Decomposing long bond returns: A decentralized modeling approach*

Peter Carr[a†], Liuren Wu[b‡]

[a]Courant Institute, New York University, 251 Mercer Street, New York, NY 10012, USA

[b]Baruch College, Zicklin School of Business, One Bernard Baruch Way, New York, NY 10010, USA

Abstract

This paper develops a decentralized theory by decomposing the return on a particular bond with respect to the variation of its own yield. The theory links the fair value of the bond to the conditional forecasts of its own yield's expected rate of change, risk, and risk premium. We apply the theory to predict returns on a long-term bond by assuming no predictability on its yield movement; to perform comparative yield curve analysis by imposing common factor structures on yield dynamics; and to form stable, decentralized butterfly portfolios with bonds of nearby maturities by introducing the concept of local commonality.

JEL Classification: C13; C51; G12; G13

Keywords: Bond returns; return attribution; yield decomposition; expectation; risk premium; convexity effect; common market price of bond risk; local commonality; butterfly trades

---

*The authors thank Zhenyu Cui, Ionut Florescu, Nicola Fusari, Francis Longstaff, Scott Joslin, Joseph Santos, and seminar participants at Baruch College, City University of New York, the Financial Engineering Seminar at Stevens Institute of Technology, the Dykhouse Proram Workshop Series in Money, Banking, and Finance at South Dakota State University, the 2016 JHU-AQR Conference on "The Role of Derivatives in Asset Pricing," the 2017 NYU Conference on "Derivatives and Volatility 2017: The State of the Art," and the 2018 American Finance Association Meetings for their comments and suggestions. Liuren Wu gratefully acknowledges the support by a grant from The City University of New York PSC-CUNY Research Award Program.

†Tel: +1 646 997 3539. *E-mail address:* petercarr@nyu.edu.

‡Tel: +1 646 312 3509. *E-mail address:* liuren.wu@baruch.cuny.edu.

1. Introduction

The literature on interest rate modeling is vast, with different approaches targeting different challenges. The traditional literature on expectation hypothesis (e.g., Fama and Bliss (1987) and Campbell and Shiller (1991)) focuses on predicting changes in short-term interest rates with the yield curve slope. The literature on dynamic term structure models (DTSMs) strives to impose cross-sectional consistency on the shape of the yield curve by specifying the instantaneous interest rate dynamics and applying the principle of no dynamic arbitrage to obtain pricing implications on the whole yield curve (e.g., Duffie, Pan, and Singleton (2000)). The literature on interest-rate options, on the other hand, often takes the observed yield curve as given and focuses on modeling the interest-rate volatility (e.g., Heath, Jarrow, and Morton (1992)) or inferring the risk-neutral interest rate distribution from the observed option prices (Breeden and Litzenberger (2013)).

All these frameworks, however, have limited success in explaining the short-term return behavior of long-dated bonds. The expectation hypothesis literature proposes to use long rates to predict short rate changes, not the other way around. The options literature focuses on interest-rate volatility and risk-neutral distribution, saying little about bond risk premium. The DTSMs can be used to fit the yield curve shape well with cross-sectional consistency, but the estimated factor dynamics tend to produce poor forecasts on future interest rate movements (Duffee (2002)); worse yet, the underlying yield curve factors may not even span the movements of bond excess returns (Duffee (2011)). Besides, modeling long rates with DTSMs often necessitates making long-term projections on the short rate. The long-run forecast is often generated based on some mean-reverting dynamics assumption; yet, mean reversion calibrated to the short end of the yield curve often implies much lower volatility for the variation of long-term rates than actually observed from data (Giglio and Kelly (2018)). In reality, long rates are neither easily predictable, nor converging to a constant any time soon. They tend to move randomly, with substantial volatility, a behavior that is difficult to be reconciled within most short rate dynamics specifications.

In this paper, we propose a new decentralized theory that provides pricing insights for a particular bond or bond portfolio based on conditional forecasts on its own yield's expected rate of change, its volatility, and the market price of risk on that particular bond. The new theory compliments and contrasts with the centralized DTSM approach. While a DTSM provides a centralized view for cross-sectional consistency and relative valuation across different bonds, the decentralized theory prices the yield of a bond based on the mean and variance forecasts on the changes of the yield itself. While DTSMs have been used to identify mispricing opportunities for statistical arbitrage trading (e.g., Duarte, Longstaff, and Yu (2007) and Bali, Heidari, and Wu (2009)), the new framework is particularly suited to lever an investor's domain expertise on a particular set of bonds, and make investment and risk management decisions based on classic risk-return tradeoff analysis.

The new theory starts by performing an instantaneous return attribution to a bond investment with respect to the variation of its own yield to maturity. Taking expectation on the return attribution under the risk-neutral measure and imposing the no-dynamic-arbitrage condition, we obtain a simple algebraic pricing relation on the bond's yield. The pricing equation decomposes the fair value of the bond yield into three components: its expected rate of change, its risk premium, and the convexity effect driven by the variance of the yield change. The three components represent three conditional forecasts. The pricing at a given point in time does not depend on where the forecasts come from and how the forecasts vary over time.

Since the yield on each bond can be analyzed on its own, there is no pricing error contagion effect from one bond to another, which can happen when estimating a centralized model. Since the pricing of the current yield level only depends on the current mean, variance, and risk premium forecasts on its own change, one does not need to make any assumptions on the long-run dynamics of this yield or any other interest rates. Since the sources and dynamics of the conditional forecasts have no bearing on the pricing, one can obtain the forecasts either from separate statistical model estimation or outside expert survey, greatly facilitating collaboration across teams with different domain expertise.

We explore applications of the new theory in predicting individual bond returns, in performing comparative yield curve analysis and forecasting yield curve slope changes, and in investing in robustly constructed butterfly bond portfolios based on classic mean-variance risk-return analysis. To predict the excess return on an individual bond, we recognize the extremely difficulty in predicting the directional movement of long-term yields (Duffee (2013)). We take this difficult-to-predict feature as our starting point by assuming that long-term constant-maturity bonds follow random walk dynamics. We then extract the market pricing of the underlying bond from the yield spread by separating out the convexity effect based on historical variance estimators on yield changes. We perform empirical analysis on US long-term swap rates and treat them as the coupon rates for par bonds. The analysis shows that the risk premium extracted from each long-term par bond can be used to predict the future excess returns on that bond. The out-of-sample forecasting performance is significantly better than the performance of standard predictive regressions in the literature.

When an investor desires to compare a selected basket of bonds, the new theory can be used for the comparative analysis by directly comparing the risk and risk premium behaviors of the underlying yields. We illustrate this application by performing comparative analysis on the US swap rate curve. To do so, we directly estimate the convexity effect on each swap rate using historical volatility estimators on its past movements; we assume a common market price of bond risk across all swap rates based on previous evidence; and we impose a smooth term structure of decay on the expected rate of swap rate change from short to medium maturities, while maintaining the no-predicability assumption on the change of long-term rates. The parsimonious structure allows us to separate a common market price of risk component from the expectation variation. We propose a state-space estimation procedure to extract the common market price of risk factor and the expectation component from the observed swap rate curve. Historical analysis shows that the extracted common market price of risk factor can be used to predict future excess returns on bonds of all maturities, and the extracted expected rate of change differences across maturities strongly predict future movements in the swap rate term structure slope.

Investors often combine bonds of different maturities to hedge away systematic interest-rate movements. Constructing butterfly portfolios with three bonds has been a staple trade in the fixed income market for this purpose. By hedging away major movements in interest-rate levels and possibly slope changes, a well-constructed butterfly portfolio can become very stable, allowing investors to achieve targeted exposures to the curvature of the yield curve, as well as potential temporary mispricings on the particular bonds. Historically, DTSMs have been used to construct butterfly portfolios for statistical arbitrage trading on the model's pricing errors. In this paper, we apply the new decentralized theory to construct stable and decentralized butterfly portfolios with bonds of nearby maturities, and propose new investment strategies on these decentralized butterflies, not based on statistical arbitrage of some model-based pricing errors, but based on classic risk-return tradeoff analysis on the conditional mean and variance forecasts on the bond portfolio's excess returns. The new decentralized conditional mean-variance forecasts based pricing theory provides the natural framework for comparing the risk-return tradeoffs of different bonds and bond portfolios. We show that the mean-variance investment strategy on the butterflies can generate more robust out-of-sample performance than traditional model-based statistical arbitrage strategies.

Representing the relative value of a bond in terms of its yield to maturity and measuring its risk exposures via the bond's duration and convexity have long been the industry standard and the starting point of most fixed income textbooks. Christensen and Sorensen (1994) and Chance and Jordan (1996) show how these risk measures can be applied to perform bond return attributions.This paper builds a new pricing theory based on a very traditional foundation by deriving the pricing implications from traditional bond return attribution analysis with respect to its own yield to maturity. The new theory integrates no-arbitrage pricing naturally into the classic risk management, return attribution, and risk-return investment analysis.

Interest-rate sensitivities, including the duration and convexity definitions, can be, and have been, measured relative to other types of risk representations, such as the parallel, slope, or other shape shifts of a yield curve, the shifts of some key rates, or the shifts of some common risk factor structures under a term structure model. The one-to-one, monotone, and unique mapping between the price of a bond and its yield to maturity makes the yield-to-maturity representation and the associated duration and convexity risk exposure definitions particularly suited for the development of a decentralized pricing theory. Developments of decentralized pricing theories for other financial securities depend crucially on identifying the corresponding relative value transformation for the particular security type that retains the one-to-one, monotone, and unique mapping to the security price.

2. A decentralized theory of bond yields

We consider an infinite-horizon continuous-time economy. Uncertainty is represented by a filtered probability space $\{\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}\}$, where $\mathbb{P}$ is the physical measure. We assume the usual conditions of right continuity and completeness with respect to the null sets of $\mathbb{P}$. We further assume the existence of a money market account associated with an instantaneous interest rate (or "short rate" for short) $r_t \geq 0$. The assumption of no dynamic arbitrage implies the existence of an equivalent martingale measure $\mathbb{Q}$, often labeled as the risk-neutral measure, associated to this money market account as the numeraire.

Let B_t denote the time-t value of a riskfree bond with a stream of N cash payments $\{\Pi_j\}_{j=1}^N$ at times $\{t+\tau_j\} \geq t$ for $j = 1, 2, \cdots, N$, with τ_j denoting the time to maturity of the jth payment. Traditional DTSMs start by modeling the dynamics of the instantaneous interest rate r_t and value the bond via the following expectation operations,

$$B_t = \sum_{j=1}^{N} \Pi_j \mathbb{E}_t^{\mathbb{P}} \left[\left(\frac{d\mathbb{Q}}{d\mathbb{P}} \right) e^{-\int_t^{t+\tau_j} r_u du} \right] \quad (1)$$

$$= \sum_{j=1}^{N} \Pi_j \mathbb{E}_t^{\mathbb{Q}} \left[e^{-\int_t^{t+\tau_j} r_u du} \right]. \quad (2)$$

where $\mathbb{E}_t^{\mathbb{P}}[\cdot]$ and $\mathbb{E}_t^{\mathbb{Q}}[\cdot]$ denote the expectation operator conditional on time-t filtration $\mathcal{F}_t$ under the physical measure $\mathbb{P}$ and the risk-neutral measure $\mathbb{Q}$, respectively, and $\frac{d\mathbb{Q}}{d\mathbb{P}}$ defines the measure change from $\mathbb{P}$ to $\mathbb{Q}$ via the specification of the market pricing of various risks. Through the expectation operation, bonds with cash flows at all times are chained together via either the centralized modeling of the instantaneous interest rate dynamics and its market pricing of risk in (1), or the centralized modeling of the instantaneous interest rate risk-neutral dynamics in (2). In particular, valuing long-dated bonds necessitates projections of the short rate dynamics over the long horizons of the bond cash flows.

2.1. Decentralized return attribution of bond investments

An investor can invest in a very long-dated bond for a very short period of time. In this case, the investor worries more about the short-term value fluctuation of the bond in question than about long-term projections of the short rate. Even for investors with a long investment horizon, managing the daily return fluctuation remains vitally important. Given these practical considerations, our new pricing framework does not rely on the long-term projection of a centralized instantaneous interest rate, but builds on the decentralized return attribution of the particular bond investment.

To decentralize the return attribution, we represent the bond value not as a function of some common factors, but in terms of its own yield to maturity. Given the price of a particular bond B_t, its yield to maturity y_t is defined via the following equality,

$$B_t \equiv \sum_{j=1}^{N} \exp\left(-y_t \tau_j\right) \Pi_j. \tag{3}$$

The yield to maturity is decentralized in the sense that it is unique and particular to the bond in question. Given the set of cash flows for the bond, the level of the yield to maturity has a unique and monotone mapping with the price of the bond. Absence of arbitrage between this bond and cash requires that the bond be priced such that its yield is positive.

With the decentralized yield to maturity as defined in (3), we can attribute the instantaneous return of the bond investment with respect to the variation of its own yield to maturity as,

$$\frac{dB_t}{B_t} = \frac{\partial B_t}{B_t \partial t} dt + \frac{\partial B_t}{B_t \partial y} dy + \frac{1}{2} \frac{\partial^2 B_t}{B_t \partial y^2} (dy)^2 + o(dt), \tag{4}$$

where $o(dt)$ denotes higher-order terms of dt when yield moves diffusively. When the yield can jump randomly, the jump can induce more significant higher-order terms. We henceforth assume that the next move for the yield is continuous, and attribute the return solely to the time effect and the first and secondorder effects from the yield movement. Since the attribution focuses on the bond return over the next instant, the continuity assumption applies only for the next instant. The results hold even if the yield can jump at any other times.

From the yield to maturity definition in (3), we can derive the three sensitivity terms in (4) as

$$\frac{\partial B_t}{B_t \partial t} = y_t, \quad -\frac{\partial B_t}{B_t \partial y_t} = \sum_{j=1}^{N} w_j \tau_j, \quad \frac{\partial^2 B_t}{B_t \partial y_t^2} = \sum_{j=1}^{N} w_j \tau_j^2, \tag{5}$$

where the weight w_j on each cash flow j is constructed based on its present value as a fraction of the total value of the bond,

$$w_j = \frac{\exp(-y_t \tau_j)\,\Pi_j}{\sum_{i=1}^{N} \exp(-y_t \tau_i)\,\Pi_i}. \tag{6}$$

The first term of the bond return decomposition in (4) measures the carry effect of the investment: Carrying the bond makes y_t per year if the yield stays constant. The second term measures the directional exposure of the bond investment to the yield to maturity variation. Bond price declines as the yield to maturity increases and the sensitivity is determined by the weighted average maturity, or *duration*, of the bond. The third term captures the nonlinearity, or *convexity*, of the price-yield relation. The positively convex relation increases the bond return when the yield varies more. We use $\mathcal{D}$ and $\mathcal{C}$ to denote the duration and convexity of the bond, respectively,

$$\mathcal{D} \equiv \sum_{j=1}^{N} w_j \tau_j, \quad \mathcal{C} \equiv \sum_{j=1}^{N} w_j \tau_j^2. \tag{7}$$

For a zero-coupon bond, the thus-defined duration is equal to the time to maturity of the bond τ, and the convexity is equal to the squared maturity τ^2. For coupon bonds, the duration and convexity are the value-weighted averages of the maturities and squared maturities, respectively, of all the cash flows.

Assuming diffusive yield moves for the next instant and with the definition of duration and convexity in (7), we can attribute the annualized instantaneous return of the bond investment as,

$$\frac{dB_t}{B_t dt} = y_t - \mathcal{D}\frac{dy}{dt} + \frac{1}{2}\mathcal{C}\frac{(dy)^2}{dt}. \tag{8}$$

Taking expectation on (8) under the statistical measure $\mathbb{P}$ attributes the expected bond investment return to three sources,

$$\mathbb{E}_t^{\mathbb{P}}\left[\frac{dB_t}{B_t dt}\right] = y_t - \mu_t \mathcal{D} + \frac{1}{2}\sigma_t^2 \mathcal{C}, \tag{9}$$

where $\mu_t = \mathbb{E}_t^{\mathbb{P}}[dy_t/dt]$ denotes the time-t expected rate of change on the yield, and $\sigma_t^2 = \mathbb{E}_t\left[(dy)^2/dt\right]$ denotes the time-t conditional variance rate of the yield change. The first term captures the expected return from carry. Bonds with a higher yield generate higher returns on average due to carry. Second, due to the negative bond-yield relation, an expected increase in yield reduces the expected bond return. Third, higher volatility on the yield movements leads to higher expected bond return due to the positive convexity effect.

