods and findings in economic research has long been undertaken. DeWald et al. (1986) points out that one reason for the failure to reproduce results using the same data in economic research is the prevalence of inadvertent errors in the original articles. McCullough et al. (2006) further documents that even when authors are required to provide data and analytical code under specified standards, many fail to supply data and code that satisfy those standards. They also show that, in an attempt to reproduce results from more than 150 empirical studies, fewer than ten percent of the papers could be successfully reproduced. On this basis, they propose improvements to archival systems that require the submission of data and analytical code upon publication in order to enhance reproducibility. In addition, Chang and Li (2022) attempts to reproduce the results of 67 macroeconomic papers, and show that only 33 percent of the studies could be reproduced without contacting the original authors, while 49 percent could be reproduced when communication with the authors was established. They further propose guidelines regarding the data items and code documentation that should be included when data and code used in the analysis are submitted together with the manuscript. Dreber and Johannesson (2025) proposes that, in addition to reproducibility, which refers to obtaining similar results when the same data are used, examination of replicability should also be undertaken as an additional criterion, which refers to obtaining similar results when different data are used.

2.2 The term structure of interest rates, equity markets, and business cycles

A substantial body of research has examined the relationship between the term structure of interest rates and the ERP in US financial markets. Fama and French (1989) shows that expected returns on bonds and equities can be predicted by the dividend yield, the default spread measured as the yield difference between high grade corporate bonds and the overall corporate bond market, and the TMS measured as the yield difference between high grade corporate bonds and one month Treasury bills. They argue that the TMS reflects relatively short business cycles, and it is plausible to consider that it predicts high expected returns during economic downturns and low expected returns during economic expansions. They also suggest that the TMS serves as an indicator capturing the ERP. Campbell (1987) demonstrates, using time series analysis, that the term structure of interest rates is related not only to short term securities and bonds but also to realized excess stock returns, which correspond to the realized ERP. McCown (2001) analyzes the relationship between the yield curve and the ERP across eight industrialized countries, including the United States, Germany, and Japan, which were major economic powers at the time. For the United States, Germany, and Japan, he finds that the ERP tends to become negative when the yield curve is inverted. In addition, for smaller industrialized economies, he shows that the domestic ERP tends to become negative when the United States or Germany experiences an inverted yield curve. Faria and Verona (2020) extracts short term, medium term, and long term cyclical components from the yield spread between 10 year and 3 month US Treasury yields using wavelet decomposition, and shows that the long term cycle predicts the ERP in the United States. They further confirm that the long term cycle forecasts the discount rate component rather than the dividend yield component, thereby explaining fluctuations in the ERP.

In addition, a substantial literature examines the ability of the term structure of interest rates to predict future economic activity. Harvey (1988), building on Fisher (1907), who argues that the one year interest rate in equilibrium reflects the relationship between the marginal value of income this year and its marginal value next year, investigates the relationship between interest rates and personal consumption. If a recession is expected in the following year, households have an incentive to sacrifice current consumption in order to purchase one year bonds that will pay off during the downturn. As demand for bonds increases, bond prices rise and interest rates decline. Harvey (1988) conducts an empirical analysis of this mechanism and confirms a comovement between the real term structure of interest rates and future personal consumption. Estrella and Hardouvelis (1991) also show that when the TMS exhibits an upward sloping shape, future personal consumption and investment tend to increase.

2.3 The effects of QE on financial markets

The effects of QE, which constitutes an unconventional monetary policy, have been analyzed in the United States, Europe, and Japan. Williams (2014) points out that during the period of QE implemented by the Federal Reserve Board (FRB), correlations increased among financial assets that had previously been uncorrelated. Other studies confirm that the Federal Reserve's ultra low interest rate policy and excess liquidity contributed to increases in commodity prices (Amatov and Dorfman 2017; Ordu-Akkaya and Soytas 2020). Selgrad (2023) provides evidence that QE lowered corporate bond yields through a portfolio rebalancing channel. She also notes that the effects of QE on asset prices are not uniform across time and market conditions. Corbet et al. (2019) analyzes the announcement effects of US QE and shows that stock market volatility increased substantially following policy announcements. Gagnon et al. (2011) indicates that large scale asset purchases reduced risk premia in US financial markets.

With respect to Europe, Hudepohl et al. (2021) examines equity markets in ten euro area countries and suggests that QE implemented by the European Central Bank (ECB) may have contributed to overheating in equity markets. Rincon and Petrova (2024) shows that the effects of the ECB's QE on equity markets differed between the European sovereign debt crisis from 2010 to 2012 and the COVID-19 pandemic period from 2020 to 2022. In Japan, the Bank of Japan (BOJ), which has implemented QE, has conducted a comprehensive review of the effects of unconventional monetary policy. The review indicates that stock prices increased under large scale monetary easing, and that the impact of monetary policy on asset prices varies across different phases of economic conditions (BOJ 2024).

3. Reproducibility of predictive power of the TMS and regional replicability

This section provides a quantitative analysis of the predictive power of the TMS for the ERP. Following Faria and Verona (2020), we extract short term, medium term, and long term cyclical components from the TMS for the United States, Europe, and Japan using wavelet decomposition. We then analyze the predictive power of these cyclical components for the ERP and examine differences across regions. By extracting cyclical components and analyzing the predictive power for the ERP in the United States, we conduct an examination of the reproducibility of prior findings. Furthermore, by extending the analysis to Europe and Japan, we conduct an examination of regional replicability.

3.1 Data

For the calculation of the TMS in each region, we use the 10 year government bond yield as the long term interest rate for the United States and Europe, while for Japan we use the 9 year government bond yield, for which a sufficiently long time series is available. As the short term interest rate, we use the 2 year government bond yield in each region[4]. For the ERP, we compute the monthly log return of the S&P 500 index for the United States, the DAX index for Europe, and the TOPIX index for Japan, and subtract the corresponding monthly converted 2 year government bond yield for each region[5]. The sample period spans from December 1979 to October 2025, resulting in 551 observations.

To examine in greater detail the predictive power of the TMS for the ERP across regions, we employ wavelet multiresolution analysis (MRA), and the maximal overlap discrete wavelet transform (MODWT), following Faria and Verona (2020). Using MODWT MRA, we do not directly use the raw TMS series, but instead decompose it into three cyclical components, namely short term (HF: high frequency), medium term (MT: mid term frequency), and long term (LF: low frequency). The short term component captures fluctuations with periodicities ranging from 2 to 16 months, and reflects speculative movements, capital flows, and market reactions to events. The medium term component captures fluctuations with periodicities ranging from 16 to 128 months, and represents business cycle dynamics, including macroeconomic cycles and fiscal and monetary policy cycles. The long term component captures fluctuations with periodicities exceeding 128 months, and reflects structural economic trends in each region, including potential growth, the natural rate of interest, and inflation trends.

Table 1 reports descriptive statistics for the ERP and the TMS in each region, as well as for the short term, medium term, and long term components obtained by decomposing the TMS using MODWT MRA. The average ERP is highest in the United States at 0.41 percent, followed by Europe at 0.40 percent and Japan at 0.19 percent. The average TMS is highest in Europe at 0.90 percent, followed by the United States at

---

[4] US government bond yields are obtained from FRED (https://fred.stlouisfed.org/)
, German government bond yields are obtained from the Bundesbank (https://www.bundesbank.de/en)
, and Japanese government bond yields are obtained from the Ministry of Finance (https://www.mof.go.jp/index.htm)
. All other data are obtained from FactSet.

[5] In calculating the historical ERP, a risk free interest rate is typically employed. Although the choice of the risk free rate varies across studies, we use the 2 year government bond yield in each region, for which long historical data are available, in order to ensure comparability across regions.

Table 1. Descriptive statistics for the ERP and the TMS in each region

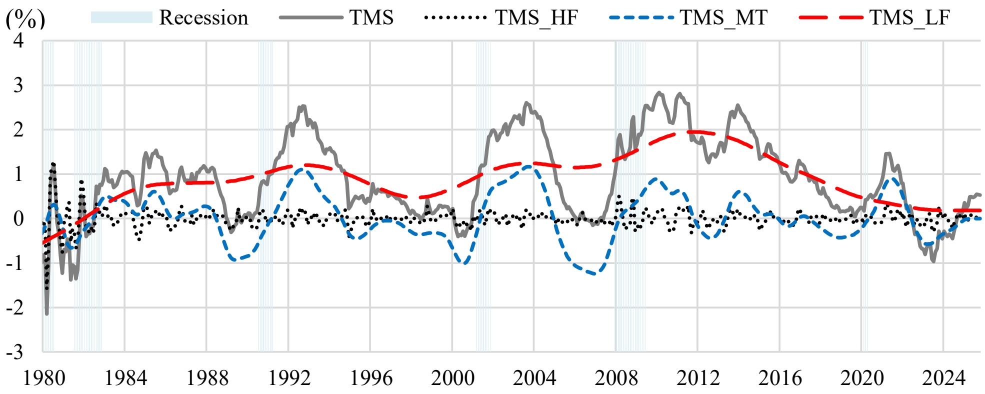
		Mean	Std. Dev.	Min	25%	50%	75%	Max	N
US									
ERP	(%)	0.41	3.65	-23.11	-1.30	0.82	2.62	11.39	551
TMS	(annualized %)	0.89	0.92	-2.14	0.22	0.00	1.52	2.84	551
TMS_{HF}	(annualized %)	0.00	0.20	-1.59	-0.09	0.00	0.08	1.30	551
TMS_{MT}	(annualized %)	0.01	0.54	-1.24	-0.35	0.87	0.41	1.17	551
TMS_{LF}	(annualized %)	0.89	0.54	-0.55	0.52	0.00	1.20	1.95	551
Europe									
ERP	(%)	0.40	4.88	-28.35	-2.08	0.96	3.41	12.80	551
TMS	(annualized %)	0.90	0.90	-2.27	0.23	0.98	1.52	2.87	551
TMS_{HF}	(annualized %)	0.00	0.17	-0.94	-0.08	0.01	0.09	0.62	551
TMS_{MT}	(annualized %)	-0.01	0.54	-1.36	-0.37	0.01	0.33	1.33	551
TMS_{LF}	(annualized %)	0.90	0.47	-0.24	0.48	1.06	1.24	1.58	551
Japan									
ERP	(%)	0.19	4.49	-24.85	-2.34	0.37	2.98	13.19	551
TMS	(annualized %)	0.69	0.60	-1.72	0.24	0.69	1.10	2.50	551
TMS_{HF}	(annualized %)	0.00	0.12	-1.05	-0.04	0.00	0.05	0.54	551
TMS_{MT}	(annualized %)	-0.01	0.30	-0.91	-0.11	-0.01	0.14	0.66	551
TMS_{LF}	(annualized %)	0.69	0.42	0.02	0.33	0.55	1.04	1.57	551

For the construction of the TMS, we use the 10 year government bond yield as the long term interest rate for the United States and Europe, and the 9 year government bond yield for Japan. As the short term interest rate, we use the 2 year government bond yield for each region. The ERP is calculated as the monthly log return on equity indices, namely the S&P 500 for the United States, the DAX for Europe, and the TOPIX for Japan, minus the corresponding 2 year government bond yield converted to a monthly frequency. The TMS is decomposed into three cyclical components using MODWT MRA: short term (HF: high frequency), medium term (MT: mid term frequency), and long term (LF: low frequency). The short term component captures fluctuations over horizons from 2 to 16 months, the medium term component captures fluctuations over horizons from 16 to 128 months, and the long term component captures fluctuations over horizons exceeding 128 months. The sample period spans from December 1979 to October 2025.

