Stripping Discount Curves Across Currencies: Transfer Learning from US Treasuries

Damir Filipović* Markus Pelger† Rose Wang‡

This draft: August 20, 2026
First draft: July 16, 2026

Abstract

We provide precise international yield curves that constitute the new reference database for the government bond markets of six major economies, estimated with a robust and flexible machine learning method regularized by economic principles. We first show that the kernel ridge regression (KR) method of Filipovic, Pelger, and Ye (2024), the most precise method for US Treasuries, generalizes to all international bond markets and uniformly dominates the Nelson–Siegel–Svensson (NSS) benchmark used by many central banks. When foreign exchange (FX) markets provide additional information, we further improve upon KR with a novel transfer learning extension, KR-Transfer (KR-T). KR-T converts US Treasuries into target-currency "synthetic bonds" using FX spot and forward rates and estimates the discount curves jointly under a *weak covered interest parity (CIP)* condition, which penalizes only the curvature of the cross-currency spread and leaves its level and slope unconstrained. Under block cross-validation that mimics real-world maturity gaps, KR-T reduces out-of-sample yield errors by up to 51% relative to KR and by 71–86% relative to NSS, while matching KR where local data are sufficient. The gains concentrate at maturities where FX forwards are observed, identifying FX forward availability as the economic channel of transfer learning. KR-T also delivers more stable forward rates and more plausible yield curve dynamics.

Keywords: yield curve estimation, international bond markets, transfer learning, kernel ridge regression, reproducing kernel Hilbert space
JEL classification: C14, E43, G12

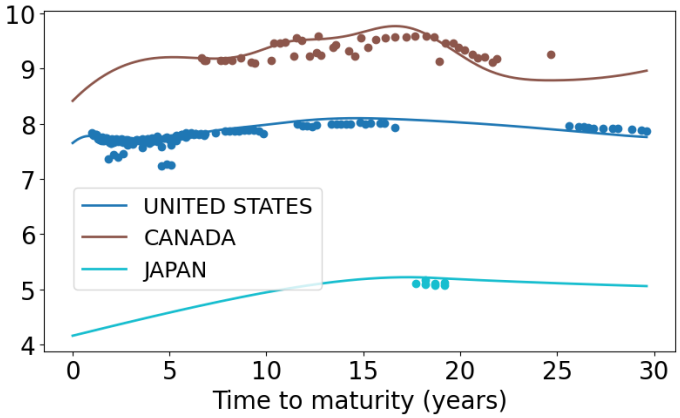
*École Polytechnique Fédérale de Lausanne and Swiss Finance Institute, Email: damir.filipovic@epfl.ch
†Stanford University, Department of Management Science & Engineering, and NBER, Email: mpelger@stanford.edu.
‡Stanford University, Department of Management Science & Engineering, Email: rosew47@stanford.edu.

1 Introduction

Every day, central banks extract monetary policy signals from yield curves, government debt offices price new bond issuances at untraded maturities, pension funds discount liabilities stretching decades into the future, and asset managers construct portfolios across global fixed-income markets. All of these decisions depend on accurate yield curve estimates. Yet outside the US Treasury market, which enjoys hundreds of actively traded securities (bonds) spanning maturities from weeks to thirty years, government bond markets around the world suffer from sparse issuance, limited maturity coverage, and noisy price quotes. When yield curves are poorly estimated due to these data limitations, the downstream consequences are substantial: mispriced government debt, flawed monetary policy signals, inaccurate pension valuations, and unreliable risk measurement across the global financial system.

Figure 1 illustrates the severity of this data problem by showing a cross-section of yield curves and observed bond maturities across different markets on a representative date. Government bond availability varies dramatically across countries, with large maturity gaps in some markets where the yield curve must be extrapolated from distant observations. Figure 2 takes a closer look at a short panel of bond availability in Japan, where daily observations are extremely sparse. The total number of observed government bonds per day ranges from 1 to 12 over this period, and thirty-year JGBs were introduced only in 1999, leaving a long-maturity gap in earlier years.

Figure 1: Yield estimates and observed bond YTM across markets on 1989-07-11

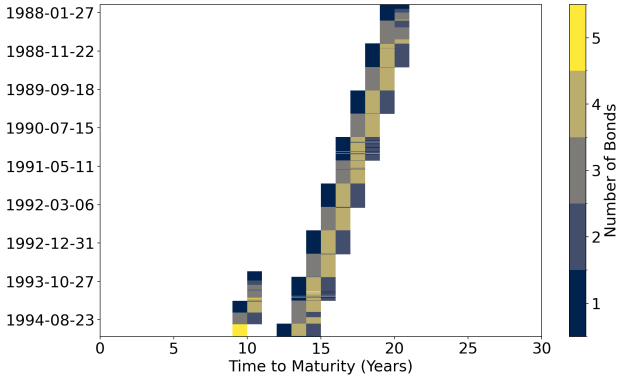


The figure shows yield curve estimates and observed government bond maturities across different markets on 1989-07-11. Each colored line represents the estimated yield curve for a country, while the corresponding colored dots show observed bond yields to maturity (YTM), indicating the actual maturities of bonds available in that market. The gaps between dots illustrate the maturity ranges where yield curve estimation relies on interpolation and extrapolation. The yield estimates are from the KR method.

This paper provides precise international yield curves that constitute the new reference database for the government bond markets of six major economies.[1] Our estimation uses a robust and flexible

[1]We provide a publicly available and regularly updated dataset of daily zero-coupon yields based on our precise estimates for the US, Germany, Japan, UK, Canada and Switzerland. Our estimated yields of zero-coupon government bonds at daily and monthly frequency are available at https://www.discount-bond-data.org.