The decomposition highlights the key risk and return sources of a bond investment. If an investor has no view on the direction of the yield movement, the investor can form duration-neutral bond portfolios with bonds of nearby maturities by assuming that yields at nearby maturities move (mostly) together. The duration-neutral portfolio has minimal exposure to common directional movements of the bond yields, and the long-short positioning of the bonds in the portfolio can be driven by the differences in the carry and convexity benefits of each bonds in the portfolio.

2.2. Decentralized no-arbitrage pricing and yield decomposition

To generate pricing implications, we take expectation under the risk-neutral measure $\mathbb{Q}$ on the bond return attribution in (8),

$$\mathbb{E}_t^{\mathbb{Q}}\left[\frac{dB_t}{B_t dt}\right] = y_t - \mu_t^{\mathbb{Q}} \mathcal{D} + \frac{1}{2}\sigma_t^2 \mathcal{C}, \tag{10}$$

where $\mu_t^{\mathbb{Q}} = \mathbb{E}_t^{\mathbb{Q}}[dy_t/dt]$ denotes the time-t expected rate of yield change under the risk-neutral measure. If we use λ_t to denote the market pricing of the bond Brownian risk, under the diffusion assumption, we can link the expected rate of yield change under the two measures by[1]

$$\mu_t^{\mathbb{Q}} = \mu_t + \lambda_t \sigma_t. \tag{11}$$

The market price of bond risk is positive $\lambda_t > 0$ if bond returns are expected to contain a positive risk premium.

The absence of dynamic arbitrage dictates that the risk-neutral expected instantaneous rate of return on any investment is equal to the instantaneous interest rate r_t. Applying this no-dynamic-arbitrage condition to the risk-neutral expectation in (10) leads to a simple algebraic pricing relation for the bond yield.

Theorem 1 *If the yield of a bond is moving continuously over the next instant, no dynamic arbitrage dictates that the fair spread of this yield over the instantaneous interest rate is linked to its expected rate of change (μ_t), its bond risk premium ($\lambda_t \sigma_t$), and its variance rate (σ_t^2) through the bond's duration $\mathcal{D}$ and convexity $\mathcal{C}$ exposures by*

$$y_t - r_t = \mu_t \mathcal{D} + \lambda_t \sigma_t \mathcal{D} - \frac{1}{2} \sigma_t^2 \mathcal{C}. \tag{12}$$

The pricing relation in (12) is highly decentralized. The fair value of the yield spread $(y_t - r_t)$ on the bond investment is determined by its own expected rate of change forecast (μ_t), its bond risk premium ($\lambda_t \sigma_t$), and its volatility forecast (σ_t). The valuation has no direct dependence on the instantaneous interest rate dynamics or the dynamics of any other yields. Indeed, there should be an indexing i on $(y_t, \mathcal{D}, \mathcal{C}, \lambda_t, \mu_t, \sigma_t^2)$ to emphasize that they are all specific to the particular bond i in question. We omit this indexing to reduce notation cluttering when no confusion shall occur.

Even for the particular bond, the pricing does not rely on the full dynamics of its own yield, but only depends on the conditional estimators of its rate of change μ_t, its volatility σ_t, and its market price of bond risk λ_t. All three estimators can change over time, so can the dynamics of the yield, but none of these changes enter the pricing of the current yield spread. One can bring in the forecasts from any outside sources and

[1] Given the inverse price-yield relation, the market price of yield risk takes the opposite sign of the market price of bond risk.

directly examine their pricing implication under the new theory. These forecasts can come from any model assumptions, algorithms, or expert opinions, allowing maximum flexibility for cross-field collaboration.

2.3. Comparison with centralized DTSMs

The new pricing relation in (12) is obtained by attributing the instantaneous bond return to variations in its own yield to maturity. Keeping the partial derivatives in the return attribution, we can write the pricing equation as a partial differential equation,

$$r_t = \frac{\partial B_t}{B_t \partial t} + \frac{\partial B_t}{B_t \partial y}\mu_t^{\mathbb{Q}} + \frac{1}{2}\frac{\partial^2 B_t}{B_t \partial y^2}\sigma_t^2, \tag{13}$$

which is the result of imposing the no-dynamic-arbitrage condition that the risk-neutral expected instantaneous return on the bond investment is equal to the instantaneous interest rate r_t.

DTSMs value bonds by imposing the same no-dynamic-arbitrage condition; however, the centralized modeling approach does not attribute bond returns with respect to its own yield to maturity, but rather with respect to a set of common factors X_t that determine the instantaneous interest rate dynamics. Assuming diffusive dynamics for the common factor X_t with μ_t^X denoting its risk-neutral drift vector and Σ_t^X its covariance matrix, we can express the same no-dynamic-arbitrage condition via an analogous partial differential equation,

$$r_t = \frac{\partial B_t}{B_t \partial t} + \left[\frac{\partial B_t}{B_t \partial X_t}\right]^{\top} \mu_t^X + \frac{1}{2} tr\left[\frac{\partial^2 B_t}{B_t \partial X X^{\top}} \cdot \Sigma_t^X\right]. \tag{14}$$

The two partial differential equations in (13) and (14) are similar in form and represent the same no-dynamic-arbitrage condition, but they have crucially different implications. First, the relation between the bond price and its own yield is explicit and definitional. As a result, the partial derivatives in (13) are known ex ante. Plugging in these partial derivatives yields the simple algebraic pricing relation in (12) that determines the fair value of the yield at time t as a function of the time-t conditional forecasts on its risk-neutral mean ($\mu_t^{\mathbb{Q}} = \mu_t + \lambda_t \sigma_t$) and variance ($\sigma_t^2$). By contrast, the bond price as a function of the common factor vector, $B(X_t)$, is unknown ex ante. The partial differential equation in (14) is used to solve for this pricing relation, starting with the boundary conditions at the cash payments dates of the bond.

Second, to solve the partial differential equation in (14) backward starting at a bond's payment date T, one must necessarily fully specify the common factor dynamics (μ_s^X, Σ_s^X) across all times s from today t to the payment date T. By contrast, our localized pricing relation only needs the current mean and variance forecasts of its yield change for its current pricing, without the need to know its future dynamics.

Finally, since the common factor dynamics in a DTSM determines the instantaneous interest rate dynamics, once the common factor dynamics are specified, the values of bonds of all maturities are set, true to the centralized nature of the modeling approach. By contrast, knowing the conditional mean and variance forecasts of the yield changes for one bond at time t determines the time-t valuation of the yield for this particular bond, but it does not pin down the conditional mean and variance of yield changes for other bonds, nor the pricing of other bonds, true to the decentralized nature of the new pricing approach.

In principle, one can centralize the new pricing relation by centralizing mean-variance forecasts on yield changes across different maturities via a common factor structure, but a DTSM provides a more natural way of achieving this centralization. On the other hand, bond valuation derived from a DTSM naturally satisfies the new pricing relation: The model-implied yield level must be consistent with the model-implied conditional mean and variance of the corresponding yield change; nevertheless, through the decentralization, one can directly make assumptions and/or statistical estimates on the mean and variance of a yield change, and derive implications on the fair valuation of that particular yield, without the need to make assumptions on any long-run dynamics of the instantaneous interest rate. Therefore, the two modeling approaches do not directly compete, but rather complement each other by starting from different perspectives.

Through centralization, DTSMs can readily achieve its main objective of imposing cross-sectional consistency on the valuation of different bonds. The consistency is achieved by using the same set of common factor dynamics to price bonds of all maturities. The achievement, however, does not come for free. First, it is not easy to specify one common dynamics that matches the behavior of all bonds. As one considers the pricing of more bonds across a wider spectrum of maturities, one is often forced to expand the dimensionality of the common factor structure to accommodate the wider range of behaviors. A model that looks sufficient for a subset of data can generate gross mispricing when the data set is expanded. Second, centralization is good for information consolidation and noisy reduction, but it can also generate contagious mis-valuation across all bonds with the introduction of a data error on a single bond.

Through decentralization, the new pricing theory cannot tell us much about cross-sectional consistency. Each bond is priced to be consistent with its own mean-variance forecasts, but there is no direct way of checking whether the mean-variance forecasts on one bond are consistent with the forecasts on another bond. On the other hand, the decentralization allows one to better lever one's domain expertise on a particular set of bonds to generate better pricing on these bonds based on potentially more insightful forecasts.

The different perspectives also lead to different starting points and accordingly different assumptions. For example, to do centralized pricing under DTSM, one must specify the full dynamics for the instantaneous interest rate. Since our knowledge about the far distant future is limited, researchers often rely on stationarity conditions to generate limiting behaviors. Seemingly innocuous stationarity assumptions, however, can generate grossly violated extrapolation results. Giglio and Kelly (2018) show that mean-reverting dynamics calibrated to the short end of the yield curve often generate variations that are too small for long-term yields. In another example, based on stationarity assumptions, Dybvig, Ingersoll, and Ross (1996) make the bold prediction that long-run forward rates can never fall in the limit. While the prediction is consistent with the assumption and the limit can always be set longer than any observation, the reality is that the observed ultra long-dated interest rates up to 50 years vary as much as rates at shorter maturities. Forward rates at these observable ultra long maturities tend to, more often than not, fall lower than forward rates at intermediate maturities.

In reality, long rates are neither easily predictable, nor converging to a constant any time soon. They tend to move randomly, with substantial volatility, a behavior that is difficult to be reconciled within most stationary short rate dynamics assumptions in DTSMs. The new decentralized theory does not require one to reconcile the current behavior of long-dated yields with any long-run assumptions of short rate dynamics. Instead, one can take the difficult-to-predict nature of long rates as the starting point, and take the high volatility estimates as direct inputs to make inference on the market pricing of that bond's risk, an application we take up in the empirical section.

3. Data and summary statistics

We explore the applications of the new theory using data on US swap rates. The financing leg of the swap contracts is the 6-month LIBOR rate. We treat the swap rates as the coupon rate of a par bond. We obtain the LIBOR and swap rate data from Bloomberg daily from January 4, 1995 to December 28, 2018, spanning for 6,041 business days. The swap maturities include 2, 3, 4, 5, 7, 10, 15, 20, and 30 years. When computing summary statistics and estimating models, we sample the data weekly every Wednesday from January 4th, 1995 to December 26, 2018, for 1,252 weeks. The weekly sampling is to mitigate week-day effects.

Nowadays, swap contracts at ultra long maturities up to 50 years are actively traded across several currencies, but their histories are much shorter and are hence excluded from our analysis.

Table 1 reports the summary statistics of the swap rate series (in percentage points), including the sample average ("Mean"), standard deviation ("Stdev"), skewness ("Skew"), excess kurtosis ("Kurt"), and weekly autocorrelation ("Auto"). Panel A reports the statistics on the swap rate levels. The mean swap rate term structure is upward sloping, while the standard deviation shows a declining term structure, suggesting that longer-term rates vary within a narrower range. The skewness estimates are small and increasingly so with increasing maturity. The excess kurtosis estimates are all negative.

[Table 1 about here.]

The weekly autocorrelation estimates for all swap rate series are close to one (0.996 to 0.998), highlighting the highly persistent (if at all stationary) nature of the interest rate series. Interestingly, while the general perception is that short-term interest rates should be more mean-reverting than long-term rates, the weekly autocorrelation estimates during our sample period are lager for shorter-term rates than for longer-term rates.

Panel B of Table 1 reports the summary statistics of the weekly changes of the swap rates. The mean and standard deviation estimates of the weekly changes are annualized. The annualized sample averages of the weekly changes are all negative, showing that interest rate levels have been trending downward over the past two decades as the economy approaches the late phase of the long credit cycle (Dalio (2012)).

The annualized standard deviation estimates of the weekly swap rate changes represent the unconditional volatility estimators for the interest rate changes. The estimators show remarkable stability across maturities. The volatility estimates range from 0.81 at two-year maturity to 0.93 at five to ten-year maturity. Long-term interest rates vary as much as short-term interest rates.

The skewness estimates of the weekly changes remain small, but the excess kurtosis estimates are all positive. The autocorrelation estimates on the weekly changes are all close to zero.

Figure 1 plots the time series of the swap rates at three selected maturities: 2-year (solid line), 10-year (dashed line), and 30-year (dash-dotted line). In line with the negative mean swap rate change estimates in Panel B of Table 1, the swap rates show a distinct downward trend during the 24-year sample period. We overlay the swap rate time series with the recession band for the US economy. During our sample period, the US economy has experienced two recessions, a minor one in 2001 and the more severe one often dubbed as the great recession in 2008-2009. At the start of a recession, the central bank tends to cut the short-term interest rate in an effort to stimulate the economy. Such actions are reflected in the sharp short rate drop during the shaded recession period. The US economy recovered quickly after the 2001 recession. The short-term swap rate also went up quickly after the initial drop. After the great recession, however, the short-term rate stayed low for a much longer period, and the long-term swap rates continued with the downward trend.

[Fig. 1 about here.]

Based on the weekly changes of the swap rates series, we construct annualized volatility estimators on each series with a one-year rolling window. Figure 2 plots the time series of the rolling volatility estimates at selected maturities, overlaid with the recession bands. The rolling volatility estimates vary strongly over time, reaching peak levels during the 2009 financial crises but having been calming down since then.

[Fig. 2 about here.]

Across maturities, the volatility estimates show both upward and downward sloping term structure patterns. The term structure tends to be downward sloping when the volatility level is high, and upward sloping during more quiet periods. Across the whole sample period, however, the variation of the long-dated 30-year swap rate has never been much less than the variation of the short-dated swap series.

The volatility estimates tend to be high during transition periods in a business cycle when the central bank is actively cutting or raising rates. During these periods, the volatility estimates for short-term rates tend to be higher than the estimates for long-term rates, leading to downward sloping volatility term structures. In the US, the Fed was actively cutting the Fed fund rate target 11 consecutive times in 2001. It was actively raising rates from June 2004 to June 2006 for 17 consecutive times. During the beginning of the great recession, the Fed cut the Fed fund rate aggressively from 4.25% in December 2017 to virtually zero by December 2008. During these periods, the volatility estimators for the short-term rates are higher than for the long-term rates.

During periods with few central bank actions on the short rate, such as during the past decade as the short rate was trapped at virtually zero, the volatility term structure becomes distinctively upward sloping. The upward sloping volatility term structure is interesting but can prove challenging for classic models that assume mean-reverting dynamics on the instantaneous interest rate. By directly taking the volatility estimators as inputs, the decentralized pricing theory has no difficulty accommodating the upward sloping volatility term structure and the large volatility estimates for long-dated interest rate series.

4. Applications

We explore practical applications of the new decentralized pricing theory from three broad angles. First, we perform separate analysis on each individual swap rates series, and propose to predict excess returns on long-dated par bonds by assuming no predictability on the corresponding constant-maturity swap rate series. Second, we perform comparative analysis of the swap rate curve via commonality assumptions on the conditional mean and variance forecasts of the yield changes and the market price of risks. We design a simple factor structure that allows us to separate a common market price of risk component from the expectation on the rate of change across the yield curve. Third, we apply the new theory to construct stable and decentralized butterfly portfolios with bonds of nearby maturities, and propose new investment strategies on these decentralized butterflies based on classic risk-return tradeoff analysis.

4.1. Predicting long bond returns without predicting interest-rate movements

It is difficult to predict long-term interest rate movements. As an application of the new pricing theory, this section takes the difficult-to-predict nature of long-term rates head on, treats it as the starting point, and infers the bond risk premium from the observed interest rate level and interest rate volatility estimators.

4.1.1. Extracting market price of bond risk with no predictability on long-dated yields

We start by assuming that the constant-maturity yield $y_t(\tau)$ at some long fixed time-to-maturity τ moves diffusively as a random walk over the next instant,

$$dy_t(\tau) = \sigma_t(\tau) dW_t^{\mathbb{P}}, \tag{15}$$

where $\sigma_t(\tau)$ denotes the time-t conditional forecast of the volatility rate of this yield. The conditional volatility can vary over time with unspecified dynamics. It is also possible that the volatility forecasts differ for yields of different maturities. The key assumption underlying (15) is the absence of predictability on the long-term interest rate as the time-t expected rate of change is assumed to be zero.