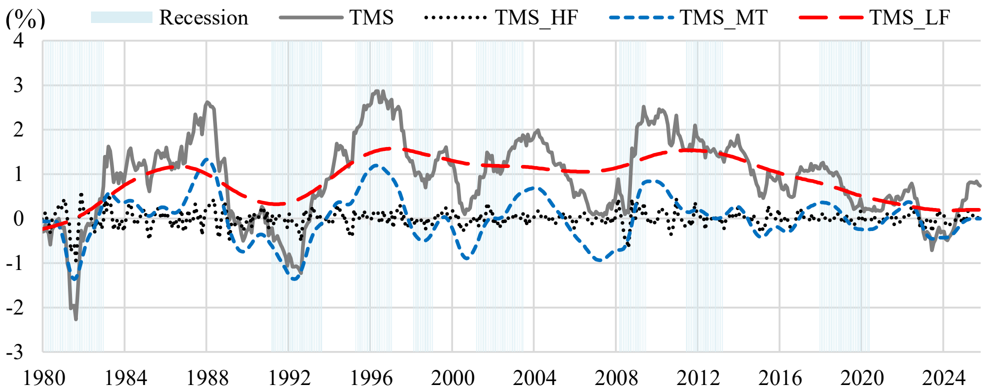
0.89 percent and Japan at 0.69 percent. The standard deviations indicate that the volatility of the TMS, TMS_{HF}, TMS_{MT}, and TMS_{LF} in Japan is smaller than that in the United States and Europe.

Figure 1 illustrates the evolution of the TMS in each region and its components obtained through decomposition using MODWT MRA. In the United States, peaks in the TMS are observed in the early 1990s, the early 2000s, around 2010, and the early 2020s. Corresponding to these episodes, TMS_{LF} generally increases. Notably, the TMS reaches its final peak within the sample period in February 2010, after which TMS_{LF} exhibits a downward trend, indicating a structural compression of the TMS. It is plausible to consider that this development reflects the global financial crisis (GFC) triggered by the subprime mortgage problem in 2007 and the subsequent implementation of QE1 by the Federal Reserve beginning in November 2008. The TMS_{MT} component, which captures business cycle dynamics, turns negative prior to the recessions in the early 1990s, the early 2000s, and the late 2000s. This pattern confirms the practical view that the TMS serves as a leading indicator of the business cycle. These findings are consistent with Faria and Verona (2020), thereby confirming the reproducibility of prior research.

(a) US



(b) Europe



(c) Japan

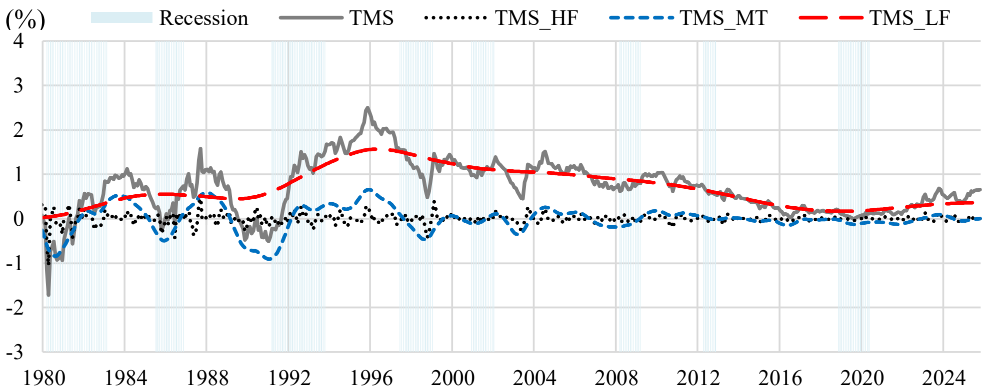


Figure 1. Decomposition of the TMS

The TMS is decomposed into three cyclical components using MODWT MRA: short term (HF: high frequency), medium term (MT: mid term frequency), and long term (LF: low frequency). The short term component captures fluctuations over horizons from 2 to 16 months, the medium term component captures fluctuations over horizons from 16 to 128 months, and the long term component captures fluctuations over horizons exceeding 128 months. Shaded areas indicate recession periods in each region.

In Europe, the timing of peaks in the TMS differs from that in the United States, but it can be confirmed that peaks in the TMS and TMS_{LF} broadly coincide. The TMS reaches its final peak within the sample period in May 2009 following the GFC, after which TMS_{LF} exhibits a downward trend. The ECB implemented government bond purchase policies under the securities markets programme (SMP) beginning in May 2010. In addition, the asset purchase programme (APP) was initiated in 2015, and it is plausible to consider that these measures contributed to the post GFC decline in the TMS.

By contrast, the peak of the TMS in Japan occurs earlier than in the United States and Europe, specifically in November 1995. The BOJ introduced the zero interest rate policy in February 1999 and initiated QE in March 2001 with the objective of overcoming deflation that had persisted since the 1990s. At the time of introducing the zero interest rate policy in 1999, the BOJ announced that it would maintain zero interest rates until deflationary concerns were dispelled, thereby influencing market expectations regarding future interest rate levels and aiming to lower long term interest rates. This approach is referred to as forward guidance.

As discussed above, the application of MODWT MRA to the United States confirms the reproducibility of prior findings. Moreover, the decomposition of the TMS using MODWT MRA can be conducted for each region, and it is plausible to consider that the methodology itself exhibits reproducibility across regions. The results of MODWT MRA indicate that TMS_{LF} declines toward the early 2020s in all regions, implying a structural compression of the TMS. It can be to considered that this development reflects the effects of QE implemented in each region.

Using these data, we conduct single variable regressions in which the ERP is the dependent variable and the TMS, TMS_{HF}, TMS_{MT}, and TMS_{LF} are each employed as explanatory variables. If all explanatory variables are included simultaneously in a single regression, multicollinearity may arise and potentially bias the estimation results. Following Faria and Verona (2020), we perform both in-sample and out-of-sample analyses. In the in-sample analysis, we apply MODWT MRA to the full sample at once and use the extracted explanatory variables to analyze their ability to explain future ERP. In contrast, in the out-of-sample analysis, we apply MODWT MRA using only the historical data available at each estimation point, and examine the ability of the explanatory variables to predict future ERP.

3.2 In-sample analysis

The estimation equation used in the in-sample analysis is given by Equation (3-1). Let r_t denote the ERP in period t, and let h denote the forecast horizon. The explanatory variable x_t corresponds to TMS, TMS_{HF}, TMS_{MT}, or TMS_{LF}, and ε_t denotes the error term. The dependent variable $r_{t:t+h}$ is defined as $r_{t:t+h} = (1/h)(r_{t+1} + \cdots + r_{t+h})$ which standardizes the ERP to a monthly basis regardless of the forecast horizon.

$$r_{t:t+h} = \alpha + \beta x_t + \varepsilon_{t:t+h}\ , t = 1, \ldots, T - h \qquad (3\text{-}1)$$

In Equation (3-1), we examine whether the coefficient β is statistically significant in order to assess whether the cyclical components of the TMS are useful in predicting the ERP. Following prior studies, we employ heteroscedasticity and autocorrelation robust t statistics, and conduct inference using p values based on a wild bootstrap procedure to test the null hypothesis ($H_0\!: \beta = 0$) against the alternative hypothesis

(H_A: $\beta > 0$). All other details of the estimation methodology follow Faria and Verona (2020). Using Equation (3-1), we analyze the ERP at forecast horizons of 1 month ($h = 1$), 3 months ($h = 3$), 6 months ($h = 6$), 12 months ($h = 12$), and 24 months ($h = 24$), as the dependent variable.

Table 2 reports the estimation results for each region. For the United States, statistical significance is observed for the TMS when the dependent variable is the cumulative ERP up to 24 months ahead, $h = 24$, and for TMS_{HF} when the dependent variable is the cumulative ERP up to 1 month ahead, $h = 1$. In Faria and Verona (2020), which analyze the period from 1973 to 2017, both the TMS and TMS_{LF} are statistically significant for all forecast horizons h. However, in our estimation results, these variables are not statistically significant[6]. For Europe, the TMS, TMS_{HF}, and TMS_{MT} are statistically significant across all forecast horizons. For $h = 12$ and $h = 24$, the t statistics for TMS_{MT} are larger than those of the other explanatory variables, suggesting that the medium term cyclical component of the TMS is informative for long horizon fluctuations in the ERP. For Japan, TMS_{MT} is statistically significant for forecast horizons from $h = 1$ to $h = 6$.

Table 2. Results of in-sample regression

	h=1		h=3		h=6		h=12		h=24	
	β	R^2	β	R^2	β	R^2	β	R^2	β	R^2
US										
TMS	0.15	0.2%	0.12	0.3%	0.09	0.3%	0.12	0.8%	0.20	5.0%
	(0.97)		(0.86)		(0.63)		(0.87)		(1.77)*	
TMS_{HF}	0.35	0.9%	0.15	0.4%	0.05	0.1%	0.00	0.0%	-0.01	0.0%
	(2.26)**		(1.08)		(0.46)		(0.06)		(-0.16)	
TMS_{MT}	0.12	0.1%	0.12	0.3%	0.12	0.4%	0.13	1.0%	0.21	5.3%
	(0.94)		(0.96)		(0.86)		(0.98)		(1.42)	
TMS_{LF}	0.01	0.0%	0.03	0.0%	0.03	0.0%	0.09	0.4%	0.14	2.6%
	(0.03)		(0.19)		(0.18)		(0.55)		(1.3)	
Europe										
TMS	0.39	0.6%	0.45	1.8%	0.47	3.7%	0.46	6.7%	0.42	11.9%
	(2.17)**		(2.49)***		(2.83)***		(2.96)***		(2.74)***	
TMS_{HF}	0.40	0.7%	0.46	2.0%	0.39	2.6%	0.21	1.4%	0.11	0.9%
	(1.87)**		(2.52)***		(2.62)***		(2.38)**		(1.85)**	
TMS_{MT}	0.36	0.5%	0.43	1.7%	0.51	4.3%	0.55	9.9%	0.54	20.3%
	(1.71)**		(2.02)**		(2.56)***		(3.11)***		(3.09)***	
TMS_{LF}	0.18	0.1%	0.18	0.3%	0.17	0.5%	0.17	0.9%	0.13	1.2%
	(0.93)		(0.96)		(0.91)		(0.95)		(0.79)	
Japan										
TMS	-0.07	0.0%	-0.07	0.1%	-0.07	0.1%	-0.08	0.2%	-0.10	0.6%
	(-0.38)		(-0.42)		(-0.41)		(-0.39)		(-0.49)	
TMS_{HF}	-0.23	0.3%	-0.22	0.5%	-0.13	0.3%	-0.05	0.1%	-0.02	0.0%
	(-1.25)		(-1.36)		(-1.05)		(-0.64)		(-0.36)	
TMS_{MT}	0.41	0.8%	0.38	1.5%	0.35	2.2%	0.31	3.1%	0.24	3.5%
	(2.12)**		(1.91)**		(1.76)**		(1.4)		(1.15)	
TMS_{LF}	-0.35	0.6%	-0.34	1.2%	-0.33	2.0%	-0.32	3.5%	-0.32	6.0%
	(-2.03)		(-1.92)		(-1.86)		(-1.87)		(-2.12)	

Values in parentheses are heteroscedasticity- and autocorrelation-robust t-statistics, and the reported p-values are obtained from wild bootstrap tests of the null hypothesis (H_0: $\beta = 0$) against the alternative hypothesis (H_A: $\beta > 0$). ***p<0.01, **p<0.05, *p<0.10

[6] It should be noted that when the sample period is restricted to 1973 to 2017, which corresponds to the period analyzed in Faria and Verona (2020), similar results are obtained, thereby confirming the reproducibility of the analysis.