Figure 2: Short panel of bond availability in Japan



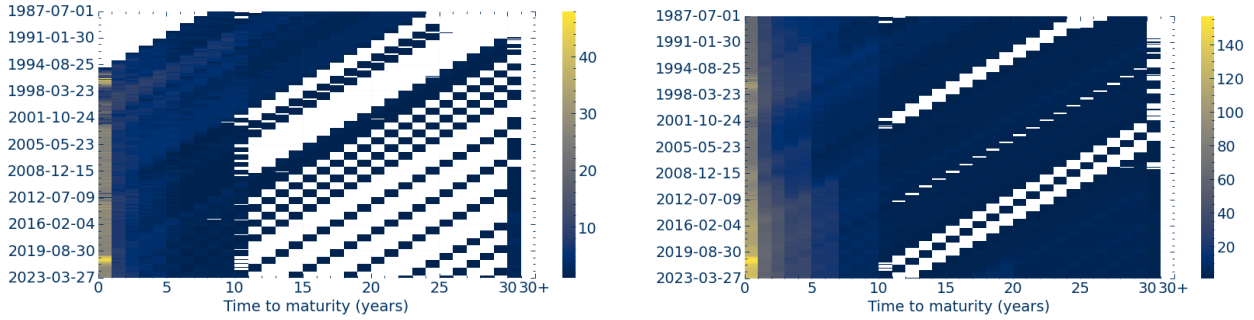
The figure shows a short panel of bond availability in Japan from 1988 to 1994, highlighting the extremely sparse daily observations. The color indicates the number of observed bonds with time to maturity within the maturity years indicated on the y-axis. The total number of observed government bonds per day ranges from 1 to 12 over the sample, leaving large gaps to extrapolate as indicated by the white area.

machine learning method regularized by economic principles. Our first key finding is that the kernel ridge regression (KR) method of Filipovic, Pelger, and Ye (2024), which provides the most precise yield curve estimates for the US, generalizes to all international bond markets and uniformly dominates the Nelson–Siegel–Svensson (NSS) parametric benchmark used by many central banks, academics and practitioners. This establishes KR as the state-of-the-art baseline for yield curve estimation worldwide.

When additional information is available through foreign exchange (FX) markets, we can further improve upon KR. We develop a novel transfer learning extension, KR-Transfer (KR-T), that exploits observed FX spot and forward rates to convert US government bonds into target-currency equivalents, creating "synthetic bonds" that augment sparse local data. Rather than imposing full covered interest parity (CIP), which is persistently violated in the data (Du, Tepper, and Verdelhan, 2018), our framework introduces a *weak CIP* condition that penalizes only the curvature of the cross-currency yield spread while explicitly permitting level and slope deviations. Economically, penalizing the spread curvature bounds the payoffs of butterfly trades on the cross-currency spread, so the cross-market link remains defensible even when CIP deviations persist. Figure 3 shows the observed bond maturities for Canada alone and after augmenting with synthetic, FX-adjusted US bonds: the synthetic bonds turn sparse maturity coverage into dense coverage, filling the maturity gaps and anchoring the curve.

KR-T is most valuable at maturities with little or no local bond information. Under a block cross-validation that withholds contiguous maturity ranges to mimic real-world data gaps, KR-T lowers out-of-sample yield errors by up to 51% relative to standalone KR and by 71–86% relative to NSS. For maturities with sufficient local data, KR-T is essentially identical to KR, so the framework does not distort local accuracy. We formulate KR-T as a pairwise vector-valued kernel ridge regression that couples the target-market discount curve to a synthetic discount curve implied by synthetic, FX-adjusted US government bonds. A single transfer parameter θ governs how strongly

Figure 3: Observed Canada maturities: standalone vs with US information sharing



(a) Canada only **(b)** Canada + synthetic (FX-adjusted US) bonds

Observed bond time-to-maturity panel for Canada alone (left) versus after augmenting with synthetic bonds obtained by converting US government bond cash flows into CAD using spot and forward FX rates (right). Information sharing via synthetic bonds markedly reduces maturity gaps (white bands), especially in the short and intermediate maturities, improving curve anchoring and extrapolation stability.

the curvatures of the two curves are pulled toward each other, and the optimization admits a closed-form solution that remains efficient even for large cross-sections of bonds. The spread penalty is precisely the weak CIP condition: the framework never forces the yields of synthetic bonds to equal local yields; it only requires that curvature dynamics be shared across markets.

Our empirical analysis covers six major government bond markets (US, UK, Japan, Canada, Germany, and Switzerland) with data spanning several decades. Crucially, we characterize when and where transfer learning is most valuable. The gains concentrate at short and intermediate maturities where FX forward rates are directly observed, which is the economic channel through which cross-market information flows. At long maturities beyond the FX forward horizon, synthetic bonds carry little new information and the benefits attenuate. The advantage is persistent over time rather than driven by a few unusual dates, and allowing the transfer parameter to adapt to market conditions yields further modest improvements.

In summary, this paper makes three contributions. First, we provide a new reference database and the most comprehensive out-of-sample comparison of yield curve estimation methods across six major government bond markets. Our KR and KR-T estimates constitute the best available international yield curves, with KR uniformly dominating NSS in every country and KR-T delivering further improvements of up to 51% versus standalone KR and 71–86% versus NSS. Second, we develop a novel machine learning framework (KR-T) for international yield curve estimation that embeds economic structure through a weak CIP condition. Our approach generalizes the KR framework of Filipovic, Pelger, and Ye (2024) to the vector-valued setting for transfer learning across countries and currencies, using synthetic, FX-adjusted bonds and a curvature-based spread penalty to carry information from data-rich to data-poor markets. The optimization retains the closed-form solution of KR, and the separable matrix-valued kernel structure provides both computational tractability and economic interpretability. Third, we identify the availability of FX forward rates as the sole economic channel driving transfer learning gains. This delivers a sharp, testable

prediction: improvements should concentrate at the maturities where FX forwards are observed. We confirm this prediction through a masked-maturity block analysis.