If we denote the market price of the risk for the corresponding bond as λ_t, we can derive the risk-neutral dynamics for the yield as

$$dy_t(\tau) = \lambda_t \sigma_t(\tau) dt + \sigma_t(\tau) dW_t. \tag{16}$$

For the fixed-expiry yield $y_t(T)$ of a particular bond contract, the time-to-maturity $\tau = T - t$ shrinks as the calendar time moves forward, and the yield slides along the term structure. The statistical and risk-neutral drifts of this fixed-expiry yield $y_t(T)$ need to be further adjusted for the sliding along the yield curve,

$$\mu_t = -y_t'(\tau), \quad \mu_t^{\mathbb{Q}} = \lambda_t \sigma_t(\tau) - y_t'(\tau). \tag{17}$$

Plugging the no-prediction assumption in (17) into the new pricing relation in (12), we have

$$y_t = r_t + \lambda_t \sigma_t(\tau) \mathcal{D} - y_t'(\tau) \mathcal{D} - \frac{1}{2} \sigma_t^2(\tau) \mathcal{C}. \tag{18}$$

Rearrange, we have,

$$y_t + y'(\tau) \mathcal{D} = r_t + \lambda_t \sigma_t(\tau) \mathcal{D} - \frac{1}{2} \sigma_t^2(\tau) \mathcal{C}. \tag{19}$$

For zero-coupon bonds, $\mathcal{D} = \tau$ and $y_t + y'(\tau)\tau = \frac{\partial(y_t \tau)}{\partial \tau} = f_t(\tau)$ is the instantaneous forward rate. For coupon bonds, we can directly estimate the yield curve slope against the bond maturity based on observed yields at nearby maturities and treat the slope-adjusted yield $y_t + y'(\tau)\mathcal{D}$ as a forward yield analog.

Equation (19) shows that in the absence of rate prediction, positive market price of bond risk drives the yield curve up with increasing duration whereas convexity drives the curve down. Based on observed yield curve time series, we can construct volatility estimators for yield changes across different maturities to generate a volatility term structure curve $\sigma_t(\tau)$ at each date. We can also use the observed yield curve to infer the yield level and yield curve slope at the corresponding duration. Combining these observations and estimates with the financing cost r_t, we can infer the market price of bond risk.

Proposition 1 *Under the assumption of no-predictability on long-term constant-maturity yields as in (15), the time-t market price of the risk on a long-term bond, λ_t, can be estimated from the time-t observed yield level y_t, the local yield curve slope $y_t'(\tau)$, and the volatility estimator σ_t of the yield change,*

$$\lambda_t = \frac{y_t(\tau) + y'(\tau)\mathcal{D} - r_t + \frac{1}{2}\sigma_t^2(\tau)\mathcal{C}}{\sigma_t(\tau)\mathcal{D}}. \tag{20}$$

4.1.2. Empirical implementation and market price of bond risk estimation

To implement Proposition 1 empirically, we assume that long-dated US swap rates are not predictable, and we take the financing rate (6-month LIBOR) as the short rate r_t. We estimate the local swap rate curve slope $y_t'(\tau)$ at each maturity using a local linear regression of the swap rates against the time to maturity. We construct historical volatility estimators on weekly changes of each swap rate series with a one-year rolling window, and we further smooth out noises in the individual estimators by performing a local linear smoothing regression of the estimators across the term structure.

We treat the swap rates as the coupon rates of par bonds, and estimate the duration and convexity of each par bond as the value-weighted maturity and squared maturity of the bond's coupons and principals. We strip the swap rate curve by assuming piece-wise constant forward rates and value the coupons and principals based on the stripped discount rate curve.

At each date t, for each par bond, we obtain its market price of risk estimate λ_t according to equation (20) with the observed yield level y_t, the estimated local yield curve slope $y_t'(\tau)$, the financing rate r_t, the smoothed historical volatility estimator $\sigma_t(\tau)$, and the estimated duration $\mathcal{D}$ and convexity $\mathcal{C}$ exposures. Figure 3 plots the time series of the extracted market price of bond risk (λ_t) from each swap series from 10 to 30 years. The market prices of bond risk extracted from different swap rate series are similar in magnitude and move closely together. The cross-correlation estimates among the different λ_t series obtained from 10-

to 30-year swap rates average 99.7%. Similar to findings in Cochrane and Piazzesi (2005), our evidence supports the existence of a common market price of risk factor across bonds of different maturties.

[Fig. 3 about here.]

On average, the market price of bond risk estimates are positive, supporting the hypothesis of positive average bond risk premium. Nevertheless, the estimates vary strongly over time. In particular, the estimates went below zero right before the start of the two recessions in 2000 and 2007, respectively. At the end of our sample in December 2018, the market price estimates dip below zero again, foretelling the possible incoming of another recession. Industry folklore tends to use the inversion of a particular segment of the yield curve as the harbinger for an incoming recession. Our market price estimates are also built on the yield curve slope, but are further adjusted for the different convexity contributions at different maturities. As a result, the market price estimates are less dependent on the particular maturity choice.

4.1.3. Predicting long bond excess returns with extracted bond risk premiums

Given the extracted market price of risk $\lambda_{i,t}$ on each long-term par bond i, we can construct the bond risk premium estimator as

$$\widehat{CW}_{i,t} = \lambda_{i,t}\sigma_{i,t}\,. \tag{21}$$

In theory, the instantaneous expected excess return on the corresponding par bond is proportional to the bond risk premium,

$$\mathbb{E}_t^{\mathbb{P}}\left[\frac{dB_{i,t}}{B_{i,t}dt}\right] - r_t = \lambda_{i,t}\sigma_{i,t}\mathcal{D}_i\,. \tag{22}$$

We test theoretical implication empirically by examining whether our bond risk premium estimator has any actual predictive power of future excess returns on the corresponding par bond. We compute the six-month ahead excess return on each par bond based on the stripped discount curve six months later. The choice of the half-year horizon matches the semi-annual coupon payment and the financing rate (LIBOR) horizon.

Since the volatility of yield changes $\sigma_{i,t}$ is similar across maturities, the variation of the bond excess returns increases with the bond duration $\mathcal{D}_i$. We annualized the excess return on each par bond and divide it by the bond duration to make the magnitude of scaled excess returns more comparable across maturities. We examine the predictability of the extracted bond risk premium via the following forecasting regression,

$$ER_{i,t+h} = c_i \widehat{CW}_{i,t} + e_{i,t+h}, \tag{23}$$

where $ER_{i,t+h}$ denotes the annualized and duration-scaled excess return on the ith par bond over the half-year future horizon.

For comparison, we consider as a benchmark the bond risk premium constructed by Cochrane and Piazzesi (2005) with a portfolio of forward rates with maturities from 1 to 5 years. They find that the excess returns on Treasury bonds across different maturities can be predicted by the same forward rate portfolio. To construct this alternative bond risk premium estimator, we strip the forward rate curve from the observed swap rates, and estimate the weight on the forward rate portfolio by performing a forecasting regression of the average future excess returns on the par bonds on the five forward rates,

$$AER_{t+h} = \beta_0 + \sum_{j=1}^{5} \beta_j f_t^j + e_{t+h}, \tag{24}$$

where $AER_{t+h} = \frac{1}{9}\sum_{i=1}^{9} ER_{i,t+h}$ denotes the average excess returns on the nine par bonds from maturities 2 to 30 years. The common bond risk premium factor is constructed from the estimated relation in (24),

$$\widehat{CP}_t = \widehat{\beta}_0 + \sum_{j=1}^{5} \widehat{\beta}_j f_t^j, \tag{25}$$

and we forecast the excess return on each par bond with this common risk premium factor via the following forecasting regression,

$$ER_{i,t+h} = c_i \widehat{CP}_t + e_{i,t+h}. \tag{26}$$

We examine the predictability of the two types of risk premium estimators through an out-of-sample exercise. Starting from January 5, 2005, on each date t, we estimate the forecasting relations from (23) to (26) with a 10-year rolling window. We compute the out-of-sample forecasting error on each date t as the difference between the future realized excess return and the forecasted excess return,

$$e_{i,k,t+h} = ER_{i,t+h} - \widehat{ER}_{i,k,t}, \tag{27}$$

where $\widehat{ER}_{i,k,t}$ denotes the time-t out-of-sample forecast from predictor k on excess return $ER_{i,t+h}$. The out-of-sample forecasting performance of each predictor is measured by the sum squared forecasting error (SSFE) over T out-of-sample observations,

$$SSFE_{i,k,T} = \sum_{t=1}^{T} \left(ER_{i,t+h} - \widehat{ER}_{i,k,t} \right)^2. \tag{28}$$

As in Welch and Goyal (2008), the forecasting performance of each risk premium estimator is compared with the 10-year rolling window historical average of the bond excess return up to that point t,

$$\overline{ER}_{i,t} = \frac{1}{L} \sum_{s=t-L-h}^{t-h} ER_{i,s+h}, \tag{29}$$

with L denoting the rolling window length. Representing the forecasting error on the historical average benchmark as

$$e_{i,0,t+h} = ER_{i,t+h} - \overline{ER}_{i,t}, \tag{30}$$

we compute the SSFE on the historical average benchmark,

$$SSFE_{i,0,T} = \sum_{t=1}^{T} \left(ER_{i,t+h} - \overline{HER}_{i,t} \right)^2, \tag{31}$$

and measure the relative performance of each predictor via an out-of-sample R^2 measure as in Rapach and

Zhou (2013),

$$R^2_{i,k} = 1 - SSFE_{i,k,T}/SSFE_{i,0,T}. \tag{32}$$

A positive R^2 estimate indicates that the predictor outperforms the historical average benchmark.

To test the statistical significance of the forecasting performance difference over the historical average benchmark, we compute the Diebold and Mariano (1995) t-statistics (DM) on the squared forecasting error difference, $\delta_{k,t+h} = e^2_{i,0,t+h} - e^2_{i,k,t+h}$,

$$\text{DM}_k = \frac{\mu_\delta}{\sigma_\delta}\left(\frac{T+1-2(h+1)+h(h+1)/T}{T}\right)^{0.5}, \tag{33}$$

where μ_δ denotes the sample mean of the difference, σ_δ denotes the Newey and West (1987) standard error estimate, computed with a lag equal to the forecasting horizon. The statistics adjust for the small-sample size bias according to Harvey, Leybourne, and Newbold (1997). Under the null hypothesis that each rolling-window estimated model and the historical average benchmark have equal finite-sample forecast accuracy,[2] Clark and McCracken (2012) find that the thus-computed DM test statistic can be compared to standard normal critical values.

Table 2 reports the out-of-sample R^2 estimates and the DM test statistics against the historical average benchmark for each par bond. Panel A reports the performance of the benchmark Cochrane-Piazzesi (CP) bond risk premium estimator constructed with a portfolio of forward rates. Panel B reports the performance of our (CW) bond risk premium estimators based on the assumption on no predictability on the changes of the underlying constant-maturity swap rates. The out-of-sample statistics are computed over 704 weekly observations from January 5th, 2005 to Jun 27th, 2018 for excess returns over the next six months.

[Table 2 about here.]

[2]See Inoue and Kilian (2004) for a discussion on the distinction between population and finite-sample predictive accuracy, and Clark and McCracken (2012) for a thorough review on forecast performance evaluation.

Our no-predictability assumption is meant for long-term swap rates. Accordingly, our bond risk premium estimator is unlikely to be valid for short-term bonds as short-term interest rate movements can in principle be predicted by the slope of the interest-rate term structure. On the other hand, the Cochrane and Piazzesi (2005) forward-rate portfolio was originally constructed for predicting excess returns on short-term bonds with maturities from 2 to 5 years. We report the forecasting performance for both short-term and long-term bonds in Table 2.

Panel A shows that the CP bond risk premium estimator generates negative out-of-sample forecasting R^2 estimates across all maturities. By contrast, Panel B shows that our CW bond risk premium estimators generate positive out-of-sample forecasting R^2 estimates on bonds of all maturities. Furthermore, the DM statistics show that the out-of-sample forecasting performance of the CW bond risk premium estimator is significantly better than the historical average benchmark for long-term bonds with maturities from 5 to 30 years, where the assumption of no interest-rate predictability is more likely to hold.

Cochrane and Piazzesi (2005) identify the forward-rate portfolio with a tent-shaped portfolio weights based on full-sample forecasting regressions from 1964 to 2003, and they show in-sample forecasting R^2 estimates around 40%. Over our full-sample period from 1996 to 2018, a full-sample regression on (24) generates portfolio weights with a qualitatively similar tent shape, and the in-sample forecasting R^2 estimates from equation (26) are universally positive across all par bonds, from 14% for the 30-year bond to 32% for the 2-year bond. The in-sample forecasting R^2 estimates from the CW bond risk premium estimators in equation (23) are lower, from 5% for the 30-year bond to 12-14% for bonds at 2-7 years. Compared to the CP bond risk premium construction with five forward rates as regressors in a forecasting regression, the CW bond risk premium construction is much more rigid as the market price of bond risk λ_t is estimated without performing any forecasting regressions. The rigidity lowers its in-sample fitting performance, but drastically enhances its out-of-sample stability.

As in Rapach and Zhou (2013), Figure 4 plots the cumulative squared forecasting error differences

(CSFED) on the excess returns of selected long-term bonds,

$$CSFED_{i,k,t} = SSFE_{0,k,t} - SSFE_{i,k,t}, \quad t = 1, 2, \cdots, T. \tag{34}$$

Panel A plots the cumulative out-of-sample performance of the CP bond risk premium, and Panel B plots the cumulative performance of the CW bond risk premium estimator. For the CP bond risk premium, the performance is worse than the historical average benchmark across all maturities, and the cumulative performance difference becomes increasingly negative. By contrast, the CW bond risk premium estimators perform uniformly well across all long maturities from 10 to 30 years, and much better than the historical average benchmark. The outperformance is also reasonably uniform across the out-of-sample period as the cumulative performance charts show a steady growth.

[Fig. 4 about here.]

With the assumption of no long-term yield predictability, our bond risk premium estimator is driven by the observed yield curve slope adjusted for the convexity effect. The convexity effect is constructed with a historical interest-rate volatility estimator. By contrast, Cochrane and Piazzesi (2005) construct the bond risk premium purely with a portfolio of forward rates. Nevertheless, the underlying drivers of the two bonds risk premium estimators may not be as far apart as they seem. The forecasting regression on (24) tends to generate tent-shaped portfolio weights. A portfolio of forward rates with tent-shaped weights provide an estimate of the curvature of the forward rate curve. The literature has long linked the curvature of the yield curve to the interest-rate volatility (Heidari and Wu (2003)). Therefore, although from different starting points, the tent-shaped forward-rate portfolio can in principle be treated as an alternative representation of the combined effect of the yield curve slope and the convexity effect. As such, the drastically different out-of-sample performance from the two sets of bond risk premium estimators probably has less to do with whether one can represent the premium in term of a portfolio of interest rates, but more to do with how the portfolio weights are estimated. By nature, forecasting regressions on security returns have low predictive power, and are likely to degenerate in out-of-sample applications. By avoiding multivariate out-of-sample forecasting regressions in constructing our bond risk premium estimator, we achieve more out-of-sample stability.