The in-sample results indicate that the finding of Faria and Verona (2020), which suggests that the long term cyclical component is most informative for predicting the ERP, does not hold uniformly across regions. The discrepancy between our results and those of the prior study for the United States can be attributed to differences in the sample period. Moreover, for Europe and Japan, the medium term component exhibits stronger explanatory power than the long term component, implying that the impact of the TMS on the ERP varies across regions and sample periods. Those results suggest that the relationship between the TMS and the ERP does not exhibit consistent regional or temporal replicability.

3.3 Out-of-sample analysis

The out-of-sample analysis enables us to examine the leading properties of the TMS for the ERP in a manner that more closely reflects actual decision making in asset management. In this analysis, we first estimate the initial regression using data from December 1979 to December 1989. The forecasted ERP at horizon h, denoted as $\hat{r}_{t:t+h}$, is obtained by estimating the coefficients as specified in Equation (3-2). We then expand the sample window by adding one month of data at a time and compute $\hat{r}_{t:t+h}$ recursively. The out-of-sample forecast period spans from January 1990 to October 2025, resulting in 430 monthly observations.

$$\hat{r}_{t:t+h} = \hat{\alpha}_t + \hat{\beta}_t x_t \tag{3-2}$$

We evaluate predictive performance using the out-of-sample R^2_{OS} proposed by Campbell and Thompson (2008), which is computed as shown in Equation (3-3). In Equation (3-3), R^2_{OS} increases when the mean squared forecast error (MSFE) of the forecast based on Equation (3-2), $\hat{r}_{t:t+h}$, relative to the realized value $r_{t:t+h}$, is smaller than the MSFE obtained using the historical average, $\bar{r}_t$. A larger R^2_{OS} therefore indicates stronger predictive power of the TMS.

$$R^2_{OS} = 100\left(1 - \frac{MSFE_{PRED}}{MSFE_{HM}}\right) = 100\left[1 - \frac{\sum_{t=t_0}^{T-h}(r_{t:t+h} - \hat{r}_{t:t+h})^2}{\sum_{t=t_0}^{T-h}(r_{t:t+h} - \bar{r}_t)^2}\right] \tag{3-3}$$

To evaluate the economic value generated by forecasts based on the TMS, we compute the certainty equivalent return (CER) using Equation (3-4). Let $\overline{RP}$ denote the sample mean of the portfolio return, and let σ^2_{RP} denote its variance. The parameter γ represents the coefficient of relative risk aversion, and following the prior literature, we set $\gamma = 3$.

$$CER = \overline{RP} - 0.5\gamma\sigma^2_{RP} \tag{3-4}$$

In this study, we compute CER gains, defined as the difference between the CER obtained from the forecasting model based on the TMS, which yields $\overline{RP}$ and σ^2_{RP}, and the CER obtained from the historical mean benchmark, which yields $\overline{RP}_{HM}$ and $\sigma^2_{RP_{HM}}$. CER gains can be interpreted as the annualized excess return attributable to active investment based on the TMS forecasting model.

$$CER\ gains = (\overline{RP} - 0.5\gamma\sigma^2_{RP}) - (\overline{RP}_{HM} - 0.5\gamma\sigma^2_{RP_{HM}}) \qquad (3\text{-}5)$$

We assume a mean variance investor who allocates wealth between equities and a risk free asset in constructing the portfolio return. At the end of month t, the investor allocates a fraction w_t to equities as specified in Equation (3-6). Let $\hat{R}_{t+h}$ denote the out-of-sample forecast of the equity return from period t to $t+h$, and let $\hat{\sigma}^2{}_{t+h}$ denote the out-of-sample forecast of the variance of equity returns. Following the prior literature, the forecast of $\hat{\sigma}^2{}_{t+h}$ is obtained using a ten year rolling window of past equity returns. The portfolio weight w_t is constrained to lie between −0.5 and 1.5. The portfolio is rebalanced at the same frequency as the forecast horizon h, and the equity allocation is updated according to the forecasted future equity return.

$$w_t = \frac{1}{\gamma}\frac{\hat{R}_{t+h}}{\hat{\sigma}^2{}_{t+h}} \qquad (3\text{-}6)$$

Table 3 reports the estimation results. For the United States, the R^2_{OS} of TMS_{LF} is larger than that of TMS, TMS_{HF}, and TMS_{MT} for all forecast horizons h. This result is consistent with Faria and Verona (2020), thereby confirming the reproducibility of the out-of-sample analysis[7]. It is plausible to consider that, for the US ERP, the long term cyclical component of the TMS possesses the strongest predictive power. For Europe, the R^2_{OS} of the raw TMS is generally larger than that of its decomposed components, except for the two year ahead forecast ($h = 24$), where the results suggest that referring to TMS_{MT} is informative. For Japan, TMS_{MT} exhibits the largest R^2_{OS} for the one year ahead forecast ($h = 12$), while TMS_{LF} yields the largest R^2_{OS} for the two year ahead forecast ($h = 24$).

CER gains represent the excess return that would be obtained by implementing the portfolio strategy based on the forecast model. Accordingly, higher R^2_{OS} tends to be associated with larger excess returns. For the United States, using TMS_{LF} yields annualized excess returns ranging from 2.38 percent at $h = 6$ to a maximum of 2.86 percent at $h = 12$. For Europe, using TMS_{MT} suggests annualized excess returns of 1.06 percent at $h = 12$ and 1.17 percent at $h = 6$. In Japan, although TMS_{LF} exhibits the largest R^2_{OS} at $h = 24$, the CER gains are larger for TMS_{MT} at TMS_{MT} than for $h = 24$.

In summary, the out-of-sample analysis indicates that, for the United States, the long term cyclical component of the TMS exhibits the strongest predictive power for the ERP across all forecast horizons. By contrast, in Europe, either the raw TMS or the medium term cyclical component demonstrates stronger predictive performance. In Japan, the predictive power of the TMS for the ERP is smaller than that observed in the other regions. It is confirmed that the out-of-sample predictive power of the TMS for the ERP does not exhibit consistent regional or temporal replicability. As illustrated in Figure 1, these differences

[7] However, compared with the results obtained for the sample period from 1973 to 2017 analyzed in Faria and Verona (2020), the R^2_{OS} values in our analysis using data from December 1979 to October 2025 decline to less than half of those reported in the prior study. It is plausible to consider that temporal replicability of the analysis may therefore not be fully ensured.

Table 3. Results of out-of-sample regression

	Predictive performance(R^2_{os})					Certainty equivalent return (CER) gains				
	Forecast horizon					Forecast horizon				
	1-month-ahead (h=1)	3-month-ahead (h=3)	6-month-ahead (h=6)	12-month-ahead (h=12)	24-month-ahead (h=24)	1-month-ahead (h=1)	3-month-ahead (h=3)	6-month-ahead (h=6)	12-month-ahead (h=12)	24-month-ahead (h=24)
US										
TMS	-1.17	-1.39	-1.28	-1.85	3.33***	-1.27	-0.20	0.02	0.15	1.18
TMS_{HF}	-0.52	-0.33	-0.42	-0.61	1.67***	-1.39	-0.27	-0.16	0.00	0.21
TMS_{MT}	-0.22	0.07	0.54**	0.67**	-0.85	0.14	0.64	0.84	0.44	-0.16
TMS_{LF}	0.38**	1.46***	3.02***	7.2***	15.21***	2.67	2.62	2.38	2.86	2.56
Europe										
TMS	0.54**	2.06***	4.37***	7.79***	13.67***	1.27	1.52	1.56	1.14	0.90
TMS_{HF}	0.21	-0.03	-0.19	-0.1	-0.29	0.42	-0.01	-0.08	-0.02	-0.07
TMS_{MT}	0.17	1.1***	2.75***	5.59***	15.83***	0.74	1.13	1.17	1.06	1.09
TMS_{LF}	0.27	0.75**	1.14**	1.45***	-0.69	0.16	0.08	0.21	0.25	0.09
Japan										
TMS	-0.4	-0.41	-0.52	-1.02	-1.38	-0.54	0.02	0.11	0.29	0.38
TMS_{HF}	-0.2	-0.27	-0.37	-0.41	-0.6	-0.40	-0.21	-0.15	-0.09	-0.06
TMS_{MT}	-0.57	-0.48	0.1*	1.13***	1.86***	-1.27	-0.46	0.09	0.45	0.58
TMS_{LF}	-0.46	-0.53	-0.21	0.93***	4.72***	-0.84	-0.35	-0.03	0.20	0.47

***p<0.01, **p<0.05, *p<0.10

may reflect variations in the cyclical behavior of the TMS and in the underlying macroeconomic environment across regions. In the United States, business cycles tend to be relatively long, approximately ten years in duration, and the results suggest that this contributes to a stronger relationship between the long term cyclical component of the TMS and the ERP. In Europe, business cycles are shorter and expansions and contractions occur more frequently than in the United States. One possible interpretation is that the prominence of manufacturing industries, particularly in Germany, which are more sensitive to cyclical fluctuations, influences the predictive performance of the TMS. In Japan, the TMS has exhibited a declining trend since its peak in 1995, and cyclical fluctuations in the TMS have remained subdued for most of the sample period. It is conceivable that this has reduced the amount of information contained in the TMS that can be used to predict the ERP.

4. Regional and temporal replicability of the predictive power of the TMS

The wavelet decomposition results in Section 3 and the out-of-sample regression analysis indicate that the evolution of interest rate cycles and their relationship with the ERP differ across regions. The results suggest that one possible explanation for these differences is the impact of large scale monetary easing implemented in each region. Accordingly, in this section, we divide the sample into periods before and after the introduction of QE in each region. This approach allows us to conduct an examination of the temporal replicability of the predictive power of the TMS for the ERP across regions.

4.1 Sub-sample analysis of structural change

In the United States, the TMS reaches its final peak within the sample period in February 2010, after which the long term cyclical component, TMS_{LF}, exhibits a downward trend. This may reflect QE1 implemented by the Federal Reserve beginning in November 2008 in response to the GFC. In Europe, government bond purchases were initiated under the SMP in May 2010. In Japan, the BOJ introduced a zero interest rate policy accompanied by forward guidance in February 1999, ahead of the United States and Europe, and initiated QE two years later. In this paper, we treat November 2008 for the United States, corresponding to QE1 by the Federal Reserve, May 2010 for Europe, corresponding to the introduction of the SMP by the ECB, and March 2001 for Japan, corresponding to the introduction of QE by the BOJ, as the starting points of QE in each region.

We conduct a sub-sample analysis of the predictive power of the TMS for the ERP. For each region, the sample is divided according to the timing of the introduction of QE[8], and the out-of-sample regression analysis described in Section 3 is implemented for each sub-sample.

Tables 4 through 6 report the results of the out-of-sample regressions for each region by sub-sample. In each table, the cells that are statistically significant in the full sample results reported in Table 3 are shaded. For the United States, the predictive power of the TMS and TMS_{LF}, which is statistically significant in the full sample, is confirmed in the pre QE period, but disappears in the post QE period. For Europe, the TMS and TMS_{MT} are statistically significant in the pre QE period, exhibiting patterns similar to those in the full sample.