More broadly, our approach illustrates a general principle for combining economic structure with machine learning: when related entities (here, countries) share underlying economic linkages but differ in data availability, domain-specific constraints (such as our weak CIP condition) can be used to transfer information from data-rich to data-poor environments in a principled way. This principle extends beyond fixed-income markets to any setting where some entities are data-rich and others are data-poor, such as transferring demand forecasts across related products, estimating risk models across business units, or learning from developed to emerging markets.

The paper is organized as follows. In Section 2, we formulate the estimation problem and develop the weak-CIP transfer learning framework. In Section 3, we describe the data. In Section 4, we establish that standalone KR dominates the NSS benchmark across all markets. In Section 5, we present the KR-Transfer results, including cross-validation, the when-and-where analysis, FX curve fitting, yield and forward rate estimates, and the factor structure. In Section 6, we discuss the economic and policy implications. In Section 7, we conclude. The appendix provides theoretical background, implementation details, and supplementary empirical results.

Related Literature

Our work connects to several strands of the literature. The first is *yield curve estimation*. The parametric approach originates with Nelson and Siegel (1987) and Svensson (1994), whose Nelson–Siegel–Svensson (NSS) model remains the standard at many central banks, including Germany and Switzerland, and is used for the publicly available yield curves of Gürkaynak, Sack, and Wright (2007), which is a standard reference in academia and industry. Christensen, Diebold, and Rudebusch (2011) embed the Nelson–Siegel structure in an arbitrage-free affine framework, while Adrian, Crump, and Moench (2013) estimate term premia via regression-based models. On the nonparametric side, Fama and Bliss (1987) introduced bootstrapping methods, and Liu and Wu (2021) develop a kernel-smoothing approach for US Treasuries. Most recently, Filipovic, Pelger, and Ye (2024) introduced kernel ridge regression (KR) in a reproducing kernel Hilbert space (RKHS), which uniformly dominates both parametric and nonparametric benchmarks for US and Swiss bonds (Camenzind and Filipovic, 2023). Filipović, Pelger, and Ye (2025) show that the same RKHS framework unifies curve estimation with cross-sectional factor modeling: the optimal basis functions that span the discount bond return curve are the smoothest basis functions in decreasing order of smoothness, explaining the familiar level, slope, and curvature factors encountered in PCA. Their finding that precise yield curve estimation is essential for discovering all relevant asset pricing factors, including a fourth "complexity" factor that carries a large risk premium, further motivates the need for accurate curve estimates across international markets. All of these methods estimate each country's yield curve in isolation. Our contribution is to show that KR generalizes to all international markets and that exploiting cross-country information through FX markets via

KR-T substantially improves estimation, particularly where single-market methods struggle most.[2]

The second strand is *international term structure modeling.* Diebold, Li, and Yue (2008) develop a dynamic Nelson–Siegel model for global yield curves and document strong cross-country co-movement in level and slope factors. Jotikasthira, Le, and Lundblad (2015) show that term structures in different currencies are linked through global risk factors and currency risk premia. Joslin, Singleton, and Zhu (2011) provide a canonical Gaussian dynamic term structure framework that has been extended to international settings. These papers model yield curve dynamics and risk premia, but do not address the estimation problem we focus on: producing accurate yield curve *estimates* for countries with sparse bond data.

The third strand is *machine learning in finance.* Gu, Kelly, and Xiu (2020) and Chen, Pelger, and Zhu (2024) demonstrate that deep learning methods can substantially improve asset pricing, while Bianchi, Büchner, and Tamoni (2021) apply ML to bond return prediction. More broadly, the transfer learning literature in computer science (Pan and Yang, 2010; Weiss, Khoshgoftaar, and Wang, 2016) develops methods for leveraging knowledge from data-rich "source" domains to improve learning in data-poor "target" domains. Hellum, Pedersen, and Rønn-Nielsen (2024) provide an example of financial transfer learning applied to cross-country return predictability. The multi-task learning framework (Caruana, 1997) is closely related to our vector-valued RKHS approach, in which graph regularization encodes task relationships. Our paper shows how these ML ideas can be combined with economic structure, specifically a weak CIP condition linking cross-currency yield curve curvature, to produce a transfer learning framework that is both theoretically principled and empirically powerful. The connection between KR and Gaussian process regression (Rasmussen and Williams, 2006) provides additional interpretability and a natural path to uncertainty quantification.

Finally, our framework is informed by the literature on *covered interest parity (CIP) and FX markets.* Du, Tepper, and Verdelhan (2018) document persistent CIP deviations in the post-2008 period, sometimes exceeding 30 basis points for major currency pairs. Our framework accommodates these deviations at two levels. First, in the FX forward estimation step, we use the CIP-implied rate as a prior and estimate deviations nonparametrically from observed FX forward quotes. Second, and more fundamentally, the spread penalty enforces only a *weak CIP* condition that penalizes curvature differences between local and synthetic discount curves while leaving their levels and slopes unconstrained. This explicitly allows for the persistent yield differentials that CIP violations produce.

[2]In a related but distinct setting, Camenzind and Filipovic (2025) study multi-curve estimation within a single currency using government bonds and interest rate swaps. Our framework develops transfer learning for international markets across currencies, which is a fundamentally different economic problem requiring FX forward curve estimation, accommodation of CIP deviations, and handling of maturity-dependent exchange rate availability.

2 Transfer Learning the Discount Curve

This section develops our framework for estimating discount curves across major international bond markets. We first formulate the single-market discount curve estimation problem and its kernel ridge regression (KR) solution. We then introduce transfer kernel ridge regression (KR-Transfer), which pairs each target market with the US government bond market as a source and transfers information via FX-adjusted synthetic bonds under a novel *weak CIP* condition. Appendix A.3 provides a summary of the notation.