While most of the traditional expectation hypothesis literature focuses on forecasting short-rate changes with the yield curve slope, several studies in that literature also explore forecasting long-term yield changes with the yield curve slope, with some variations of the following regression specification,[3]

$$\frac{y_{t+h}(T) - y_t(T)}{h} = a + b\frac{(y_t(T) - r)}{\mathcal{D}} + e_{t+h}. \tag{35}$$

We can interpret this regression within our decentralized pricing relation in equation (12) of Theorem 1: The dependent variable represents the annualized realization of the yield changes, the expectation of which gives us the expected rate of change on the fixed-expiry yield, μ_t. The null hypothesis of $a = 0, b = 1$ amounts to the hypothesis of $\mu_t = (y_t - r_t)/\mathcal{D}$, which can be obtained from our pricing relation by assuming zero risk ($\sigma_t = 0$), resulting in both zero risk premium ($\lambda_t \sigma_t = 0$) and zero convexity effect ($\sigma_t^2 \mathcal{C} = 0$). The empirical literature strongly rejects the null hypothesis and instead obtains mostly negative slope estimates from the regression. While this traditional literature starts with the zero risk assumption and strives to predict future rate changes from the yield curve slope, our application in this section takes the estimated risk levels of yield changes as given, and strives to identify the market price on this risk with the assumption of no predictability on the changes of constant-maturity interest rates. Our no-predictability assumption on constant-maturity interest rates attributes the expected rate of change on the corresponding fixed-expiry yield to purely sliding along the yield curve:

$$\mu_t = -y'(\tau) \tag{36}$$

When the yield curve slope is monotonic, the local yield curve slope $y'(\tau)$ shares the same sign as the average slope $(y_t(T) - r)/\mathcal{D}$ used in the regression in (35). The implication of our no-predictability assumption on

[3]Prominent examples include Roll (1970), Campbell and Shiller (1991), Evans and Lewis (1994), and Bekaert, Hodrick, and Marshall (1997). The regressions are usually run on stripped continuously compounded spot rates of zero-coupon bonds.

constant-maturity long-term interest rates in (36) is consistent in sign with the negative slope estimates generally obtained from this traditional literature, suggesting that the yield curve sliding indeed represents a major predictable component of the expected change on the fixed-expiry yield.

Performing a predictive regression can calibrate the magnitude (b) of the slope dependence to a more exact number over an in-sample period, but with no guarantee that the in-sample estimate works out out of sample. The Cochrane and Piazzesi (2005) regression accommodates an even more flexible dependence structure on the yield curve, and thus generates even higher in-sample explanatory power, but the flexible specification also leads to more dramatic out-of-sample degeneration.

The literature has also identified a long list of macroeconomic variables that show predictability on bond returns. Prominent examples include Cooper and Priestley (2009), who link the bond risk premium to the output gap; Ludvigson and Ng (2009), who extract factors from a large set of macro variables to predict bond returns; Greenwood and Vayanos (2014), who link the Treasury bond return to Treasury bond supply; Joslin, Priebsch, and Singleton (2014), who include measures of economic growth and inflation into dynamic term structure modeling; and Cieslak and Povala (2015), who decompose Treasury yields into long-horizon inflation expectations and maturity related cycles. While these studies are informative about the bond risk premium behavior, how to accurately identify bond risk premium in an out-of-sample setting remains a challenging task (Gargano, Pettenuzzo, and Timmermann (2017)). Our market price of bond risk estimator is constructed from the yield curve with highly stylized assumptions, but without performing predictive regressions in its identification, thus allowing it to retain superior out-of-sample performance, and making it a simple and robust candidate in the long list of bond return predictors.

Although our no-predictability assumption on constant-maturity interest rates leads to market price of risk estimators that can predict future bond excess returns out of sample, the assumption is made based less on a strong belief of its validity, but more on the lack of robust out-of-sample predictors on changes of constant-maturity interest rates. When one can identify robust predictors, one can readily incorporate them into our pricing relation to obtain potentially sharper estimates on the bond risk premium and better predictions of future bond returns. A potentially promising source for such predictors is survey forecasts from economists. Kim and Orphanides (2012) propose to use survey consensus to to pin down interest-rate expectation for better identification of the bond risk premium in a DTSM setting. Similar survey data can be used to generate rate change predictions, which can be incorporated into our decentralized pricing relation for bond risk premium identification. We leave the exploration of survey data for future research.

4.2. Comparative yield curve analysis via commonality assumptions on conditional forecasts

The new theory is built to analyze the relation between the yield of a particular bond and the conditional forecasts on its expected rate of change, its volatility, and the market price of its risk. To perform comparative analysis of yields across different maturities, we can make commonality assumptions on the corresponding conditional forecasts. For illustration, this section considers a particularly simple common factor structure on the conditional forecasts of the US swap rate curve and explores its implications.

4.2.1. Common factor structures on the swap rate term structure

Motivated by the evidence from the previous section, we make the following common factor structure assumptions on the interest-rate risk, the market price of risk, and the expected rate of changes on the US swap rates across different maturities:

Assumption 1 (Volatility). *The annualized volatility rates of swap rate changes are equal to their corresponding historical estimators* $\sigma_t(\tau)$.

As in the previous section, we take the historical volatility estimators as our starting points without making explicit assumptions on their dynamics.

Assumption 2 (Market price of risk). *The market price of bond risk is identical across maturities,*

$$\lambda_t(\tau) = \lambda_t. \tag{37}$$

The one-factor structure on the market price of risk is in line with the empirical evidence from the previous section and the literature findings.

Assumption 3 (Expected rate of change term structure). *The expected rates of change on the constant-maturity swap rates vary smoothly across maturities via the following functional form,*

$$\mu_t(\tau) = \mu_t \eta_t(\tau), \quad \eta(\tau) = (\max(0, 1 + \ln(2/\tau))^{\kappa_t}. \tag{38}$$

The assumption builds the expected rate of change term structure through a multiplicative form, with μ_t capturing the time variation of the expected rate of change and $\eta_t(\tau)$ capturing the variation of the term structure shape, which is normalized to be 1 at the short end of the swap curve at two-year maturity and flattened at zero for maturities at 5.44 years or longer. The structure assumes that the expected rate of change across the term structure shares the same sign and its absolute magnitude decays smoothly with increasing maturity from one at two-year maturity to zero at maturities 5.44 years. The power coefficient κ_t controls the curvature of the decay speed within the maturity span.

Out of the three components of the yield decomposition, the convexity effect can be constructed based on the historical volatility estimators, and the market price of bond risks has been found to share a one-factor structure. This leaves the yield change expectation the most variable and the hardest-to-identify component. The previous section makes the identification by focusing on the long end of the yield curve and assuming no-predictability on the changes of long-term constant-maturity swap rate series. This section maintains the same no-predictability assumption on the changes of long-term swap rates and sets the cutoff point around five years based on evidence from the previous section. In addition, the assumption in (38) allows non-zero expected rate of change in the short end of the swap rate curve, and assumes that the absolute magnitude of the expected rate of change decays smoothly with increasing maturity between two to five years.

With assumptions (1) to (3), we can write the term structure of the yield curve as

$$y_t(\tau) = r_t - y'(\tau)\mathcal{D} - \frac{1}{2}\sigma_t^2(\tau)\mathcal{C} + \mu_t\eta_t(\tau)\mathcal{D} + \lambda_t\sigma_t(\tau)\mathcal{D} + e_t, \tag{39}$$

with e_t capturing data measurement noise and pricing errors.

At each date t, in addition to the variation of the observed and estimated quantities, e.g., $r_t, y'(\tau), \sigma_t(\tau)$, the specification governs the yield curve term structure via three variables $(\mu_t, \kappa_t, \lambda_t)$. Intuitively, a positive market price of bond risk λ_t contributes to a positive slope to the term structure. The expected rate of change on the two-year swap rate μ_t further adjusts the slope through the expectation difference across maturities, and the power coefficient κ_t controls the curvature of the slope and the speed by which the expectation contribution declines as maturity increases.

In the DTSM literature, researchers have learned to specify models with arbitrarily many factors with parsimonious factor structures to achieve near-perfect fitting of the yield curve, e.g., Calvet, Fisher, and Wu (2018). Under the new framework, we can specify similarly flexible functionals on the term structure of the expected rate of change to achieve near-perfect fitting of the yield curve. Nevertheless, in the pursuit of fitting the observed yield curve increasingly better, one often loses identification of true market expectation, but ends up with a flexible functional form that only serves as a smooth interpolation of the yield observation. Thus, the flexibility choice on the specification depends crucially on the particular application. In this section, we choose a particularly parsimonious specification with strong identification restrictions for better separation of the risk premium component from the expectation component.

4.2.2. Identifying common factor structures from the observed swap rates

Under the common factor structure assumed in this section, the swap rate curve at time t, as shown in equation (39), depends on the levels of the three state variables at time t, $(\mu_t, \kappa_t, \lambda_t)$, but does not depend on the particular dynamics specification for these variables. Therefore, the emphasis of the empirical analysis involves the extraction of the state variables from the swap rates, without knowing their actual dynamics.

Based on this unique feature, we cast the model into a state-space form, where we treat the three variables as the hidden states and treat the observed swap rates y_t as measurements with errors. We define the state vector as X_t,

$$X_t \equiv [\mu_t, \ln(\kappa_t), \lambda_t]^\top, \tag{40}$$

where the logarithm transformation on κ_t guarantees that it stays strictly positive and the state vector can take values on the whole real line. Since how the state vector varies over time does not affect the pricing, we can specify the state propagation equation without worrying about its pricing implications. We make the particularly simple assumption of a random walk dynamics,

$$X_t = X_{t-1} + \sqrt{\Sigma_x}\varepsilon_t. \tag{41}$$

where the standardized error vector ε_t is assumed to be normally distributed with zero mean and unit variance. We further assume that the covariance matrix of the state vector Σ_x is a diagonal matrix with distinct diagonal values.

We define the measurement equation on the observed swap rates,

$$y_t = h(X_t, \tau) + e_t, \tag{42}$$

where $h(X_t, \tau)$ denotes the value of the swap rate as a function of the state vector X_t and the time to maturity τ as specified in equation (39), and e_t denotes the measurement error. We assume that the errors across different series are iid normally distributed with error variance σ_e^2.

When the state-space model is Gaussian linear, the Kalman (1960) filter provides efficient forecasts and updates on the mean and covariance of the state and observations. Our state-propagation equations are constructed to be Gaussian and linear, but the measurement functions $h(X_t)$ are not linear in the state vector. We use the unscented Kalman filter (Wan and van der Merwe (2001)) to handle the nonlinearity.

The setup introduces four auxiliary parameters that define the covariance matrices of the state propagation errors and the measurement errors. The relative magnitude of the state propagation error variance versus the measurement error variance controls the speed with which the state vector is updated based on new observations. We estimate these auxiliary parameters by minimizing the sum of squared forecasting errors in a quasi maximum likelihood setting (Duffee and Stanton (2012)).

In comparing the swap rates across different maturities, we assume some common structures on their rate of change and market price of bond risk. The assumed common structures lead to common pricing on the swap rate curve; nevertheless, it is important to understand that the pricing is consistent with the assumed dynamics, but there is no direct guarantee that the assumed common factor structure is fully consistent by itself. Hence, to achieve internal consistency across the whole term structure, the traditional approach of DTSM remains the most convenient approach. The common structure assumption in this section is not for achieving cross-sectional consistency, but for enhanced identification and separation of the market pricing of risk component from the expectation component.

4.2.3. Predicting bond excess returns with the common market risk of bond risk

Figure 5 plots in the solid line the time series of the common market price of bond risk λ_t extracted from the US swap rate curves based on the common factor structure assumptions. For comparison, we overlay in dotted lines the four time series of market price of bond risk estimates from the previous section based on no-predictability assumptions on each individual swap rates with maturities from 10 to 30 years. The two sets of estimates follow each other closely. The cross-correlation estimates between the common market price of risk time series and the four series derived from each swap rate series from the previous section are all over 99%.

[Fig. 5 about here.]

With the common market price of bond risk λ_t, we can construct a bond risk premium estimator for each par bond series as $\widehat{CCW}_{i,t} = \lambda_t \sigma_{i,t}$, where the market price of risk is common across all par bonds while the volatility rate $\sigma_{i,t}$ is separate for each swap rate series. We extend the out-of-sample exercise in the previous section to include this common risk premium estimator as another predictor of future par bond excess returns. For the out-of-sample exercise, we estimate the auxiliary model parameters that control the common factor structure using data up to 2004. Then, starting from January 5, 2005, on each date t, we keep the estimated model parameters fixed, run through the unscented Kalman filter to extract the state variables up to that date, and construct the common risk premium estimator $\widehat{CCW}_{i,t}$ for each swap rate series i for that date t. We repeat the out-of-sample exercise and use this common risk premium estimator to predict future par bond excess returns. The results are reported in Panel C of Table 2. The common risk premium estimator can predict bond excess returns of both long and short maturities. The DM test statistics show that the out-of-sample forecasting performance of the CCW bond risk premium estimator is significantly better than the historical average benchmark for bonds of all maturities.

4.2.4. Predicting changes in yield curve slope with the extracted rate of change

With the common factor structure, not only can we extract a common market price of bond risk that predicts bond excess returns across all maturities, but we can also separate out an expectation component for the swap rate curve canges. Figure 6 plots the time series of the extracted state variable μ_t, which measures the expected annualized rate of change at two-year maturity. The estimated rate of change switches signs several times during the sample period. The time series show both large variations following the business cycle and shorter-term temporal variations.

[Fig. 6 about here.]

For identification, we set the expected rate of change for swap rates with maturities longer than five years to zero. As such, μ_t captures more of the relative component of the expected rate of change between short- and long-maturity swap rates, and thus the swap curve slope changes. To examine whether the extracted rate of change is informative about future swap curve slope changes, Table 3 reports the forecasting correlation estimates between the expected rate of change differentials $(\mu_t(\tau_0) - \mu_t(\tau))$ and changes in the swap curve slope over different horizons (h, in weeks), from one month ($h = 4$ weeks), to one quarter ($h = 13$ weeks), and to half a year ($h = 26$ weeks). We measure the expected rate of change differential and swap curve slope between two year ($\tau_0 = 2$) and other maturities (τ) from three to 30 years. The forecasting correlation estimates are all strongly positive. For example, the forecasting correlation on changes of the 2-3 swap slope is 25% at one-month horizon, 44% at quarterly horizon, and 54% at half-year horizon. The forecasting correlation estimates are similar on other slope measures. The high correlation estimates suggest that the simple common factor structure can effectively separate out an expectation component that is highly predictive of future swap curve slope changes.

4.3. Investing in decentralized butterfly bond portfolios

Butterfly portfolios constructed with three bonds have been a staple trade in the fixed income market. By hedging away systematic movements in the interest-rate levels and possibly also slope changes, a well-constructed butterfly portfolio can become very stable, allowing investors to achieve targeted exposures to the curvature of the yield curve, as well as potential temporary mispicings of the particular bonds.

Duarte, Longstaff, and Yu (2007) examine the profitability of the staple trade by constructing butterfly portfolios based on a two-factor Gaussian affine model. Two-factor Gaussian affine models have been the anchor for many quantitative fixed income desks since the 1980s. A three-leg butterfly portfolio can be constructed to neutralize the two common factors. If the two-factor structure is well-specified and the pricing residuals are mean-reverting, the constructed butterfly portfolios can become much more mean-reverting than the individual interest-rate series, and one can treat the negative of the pricing residual as the alpha source. The stability and profitability of such portfolios depend on how well the model specification captures the true underlying factor structure and how successful the model can separate persistent interest-rate movement from the more mean-reverting residual movements due to yield curvature changes and/or temporary mispricings on individual series. Given the centralized nature of a dynamic term structure model, the portfolio construction is in theory agnostic to the particular bond maturity choice: One can in principle hedge away the two interest-rate factors with combinations of any three distinct interest-rate series. Nevertheless, Bali, Heidari, and Wu (2009) find that different maturity combinations can lead to butterfly portfolios with very different behaviors. Some portfolio are very stable and highly mean-reverting, while others are not stable at all.

In this section, based on our new decentralized pricing theory, we propose to construct decentralized butterfly portfolios that do not rely on factor structure assumptions but focus much more on the particular choice of the maturity combination, and we propose to invest in these decentralized butterflies, not based on statistical arbitrage trading of the pricing errors of a term structure model, but based on classic risk-return analysis of the portfolio's excess returns.

4.3.1. Local commonality of interest-rate movements

To motivate our emphasis on the maturity combination choice, we start with the intuitive observation that bonds of nearby maturities tend to behave similarly and co-move strongly. The closer the maturities between two bonds, the closer their behaviors are and the stronger their co-movements. We label this behavior as *local commonality*. It is driven less by any model assumptions, but more by the fact that nearby contracts have similar payoffs and should hence behave similarly regardless of any dynamics assumptions.