Table 4. Comparison of out-of-sample predictive power across sub-samples in the US

	Predictive performance(R^2_{os})					Certainty equivalent return (CER) gains				
	Forecast horizon					Forecast horizon				
	1-month-ahead (h=1)	3-month-ahead (h=3)	6-month-ahead (h=6)	12-month-ahead (h=12)	24-month-ahead (h=24)	1-month-ahead (h=1)	3-month-ahead (h=3)	6-month-ahead (h=6)	12-month-ahead (h=12)	24-month-ahead (h=24)
from December 1979 to October 2008										
TMS	-2.23	-3.03	-2.77	-2.35	4.03**	-2.69	-0.62	0.05	0.54	1.41
TMS_{HF}	-1.41	-0.96	-0.59	-1.3	0.83	-3.44	-0.78	-0.27	-0.20	0.12
TMS_{MT}	-0.07	-0.23	-0.75	-2.04	-1.65	0.96	0.81	0.70	0.13	-0.38
TMS_{LF}	0.77**	2.81***	5.34***	11.55***	20.01***	3.46	3.53	3.13	3.57	3.28
from November 2008 to October 2015										
TMS	-0.85	-0.55	-0.19	-0.5	-16.51	-2.14	-0.38	0.51	1.29	-9.86
TMS_{HF}	-2.37	-2.59	-3.9	-13.27	2.06	-4.52	-2.25	-2.54	-5.23	0.48
TMS_{MT}	-1.15	-1.7	-2.15	2	-9.69	-4.10	-2.94	-2.82	-0.23	-2.05
TMS_{LF}	-1.34	-2.82	-5.21	-7.24	6.46*	-1.77	-1.41	-3.13	-7.47	-25.11

In the pre QE out-of-sample analysis, the initial regression is estimated using data from December 1979 to December 1989. The coefficients are estimated as in Equation (3-2) for the h-step-ahead forecast of the ERP, $\hat{r}_{t:t+h}$. The sample window is then expanded recursively by one month at a time to compute $\hat{r}_{t:t+h}$. In the post QE analysis, the estimation is conducted using the same procedure, with the initial sample covering the period from October 2007 to October 2008. Shaded cells indicate cases where the out-of-sample results for the full sample are statistically significant. ***p<0.01, **p<0.05, *p<0.10

[8] Appendix A confirms, using the Chow test, that the mean and variance of the TMS in each region exhibit statistically significant differences before and after the introduction of QE.

Table 5. Comparison of out-of-sample predictive power across sub-samples in Europe

	Predictive performance(R^2_{os})					Certainty equivalent return (CER) gains				
	Forecast horizon					Forecast horizon				
	1-month-ahead (h=1)	3-month-ahead (h=3)	6-month-ahead (h=6)	12-month-ahead (h=12)	24-month-ahead (h=24)	1-month-ahead (h=1)	3-month-ahead (h=3)	6-month-ahead (h=6)	12-month-ahead (h=12)	24-month-ahead (h=24)
from December 1979 to April 2010										
TMS	1.76***	4.63***	8.52***	13.72***	19.22***	3.81	3.35	3.03	2.29	1.59
TMS_{HF}	0.2	-0.08	-0.22	-0.04	-0.42	0.44	-0.04	-0.08	0.00	-0.05
TMS_{MT}	0.83**	2.63**	5.55***	10.74***	27.52***	2.20	2.26	2.22	2.06	2.40
TMS_{LF}	0.79*	1.57	2.04*	1.9	-1.13	1.24	0.73	0.66	0.42	0.06
from May 2010 to October 2025										
TMS	-2.24	-8.49	-21.21	-44.06	-98.39	-4.91	-6.86	-8.89	-11.09	-20.21
TMS_{HF}	-0.49	0.26	-0.71	0.11	3.05**	-0.06	0.15	-0.28	0.06	0.06
TMS_{MT}	-0.45	-1.31	-2.62	-5.69	-35.77	-1.05	-1.01	-1.06	-1.39	-6.01
TMS_{LF}	-8.77	-21.4	-42.32	-79.38	-95.45	-21.79	-19.53	-22.64	-27.02	-47.32

In the pre QE out-of-sample analysis, the initial regression is estimated using data from December 1979 to December 1989. The coefficients are estimated as in Equation (3-2) for the h-step-ahead forecast of the ERP, $\hat{r}_{t:t+h}$. The sample window is then expanded recursively by one month at a time to compute $\hat{r}_{t:t+h}$. In the post QE analysis, the estimation is conducted using the same procedure, with the initial sample covering the period from April 2000 to April 2010. Shaded cells indicate cases where the out-of-sample results for the full sample are statistically significant. ***p<0.01, **p<0.05, *p<0.10

Table 6. Comparison of out-of-sample predictive power across sub-samples in Japan

	Predictive performance(R^2_{os})					Certainty equivalent return (CER) gains				
	Forecast horizon					Forecast horizon				
	1-month-ahead (h=1)	3-month-ahead (h=3)	6-month-ahead (h=6)	12-month-ahead (h=12)	24-month-ahead (h=24)	1-month-ahead (h=1)	3-month-ahead (h=3)	6-month-ahead (h=6)	12-month-ahead (h=12)	24-month-ahead (h=24)
from December 1979 to February 2001										
TMS	-0.96	-0.96	-1.26	-2.73	-6.22	-1.21	0.15	0.39	0.88	1.40
TMS_{HF}	-0.47	-0.65	-0.83	-0.68	-2.31	-0.96	-0.53	-0.37	-0.19	-0.33
TMS_{MT}	-1.31	-0.85	0.54	1.82	2.91	-2.96	-0.74	0.61	1.31	2.12
TMS_{LF}	-1.71	-2.42	-2.59	-2.33	0.56	-3.38	-1.77	-0.92	-0.47	-0.01
from March 2001 to October 2025										
TMS	-1.19	-1.8	-1.76	-0.52	8.09***	-3.00	-2.26	-1.38	-0.44	0.93
TMS_{HF}	-0.17	-0.12	-0.29	-1.34	-0.21	-0.22	-0.10	-0.11	-0.27	-0.02
TMS_{MT}	-0.22	-0.22	-0.14	-0.08	0.03	-0.49	-0.23	-0.08	0.03	0.03
TMS_{LF}	-0.57	-1.09	-1.15	-0.59	7.58***	-1.36	-0.99	-0.58	-0.18	0.81

In the pre QE out-of-sample analysis, the initial regression is estimated using data from December 1979 to December 1989. The coefficients are estimated as in Equation (3-2) for the h-step-ahead forecast of the ERP, $\hat{r}_{t:t+h}$. The sample window is then expanded recursively by one month at a time to compute $\hat{r}_{t:t+h}$. In the post QE analysis, the estimation is conducted using the same procedure, with the initial sample covering the period from February 2000 to February 2001. Shaded cells indicate cases where the out-of-sample results for the full sample are statistically significant. ***p<0.01, **p<0.05, *p<0.10

However, in the post QE period, predictive power for the ERP is observed only for some cases of TMS_{HF}. For Japan, among TMS_{MT} and TMS_{LF}, which display predictive power in the full sample, neither is statistically significant in the pre QE period, and only part of TMS_{LF} remains significant in the post QE period.

Taken together, these results indicate that in the United States and Europe, the predictive power of the TMS for the ERP declines after the introduction of QE. In contrast, in Japan, predictive power is weak in both the pre and post QE periods, suggesting that factors other than QE may also influence the time variation in predictive performance. Accordingly, the next section examines the time variation in the coefficients associated with the medium term and long term cyclical components of the TMS, for which predictive power has been identified in each region.

4.2 Analysis of structural change using a time varying parameter regression

The sub-sample analysis confirms that in the United States and Europe, the predictive power of the TMS for the ERP declines after the introduction of QE. In contrast, in Japan, predictive power is weak in both the pre QE and post QE periods, suggesting that factors other than QE may also contribute to the observed decline in predictive performance. In this section, we estimate the time variation in the relationship between the TMS and the ERP using a time varying parameter (TVP) regression framework. The estimation methodology follows Nakajima (2011). Specifically, the model is formulated as in Equations (4-1) and (4-2).

$$y_t = x_t'\beta + z_t'\alpha + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \ \ t = 1, \dots, n \qquad (4\text{-}1)$$

$$\alpha_{t+1} = \alpha_t + u_t, \quad u_t \sim N(0, \Sigma), \ \ t = 0, \dots, n-1 \qquad (4\text{-}2)$$

Let y_t denote the scalar dependent variable, which in this analysis corresponds to the ERP defined as $r_{t:t+h} = (1/h)(r_{t+1} + \cdots + r_{t+h})$. As in Section 3.2, we conduct an in-sample analysis and set $h = 12$ in order to estimate the time variation in the predictive power of the TMS for the cumulative ERP over the subsequent 12 months. The vectors x_t' and z_t' denote $(k \times 1)$ and $(p \times 1)$ covariate vectors, respectively. In this analysis, x_t consists only of a constant term, so $k = 1$. The vector z_t is specified as $z_t = (TMS_{MT,t}, TMS_{LF,t})$, so $p = 2$. The parameter β is a $(k \times 1)$ coefficient associated with the constant term, while α is a $(p \times 1)$ vector of time varying parameters[9]. Nakajima (2011) estimates the model under both a stochastic volatility specification for σ and a specification in which σ is constant within the sample. In this paper, we assume σ to be constant in order to reduce computational cost. Following Nakajima (2011), the parameters are estimated using the Markov Chain Monte Carlo (MCMC) method with 20,000 sampling iterations.

Table 7 reports the posterior means, standard deviations, and 95 percent credible intervals of the estimated parameters. The convergence diagnostic (CD) corresponds to the statistic proposed by Geweke (1992). The

[9] We compute the correlation coefficients between TMS_{MT} and TMS_{LF} for each of the United States, Europe, and Japan, and find that the absolute values are all below 0.2. These results suggest that the influence of multicollinearity is sufficiently small.

Table 7. Results of TVP regression

Parameter	Mean	Std. Dev.	95% credible interval		CD	Inef.
US						
Σ_{11}	0.196	0.042	[0.121	,0.29]	0.712	22.670
Σ_{22}	0.239	0.028	[0.189	,0.297]	0.839	10.470
σ	0.196	0.018	[0.161	,0.232]	0.392	17.830
Europe						
Σ_{11}	0.108	0.026	[0.063	,0.165]	0.658	30.140
Σ_{22}	0.351	0.030	[0.296	,0.414]	0.662	5.740
σ	0.136	0.021	[0.098	,0.177]	0.549	51.730
Japan						
Σ_{11}	0.685	0.148	[0.435	,1.021]	0.700	20.360
Σ_{22}	0.607	0.072	[0.477	,0.76]	0.077	11.580
σ	0.283	0.024	[0.239	,0.331]	0.298	14.390

The convergence diagnostic (CD) corresponds to the statistic proposed by Geweke (1992). The CD statistic tests whether there is a statistically significant difference between the mean of the first 10 percent and the mean of the last 50 percent of the 20,000 sampled draws of each parameter. The inefficiency factor, Inef., reflects the degree of autocorrelation contained in the MCMC samples and indicates the extent to which the effective sample size is reduced. Values closer to one imply higher sampling efficiency.

CD statistic tests whether there is a statistically significant difference between the mean of the first 10 percent and the mean of the last 50 percent of the 20,000 sampled draws of each parameter. By examining whether such differences are present, we assess whether the Markov chain has converged. The inefficiency factor, Inef., reflects the degree of autocorrelation contained in the MCMC samples and indicates the extent to which the effective sample size is reduced. Values closer to one imply higher sampling efficiency. The estimation results show that the null hypothesis of convergence is not rejected at the 5 percent level for any parameter based on the CD statistic, indicating that the estimates have converged with 20,000 sampling iterations. The values of Inef. are also smaller than those reported in prior studies, confirming that the sampling procedure is efficient.