2.1 Estimation Problem

We consider the government bond markets in the six major currencies

$$\mathcal{A} \coloneqq \{\text{US}, \text{Germany}, \text{Japan}, \text{UK}, \text{Canada}, \text{Switzerland}\}.$$

Since we consider only government bonds, we simply refer to them as bonds throughout. On a given trading day, market $a \in \mathcal{A}$ contains M_a quoted bonds with common cash-flow dates $0 < x_1 < \cdots < x_N$, collected in the column vector $\boldsymbol{x} \coloneqq [x_1, \ldots, x_N]^\top$.[3] We observe their noisy ex-coupon prices $P_a \coloneqq [P_{a,1}, \ldots, P_{a,M_a}]^\top$. The associated $M_a \times N$ cash-flow matrix is denoted by $C_a = [C_{a,ij}]$, where $C_{a,ij}$ is the cash flow of bond i occurring at time x_j. Both prices P_a and cash flows C_a are denominated in their respective local currency. For any function f, we write $f(\boldsymbol{x}) \coloneqq [f(x_1), \ldots, f(x_N)]^\top$ for the corresponding vector of function values, and $C_{a,i}$ for the ith row of C_a.

Following the law of one price, we assume that each market $a \in \mathcal{A}$ admits a unique discount curve $g_a : [0, \infty) \to \mathbb{R}$ with $g_a(0) = 1$, such that the observed price of bond i satisfies

$$P_{a,i} = \underbrace{C_{a,i}\, g_a(\boldsymbol{x})}_{\text{fundamental value}} + \underbrace{\epsilon_{a,i}}_{\text{pricing error}}, \tag{1}$$

where $\epsilon_{a,i}$ captures pricing errors due to market microstructure effects, liquidity, and data imperfections.

The natural starting point to estimate the unknown discount curve g_a would be to minimize the weighted mean squared pricing errors of the observed securities,

$$\min_{g_a} \left\{ \sum_{i=1}^{M_a} \omega_{a,i} \left(P_{a,i} - C_{a,i}\, g_a(\boldsymbol{x}) \right)^2 \right\}, \tag{2}$$

for some exogenous weights $\omega_{a,i} > 0$. However, solving this problem requires further structure due to two fundamental issues. First, problem (2) is under-determined. The number of observed prices M_a is substantially smaller than the number of cash flow dates N. Even for the data-rich US government bond market, on a typical trading day we observe around $M_a \approx 250$ different

[3] Cash flow dates $\boldsymbol{x}$ are assumed to be common across all markets without loss of generality.

US bond prices, while a discount curve for 30 years requires estimates of around $N \approx 10{,}000$ daily zero-coupon bond prices. The problem is substantially more severe for markets like Japan or Switzerland, where on some days we observe only five to ten different bonds. Second, the goal is to estimate the ground truth curve and not to overfit the noise. Flexible discount curves can match the observed prices in-sample exactly, but tend to perform poorly out-of-sample and produce inconsistent curve shapes and dynamics. Therefore, any estimation approach for the discount curve imposes some form of regularization to limit the degrees of freedom in (2).

This leads us to two key questions. First, what is the class of plausible functions for the discount curve? Second, what is the appropriate regularization for data that are only moderately large and noisy?

2.2 Kernel Ridge Regression (KR)

We first recap the kernel ridge regression (KR) approach of Filipovic, Pelger, and Ye (2024). Its key idea is to trade off the pricing errors of the observed securities against the smoothness of the discount curve, over a very general class of functions. Filipovic, Pelger, and Ye (2024) show that the answer to both questions posed above, the class of plausible curves and the appropriate regularization, rests on the principle of smoothness.

To answer the first question, Filipovic, Pelger, and Ye (2024) show that subject to some mild technical conditions any arbitrage-free discount curve is twice differentiable, and therefore a smooth curve. Hence, it is essentially without loss of generality to study the extremely large space of discount curves that are twice differentiable.

To answer the second question, following Filipovic, Pelger, and Ye (2024), we measure the smoothness of a function $f : [0, \infty) \to \mathbb{R}$ by the weighted curvature metric

$$\|f\|_\alpha^2 := \int_0^\infty f''(x)^2 \, e^{\alpha x} \, dx, \tag{3}$$

for a fixed maturity weight $\alpha > 0$. Penalizing large values of this metric reduces excessive oscillations and curvature of the estimated discount curve.[4]

Rewarding smoothness is motivated by economic principles. A smooth discount curve implies that bonds with similar cash flows and maturities have similar prices. Conversely, large and sudden changes in the curve would permit trading strategies among bonds of nearby maturities with extreme payoffs that are unlikely to be feasible out-of-sample, and rewarding smoothness bounds such payoffs. In fact, most discount curve estimation methods impose smoothness, whether implicitly or explicitly.

[4]The second derivative is the appropriate object to penalize rather than the first derivative. Penalizing the first derivative would overly constrain the level and slope of the discount curve, which are economically meaningful and should be determined by the data. The second derivative captures curvature, which is the dimension along which regularization is most beneficial. This was studied and formally justified by Filipovic, Pelger, and Ye (2024), who show that the curvature-only penalty yields the optimal bias-variance trade-off for discount curve estimation. We have also verified empirically for all international markets in our sample that penalizing only the second derivative produces superior out-of-sample performance; detailed results are available upon request.

To formalize this principle, we define the large infinite-dimensional affine space of candidate discount curves by $\mathcal{G} \coloneqq 1 + \mathcal{H}$, where $\mathcal{H}$ denotes the weighted Sobolev space of differentiable functions $h : [0, \infty) \to \mathbb{R}$ satisfying $h(0) = 0$ and $\|h\|_\alpha < \infty$, which is formally defined in Appendix A.1. Hence, every curve $g_a \in \mathcal{G}$ has a unique representation $g_a = 1 + h_a$ with $h_a \in \mathcal{H}$, and $\|g_a\|_\alpha = \|h_a\|_\alpha$.