To illustrate this local commonality behavior, we choose a set of reference maturities and measure the cross-correlations of swap rate changes between these reference maturities and other maturities. In the US, the most actively traded maturities on swap rates and Treasury bonds center around the maturities of the Treasury bond futures, which are written on bonds at six segments of the maturity spectrum around 3, 5, 7, 10, 15, and 20 years, respectively.[4] We take these maturities as reference maturities and show how their correlations of weeky changes with other maturities decline as the maturity gap increases.

Figure 7 plots the correlation estimates against the maturity gap with the reference maturity. Each line plots the correlation of one reference maturity with other maturities. Our data sample includes swap rates at nine maturities from two to 30 years. The maturity gap measures the gap between the nine maturities and the reference maturity. For example, for the line with 10-year as the reference maturity, the maturity gaps with 5-, 7-, 10-, 15-, and 20-year swaps are -2, -1, 0, 1, and 2, respectively.

[Fig. 7 about here.]

The correlation estimates decline monotonically as the absolute maturity gap increases. The correlations of weekly changes of swap rates at adjacent maturities (with maturity gaps of ± 1) are all higher than 97.78%, with an average of 98.58%. The estimates remain high at the next level (with gaps of ± 2) at no less than 95.27%, with an average of 96.32%. As the gap increases further, the average correlation estimates decline from an average of 92.7% with gaps of ± 3 to an average of 73.86% with gaps of ± 7.

Local commonality is not an exact statement. It is a qualitative observation of an approximate but robust and universal phenomenon. By comparison, when one specifies a global factor structure, its implications become exact, but the exact implications are not necessarily robust in the sense that the exact implications can be wrong when the model assumptions are violated.

To design global factor structures for the term structure of interest rates, researchers often start with

[4]For example, futures with September 2017 delivery include TUU7 (2-year note), FVU7 (5-year note), TYU7 (the old 10-year contract with maturities around 7 years), UXYU7 (the new 10-year ultra, with maturities around 10 year), USU7 (long bond with maturities around 18 years), and WNU7 (ultra bond with maturities over 25 years). Based on the swap rate data we have, we choose the second shortest maturity at 3 years as the first reference maturity, and choose the second to last maturity at 20 years as the last reference maturity.

principal component analysis on the interest-rate series to determine how many principal components are necessary to capture the variation of the term structure. The findings of such analysis, however, depend crucially on the number of interest-rate series and the maturity span of the series used in the analysis. In general, the more series one includes in the analysis and the wider is the maturity span, the more principal components one finds necessary to explain a sufficiently high percentage of the variation of the term structure. To illustrate this point, we perform principal component analysis repeatedly, each time on a different subset of the swap rate series, and report the percentage explained variation of the top four principal components in Table 4. We define the subset based on the maturity range, reported in the first column of the table. For each chosen subset of the swap rate series, we measure the cross-correlation matrix of the weekly changes in the swap rates, and compute the eigenvalues of the correlation matrix. We normalize the eigenvalues to sum to 100%, and interpret each normalized eigenvalue as the percentage explained variation by that principal component.

[Table 4 about here.]

The first subset focuses on the short end of the term structure, including four swap rate series with maturities from two to five years. The historical literature on expectation hypothesis, including the bond risk premium analysis by Cochrane and Piazzesi (2005), often focuses on the short end of the term structure, with maturities usually below five years, where the convexity effect is negligible and the expectation of future interest rate movements dominates the term structure shape. The principal component analysis on this subset shows that one principal component can explain 97.3% of the variation. Therefore, a simple one-factor model, such as the Vasicek (1977) model or the Cox, Ingersoll, and Ross (1985), can already do a reasonably good job explaining the term structure variation within this narrow maturity range.

When the maturity range expands to include swap rates from two to 10 years, the first principal component explains 94.8% of the variation. The role of the second principal component becomes more prominent and explains an additional 4.6% of the variation. Quantitative researchers in the industry started to examine a similarly extended maturity range in the mid 1980s and started to implement two-factor term strucure models to capture variations in both the interest-rate level and the slope. Butterfly trades are a natural result of such two-factor structures.

Longer-term Treasury bonds and swap rates with maturities up to 30 years become more actively traded since the early 1990s. The third subset includes maturities from two to 30 years. In this case, the first principal component explains 91.5% of the variation, the second component explains 7.3% of the variation, and the third component also starts to become somewhat important by explaining 0.8% of the variation. The factor analysis of Litterman and Scheinkman (1991) over this maturity range becomes the watershed evidence for establishing the existence of a three-factor structure and for popularizing the level-slope-curvature interpretation of the three factors. This finding has also become the starting point for subsequent specification analysis of dynamic term structure models (e.g., Dai and Singleton (2000)).

During the past decade, as the levels of long-term interest rates become increasingly lower across the world, federal governments race to issue ultra long-dated bonds, with maturities extending to 50-60 years. The corresponding swap market has also started trading contracts with ultra long maturities up to 50 years. Including such ultra long maturities in the analysis can further reduce the explained variation of the first principal component, and ask for even more principal components.

In fact, a standard three-factor Gaussian affine model has been shown to become increasingly inadequate in capturing the full term structure variation over such a wide maturity span. The inadequacy happens at both the very short end and the very long end. At the short end, Fed policy inertia (Woodford (1999)) reduces the short-rate volatility and induces a deterministic trending component that is difficult to be accommodated within a standard three-factor structure (Heidari and Wu (2010)). Short rate hitting the zero lower bound creates another layer of complication and tension for standard affine term structure models, asking for option-like modeling approaches (e.g. Black (1995) and Krippner (2012)). At the very long end, 50-year rates move with as much volatility as interest rates at 10 to 30-year maturities, adding tension to the standard short-rate stationarity assumptions (Giglio and Kelly (2018)).

The principal component analysis on different maturity ranges and the history of data expansions and model extensions show that any global factor structure assumptions are likely to be only *locally true* to the scope of data under analysis. The only thing that is truly global is the local commonality of interest rates with nearby maturities, regardless of how far the data extend and what the true underlying dynamics are.

It is important to point out that our documented local commonality of interest-rate movements is not inconsistent with a global factor structure per se. In fact, it is an intuitive observation that is consistent with the implications of most commonly specified global factor structures. Therefore, local commonality is not evidence against a global factor structure, but rather a more universally robust observation that does not depend on the exact specification of any particular global factor structure.

By relying more on the universal robustness of the local commonality observation and less on the accuracy of a particular global factor structure specification, we can achieve more robustness in butterfly construction and investment performance.

4.3.2. Constructing decentralized butterflies with nearby maturities

With local commonality in mind, we propose to build decentralized butterfly portfolios with bonds of nearby maturities. Regardless of the underlying dynamics, since interest rates at nearby maturities show stronger co-movements, butterflies formed with nearby-maturity bonds are likely to become more stable, with only a small proportion of idiosyncratic movements left. Butterflies constructed with nearby maturities are also less intertwined with other butterflies and are therefore more decentralized, allowing larger diversification benefits when investing in multiple butterflies across different reference maturities.

To construct a decentralized butterfly, we normalize the weight at the middle maturity to one, and determine the weights on the two maturities at the wings. We do not make any factor dynamics assumptions, but directly estimate the covariance matrix Σ_t of changes in the three swap rate series in the butterfly portfolio, and choose the butterfly weights to minimize the portfolio variance,

$$\min_{\beta_t} \beta_t^\top \Sigma_t \beta_t, \quad \text{subject to} \quad \beta_t^\top i_2 = 1, \tag{43}$$

where i_2 is an indicator vector with all elements being zero except the second element, as a way of normalizing the weight at the middle maturity to one. The solution is

$$\beta_t = \Sigma_t^{-1} i_2 / (i_2^\top \Sigma_t^{-1} i_2). \tag{44}$$

Instead of inverting the covariance matrix Σ_t, one can also directly regress rate changes at the middle maturity against the two rate changes on the wings. The slope estimates would be for $(-\beta_1, -\beta_3)$.

In the extreme case of perfect parallel movements, the covariance matrix Σ_t becomes singular and the regression can experience multi-collinearity issue. In this case, we can simply set $\beta_1 = \beta_3 = -0.5$ to construct a symmetric butterfly.

We treat each swap rate as the coupon of a par bond and solve the weight for the par bond by adjusting the interest-rate exposure estimates β_t for the duration difference,

$$w_i = \beta_i \mathcal{D}_2 / \mathcal{D}_i, \tag{45}$$

where we normalize the weight on the middle-maturity bond to one ($w_2 = 1$) and convert the wing weights on the interest rates to the weights on the corresponding par bonds based on their duration differences.

Butterfly portfolios constructed with the above weights have minimum variation left. Recall the bond return decomposition in (8), where the carry term y_t is known ex ante, and the convexity term $\frac{1}{2}\mathcal{C}(dy_t)^2/dt$ also becomes deterministic under diffusive moves as the term $(dy_t)^2/dt$ converges to the annualized variance rate σ_t^2 of the yield changes. The bond return variation is thus chiefly driven by the variation of the bond yield dy_t. With the butterfly construction, the bond portfolio return attribution can be written as,

$$w_t^\top \frac{dB_t}{B_t dt} = w_t^\top y_t + \frac{1}{2} w_t^\top \left(\mathcal{C} \frac{(dy_t)^2}{dt} \right) - \mathcal{D}_2 \beta_t^\top \frac{dy_t}{dt}, \tag{46}$$

where in the last term we replace the bond portfolio weight w_t with the corresponding interest-rate portfolio weight β_t via the relation in (45). Since the interest-rate portfolio weight β_t is chosen in (43) to minimize the variance of the yield combination $\beta_t^\top dy_t/dt$, the more highly correlated the movements of the three yields are, the less remaining variation is left in the butterfly portfolio.

4.3.3. Maturity gaps and stability of butterfly portfolios

To examine how effective the butterfly construction is in removing systematic interest-rate risk, we estimate the butterfly weights based on full sample regression on weekly changes of the associated swap rates. The regression residual represents the residual risk of the butterfly portfolio $\beta^\top dy$. At each reference maturity, we construct butterflies both with adjacent maturities (with maturity gap of ± 1) and with increasingly larger maturity gaps. We compare how the behaviors of the butterflies vary with the maturity gap.

Table 5 reports in Panel A the annualized volatility of the weekly changes of the reference swap rates, as well as the butterflies constructed around the reference maturity with increasing maturity gaps, all in percentage points. The data contain 9 swap series with maturities from 2 to 30 years, with 7-year maturity being the median maturity. Thus, with 7 year as the reference maturity, the wings of the butterfly can be constructed with maturity gaps from 1 to 4. At other reference maturities, when one side of the wing maturity reaches the edge of the maturity span, we only increase the maturity gap on the other side. The label shows the absolute maximum maturity gap from 1 to 4.

[Table 5 about here.]

Weekly changes in the reference swap rate series have annualized volatility estimates from 0.866% to

0.934%. When we construct butterflies with adjacent maturities, the variance reduction is drastic. The volatility estimates of the weekly changes in the butterflies range from 0.064% to 0.109%. The variance ratios of the butterflies to the reference swap rate are merely 0.5-1.4%. Therefore, the butterfly construction with adjacent maturities removes 98.6-99.5% of the swap rate change variance, resulting in remarkably stable interest-rate portfolios.

As the maturity gap increases, the volatility estimates for the butterflies increase, a sign of deterioration in the effectiveness of variance reduction. When the maturity choice of the butterflies becomes the widest apart at the 7-year reference maturity with a maturity gap of 4, the volatility estimate reaches as high as 0.242%. The variance ratio to the reference swap rate increases to 6.7%.

Another way of examining the stability of the butterflies is to examine their mean-reversion behavior. Panel B of Table 5 compares the autocorrelation of the weekly changes of the reference swap rates to the autocorrelation of weekly changes of the butterflies. The autocorrelation estimates for weekly changes of the reference swap rates are close to zero, reflecting the high persistence of the swap rate series and the hard-to-predict nature of their weekly changes. By contrast, the autocorrelation estimates for the weekly changes of butterflies constructed with adjacent maturities become much more negative, ranging from -39.5% to -52.4%. A well-constructed butterfly can remove the most persistent component of the interest-rate movements and leave a portfolio that becomes much more mean reverting. Butterfly construction with adjacent maturities is very effective in removing the persistent movements.

As the maturity gap increases, the effectiveness declines. When the reference maturity is at 7-year and the butterfly is constructed with the widest apart maturities, the autocorrelation estimate for weekly changes of the butterfly becomes very close to zero at -0.029, not much different from the autocorrelation estimate for the weekly changes of the reference swap rate series at -0.025. Thus, when the maturity gap is wide apart, the butterfly construction can completely lose its effectiveness in removing the persistent component of the swap rate movement.

The statistics in Table 5 highlight why butterfly construction is such a common practice in the fixed-income market. By forming a butterfly with nearby maturities, one can effectively remove 99% of the variance in the reference swap rate movements. The construction also effectively removes the most persistent component of the swap-rate movements, resulting in an interest-rate portfolio that is much more mean-reverting than any single interest-rate series in the portfolio. This high mean-reverting behavior dictates that the value of the butterfly portfolio shows long-run stability as the variance of the butterfly movements does not increase with the holding horizon as fast as for the individual swap rate series. Therefore, when there are opposite demands at nearby maturities, broker dealers often use butterfly construction as a cheaper and more flexible way of managing their interest-rate risk. Some investment firms with good broker relations can also share the liquidity provision business with their brokers by timing the strong mean-reversion behavior of the butterflies.

The analysis also shows the crucial importance of shrewd maturity choice. While butterflies constructed with nearby maturities show remarkable stability, butterflies constructed with far-apart maturities can be as unpredictable as the individual interest-rate series. The importance of maturity choice is an aspect often neglected in traditional model-centric analysis. If the yield curve variation can be well-captured by a two-factor structure, the traditional thinking goes, one can hedge away the variation of the two factors with the combination of *any* three distinct yields. In this mindset, model design is of top importance while maturity choice is irrelevant. When a two-factor structure looks insufficient, the researcher will attempt a three-factor or an even higher dimensional structure, holding on to the mathematical trusim that K factors can always be neutralized with the combination of *any* $(K+1)$ yields, rendering maturity choice irrelevant again. The issue with such an approach is that the number of common factors itself depends crucially on the maturity choice for the analysis. This dependence makes any assumed global factor structures decidedly local to the data set used for the analysis. Furthermore, as the number of factors becomes large, the portfolio construction with increasingly many legs loses the practical simplicity and flexibility of butterfly portfolios for liquidity provision.

The decentralized nature of our new theory prompts us to think from a different perspective, and allows us to pay more attention to maturity choice than dynamics assumptions. Our analysis shows that as long as one chooses nearby maturities, butterfly construction can be very effective in removing persistent interest-rate risk, *regardless of* what the true underlying factor structure is. Indeed, by relying on the local commonality of payoff structures, our new decentralized theory ignores the global factor structure all together. When three maturities are chosen to be very close to each other, a local two-factor structure is always more than sufficient to capture their co-movements, and a butterfly portfolio constructed on the three nearby maturities becomes naturally stable, no matter how high the dimensionality is globally.

In principle, our butterfly construction approach can fully match the traditional construction approach based on a term structure model: If one estimates a two-factor term structure model on three interest-rate series and construct the butterfly portfolio to neutralize the two interest-rate factors, the portfolio weights in principle should be similar to our construction approach in (43), where we skip the step to estimate a two-factor model and instead directly estimate the covariance matrix relevant for the portfolio construction. The key difference lies in perspectives. By relying on local commonality, we pay little attention to global factor structures but focus on choosing the interest-rate series shrewdly with nearby maturities and estimate their covariance matrix directly from data without going through the middle step of a structural model. This approach allows us to obtain stable butterflies that are independent of any structural model assumptions.

By contrast, when one starts with a global factor structure assumption in building portfolios, the perspective change alters priorities in one's considerations. First, one would pay a lot more attention in trying to specify and estimate a global factor structure model that matches the data behavior the best, while potentially neglecting the important consideration in maturity choice. Second, fitting a global factor structure to interest rates across a wide span of maturities can lead to distortions when the factor structure is insufficient in capturing the full term structure variation. The distortions can bias the portfolio construction weights, making the global portfolio construction process not as robust as the local covariance matrix estimation. Therefore, while it is possible to achieve similar results from the two approaches, the different perspectives can lead to different behaviors in maturity choice, portfolio weight estimation, and ultimately portfolio investment performance.