For the United States, the posterior mean of Σ_{11}, which represents the time variation in the predictive coefficient α_1 associated with TMS_{MT}, is 0.196, while the posterior mean of Σ_{22}, which represents the time variation in the predictive coefficient α_2 associated with TMS_{LF}, is 0.239. This suggests that the time variation in TMS_{LF} appears relatively larger than that in TMS_{MT}. The parameter σ, which captures the time variation in y_t not explained by the model parameters, is estimated at 0.196. However, the credible intervals of Σ_{11}, Σ_{22}, and σ overlap, indicating that the differences in the magnitude of time variation across these parameters are not statistically significant.

For Europe, the posterior mean of Σ_{11} is 0.108 and that of Σ_{22} is 0.351, while σ is estimated at 0.136.

Comparing the credible intervals, the 2.5 percentile value of Σ_{22} is 0.296, which exceeds the 97.5 percentile values of Σ_{11} and σ, which are 0.165 and 0.177, respectively. These results suggest that the time variation in TMS_{LF} is statistically larger than that in TMS_{MT} and the unexplained variation in y_t.

For Japan, the posterior mean of Σ_{11} is 0.685 and that of Σ_{22} is 0.607, while σ is 0.283. A comparison of the credible intervals shows that the 2.5 percentile values of Σ_{11} and Σ_{22}, which are 0.435 and 0.477, respectively, exceed the 97.5 percentile value of σ, which is 0.331. The results suggest that the time variation in both TMS_{MT} and TMS_{LF} is statistically larger than the unexplained variation in y_t. Moreover, the 2.5 percentile values of Σ_{11} and Σ_{22} for Japan exceed the 97.5 percentile values of Σ_{11} and Σ_{22} for the United States and Europe, and the results suggest that the time variation in TMS_{MT} and TMS_{LF} is statistically larger in Japan than in the other regions. These differences in the time varying properties of the coefficients may reflect structural differences across regions and may contribute to the regional heterogeneity in predictive performance documented in the previous section. It is plausible to consider that the relationship between TMS_{MT} or TMS_{LF} and the ERP is more unstable in Japan, which may account for the relatively weaker predictive power compared with the United States and Europe.

Figures 2 through 4 present the estimated time-varying parameters for each country. For the United States, we focus on the parameter α_2 associated with TMS_{LF}, for which predictive power is shown to disappear in the post QE period in Table 4. Based on Faria and Verona (2020) and the estimation results in Table 2 of this paper, the coefficient of the TMS with respect to the ERP is expected to be positive, and if this sign remains stable over time, temporal replicability of predictive power would also be high. The estimation results show that from 1980 to 2000, α_2 is generally positive, except during the period from 1980 to March 1983, which corresponds to the second oil shock, and around October 1987, when the US stock market experienced a sharp decline, known as Black Monday. Around the collapse of the IT bubble in 2001 and the onset of the GFC in 2007, the coefficient moves sharply in the negative direction, and subsequently returns to positive territory. This may reflect the fact that during episodes of sharp stock market declines, the realized ERP used in this paper becomes negative, while TMS_{LF} remains positive, as shown in Figure 1. In addition, during the COVID-19 pandemic in 2020, the coefficient initially moves in the positive direction and subsequently turns negative toward the end of the sample period. This may reflect the substantial negative realizations of the ERP during the pandemic period. Accordingly, the credible intervals indicate that the coefficient is temporarily and significantly negative.

The analysis in Section 4.1 shows that the predictive power of the TMS declines after 2008, and that temporal replicability is not ensured. The time varying parameter analysis further indicates that predictive power weakens not only after the introduction of QE but also during periods of major economic and market shocks. However, even after 2012, there are periods in which the coefficient of TMS_{LF} with respect to the ERP remains positive, and the results suggest that some degree of predictive power may have been present during relatively stable market conditions in the post QE period.

For Europe, we examine the parameter α_1 associated with TMS_{MT}, for which predictive power is shown to weaken in the post QE period in Table 5. The coefficient α_1 exhibits substantial fluctuations during the second oil shock, from 1980 to March 1983, the currency crisis within the European Monetary System (EMS) from 1992 to 1993, the GFC in the late 2000s, and the COVID-19 pandemic in 2020. In other periods,

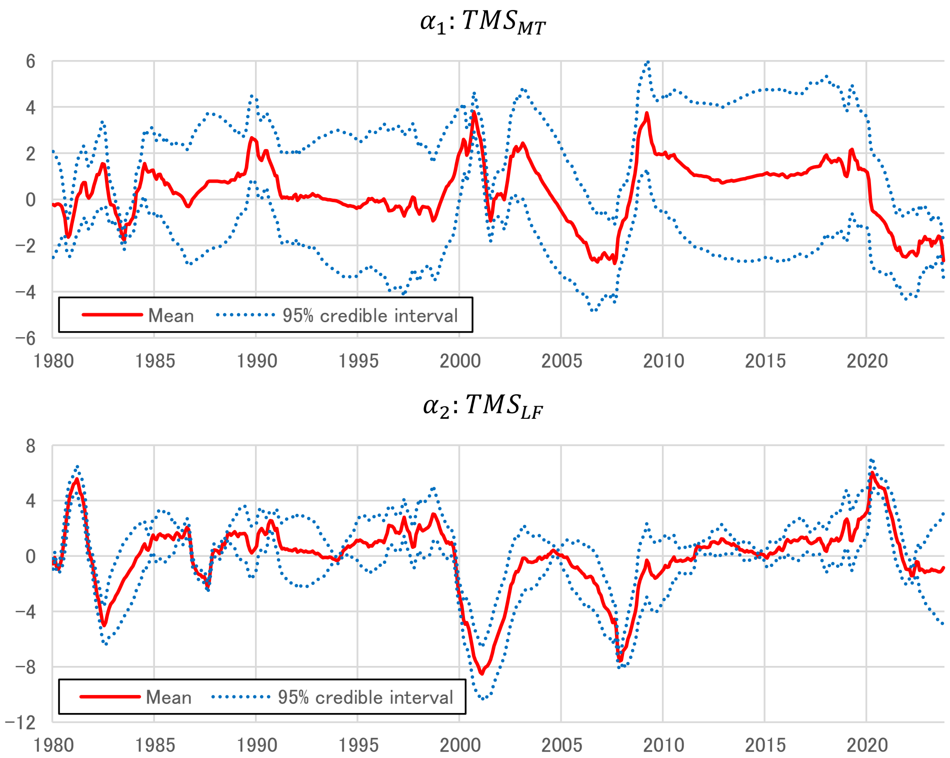


Figure 2. Time variation in the coefficients of TMS_{MT} and TMS_{LF} in the US

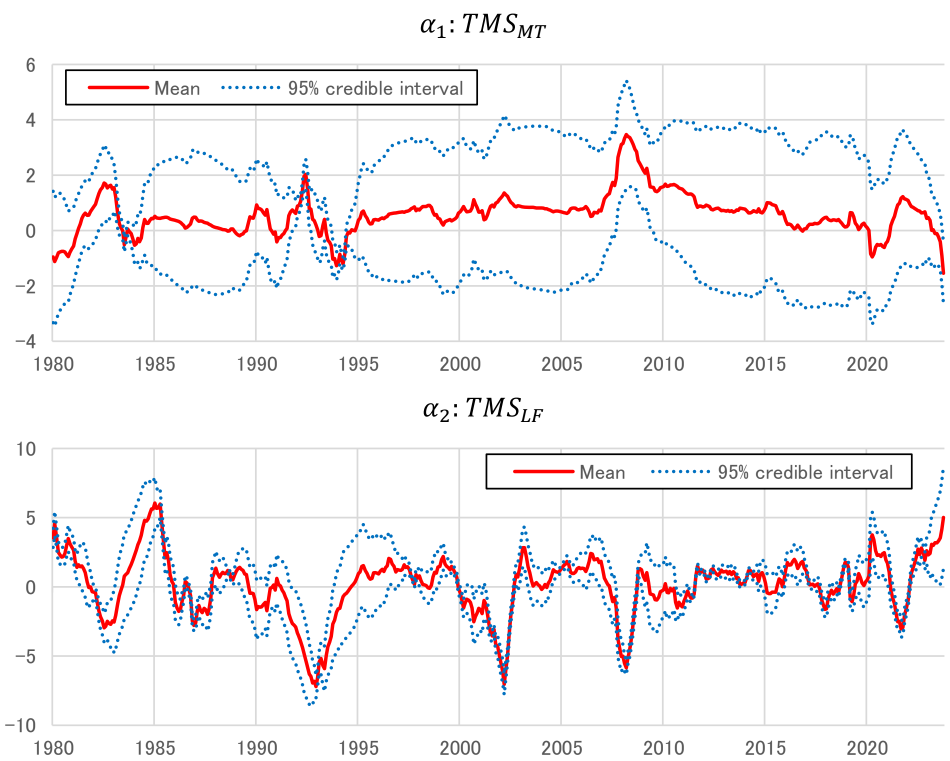


Figure 3. Time variation in the coefficients of TMS_{MT} and TMS_{LF} in Europe

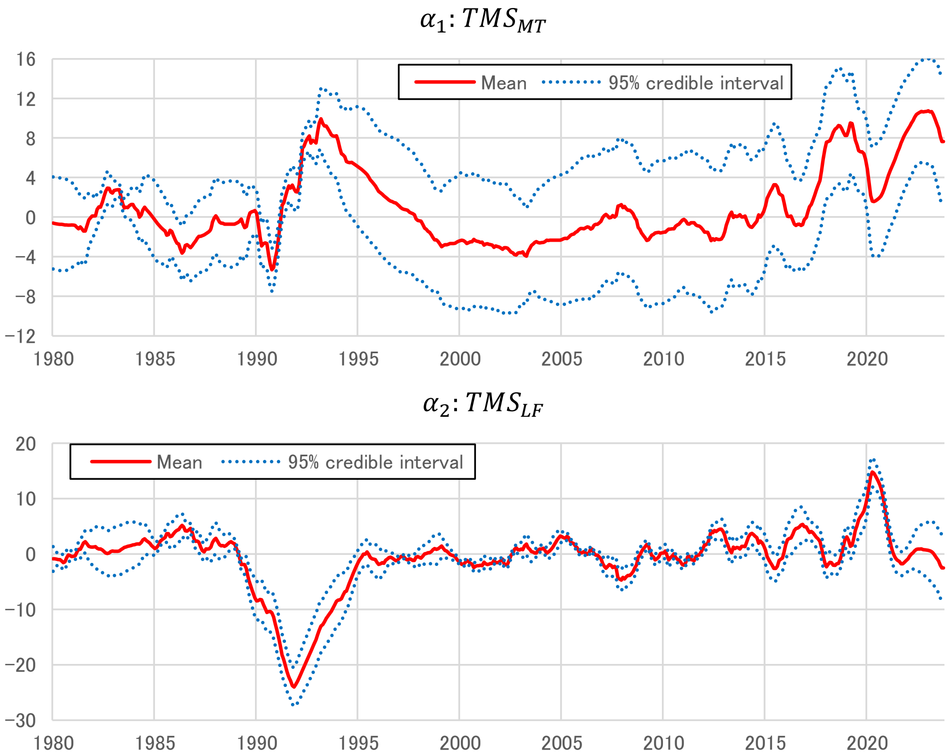


Figure 4. Time variation in the coefficients of TMS_{MT} and TMS_{LF} in Japan

however, the parameter generally remains within a range between 0 and 1. Even after May 2010, which marks the structural break examined in Table 5, α_1 declines gradually but remains positive until 2020. The negative coefficient observed toward the end of the sample period may reflect, as in the United States, the substantial negative realizations of the ERP and the occurrence of an inverted yield curve.