We then estimate g_a by solving the kernel ridge regression (KR) problem over $\mathcal{G}$, which trades off fitting error against smoothness,

$$\min_{g_a \in \mathcal{G}} \left\{ \underbrace{\sum_{i=1}^{M_a} \omega_{a,i} \big(P_{a,i} - C_{a,i}\, g_a(\boldsymbol{x})\big)^2}_{\text{pricing error}} + \underbrace{\lambda_a \, \|g_a\|_\alpha^2}_{\text{smoothness}} \right\}, \tag{4}$$

where $\lambda_a > 0$ is a market-specific regularization parameter. Increasing the regularization parameter has three effects. First, it enforces smoothness by reducing excessive curvature of the estimated curve. Second, it makes the estimates more robust by damping the influence of outlier prices. And third, it regularizes the problem and renders it well posed and tractable with a remarkably simple closed-form solution, as we show below. Given exogenous weights $\omega_{a,i}$, the estimation problem is completely determined up to the tuning parameters α and λ_a. We will select these parameter values empirically from the data via cross-validation to minimize out-of-sample pricing errors.

This closed-form solution leverages the structure of reproducing kernel Hilbert spaces (RKHS), function spaces that are central to machine learning because of the celebrated representer theorem. The theorem states that any function in such a space that minimizes an empirical objective can be written as a linear combination of the reproducing kernel evaluated at the training points (here the cash flow dates x_j), which reduces the infinite-dimensional problem to a finite-dimensional one. Given our objective (pricing error plus smoothness) and feasible set (twice differentiable curves), the representer theorem pins down the optimal basis functions, and the solution is linear in them, that is, a kernel ridge regression. We recap the formal definition and basic properties of RKHS in Appendix A.1.

The following theorem, a restatement of (Filipovic, Pelger, and Ye, 2024, Theorem 2), formalizes all this: it asserts that the problem (4) is well posed and provides the solution in closed form. It builds on the fact, shown in Filipovic, Pelger, and Ye (2024), that $\mathcal{H}$ endowed with norm $\|\cdot\|_\alpha$ is a separable RKHS with reproducing kernel

$$k(x, y) = -\frac{\min\{x, y\}}{\alpha^2}\, \mathrm{e}^{-\alpha \min\{x,y\}} + \frac{2}{\alpha^3}\big(1 - \mathrm{e}^{-\alpha \min\{x,y\}}\big) - \frac{\min\{x, y\}}{\alpha^2}\, \mathrm{e}^{-\alpha \max\{x,y\}}. \tag{5}$$

Henceforth, we write $\mathbf{1}$ for the column vector of ones, and define the $N \times N$ kernel matrix K with elements $K_{ij} \coloneqq k(x_i, x_j)$, for the kernel function given in (5).

Theorem 1 (KR Solution)

The KR problem (4) *admits a unique solution, which is given by*

$$g_a^{\mathrm{KR}} = 1 + \sum_{j=1}^{N} k(\cdot, x_j)\,\beta_j, \tag{6}$$

where $\beta = [\beta_1, \ldots, \beta_N] \in \mathbb{R}^{1\times N}$ *takes the form*

$$\beta^\top = C_a^\top \big(C_a\, K\, C_a^\top + \lambda_a\, J_a\big)^{-1} \big(P_a - C_a \mathbf{1}\big),$$

for the diagonal inverse weight matrix $J_a = \mathrm{diag}(\omega_{a,:})^{-1}$.

The kernel basis functions $k(\cdot, x_j)$, which span the estimated discount curve (6), are explicitly given in (5) and entirely derived from the economically motivated smoothness reward, with hyper-parameters learned from market data. This renders our approach fully non-parametric and data-driven.

Being the solution of a kernel ridge regression, the KR curve (6) is linear in the observed prices. It is fast to compute and fully reproducible by any researcher: the main computational task is a single left matrix division in dimension M_a together with a few matrix multiplications, giving an overall complexity of order $O(M_a^3 + M_a N^2)$. The discount bonds in (6) are replicated by a portfolio of observed coupon bonds and a cash investment, which is of independent interest because it allows one to construct discount bonds of any desired maturity from the traded coupon bonds. The KR framework is also transparent in its choices and fundamentally different from most yield curve estimation methods, which first specify an ad-hoc set of basis functions and then study the properties of the fitted curve. KR reverses this order: it derives the estimation problem from first principles and the properties of the discount curve, which in turn imply the optimal, data-selected basis functions.

We follow the standard approach and set the pricing error weights $\omega_{a,i}$ inversely proportional to the squared duration $D_{a,i}$ of bond i in market a, that is,

$$\omega_{a,i} = \frac{1}{M_a}\,\frac{1}{(D_{a,i}\,P_{a,i})^2}. \tag{7}$$

With this choice, the weighted mean squared pricing error in (4) is approximately equal to the mean squared yield-to-maturity fitting error, consistent with standard practice such as in Gürkaynak, Sack, and Wright (2007) and Filipovic, Pelger, and Ye (2024).

Remark 1 (Gaussian Process Interpretation)
The KR estimator in Theorem 1 coincides with the posterior mean of a Gaussian process prior with covariance kernel k *and observation noise variance proportional to* $\lambda_a/\omega_{a,i}$ *(Filipovic, Pelger, and Ye, 2024, Proposition 1). The posterior variance provides natural uncertainty bands, with wider bands at maturities where data are sparse.*

2.3 Transfer Kernel Ridge Regression (KR-Transfer)

We now introduce our novel transfer learning framework. We adopt a pairwise design in which, for each target market $a \in \mathcal{A} \setminus \{\text{US}\}$, the US bond market serves as the source market. The central idea is to exploit observed FX spot and forward rates to convert US bonds into target-currency equivalents, and then estimate the target discount curve from both local bonds and these synthetic bonds using transfer learning.