4.3.4. Mean-variance investments in decentralized butterflies

The previous section highlights the importance of maturity choice in forming stable butterfly portfolios and shows that decentralized butterflies constructed with nearby maturities are naturally stable due to the local commonality of their payoff structures. This section uses an out-of-sample investment exercise to elaborate how to invest in these stable butterflies based on classic mean-variance analysis of the portfolio's excess returns.

Starting on January 3rd, 1996, at each Wednesday t, for each butterfly maturity combination, we use the past 1-year history of weekly changes of the swap rates to estimate the annualized variance rate on each series σ_t^2. We use the variance rate estimate to construct the convexity contribution for each series, $C_t = \frac{1}{2}\mathcal{C}\sigma_t^2$.

We also use the 1-year history to regress weekly changes of the reference swap rate against weekly changes of the two swap rate series on the wings of the butterfly. The regression slope estimates represent estimates for $(-\beta_1, -\beta_3)$. From the regression residual series, we obtain an annualized variance rate estimator ζ_t^2 for the butterfly interest-rate portfolio $\beta^\top dy_t/dt$. According to (46), the variance of the par bond portfolio excess return is related to the variance of the interest-rate portfolio $\beta^\top dy_t/dt$ by

$$V_t^f = \mathcal{D}_2^2 \zeta_t^2. \tag{47}$$

With the weight on the reference bond at the middle maturity normalized to one, the portfolio variance is proportional to the residual variance of the yield portfolio ζ_t^2, scaled by the squared duration of the reference bond $\mathcal{D}_2^2$.

To determine the expected excess return on the butterfly bond portfolio, we take expectation on the bond return decomposition equation in equation (8) and deduct the financing rate from it,

$$EER_t^f = w_t^\top (y_t - r_t) + w_t^\top \left(\frac{1}{2} \mathcal{C} \sigma_t^2 \right) - \mathcal{D}_2 \mu_t^f, \tag{48}$$

where the first term captures the carry contribution over the financing rate, and the second term captures the contribution from convexity. Since the expected rate of change on each individual interest rate series $\mu_t = \mathbb{E}_t[dy_t/dt]$ is inherently difficult to predict, we avoid predicting the expected rate of change on each individual series, but exploit the high mean-reverting behavior of the butterfly interest-rate portfolio and directly predict the expected rate of change on the interest-rate portfolio, $\mu_t^f = \mathbb{E}_t \left[\beta_t^\top dy_t/dt \right]$. For this purpose, we take the last day of the regression residual e_t, assume that it will converge to zero in a week, and set the expected rate of change to the negative of this residual annualized to 52 weeks, $\mu_t^f = -52 e_t$.

Given the conditional mean excess return estimate in (48) and the conditional variance estimate in (47), we perform classic mean-variance investing by making the investment weight on the butterfly portfolio proportional to the mean-variance ratio,

$$n_t = c\, EER_t^f / V_t^f. \tag{49}$$

where c denotes a constant proportional coefficient that we use to scale the weights. We apply the same proportionality coefficient c to different butterfly portfolios. While the choice of the scaling coefficient c is immaterial to our analysis, we choose a scaling level so that the variations of the weights are within the usual confines of the leverage levels employed by institutional fixed income investors.

Table 6 reports the summary statistics of the time-varying allocation weights to butterflies constructed with adjacent maturities around each of the six reference maturities. The statistics include percentile values of the allocation weights at 10, 25, 50, 75, and 90 percentiles. The median allocation weights are close to zero and the percentile values reveal a reasonably symmetric distribution. The variations of the allocation weights are larger at shorter maturities than at longer maturities.

[Table 6 about here.]

To examine how the allocation weights to different butterflies co-move, the last row of Table 6 reports the average cross-correlation estimates of each allocation weight series with the other allocation weight series. The average cross-correlation estimates are close to zero, suggesting that the variations of the estimated expected excess returns from these decentralized butterfly portfolios are largely independent of one another. This behavior forms a sharp contrast with the extremely high cross-correlation between the market price of risk estimates extracted from different par bonds in Section 4.1. While the expected excess returns on different par bonds tend to move strongly together, the expected excess returns on our decentralized butterfly portfolios are largely independent of one another.

Table 7 reports the summary statistics of the weekly excess returns from investing in the decentralized butterflies constructed with adjacent maturities. The investment performances are similar on butterflies with different reference maturities. The annualized mean excess returns range from 7% to 14%. The annualized standard deviations range from 9% to 12%. The annualized information ratios range from 0.84 to 1.18. The excess returns show large excess kurtosis, and positive skewness.

[Table 7 about here.]

The last row of Table 7 reports the average cross-correlation estimates of each excess return series with other excess return series. Just as the allocation weights across different butterflies are largely independent of one another, so are the realized excess returns from the different butterfly portfolios. The near-independent nature of the decentralized butterflies offers strong diversification benefits when one simultaneously invests in multiple butterflies at different reference maturities. The last column reports the excess return statistics from an equal weighted portfolio of the six butterflies. While the mean excess return retains the average magnitude of 10%, the standard deviation of the aggregate portfolio becomes much smaller than that of the individual butterflies at merely 5%. The annualized information ratio nearly doubles that of the individual butterfly investment, reaching 2.14 for the equal-weighted butterfly portfolio.

We also repeat the investment exercise on butterflies constructed with wider maturity gaps. Table 8 reports the summary statistics of the excess returns from investing in butterflies constructed with maximum maturity gaps from two in Panel A to four in Panel C. Compared to butterflies constructed with adjacent maturities in Table 7, across all reference maturities, the annualized information ratios of the investments on butterflies with wider maturity gaps become significantly lower, and increasingly so as the maturity gap becomes wider. The performance deterioration comes from both lowered mean excess return and increased return volatility. With larger maturity gaps, the excess returns also show less positive or even negative skewness. Thus, the investment not only shows worse average performance, but the performance is also more likely to experience large negative realizations.

[Table 8 about here.]

Worse yet, as the maturity gap becomes wider, butterflies with different reference maturities become intertwined on the wings and lose their decentralized nature. As a result, the investment returns become more correlated with one another. The average cross-correlation among the excess return series is merely 9% when the maturity gap is 1, but increases to 23% when the maximum maturity gap is 2, to 37% when the maximum maturity gap is 3, and reaches as high as 47% when the maximum maturity gap is 4. The increased correlation reduces the benefit of diversification when one invests in multiple butterfly portfolios across different reference maturities. With the combined effects of deteriorated individual butterfly investment performance and reduced diversification benefits across different butterfly portfolios, the annualized information ratio for the aggregate butterfly portfolio declines quickly as the maturity gap increases. With a maximum maturity gap of 4, the annualized information ratio for the aggregate portfolio reduces to 0.66, largely losing the allure of the staple trade.

4.3.5. Statistical arbitrage investments with DTSMs

For comparison, we also perform an analogous out-of-sample investment analysis on butterflies constructed to benefit from the statistical arbitrage opportunities identified from estimating a term structure model.

We estimate both a two-factor term structure model as in Duarte, Longstaff, and Yu (2007) and a three-factormodel as in Bali, Heidari, and Wu (2009). With the two-factor model, we construct butterflies to neutralize both interest-rate factors. With the three-factor model, we construct butterflies to neutralize the two more persistent factors. We adopt the parsimonious dimension-invariant cascade term structure specification of Calvet, Fisher, and Wu (2018), and follow their procedure in estimating the model parameters with quasi-maximum likelihood and extracting the factors with unscented Kalman filter. The parsimony of the specification enhances the parameter identification. To achieve better pricing performance, we re-estimate the model parameters once a year using data from the previous year.

The out-of-sample exercise follows a similar procedure as in the previous subsection. Starting on January 3rd, 1996, at each Wednesday t, we take the model parameter estimates from the previous year, run the unscented Kalman filter through the history up to date t to generate the history of the interest-rate factors. Based on the factor values at time t, we compute the sensitivity of the swap rates to the two factors and compute the butterfly portfolio weight β_t to neutralize the portfolio's exposure to these two factors while normalizing the weight of the middle maturity to one. As in (45), we map the swap portfolio weight β_t to the corresponding par bond portfolio weight w_t by adjusting for the par bond duration.

As in Duarte, Longstaff, and Yu (2007) and Bali, Heidari, and Wu (2009), we determine the allocation weight to a butterfly at each date to benefit from the statistical arbitrage opportunities as manifested by the pricing errors. We measure the difference between the observed value of the butterfly portfolio of swap rates and the corresponding model value, and treat the pricing error as temporary market mispricing. The more positive the pricing error, the higher the coupon payments of the bar bond portfolio relative to the fair model value. To mitigate the impact of persistent mispricing due to model misspecification, we demean the pricing error by its historical average over the past year, and we further scale the demeaned pricing error by the historical variance estimator of the pricing error series over the past year, multiplied by the squared duration of the reference swap contract.

The reference maturity choice, the maturity combination, and the excess return calculation all follow the steps in the previous subsection. Table 9 summarizes the out-of-sample annualized information ratio for the statistical arbitrage investment in each butterfly portfolio. When the butterflies are constructed with adjacent maturities, the statistical arbitrage trading generates reasonable out-of-sample performance. The information ratio for the aggregate portfolio reaches 1.54 when the butterflies are constructed with the two-factor model, and 1.84 when they are constructed with the three-factor model.

[Table 9 about here.]

Even with a centralized dynamic term structure model, the investment performance remains highly dependent on maturity choice. Table 9 shows that as the maturity gap of the butterflies becomes wider, the investment performance deteriorates quickly. Indeed, when the butterfly is constructed around the 7-year maturity with the widest maturity gap using the two-factor model, the out-of-sample investment generates slightly negative average excess returns. Model centralization does not replace the need for careful contract consideration.

Given our evidence from the principal component analysis and the literature finding, a three-factor model should price the term structure from two to 30 years better than a two-factor model; nevertheless, a butterfly construction can only neutralize two factors. Thus comes the dilemma for model choice in constructing butterfly portfolios: A two-factor structure is bound to be insufficient to capture the whole term structure variation, and the pricing errors may reflect more of model misspecification than market mispricing; nevertheless, when estimating a higher-dimensional term structure model, we are faced with the difficult task of determining which factors to neutralize and which factors to be left unhedged in the butterfly construction. We choose to construct butterflies to neutralize the two more persistent factors while leaving the third factor unhedged, in the hope that the unhedged components are more mean-reverting than the hedging components. Nevertheless, the persistence of the factors also dictate their loading pattern across the term structure. The two persistent factors have higher loadings on long-term swap rates while the third, more transient, factor has higher loadings at shorter maturities. As a result, neutralizing the two persistent factors may lead to different stabilities for butterflies constructed at different reference maturities.

In a potentially high or even infinite dimensional world, the effort to build a robust and universally applicable global factor structure is likely futile. Regardless of the global factor structure, however, one can still construct locally stable butterflies by choosing nearby maturities and hedging the two sources of variations that are most relevant for the chosen maturities.

The elusive nature of a global factor structure also puts in question the true nature of the "arbitrage opportunities" identified from a no-arbitrage dynamic term structure model. Unless under certain special circumstances when a market maker can lock in a nonnegative future payoff without a cost, "arbitrage" opportunities identified from any assumed dynamic term structure models are mostly just a mirage. Investing in them as if they were truly arbitrage opportunities can be dangerous and can lead to financial ruins, as shown by the repeated blowups of certain fixed-income funds performing arbitrage trading based on no-arbitrage models. Instead, owning up to the reality that they are not truly arbitrage opportunities does not preclude us from investing in them simply as an investment opportunity with potentially attractive risk-return trade-offs. Carefully estimating their expected returns and their risks can potentially lead to more honest assessments, more prudent investments, and long-run sustainability. Our new decentralized theory provides such a framework for doing mean-variance analysis on the excess returns of bonds and bond portfolios.

5. Concluding remarks

In this paper, we propose a new modeling framework that is particularly suited for analyzing returns on a bond or bond portfolio. The framework does not try to model the full dynamics of an instantaneous short rate, but focuses squarely on the behavior of the yield of the particular bond in question. It does not even ask for the full dynamics specification of this bond yield, but only needs conditional forecasts on its expected rate of change, its risk, and its market price of risk.

It is well-known that long-dated interest-rate series are highly persistent and their changes are inherently difficult to predict. We take this hard-to-predict nature of long-term interest rate changes as our starting point, and propose to infer the market price of bond risk from our theory. We show that this approach allows us to generate better out-of-sample predicting performance on long-term bond excess returns than do commonly specified predictive regressions.

With a common market price of risk assumption across all swap rates and a smooth functional form on the expected rate of change for swap rates from short to intermediate maturities, we also apply the theory to perform comparative analysis on the swap rate curve and separate out the common market price of risk component from the expectation variation. The identified common market risk premium component strongly predicts future bond excess returns across both short and long maturities.

Finally, we apply the theory to the construction and investment of butterfly bond portfolios. The decentralized nature of our theory prompts us to pay more attention to maturity choice than dynamics assumptions in the butterfly construction. We propose the concept of local commonality and show that decentralized butterfly portfolios constructed with nearby maturities are much more stable than those constructed with maturities far apart. Global factor structure specification does not replace the need for shrewd maturity choice, but with shrewd maturity choice, we can build stable decentralized butterfly bond portfolios and make profitable investments in them based on classic mean-variance analysis, without specifying any global factor structures.

For future research, separating risk premium from expectation via local commonality assumptions and/or new information sources remains a challenging, but fruitful endeavor. The new decentralized theory allows one to consider the separation based on localized information, potentially offering a more convenient way of exploring domain expertise.

References

Bali, T., Heidari, M., Wu, L., 2009. Predictability of interest rates and interest-rate portfolios. Journal of Business and Economic Statistics 27, 517–527.

Bekaert, G., Hodrick, R. J., Marshall, D. A., 1997. On biases in tests of the expectations hypothesis of the term structure of interest rates. Journal of Financial Economics 44, 309–348.

Black, F., 1995. Interest rates as options. Journal of Finance 50, 1371–1376.

Breeden, D. T., Litzenberger, R. H., 2013. Central bank policy impacts on the distribution of future interest rates on the distribution of future interest rates. Working paper. Duke University and University of Pennsylvania.

Calvet, L. E., Fisher, A. J., Wu, L., 2018. Staying on top of the curve: A cascade model of term structure dynamics. Journal of Financial and Quantitative Analysis 53, 937–963.

Campbell, J. Y., Shiller, R., 1991. Yield spreads and interest rate movements: A bird's eye view. Review of Economic Studies 58, 494–514.

Chance, D. M., Jordan, J. V., 1996. Duration, convexity, and time as components of bond returns. Journal of Fixed Income 6, 88–96.

Christensen, P. O., Sorensen, B. G., 1994. Duration, convexity and time value. Journal of Portfolio Management 20, 51–60.

Cieslak, A., Povala, P., 2015. Expected returns in treasury bonds. Review of Financial Studies 28, 2859–2901.

Clark, T. E., McCracken, M. W., 2012. Advances in forecast evaluation. In: Timmermann, A., Elliott, G. (Eds.), Handbook of Economic Forecasting. Elsevier, Amsterdam.

Cochrane, J. H., Piazzesi, M., 2005. Bond risk premia. American Economic Review 95, 138–160.

Cooper, I., Priestley, R., 2009. Time-varying risk premiums and the output gap. Review of Financial Studies 22, 2801–2833.

Cox, J. C., Ingersoll, J. E., Ross, S. R., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–408.

Dai, Q., Singleton, K., 2000. Specification analysis of affine term structure models. Journal of Finance 55, 1943–1978.

Dalio, R., 2012. Economic Principles. Bridgewater.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.

Duarte, J., Longstaff, F. A., Yu, F., 2007. Risk and return in fixed income arbitrage: Nickels in front of a streamroller. Review of Financial Studies 20, 769–811.

Duffee, G. R., 2002. Term premia and interest rate forecasts in affine models. Journal of Finance 57, 405–443.

Duffee, G. R., 2011. Information in (and not in) the term structure. Review of Financial Studies 24, 2895–2934.