The results suggest that, similar to the findings for TMS_{LF} in the United States, TMS_{MT} in Europe is influenced not only by QE but also by major economic and market shocks. Although the relationship between the TMS and the ERP weakens in the post QE period, the results suggest that a certain degree of predictive power is maintained.

For Japan, we focus on the parameter α_1 associated with TMS_{MT}, for which predictive power disappears once the sample is divided, as reported in Table 6. The posterior mean of α_1 fluctuates around zero in an unstable manner from 1980 to 1990, and its positive magnitude increases in the early 1990s following the burst of the asset price bubble. Thereafter, the positive magnitude declines, and from 1995 onward the posterior mean again fluctuates around zero. These patterns suggest that, for most of the sample period, the coefficient of TMS_{MT} with respect to the ERP remains close to zero, implying that predictive power is extremely weak. These findings are consistent with the results in Table 6, which show that the number of statistically significant coefficients of the TMS for the ERP, both before and after QE, is smaller in Japan than in the United States and Europe. The results suggest that, in Japan, the predictive power of the TMS for the ERP remains weak throughout the sample period, irrespective of QE.

In summary, the relationship between the TMS and the ERP exhibits regional differences in the variability of TMS_{MT} and TMS_{LF}, and the time varying patterns of these coefficients also differ across regions. Although it is plausible to consider that some degree of predictive power remains in the post QE period in the United States and Europe, the results suggest that this relationship is frequently affected by major economic and market shocks. In Japan, predictive power appears unstable throughout the sample period, regardless of QE. Taken together, the findings in Section 4.1 and this section suggest that regional and temporal replicability of the predictive power of the TMS is not confirmed[10].

5. Conclusion and future research

This paper conducts an examination of the predictive power of the TMS for the ERP from the perspectives of reproducibility and replicability. Section 2 reviews the prior literature and confirms that economic analyses have primarily focused on reproducibility, and that the term structure of interest rates has been widely used as an explanatory variable in forecasting economic and financial conditions. We also confirm that the ERP has frequently been adopted as the forecasting target and that the United States has been the principal region of analysis. Section 3 analyzes the predictive power of the TMS for the ERP in the United States, Europe, and Japan, and examines the reproducibility of prior studies. The results confirm that the empirical findings obtained using the methodology of prior research are reproducible in all three regions. However, the estimation results suggest that regional and temporal replicability may not be ensured. Section 4 investigates the time variation in predictive power across regions. Sub-sample analyses based on the introduction of QE indicate that predictive power weakens in the post QE period in the United States and Europe, while predictive power remains weak in Japan both before and after QE. At the same time, the time varying parameter regression results suggest that some degree of predictive power persists in the United States and Europe even after the introduction of QE, whereas predictive power in Japan remains unstable throughout the sample period.

The contributions of this paper are as follows. First, we confirm the reproducibility of prior findings regarding the predictive power of the TMS for the ERP. Second, through regional and temporal analyses, we show that replicability, defined as reproducing similar results using different data from those employed in prior studies, is not confirmed. The results suggest that in economic and financial market analysis, structural differences in macroeconomic conditions across regions, time specific economic and market shocks, and the introduction of policy measures in response to such conditions may materially affect the relationship between the TMS and the ERP. As a consequence, regional and temporal heterogeneity may arise in the predictive relationship between the TMS and the ERP.

Several challenges for future empirical research in economics and financial markets can be identified. First, it is important to examine replicability to the greatest extent possible within the range of available data. With respect to the regions under analysis, it is desirable to obtain data covering a broader set of countries. In addition, for a given region, it is preferable to conduct analyses using multiple sub-sample splits across

[10] In addition, Appendix B conducts a time varying parameter analysis of the effects of real GDP growth and inflation, which are key indicators of economic fundamentals, on the TMS. This analysis allows us to examine the stability of the TMS itself as an indicator.

different time periods. Employing time varying parameter models is also a useful option in this context. Second, when assessing replicability across diverse datasets, it is important to investigate cases in which the pre specified model does not perform well, to attempt to improve its explanatory power, and to identify the conditions under which the model does or does not function effectively. Such efforts would enable other researchers to apply the model with a clearer understanding of its properties and limitations. Of course, the prior studies examined in this paper all document statistically significant relationships among the relevant variables and have made important contributions to the literature on forecasting based on the TMS. In addition to these contributions, it is plausible to consider that future empirical research in economics and financial markets should explicitly address situations in which previously significant results become insignificant due to structural changes in the economy or financial markets. By doing so, researchers can enhance replicability while further advancing the development of the literature.

References

[1] Amatov, A., Dorfman, J. H. (2017) The effects on commodity prices of extraordinary monetary policy. Journal of Agricultural and Applied Economics, 49(1), 83–96. https://doi.org/10.1017/aae.2016.34

[2] Bank of Japan (2024) "Review of Monetary Policy from a Broad Perspective"

[3] Campbell, J. Y. (1987) Stock returns and the term structure. Journal of Financial Economics, 18(2), 373-399. https://doi.org/10.1016/0304-405X(87)90045-6

[4] Campbell, J. Y., Thompson, S. B. (2008) Predicting excess stock returns out of sample: Can anything beat the historical average? Review of Financial Studies, 21(4), 1509–1531. https://doi.org/10.1093/rfs/hhm055

[5] Chang, A.C., Li, P. (2022) Is economics research replicable? Sixty published papers from thirteen journals say "often not". Critical Finance Review, 11(1), 185-206. https://doi.org/10.1561/104.00000053

[6] Corbet, S., Dunne, J. J., Larkin, C. (2019) Quantitative easing announcements and high-frequency stock market volatility: evidence from the United States. Research in International Business and Finance, 48, 321–324. https://doi.org/10.1016/j.ribaf.2019.01.00

[7] DeWald, W., Thrsby, J.G., Anderson, R. (1986) Replication in empirical economics: the journal of money, credit and banking project. The American Economic Review, 76, 587-503.

[8] Dreber, A., Johannesson, M. (2005) A framework for evaluating reproducibility and replicability in economics. Economic Inquiry, 63(2), 338-356. https://doi.org/10.1111/ecin.13244

[9] Engstrom, E. C., Sharpe, S. A. (2019) The near-term forward yield spread as a leading Indicator: a less distorted mirror. Financial Analyst Journal, 75(4), 37-49. https://doi.org/10.1080/0015198X.2019.1625617

[10] Estrella, A., Hardouvelis, G. A. (1991) The term structure as a predictor of real economic activity. The Journal of Finance, 46(2), 555-576. https://doi.org/10.1111/j.1540-6261.1991.tb02674.x

[11] Fama, E. F., French, K. R. (1988) Dividend yields and expected stock returns. Journal of Financial Economics, 22(1), 3-25. https://doi.org/10.1016/0304-405X(88)90020-7

[12]Fama, E. F., French, K. R. (1989) Business conditions and expected returns on stocks and bonds. Journal of Financial Economics, 25(1), 23-49. https://doi.org/10.1016/0304-405X(89)90095-0
[13]Faria, L., Verona, F. (2020) The yield curve and the stock market: mind the long run. Journal of Financial Markets, 50, 100508. https://doi.org/10.1016/j.finmar.2019.100508
[14]Fernandez, P. (2023) Equity Premium: Historical, Expected, Required and Implied. Available at SSRN. http://dx.doi.org/10.2139/ssrn.933070
[15]Fisher, I. (1907) The Rate of Interest. The Macmillan Company.
[16]Gagnon, J., Raskin, M., Remache, J., Sack, B. (2011) The financial market effects of the federal reserve's large-scale asset purchases. International Journal of Central Banking, 7(1).
[17]Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. in J. M. Bernardo et al., eds. Bayesian Statistics, 4, New York: Oxford University Press, 169–188.
[18]GPIF (2024) "FY2023 ESG report"
[19]GPIF (2025) "Annual Report Fiscal Year 2024"
[20]Harvey, C. R. (1988) The real term structure and consumption growth. Journal of Financial Economics, 22(2), 305-333. https://doi.org/10.1016/0304-405X(88)90073-6
[21]Hudepohl, T., van Lamoen, R., de Vette, N. (2021) Quantitative easing and exuberance in stock markets: evidence from the Euro area. Journal of International Money and Finance, 18, 102471. https://doi.org/10.1016/j.jimonfin.2021.102471
[22]Nakajima, J. (2011) Time-varying parameter VAR model with stochastic volatility: an overview of methodology and empirical applications. Monetary and Economic Studies, 29, 107-142.
[23]McCown, J.R. (2001) Yield curves and international equity returns. Journal of Banking and Finance, 25(4), 767-788. https://doi.org/10.1016/S0378-4266(00)00099-6
[24]McCullough, B.D., McGeary K.A., Harrison T.D. (2006) Lessons from the JMCB archive. 38(4), 1093-1107. https://www.jstor.org/stable/3838995
[25]Ordu-Akkaya, B.M., Soytas, U. (2020) Unconventional monetary policy and financialization of commodities. North American Journal of Economics and Finance, 51, 100902. https://doi.org/10.1016/j.najef.2018.12.014
[26]Primiceri, G.E., (2005) Time varying structural vector autoregressions and monetary policy. Review of Economic Studies, 72 (3), 821–852. https://doi.org/10.1111/j.1467-937X.2005.00353.x
[27]Rincon, C.J., Petrova, A.V. (2024) Assessing the impact of the ECB's unconventional monetary policy on the European stock markets. Journal of Risk and Financial Management, 17(9), 425. https://doi.org/10.3390/jrfm17090425
[28]Selgrad, J. (2023) Testing the portfolio rebalancing channel of quantitative easing. New York University Working Paper.
[29]Williams, M. (2014) The impact of quantitative easing on asset price comovement. International Finance Markets, Frontiers of Economics and Globalization, 13, 139–163. https://doi.org/10.1108/S1574-8715(2013)0000013013

Appendix A. Chow test for structural change in the TMS

In this analysis, we split the data sample for each region into pre QE and post QE periods. We then conduct an examination of structural change in the TMS using the Chow test. When implementing the Chow test, we consider the level of the TMS, denoted by y_t, and specify Equations (A-1) and (A-2) as follows.

$$y_t = \mu + \varepsilon_t \tag{A-1}$$

$$y_t = m + \phi y_{t-1} + u_t \tag{A-2}$$

In Equation (A-1), μ denotes the mean and ε_t denotes the error term. In Equation (A-2), m denotes the mean, ϕ denotes the coefficient on the lagged level of the TMS, y_{t-1}, and u_t denotes the AR(1) error term. The Chow test examines whether the estimated parameters differ across sub-samples. Accordingly, in this paper we test for structural change in μ from Equation (A-1), ϕ from Equation (A-2), and the squared residual u_t^2. In addition, when y_t is squared and used as a proxy for volatility, we test for structural change in the mean μ' from Equation (A-1) and the coefficient ϕ' from Equation (A-2).

Table A-1 reports the results of the Chow test. For the United States and Japan, μ, u_t^2, and μ' differ significantly between the pre QE and post QE periods. For Europe, u_t^2 and μ' differ significantly across sub-samples. The results suggest that, around the introduction of QE, changes occur in the level and volatility of the TMS in all three regions.