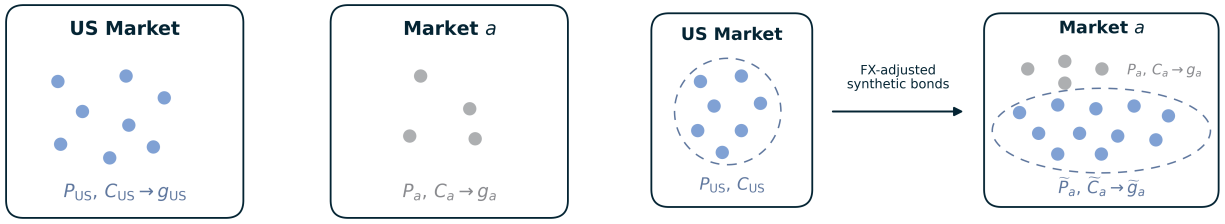
2.3.1 Synthetic Bonds

Let $a \in \mathcal{A} \setminus \{\text{US}\}$ denote a target market. We denote by $S_a = F_a(0)$ and $F_a(x)$ the spot and time-x forward price of one US dollar in units of currency a. We convert a US bond with USD-denominated price $P_{\text{US},i}$ and cash flows $C_{\text{US},ij}$ into a target-currency equivalent by buying US dollars at spot and forward-selling US dollars at the coupon payment dates. This yields synthetic bonds with transformed prices and cash flows in the local currency of market a given by

$$\widetilde{P}_{a,i} \coloneqq P_{\text{US},i}\, S_a, \qquad \widetilde{C}_{a,ij} \coloneqq C_{\text{US},ij}\, F_a(x_j). \tag{8}$$

Figure 4 illustrates the construction. Panel (a) shows bonds observed in two markets operating in different currencies. Panel (b) shows the FX-adjusted synthetic bonds $(\widetilde{P}_a, \widetilde{C}_a)$, which behave as if they had been issued in the target currency. They can substantially enlarge the data set for estimating the discount curve in the target market.

Figure 4: Two-Market Transfer Learning Framework



(a) Observed bonds in two markets **(b)** Synthetic bonds constructed via FX forwards

Illustration of target and synthetic bonds in two-market transfer learning under the weak CIP framework. (a) Observed bonds in the source market (US) with discount curve $g_{\text{US}}(x)$ and in the target market a with discount curve $g_a(x)$. (b) Synthetic bonds $(\widetilde{P}_a, \widetilde{C}_a)$ obtained by converting US bond cash flows into target currency via FX spot S_a and forwards $F_a(x)$. Under a possible violation of CIP, these synthetic bonds are priced by a synthetic discount curve $\widetilde{g}_a(x) = g_a(x)\, \mathrm{e}^{-\eta_a(x)}$, where $\eta_a(x)$ is the log-FX spread that captures deviations from CIP.

2.3.2 Weak CIP and the Synthetic Discount Curve

Under covered interest parity (CIP), the FX forward curve satisfies $F_a(x) = S_a\, g_{\text{US}}(x)/g_a(x)$. In that case, the target discount curve prices the synthetic bonds, $\widetilde{P}_{a,i} = \widetilde{C}_{a,i}\, g_a(\boldsymbol{x})$, up to FX-

transformed pricing errors.[5] In practice, however, CIP deviations are persistent and well documented (Du, Tepper, and Verdelhan, 2018). We therefore introduce a synthetic discount curve $\widetilde{g}_a$, defined as the solution to the CIP equation $F_a(x) = S_a\, g_{\mathrm{US}}(x)/\widetilde{g}_a(x)$. It prices the synthetic bonds, $\widetilde{P}_{a,i} = \widetilde{C}_{a,i}\, \widetilde{g}_a(\boldsymbol{x})$, up to pricing errors, and possibly differs from g_a.

While neither g_a nor $\widetilde{g}_a$ is directly observable, we assume that the spread $g_a - \widetilde{g}_a$ is smooth in curvature, measured in terms of the metric (3). That is, we penalize large values of

$$\|g_a - \widetilde{g}_a\|_\alpha^2 = \int_0^\infty \left(g_a''(x) - \widetilde{g}_a''(x)\right)^2 \mathrm{e}^{\alpha x}\, dx,$$

reflecting the economic idea that CIP deviations should vary gradually across maturities.

Penalizing the spread curvature leaves the level of the spread unconstrained: the spread curvature is zero whenever the spread is flat in maturity. Our framework therefore does not require that yields of synthetic bonds equal yields of actual local bonds; it explicitly allows for persistent yield differentials between the two curves, accommodating the well-documented CIP violations. What the penalty does enforce is that changes in curvature are shared: if the synthetic yield curve changes rapidly in the maturity dimension, the actual yield curve must also change, and vice versa.

Economically, this curvature linkage has a natural interpretation as a butterfly strategy on the spread: a butterfly trade isolates the curvature of a curve, so the spread penalty bounds the payoff of a butterfly trade applied to the spread $g_a - \widetilde{g}_a$. Large values of $(g_a'' - \widetilde{g}_a'')^2$ correspond to profitable butterfly trades on the cross-currency spread, which should be bounded under reasonable market efficiency.

We therefore interpret the spread penalty as imposing a weak form of CIP: rather than requiring yield equalization across markets (standard CIP), we require only that the curvature of the spread be small. This weak CIP condition is both economically defensible (it is robust to persistent CIP deviations) and empirically effective, as we demonstrate in Section 5.