Duffee, G. R., 2013. Forecasting interest rates. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Elsevier, Amsterdam.

Duffee, G. R., Stanton, R. H., 2012. Estimation of dynamic term structure models. Quarterly Journal of Finance 2, 1–51.

Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and asset pricing for affine jump diffusions. Econometrica 68, 1343–1376.

Dybvig, P. H., Ingersoll, J. E., Ross, S. A., 1996. Long forward and zero-coupon rates can never fall. Journal of Business 69, 1–25.

Evans, M., Lewis, K., 1994. Do stationary risk premia explain it all? evidence from the term structure. Journal of Monetary Economics 33, 285–318.

Fama, E. F., Bliss, R. R., 1987. The information in long-maturity forward rates. American Economic Review 77, 680–692.

Gargano, A., Pettenuzzo, D., Timmermann, A., 2017. Bond return predictability: Economic value and links to the macroeconomy. Management Science 65, 508–540.

Giglio, S., Kelly, B. T., 2018. Excess volatility: Beyond discount rates. Quarterly Journal of Economics 133, 71–127.

Greenwood, R., Vayanos, D., 2014. Bond supply and excess bond returns. Review of Financial Studies 27, 663–713.

Harvey, D. I., Leybourne, S. J., Newbold, P., 1997. Testing the equality of prediction mean squared errors. International Journal of Forecasting 13, 281–291.

Heath, D., Jarrow, R., Morton, A., 1992. Bond pricing and the term structure of interest rates: A new technology for contingent claims valuation. Econometrica 60, 77–105.

Heidari, M., Wu, L., 2003. Are interest rate derivatives spanned by the term structure of interest rates?. Journal of Fixed Income 13, 75–86.

Heidari, M., Wu, L., 2010. Market anticipation of federal reserve policy changes and the term structure of interest rates. Review of Finance 14, 313–342.

Inoue, A., Kilian, L., 2004. In-sample or out-of-sample tests of predictability: Which one should we use?. Econometric Reviews 23, 371–402.

Joslin, S., Priebsch, M., Singleton, K. J., 2014. Risk premiums in dynamic term structure models with unspanned macro risks. Journal of Finance 69, 1197–1233.

Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. Transactions of the ASME–Journal of Basic Engineering 82, 35–45.

Kim, D. H., Orphanides, A., 2012. Term structure estimation with survey data on interest rate forecasts. Journal of Financial and Quantitative Analysis 47, 241–272.

Krippner, L., 2012. Measuring the stance of monetary policy in zero lower bound environments. Economics Letters 118, 135–138.

Litterman, R., Scheinkman, J., 1991. Common factors affecting bond returns. Journal of Fixed Income 1, 54–61.

Ludvigson, S. C., Ng, S., 2009. Macro factors in bond risk premia. Review of Financial Studies 22, 5027–5067.

Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Rapach, D. E., Zhou, G., 2013. Forecasting stock returns. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Elsevier, Amsterdam.

Roll, R., 1970. The Behavior of Interest Rates. Basic Books, New York.

Vasicek, O. A., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177–188.

Wan, E. A., van der Merwe, R., 2001. The unscented Kalman filter. In: Haykin, S. (Eds.), Kalman Filtering and Neural Networks. Wiley & Sons Publishing, New York.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455–1508.

Woodford, M., 1999. Optimal monetary policy inertia. The Manchester School 67, 1–35.

Table 1
Summary statistics of swap rates.
Entries report the summary statistics of the US swap rates (in percentages) both in levels (Panel A) and weekly differences (Panel B). Data are weekly from January 4, 1995 to December 26th, 2018, for 1,252 weeks. The statistics include the sample average ("Mean"), the standard deviation ("Stdev"), the skewness ("Skew"), the excess kurtosis ("Kurt"), and weekly autocorrelation ("Auto"). The mean and standard deviation estimates for the weekly differences are annualized.

Maturity	*Panel A. Statistics on swap rate levels*					*Panel B. Statistics on weekly differences*				
Yrs	Mean	Stdev	Skew	Kurt	Auto	Mean	Stdev	Skew	Kurt	Auto
2	3.22	2.22	0.27	-1.39	0.998	-0.22	0.81	0.18	3.58	0.014
3	3.46	2.14	0.22	-1.35	0.997	-0.22	0.88	0.24	2.52	-0.001
4	3.67	2.06	0.18	-1.31	0.997	-0.23	0.91	0.25	1.94	-0.008
5	3.84	1.99	0.15	-1.29	0.997	-0.23	0.93	0.22	1.80	-0.018
7	4.12	1.87	0.12	-1.25	0.996	-0.23	0.93	0.23	1.78	-0.025
10	4.39	1.78	0.07	-1.22	0.996	-0.22	0.93	0.20	2.05	-0.034
15	4.65	1.72	0.01	-1.21	0.996	-0.22	0.89	0.22	1.87	-0.050
20	4.75	1.69	-0.02	-1.22	0.996	-0.22	0.87	0.19	1.65	-0.049
30	4.81	1.66	-0.03	-1.20	0.996	-0.23	0.84	0.19	1.87	-0.034

Table 2

Out-of-sample forecasting performance on bond excess returns.

Entries report the out-of-sample forecasting performance on six-month excess returns on par bonds with maturities from 2 to 30 years. Panel A reports the benchmark performance of the Cochrane-Piazzesi (CP) bond risk premium estimator based on a portfolio of forward rates. Panel B reports the performance of our bond risk premium estimators (CW) based on the no rate predictability assumption. Panel B reports the performance of our common bond risk premium estimators (CCW) based on the estimation of a common factor structure on the yield curve. The performance measures include both a forecasting R^2 measure and the Diebold–Mariano (DM) (1995) t-statistic, both defined on the squared forecasting error differences against the historical average benchmark. The DM statistics have standard normal critical values. The risk premiums are estimated with a 10-year rolling window. The out-of-sample statistics for each series are computed on 704 weekly observations from January 5th, 2005 to Jun 27th, 2018.

	Panel A. CP		*Panel B. CW*		*Panel C. CCW*	
Maturity	R^2	DM	R^2	DM	R^2	DM
2	-0.35	-0.96	0.13	0.57	0.30	2.16
3	-0.43	-1.61	0.14	1.15	0.18	1.95
4	-0.41	-1.61	0.18	1.88	0.17	2.22
5	-0.39	-1.51	0.19	2.26	0.17	2.42
7	-0.31	-1.30	0.18	2.59	0.17	2.73
10	-0.21	-1.02	0.15	2.73	0.15	2.93
15	-0.13	-0.76	0.12	2.75	0.13	2.94
20	-0.08	-0.59	0.10	2.78	0.10	2.79
30	-0.06	-0.47	0.09	2.82	0.09	2.58

Table 3
Predicting changes in yield curve slope with the extracted rate of change.
Entries report the forecasting correlation between the extracted rate of change differential $\mu_t(\tau_0) - \mu_t(\tau)$ between two swap maturities and future changes in the corresponding swap rate slope over different horizons h (measured in number of weeks). The swap curve slopes are measured as swap rate difference between the two-year maturity ($\tau_0 = 2$) and other maturities (τ) from three to 30 years.

$h\backslash\tau$	3	4	5	7	10	15	20	30
4	0.25	0.24	0.23	0.22	0.20	0.19	0.18	0.16
13	0.44	0.40	0.38	0.36	0.33	0.30	0.28	0.26
26	0.54	0.51	0.48	0.47	0.43	0.39	0.37	0.35

Table 4
Principal component analysis on different subsets of swap rate changes.
Entries report the percentage explained variation of the first four principal components on different subsets of the swap rate weekly changes. Each row reports the results for one subset, with the first column reporting the maturity range of the subset.

Subset	Explained percentage variation of principal components			
Maturity range	1	2	3	4
2-5	97.3	2.4	0.2	0.1
2-10	94.8	4.6	0.4	0.1
2-30	91.3	7.3	0.8	0.2

Table 5
Effectiveness of variance reduction via butterfly construction.
Entries report the annualized volatility in panel A and weekly autocorrelation in panel B of weekly changes on each reference swap rate series, as well as the butterflies constructed around the reference maturity with increasing maturity gaps. When one side of the wing maturity reaches the edge of the maturity choice, the maturity gap increase reflects the maturity choice of the other wing.

Reference	Reference	Butterflies with maximum maturity gap of			
Maturity	Swap	1	2	3	4
		Panel A. Annualized volatility of weekly changes			
3	0.876	0.078	0.089	0.106	0.122
5	0.928	0.079	0.115	0.192	0.213
7	0.934	0.078	0.123	0.169	0.242
10	0.928	0.109	0.147	0.191	0.213
15	0.893	0.064	0.108	0.124	0.135
20	0.866	0.070	0.102	0.107	0.113
		Panel B. Autocorrelation of weekly changes			
3	-0.001	-0.395	-0.326	-0.231	-0.179
5	-0.018	-0.524	-0.219	-0.073	-0.036
7	-0.025	-0.442	-0.247	-0.097	-0.029
10	-0.033	-0.412	-0.318	-0.243	-0.200
15	-0.050	-0.439	-0.316	-0.262	-0.232
20	-0.049	-0.443	-0.414	-0.401	-0.384

Table 6

Summary statistics of the time-varying allocation weights to the butterfly portfolios.

Entries report the summary statistics of the allocation weights to butterflies constructed with adjacent maturities around each of the six reference maturity. The statistics include the values of the allocation weights at 10, 25, 50, 75, and 90 percentiles. The last row reports the average cross-correlation of the allocation weights of each butterfly with other butterflies.

Maturity	3	5	7	10	15	20
10	-42.89	-23.80	-24.92	-15.36	-19.15	-19.35
25	-16.51	-9.29	-7.97	-5.14	-5.51	-5.38
50	-0.43	-0.15	-0.25	0.05	-0.05	0.00
75	14.94	7.24	7.55	4.67	6.73	5.29
90	40.44	21.07	23.21	15.31	24.57	18.98
Corr	-0.04	0.01	0.07	0.09	-0.05	-0.06

Table 7

Investment excess returns on localized butterfly portfolios with adjacent maturities.

Entries report the summary statistics of excess returns from investments on butterfly portfolios constructed with adjacent maturities. The statistics include the annualized mean excess return (Mean), the annualized standard deviation (Stdev), the annualized information ratio (IR), the skewness (Skewness), the excess kurtosis (Kurtosis), and the average correlation of each excess return series with the other series (Corr). The last column reports the statistics of an aggregate portfolio constructed with equal weighting on the six butterfly portfolios, and the correlation value in the last column represents the grand average of the cross-correlations among the six butterfly investment returns.

Maturity	3	5	7	10	15	20	Aggregate
Mean	0.11	0.11	0.07	0.08	0.09	0.14	0.10
Stdev	0.10	0.09	0.09	0.09	0.09	0.12	0.05
IR	1.11	1.18	0.84	0.92	1.02	1.14	2.14
Skewness	2.53	1.15	9.47	0.93	1.55	1.03	0.78
Kurtosis	29.95	31.87	189.46	45.97	62.31	25.71	25.11
Corr	0.02	0.09	0.11	0.12	0.08	0.09	0.09

Table 8

Investment excess returns on butterfly portfolios with wider maturity gaps.

Entries report the summary statistics of excess returns from investments on butterfly portfolios constructed with wider maturity gaps, with each panel representing butterflies with a different maximum maturity gap. The statistics include the annualized mean excess return (Mean), the annualized standard deviation (Stdev), the annualized information ratio (IR), the skewness (Skewness), the excess kurtosis (Kurtosis), and the average correlation of each excess return series with the other series (Corr). The last column reports the statistics of an aggregate portfolio constructed with equal weighting on the six butterfly portfolios, and the correlation value in the last column represents the grand average of the cross-correlations among the six butterfly investment returns.

Maturity	3	5	7	10	15	20	Aggregate
			Panel A. Maximum maturity gap = 2				
Mean	0.10	0.08	0.05	0.05	0.08	0.11	0.08
Stdev	0.11	0.11	0.10	0.10	0.12	0.14	0.07
IR	0.87	0.78	0.53	0.54	0.65	0.79	1.18
Skewness	2.02	0.96	-1.70	-4.07	-0.30	0.39	-2.00
Kurtosis	22.53	20.94	35.02	80.67	35.20	28.16	38.94
Corr	0.01	0.13	0.34	0.32	0.35	0.23	0.23
			Panel B. Maximum maturity gap = 3				
Mean	0.09	0.07	0.05	0.06	0.06	0.11	0.07
Stdev	0.14	0.15	0.11	0.13	0.12	0.14	0.09
IR	0.68	0.46	0.42	0.44	0.51	0.81	0.82
Skewness	1.24	1.04	-3.51	-4.25	-1.30	-0.17	-2.74
Kurtosis	18.52	23.20	63.53	85.68	49.94	32.65	47.18
Corr	0.11	0.32	0.48	0.49	0.46	0.37	0.37
			Panel C. Maximum maturity gap = 4				
Mean	0.09	0.06	0.05	0.06	0.06	0.10	0.07
Stdev	0.14	0.15	0.14	0.13	0.13	0.14	0.10
IR	0.62	0.40	0.33	0.45	0.45	0.72	0.66
Skewness	0.98	0.86	-1.86	-4.45	-2.25	-0.14	-2.47
Kurtosis	13.80	24.55	37.24	88.69	56.92	34.10	43.15
Corr	0.22	0.48	0.61	0.57	0.52	0.43	0.47

Table 9
Statistical arbitrage Investment performance based on DTSMs.
Entries report the annualized information ratio for investing in butterfly portfolios to benefit from statistical arbitrage opportunities identified from the pricing errors of a two-factor (Panel A) and a three-factor (Panel B) dynamic term structure model, respectively. Each column denotes a butterfly with a different reference maturity. Each row denotes the maximum maturity gap the butterfly is constructed around the reference maturity. The last column reports the information ratio on the aggregate portfolio constructed with equal weighting on the six butterfly portfolios.

Reference:	3	5	7	10	15	20	Aggregate
Maturity gap:							
Panel A. Butterflies constructed with a two-factor model							
1	0.65	1.12	0.86	0.57	0.83	0.74	1.54
2	0.66	0.52	0.29	0.33	0.51	0.74	0.86
3	0.50	0.16	0.16	0.29	0.39	0.81	0.56
4	0.40	0.07	-0.02	0.24	0.34	0.66	0.35
Panel B. Butterflies constructed with a three-factor model							
1	0.59	1.25	0.70	1.12	0.90	0.84	1.84
2	0.60	0.67	0.34	0.78	0.56	0.79	1.10
3	0.50	0.08	0.14	0.51	0.63	0.66	0.68
4	0.46	0.06	0.19	0.45	0.59	0.66	0.60

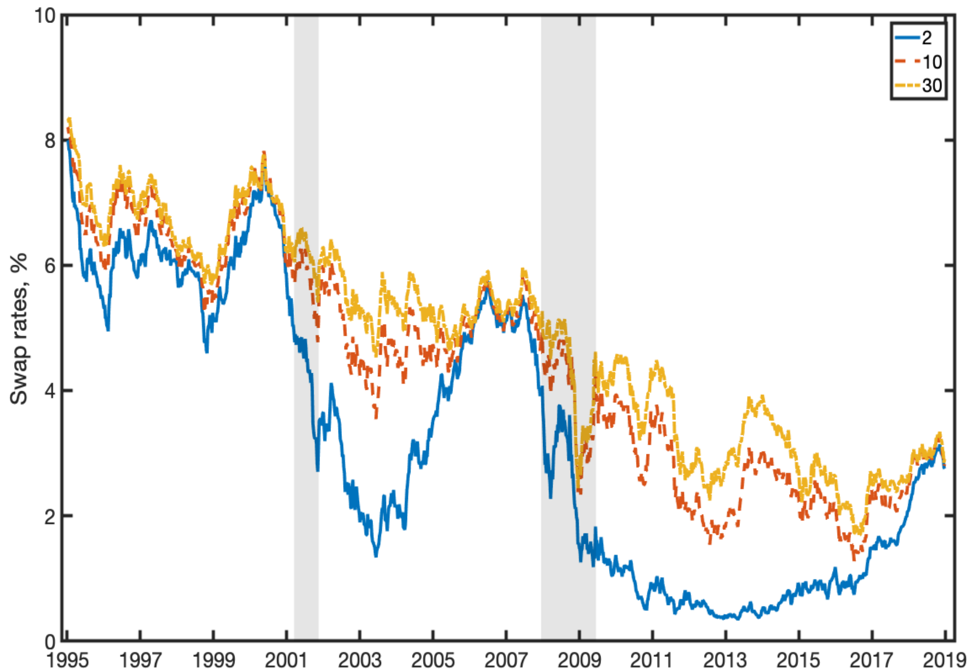

Fig. 1. The time-series variation of swap rates. Lines plot the time series of swap rates at three selected maturities: 2-year (solid line), 10-year (dashed line), and 30-year (dash-dotted line), overlaid with the recession band of the US economy.