Table A-1. Results of structural change analysis using the Chow test

			US (since November 2008)	Europe (since May 2010)	Japan (since February 1999)
Level (y_t)	Mean	μ	11.93 ***	2.44	4.76 **
	AR(1)	φ	0.92	0.83	0.54
	Variance	u^2_t	3.83 *	14.27 ***	28.37 ***
Vol (y^2_t)	Mean	μ'	18.30 ***	23.74 ***	57.78 ***
	AR(1)	φ'	0.17	2.23	1.37

Each value reports the F-statistic from the Chow test. ***p<0.01, **p<0.05, *p<0.10

Appendix B. TVP-VAR analysis on the determinants of TMS Fluctuations

The analysis in the main text examines the predictive power of the TMS for the ERP. In this section, we conduct a quantitative analysis of the macroeconomic shocks that may drive fluctuations in the TMS. Specifically, we estimate impulse response functions (IRFs) of the TMS to shocks in real gross domestic product (GDP) and the consumer price index (CPI). We allow for the possibility that the effects of macroeconomic shocks vary across policy regimes and market environments. Accordingly, rather than employing a conventional vector autoregression (VAR) model with fixed coefficients, we adopt a specification that permits both coefficients and error variances to evolve over time. In particular, we use a time varying parameter VAR (TVP-VAR) model. Within the TVP-VAR framework, estimation methods based on a mixture sampler approximate the distribution of the parameters of interest, whereas a block sampler avoids such approximation and yields a more accurate characterization of the parameter distribution, as discussed in Nakajima (2011). In this section, we follow Nakajima (2011) and estimate time specific impulse response functions in order to obtain more precise inference.

B.1 Methodology

B.1.1 Specification of the TVP-VAR model

Much of the existing literature on TVP-VAR models builds on the framework of Primiceri (2005). In this paper, following Nakajima (2011), we adopt the specification proposed in Primiceri (2005). Let $y_t = (y_{1t}, \cdots, y_{kt})$ denote a k dimensional vector of endogenous variables at time t, and let s denote the lag length. The TVP-VAR model is then given by Equation (B-1).

$$y_t = c_t + \sum_{i=1}^{s} B_{i,t} y_{t-i} + u_t, \quad u_t \sim N(0, \Omega_t) \tag{B-1}$$

Here, c_t denotes a $k \times 1$ vector of time varying intercepts, $B_{i,t}$ denotes a $k \times k$ matrix of time varying lag coefficients, and Ω_t denotes a time varying variance–covariance matrix. In this paper, we specify $y_t = (GDP_t, CPI_t, TMS_t)'$. Let β_t denote the vector of time varying coefficients obtained by stacking the intercepts and lag coefficients vertically. Defining the regressor matrix as $X_t = I_k \otimes (1, y'_{t-1}, \cdots, y'_{t-s})$, the observation equation is given by Equation (B-2), where I_k denotes the k dimensional identity matrix and $\otimes$ denotes the Kronecker product.

$$y_t = X_t \beta_t + u_t \tag{B-2}$$

For the variance–covariance matrix Ω_t of the error term u_t, we apply a Cholesky decomposition in order to separate contemporaneous interdependence from shock variances. Specifically, Ω_t is decomposed as follows.

$$\Omega_t = A_t^{-1}\Sigma_t\Sigma_t' A_t^{-1\prime} \quad \text{(B-3)}$$

Here, A_t is a lower triangular matrix with ones on the diagonal, representing contemporaneous recursive relationships. Σ_t is a diagonal matrix whose diagonal elements, σ_{it}, denote the standard deviations of the structural shocks. Introducing the orthogonalized structural shocks e_t, Equation (B-4) follows. The contemporaneous impact matrix is given by $S_t = A_t^{-1}\Sigma_t$.

$$y_t = X_t\beta_t + A_t^{-1}\Sigma_t e_t, \qquad e_t \sim N(0, I_k) \quad \text{(B-4)}$$

B.1.2 Time varying parameter dynamics and Bayesian estimation

For estimation, we assume that the lag coefficients, contemporaneous correlations, and shock variances evolve gradually over time. Specifically, for the time varying lag coefficients β_t, the vector a_t obtained by stacking the lower triangular elements of A_t, and the log shock variances $h_t = (log\ \sigma_{1t}^2, \cdots, \log\sigma_{kt}^2)'$, we assume the following state equations.

$$\beta_{t+1} = \beta_t + \varepsilon_{\beta,t}, \qquad a_{t+1} = a_t + \varepsilon_{a,t}, \qquad h_{t+1} = h_t + \varepsilon_{h,t} \quad \text{(B-5)}$$

Each error term is assumed to be mutually independent and to follow a zero mean normal distribution, with variance–covariance matrices denoted by Σ_β, Σ_a, and Σ_h, respectively. These variance matrices are fixed and serve to control the smoothness of parameter variation over time. The model described above can be represented as a state space model with latent variables (β_t, a_t, h_t). In this paper, we specify a prior distribution $p(\beta_{s+1}, a_{s+1}, h_{s+1}, \Sigma_\beta, \Sigma_a, \Sigma_h)$ and conduct inference based on the posterior distribution $p(\beta_{s+1:T}, a_{s+1:T}, h_{s+1:T}, \Sigma_\beta, \Sigma_a, \Sigma_h |\ y_{1:T})$ given the observed data $y_{1:T}$. To draw samples from the posterior distribution, we employ the MCMC method and iteratively update the following blocks based on their conditional posterior distributions.

1. $\beta_{s+1:T} \mid y, a, h, \Sigma_\beta, \Sigma_a, \Sigma_h$
2. $a_{s+1:T} \mid y, \beta, h, \Sigma_\beta, \Sigma_a, \Sigma_h$
3. $h_{s+1:T} \mid y, \beta, a, \Sigma_\beta, \Sigma_a, \Sigma_h$
4. $\Sigma_\beta, \Sigma_a, \Sigma_h \mid \beta, a, h$

Because the conditional posterior distributions of β_t and a_t reduce to a linear Gaussian state space model, the entire time series of these parameters is jointly updated using a simulation smoother. In contrast, h_t corresponds to a stochastic volatility specification and is updated using a block sampler. Based on the estimated TVP-VAR model, we compute impulse response functions at each point in time t. Using the time specific parameters (β_t, A_t, Σ_t), the VAR is transformed into its vector moving average (VMA) representation. The h period ahead impulse response is then given by Equation (B-6).

$$IRF_t(h) = C_{t,h} A_t^{-1} \Sigma_t, \qquad \text{(B-6)}$$

Here, $C_{t,h}$ denotes the moving average coefficient matrix constructed from the lag coefficients at time t.

B.2 Estimation results based on the TVP-VAR model

Figures B-1 through B-3 report the estimated IRFs for the United States, Europe, and Japan, respectively. The data used for estimation are quarterly, covering the period from 1979:Q4 to 2025:Q3, with 184 observations. For each region, we estimate a TVP-VAR model that includes TMS, GDP, and CPI. In the figures, we plot the responses one year, two years, and three years after the occurrence of a shock, based on the parameters estimated at each point in time. Unlike a conventional fixed coefficient VAR model, which produces a single IRF for the entire sample, the TVP-VAR model allows the IRF to be computed at each time t. For ease of exposition, however, we present only the responses at three horizons. Our primary focus is on

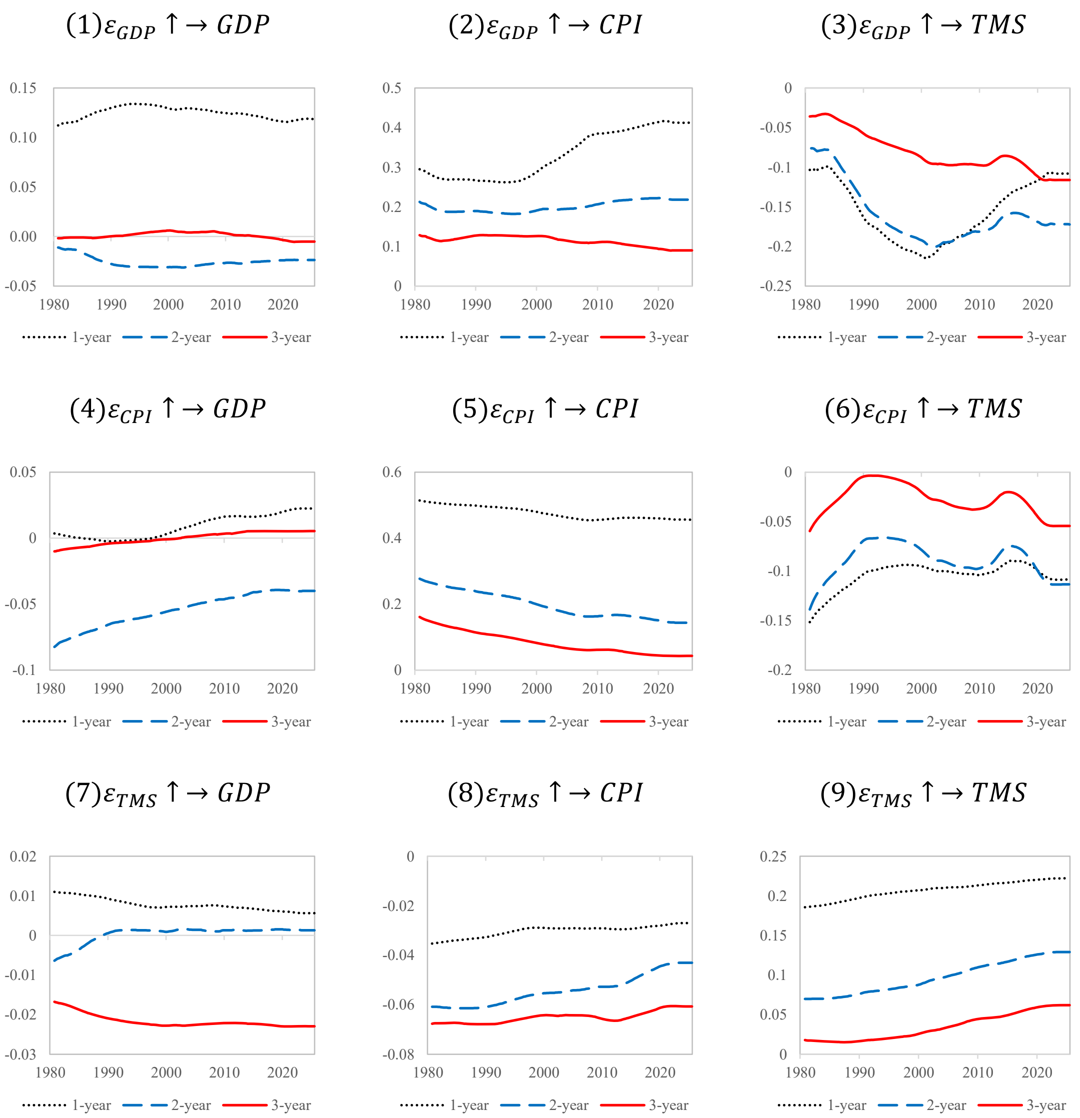


Figure B-1. Posterior mean of impulse response functions for the US the response of TMS to shocks in GDP and CPI.

We begin with the United States in Figure B-1. Panels (3) and (6) show that, throughout the sample period, the response of TMS to positive shocks in GDP and CPI is generally negative. This may reflect a mechanism whereby upward shocks to either variable increase expectations of policy rate hikes by the central bank, the FRB. As a result, upward pressure is exerted primarily on short and medium term yields along the yield curve, leading to a compression of the term spread.

An important finding is that the impulse response of TMS_t to both GDP_t and CPI_t shocks varies substantially over time. Such time variation cannot be captured by a standard fixed coefficient VAR model. The results suggest that incorporating time varying parameters through a TVP-VAR framework is essential to reveal these dynamics.