2.3.3 KR-Transfer Formulation

We implement transfer learning by combining, in a single objective, the pricing errors of the target and synthetic bonds, their respective curvature penalties, and a new penalty on the curvature of

[5] Combining (1) and (8), and writing $\widetilde{\epsilon}_{a,i} \coloneqq \epsilon_{\mathrm{US},i}\, S_a$ for the FX-transformed pricing error, gives

$$\widetilde{P}_{a,i} = P_{\mathrm{US},i}\, S_a = \sum_{j=1}^{N} C_{\mathrm{US},ij}\, S_a\, g_{\mathrm{US}}(x_j) + \epsilon_{\mathrm{US},i}\, S_a = \sum_{j=1}^{N} C_{\mathrm{US},ij}\, F_a(x_j)\, g_a(x_j) + \widetilde{\epsilon}_{a,i} = \sum_{j=1}^{N} \widetilde{C}_{a,ij}\, g_a(x_j) + \widetilde{\epsilon}_{a,i}.$$

their spread. This leads to our transfer kernel ridge regression (KR-Transfer) problem

$$\min_{g_a,\,\widetilde{g}_a\in\mathcal{G}}\Bigg\{\underbrace{\sum_{i=1}^{M_a}\omega_{a,i}\big(P_{a,i}-C_{a,i}\,g_a(\boldsymbol{x})\big)^2}_{\text{target bond pricing error}}+\underbrace{\sum_{i=1}^{M_{\mathrm{US}}}\widetilde{\omega}_{a,i}\big(\widetilde{P}_{a,i}-\widetilde{C}_{a,i}\,\widetilde{g}_a(\boldsymbol{x})\big)^2}_{\text{synthetic bond pricing error}}$$
$$+\underbrace{\lambda_a\,\|g_a\|_\alpha^2}_{\text{target curve smoothness}}+\underbrace{\widetilde{\lambda}_a\,\|\widetilde{g}_a\|_\alpha^2}_{\text{synthetic curve smoothness}}+\underbrace{\theta\,\|g_a-\widetilde{g}_a\|_\alpha^2}_{\text{spread curve smoothness}}\Bigg\},\quad(9)$$

where $\lambda_a,\widetilde{\lambda}_a>0$ are regularization parameters for the target and synthetic curves, respectively, and $\theta\geq 0$ is the transfer parameter controlling the degree of information sharing. When $\theta=0$, the problem decouples into two standalone KR problems of the form (4): one for g_a from local bonds, and one for $\widetilde{g}_a$ from synthetic bonds. Setting $\theta>0$ trades off weighted pricing errors against the curvature of $g_a-\widetilde{g}_a$, thereby enabling transfer learning from the source (US) to the target market under the weak CIP condition.

The following theorem extends Theorem 1 to the transfer learning setting and provides the closed-form solution for the pairwise configuration with two curves.

Theorem 2 (KR-Transfer Solution)
The KR-Transfer problem (9) *admits a unique solution, given by*

$$\begin{bmatrix}g_a^{\mathrm{KR-T}}\\ \widetilde{g}_a^{\mathrm{KR-T}}\end{bmatrix}=\mathbf{1}+\sum_{j=1}^{N}k(\cdot,x_j)\,B_a\,\boldsymbol{\beta}_j,\quad(10)$$

where $\boldsymbol{\beta}=[\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_N]\in\mathbb{R}^{2\times N}$ *takes the form*

$$\mathrm{vec}(\boldsymbol{\beta}^\top)=\boldsymbol{C}^\top\big(\boldsymbol{C}\,(B_a\otimes K)\,\boldsymbol{C}^\top+\boldsymbol{J}\big)^{-1}(\boldsymbol{P}-\boldsymbol{C}\mathbf{1}),$$

with the stacked price vector, block-diagonal cash flow matrix, and block-diagonal inverse weight matrix

$$\boldsymbol{P}\coloneqq\begin{bmatrix}P_a\\ \widetilde{P}_a\end{bmatrix},\qquad\boldsymbol{C}\coloneqq\begin{bmatrix}C_a & 0\\ 0 & \widetilde{C}_a\end{bmatrix},\qquad\boldsymbol{J}\coloneqq\begin{bmatrix}\mathrm{diag}(\omega_{a,:})^{-1} & 0\\ 0 & \mathrm{diag}(\widetilde{\omega}_{a,:})^{-1}\end{bmatrix},$$

and the 2×2 *transfer matrix*

$$B_a\coloneqq\frac{1}{\lambda_a\widetilde{\lambda}_a+\theta(\lambda_a+\widetilde{\lambda}_a)}\begin{bmatrix}\widetilde{\lambda}_a+\theta & \theta\\ \theta & \lambda_a+\theta\end{bmatrix}.$$

The matrix B_a governs the strength of information sharing between the two curves. When $\theta=0$, B_a is diagonal, and the solution $g_a^{\mathrm{KR-T}}$ in (10) coincides with the standalone KR solution g_a^{KR} in (6). A larger θ ties the curvatures of $g_a^{\mathrm{KR-T}}$ and $\widetilde{g}_a^{\mathrm{KR-T}}$ more tightly together. This is the transfer learning effect.

2.3.4 Information Content of Synthetic Bonds

It is important to understand what makes synthetic bonds informative. The genuinely new information enters through the FX spot S_a and forward rates $F_a(x_j)$ in (8). They establish the exchange rate link between the two currencies at each maturity x_j, and it is through this link that US bond data become informative about the target discount curve g_a. At maturities where FX forwards are directly observed, typically up to ten years for major currency pairs, though shorter in earlier parts of our sample (see Figure A.1), the synthetic bond cash flows $\widetilde{C}_{a,ij}$ are computed using actual market prices, and the resulting synthetic bonds carry genuine cross-market information. At maturities beyond the FX forward horizon, however, the exchange rate link must be extrapolated, and the synthetic bonds largely reflect the extrapolation assumptions built into the FX forward curve rather than independent market information. Consequently, transfer learning provides the greatest value at maturities where FX forwards are observed, with diminishing returns at longer horizons. This prediction is strongly confirmed in our empirical analysis. To make this discussion more specific, we propose a KR-based estimation and extrapolation method for the FX forward curve in the following section.

2.4 Estimating the FX Forward Curve

To operationalize the KR-Transfer framework and construct the synthetic bonds (8), we need an estimate of the FX forward curve $F_a(x)$ on a dense and wide maturity grid, but we only observe a limited set of FX forward quotes.