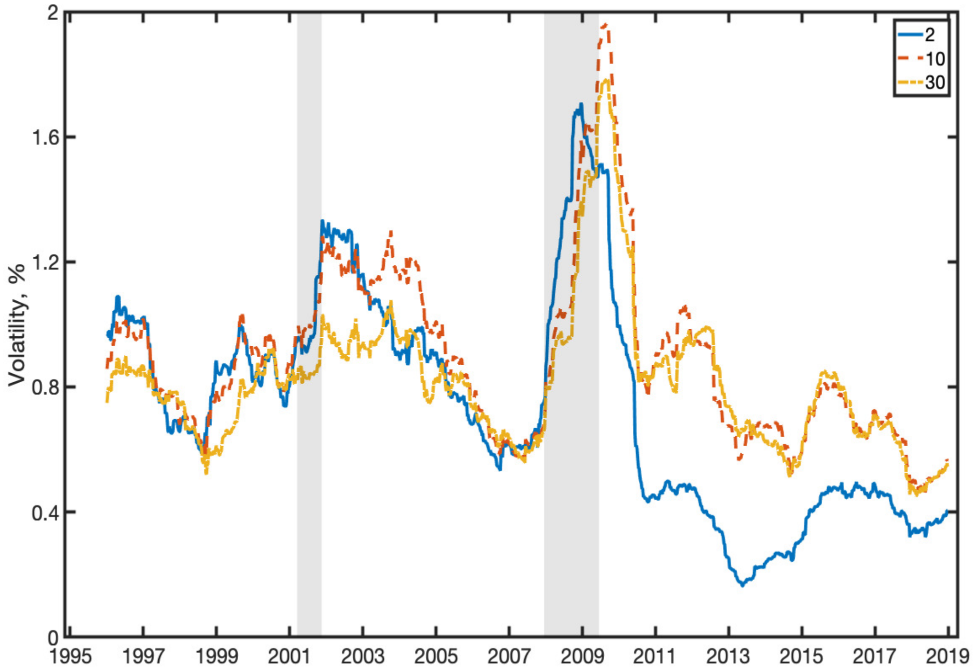

Fig. 2. The time series of volatility estimators on swap rates. Lines plot the time series of the one-year rolling volatility estimators on the weekly changes of the swap rate series at selected maturities: 2-year (solid line), 10-year (dashed line), and 30-year (dash-dotted line), overlaid with the recession band of the US economy.

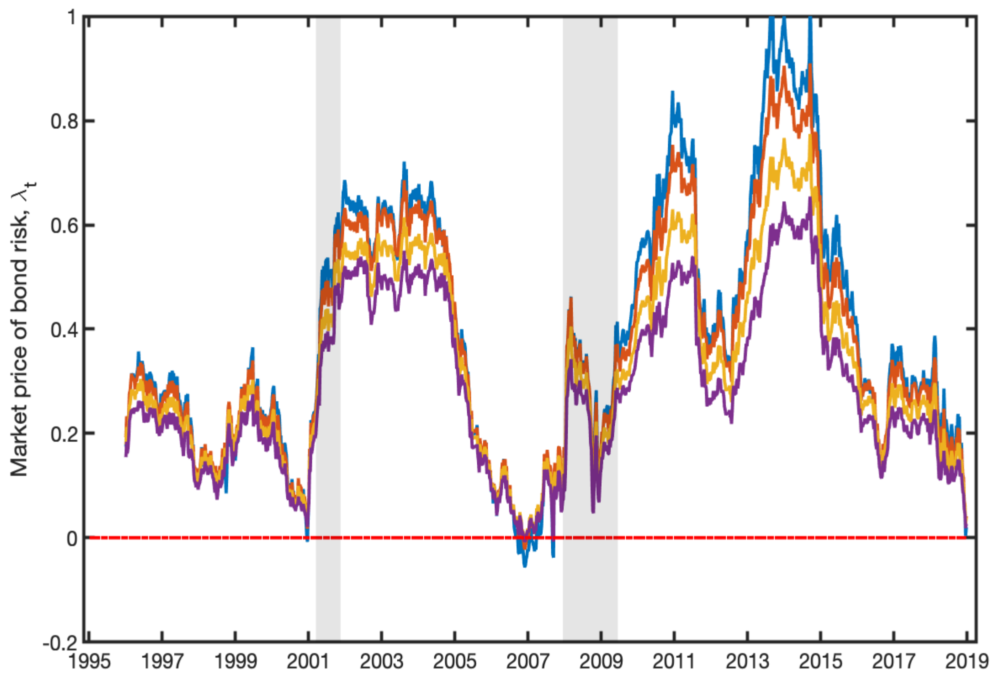

Fig. 3. Market price of bond risk extracted from long-dated swap rates while assuming no rate predictability. Lines denote the time series of the market price of bond risk extracted from swap rates with maturities 10 years and longer, with the assumption of no rate predictability, overlaid with the recession bands of the US economy.

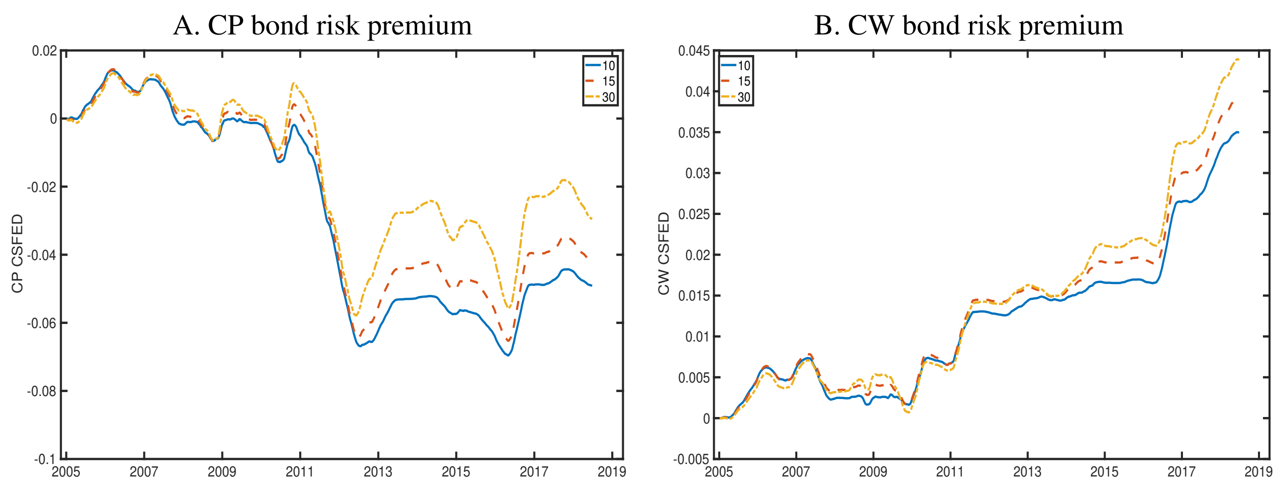

Fig. 4. Cumulative squared forecasting error difference on long-term par bond excess returns. Lines in each panel are the cumulative squared out-of-sample forecasting error difference between each method and the historical average benchmark, with the three lines representing three selected maturities: 10-year (solid line), 15-year (dashed line), and 30-year (dash-dotted line). Panel A represents the performance of the Cochrance-Piazzesi (CP) bond risk premium. Panel B represents the forecasting performance of our (CW) bond risk premium.

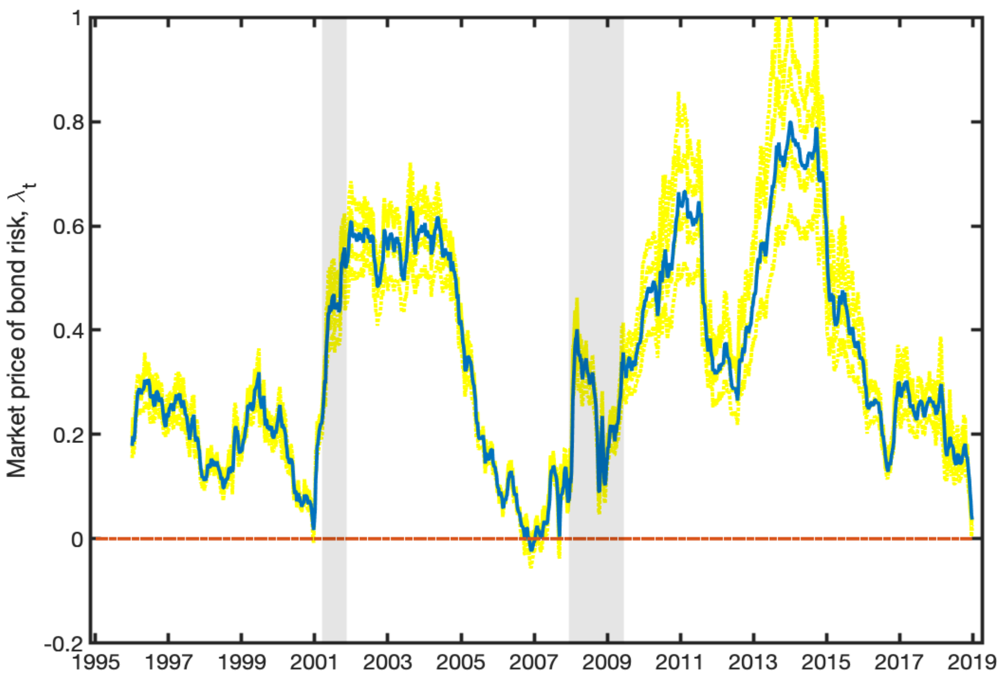

Fig. 5. The common market price of bond risk extracted from a common factor structure. The solid lines plot the time series of the common market price of bond risk (λ_t) extracted from the common factor structure on the swap rate curve. The dotted lines are market price of bond risk estimates based on no-predictability assumption on swap rates from 10 to 30 years. The plot is overlaid with the recession bands of the US economy.

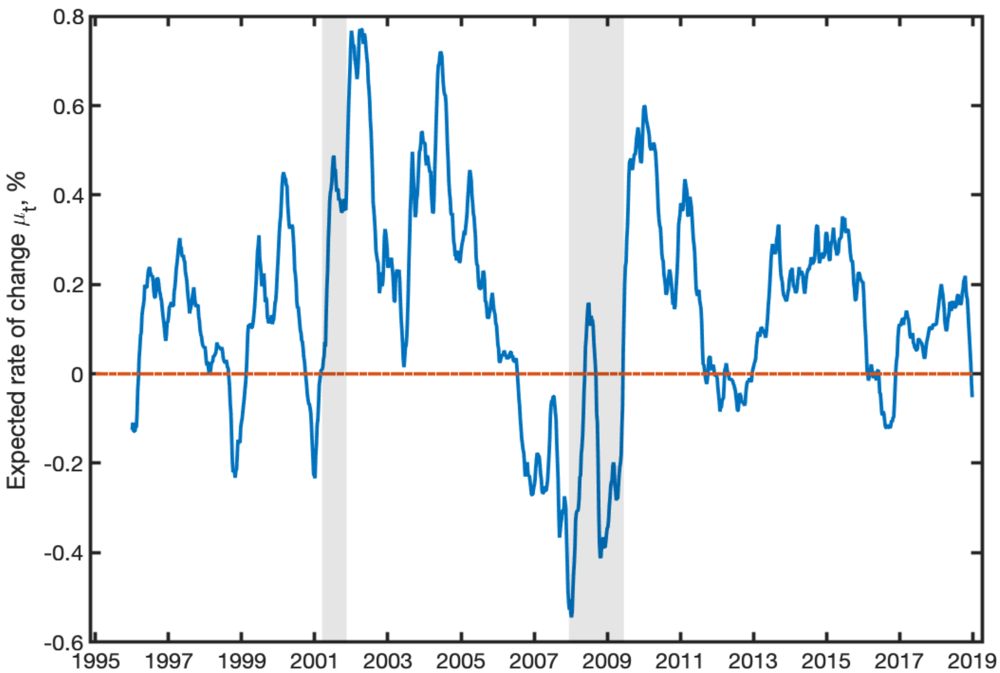

Fig. 6. The expected rate of change on two-year swap rates extracted from a common factor structure. The solid lines plot the time series of the expected rate of change (μ_t) on the two-year swap rates, extracted from the common factor structure on the swap rate curve, and overlaid with the recession bands of the US economy.

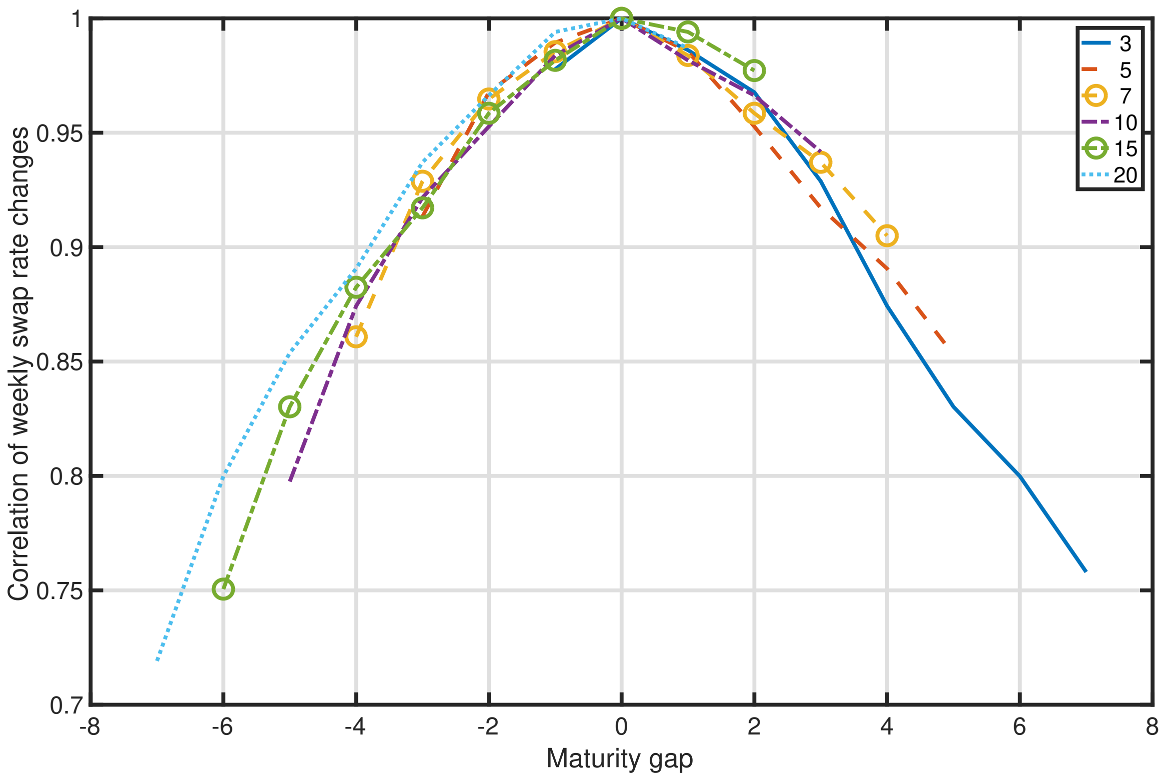

Fig. 7. Cross-correlations of weekly changes of swap rates at different maturity gaps
The six lines denote correlation estimates of weekly swap rate changes between each of the six reference maturities at 3, 5, 7, 10, 15, and 20 years and other maturities. The correlation estimates are plotted against the maturity gap between the other maturities and the reference maturity.