From 1980 through the 1990s, the negative response of the TMS to CPI shocks becomes smaller in magnitude, whereas the response to GDP shocks moves in the opposite direction. In the early 1980s, the United States experienced persistently high inflation. In such an environment, an upward shock to CPI may have heightened concerns about sustained inflationary pressures, generating upward pressure across the entire yield curve. As a result, when measured in terms of the term spread, the net impact may have been relatively limited.

After 2000, the negative impulse response of the TMS to CPI shocks narrows in the 2010s but gradually widens over the broader post 2000 period. During the 2010s, inflation remains relatively stable around the FRB's 2 percent target, and the FRB implements accommodative monetary policy. In this context, upward inflation shocks may not have led to a substantial strengthening of market expectations of policy rate hikes.

With respect to GDP shocks, the one year ahead impulse response becomes progressively less negative toward the end of the sample after 2000. Moreover, after 2020, the magnitude of the negative response does not deepen further. It is also noteworthy that panel (2) indicates that the effect of GDP shocks on CPI remains broadly stable over time. Following the COVID-19 shock, temporary price increases, initially concentrated in goods, spillover into more persistent service prices, and the FRB requires time to contain inflationary pressures. Based on this experience, it could be possible to consider that after 2020, demand side GDP shocks exert a stronger influence on CPI, thereby reinforcing expectations of policy tightening and compressing the TMS. Contrary to this possibility, however, our empirical results suggest that, when measured in terms of the effect of a one unit shock, the magnitude of the response remains broadly stable over time.

For Europe, the response of the TMS to GDP shocks is negative throughout the sample period, similar to the United States. Moreover, the responses at the two year and three year horizons exhibit relatively limited variation over time. In contrast, the negative magnitude of the one year ahead response becomes smaller during the 2010s. This may reflect the fact that, amid a series of accommodative monetary policies implemented by the ECB, including the introduction of negative interest rates, upward GDP shocks do not substantially strengthen market expectations of policy rate hikes. Regarding CPI shocks, the one year ahead response of the TMS gradually becomes less negative from 1980 through the 2010s. However, after 2020, the negative responses at all horizons from one to three years increase in magnitude. Following the surge in inflation after the COVID-19 shock, CPI shocks may have become more likely to translate into expectations of policy tightening by the ECB. This may exert upward pressure on short term yields, particularly the two

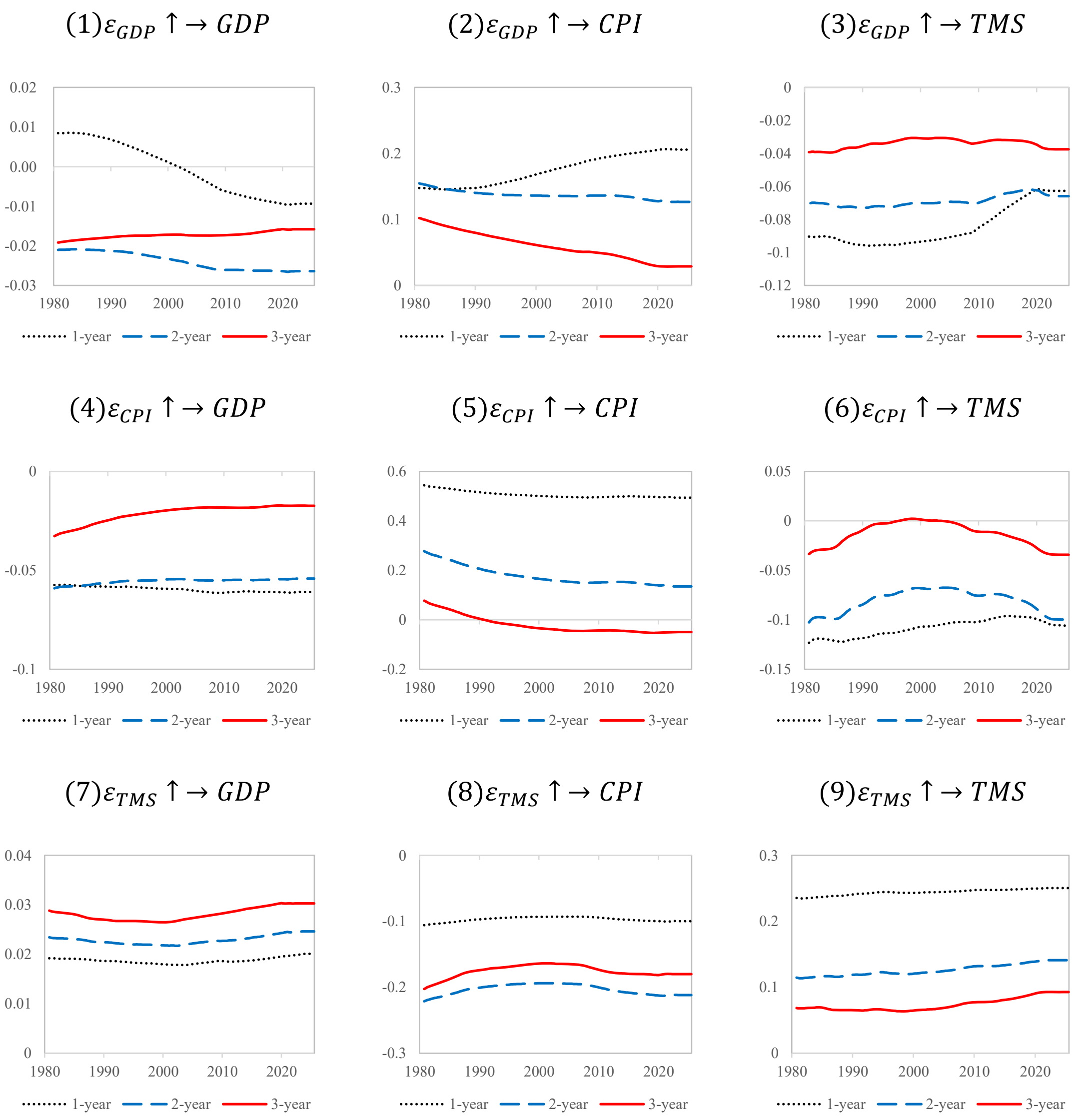


Figure B-2. Posterior mean of impulse response functions for Europe

year yield, thereby contributing to a compression of the TMS.

For Japan, the effects of both GDP and CPI shocks vary substantially over time. The response of the TMS to GDP shocks gradually becomes less negative after 1980 and turns positive after 2020. A positive response is not observed in the United States or Europe and appears unique to Japan. Following the collapse of the asset price bubble in 1990s, Japan's potential growth rate trends downward, and the BOJ maintains accommodative monetary policy for an extended period. In this environment, GDP shocks increasingly have a limited effect on expectations of policy rate hikes, particularly through the 2010s, and the response of the TMS moves close to zero. After 2020, Japan experiences a structural shift toward inflationary period accompanied by wage growth. In this context, GDP shocks may generate not only expectations of short term policy tightening but also higher long term growth expectations. This may lead to an expansion of the TMS rather than a compression. With respect to CPI shocks, the response peaks around the high inflation period near 1990. However, considering the underlying trend, the response after 2020 resembles that of GDP shocks, in that rising expectations of more persistent inflation may contribute to an expansion of the TMS.

The results suggest that, in the formation of the TMS reflecting economic fundamentals, upward pressure on interest rates stemming from stronger GDP growth and higher CPI inflation is increasingly restrained by accommodative monetary policy in the United States and Europe, leading to a gradual attenuation of the TMS response over time. In contrast, in Japan, prolonged monetary easing appears to shift the response of the TMS to GDP and CPI shocks from the short end toward the long end of the yield curve. This may reflect a transformation in which the TMS is formed with greater emphasis on forward looking expectations. Taken together, the findings suggest that once the effects of monetary easing become sufficiently pervasive, market expectations regarding long term growth and inflation may play a more prominent role in shaping the TMS.

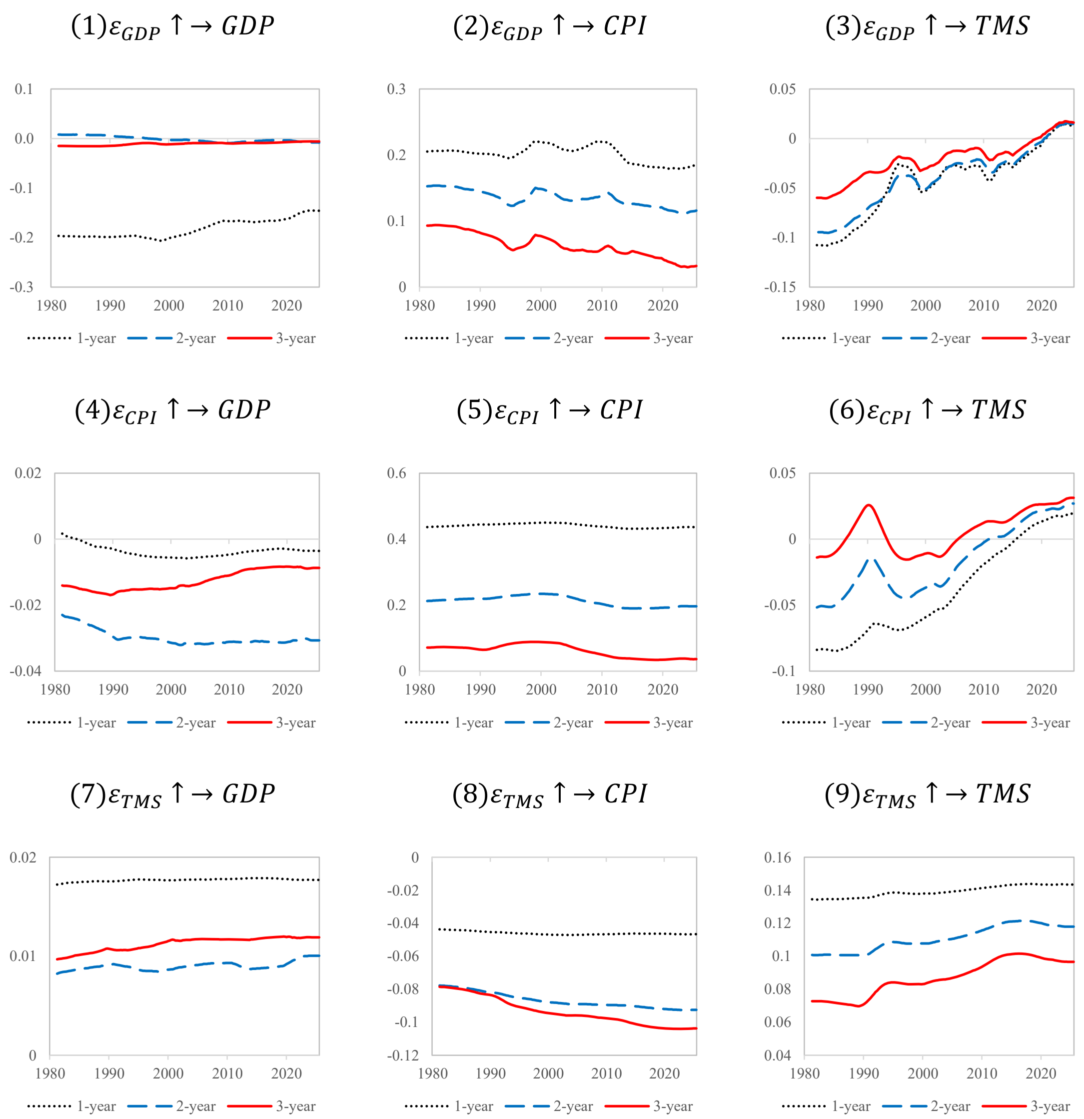


Figure B-3. Posterior mean of impulse response functions for Japan