To this end, we capture deviations from CIP through the (non-annualized) log-FX spread curve $\eta_a(x)$, defined as the difference between the log-forward premium and the CIP-baseline curve,

$$\eta_a(x) \coloneqq \underbrace{\log \frac{F_a(x)}{S_a}}_{\text{log-forward premium}} - \underbrace{\log \frac{g_{\mathrm{US}}(x)}{g_a(x)}}_{\text{CIP-baseline}}.$$

In terms of the synthetic discount curve, this can equivalently be written as $\eta_a(x) = \log g_a(x) - \log \widetilde{g}_a(x)$. By the same argument as above, we assume that market forces keep the log-FX spread curve $\eta_a(x)$ smooth in x, and we penalize large values of its curvature metric $\|\eta_a\|_\alpha^2$.

Given individual estimates for the US and target discount curves, $g_{\mathrm{US}}^{\mathrm{KR}}$ and g_a^{KR}, and observed FX spot S_a and forwards $F_a(\tau_{a,i})$ with maturities $0 < \tau_{a,1} < \cdots < \tau_{a,m_a}$, we thus estimate η_a by solving the KR problem

$$\min_{\eta_a \in \mathcal{H}} \left\{ \underbrace{\sum_{i=1}^{m_a} \frac{1}{m_a\, \tau_{a,i}^2} \Big(\log \frac{F_a(\tau_{a,i})}{S_a} - \log \frac{g_{\mathrm{US}}^{\mathrm{KR}}(\tau_{a,i})}{g_a^{\mathrm{KR}}(\tau_{a,i})} - \eta_a(\tau_{a,i}) \Big)^2}_{\text{log-forward premium fitting error}} + \underbrace{\lambda_a^{\mathrm{FX}}\, \|\eta_a\|_\alpha^2}_{\text{log-FX spread smoothness}} \right\}, \tag{11}$$

where $\lambda_a^{\mathrm{FX}} > 0$ is a regularization parameter.[6] The duration-type weights $1/(m_a\, \tau_{a,i}^2)$ transform

[6] By definition we have $\eta_a(0) = 0$, so we can assume that $\eta_a \in \mathcal{H}$.

the weighted mean squared error in (11) into the mean squared annualized log-forward-premium error, analogous to the duration weighting (7) used in the bond KR problem. The estimated full FX forward curve is then given in terms of the minimizer η_a^{KR} of (11) as

$$F_a^{\mathrm{KR}}(x) \coloneqq S_a \, \frac{g_{\mathrm{US}}^{\mathrm{KR}}(x)}{g_a^{\mathrm{KR}}(x)} \, \mathrm{e}^{\eta_a^{\mathrm{KR}}(x)}. \tag{12}$$

The argument about the information content of synthetic bonds in Section 2.3.4 can now be made more precise. As the KR estimate η_a^{KR} lies in $\mathcal{H}$, it is asymptotically flat, so that $\eta_a^{\mathrm{KR}}(x)/x \to 0$ for $x \to \infty$. Therefore, the annualized log-forward premium satisfies $(1/x)\log(F_a^{\mathrm{KR}}(x)/S_a) \sim y_a^{\mathrm{KR}}(x) - y_{\mathrm{US}}^{\mathrm{KR}}(x)$ for $x \to \infty$: it converges to the estimated baseline yield spread, exactly as it would under CIP. As a consequence, for large cash flow dates x_j at which no FX forward rate is observed, the synthetic bonds provide no genuinely new information beyond what is already embedded in the independently estimated curves $g_{\mathrm{US}}^{\mathrm{KR}}$ and g_a^{KR}. By contrast, at maturities where FX forwards are observed, the estimated $F_a^{\mathrm{KR}}(x)$ incorporates independent market prices, and the synthetic bond cash flows carry information that is not already captured by the standalone discount curve estimates.

3 Data and Implementation

3.1 Data

We construct our own comprehensive data set of government bonds from Bloomberg for the UK, Japan, Germany, Canada, and Switzerland. We query all available government bonds in these markets from 1960 to the end of March 2023, together with their static features. For the US market, we use the standard CRSP Treasuries Time Series, which contains a fuller universe of US Treasuries and underlies numerous studies of the US market. We take the CRSP sample from June 1961 to March 2023. We have verified that using Bloomberg data instead of CRSP for the US market produces essentially identical yield curve estimates, so the choice of data source does not affect our results.

We apply standard filters to the CRSP US data: we include only fully taxable, non-callable, and non-flower bond issues; we exclude on-the-run issues due to their liquidity and specialness. These filters follow Gürkaynak, Sack, and Wright (2007), Liu and Wu (2021) and Filipovic, Pelger, and Ye (2024). On-the-run issues refer to the two most recently issued securities at each of the 2, 3, 4, 5, 7, 10, 20, and 30-year maturities, among those issued since 1980.[7] The filters for the other markets are less standardized, and different studies and central banks apply different filters to remove bonds that may distort yield curve estimation.[8] We propose a set of rather general and standard filters that can

[7] The US data are the same as in Filipovic, Pelger, and Ye (2024) with the filter of excluding bonds with less than 90 days to maturity.

[8] For example, Ioannides (2003) advocates excluding bonds with less than one year to maturity because of liquidity concerns, and Bolder and Gusba (2002) and Bolder, Johnson, and Metzler (2004) suggest various filters to remove outliers and illiquid bonds. The Swiss National Bank excludes bonds with a residual maturity of less than 1 year.

be applied to all markets in our sample. For the set of bonds that are active (that is, whose pricing date lies between the bond's issue date and maturity) and quoted on the pricing date, we exclude: 1) embedded-optionality and non-standard-maturity instruments (specifically callable, sinkable, puttable, perpetual, and convertible structures); 2) floating rate notes and fungible securities; 3) bonds with less than 3 months to maturity because of liquidity concerns and their specialness; and 4) bonds with extreme yields relative to the cross-sectional median yield on each date. Specifically, we identify outliers as price observations whose YTM deviates from the cross-sectional median by more than 0.1. Figure A.3 shows the number of outliers identified over time in each market. This rule flags very few prices: on most days there are no outliers, and when any are detected they never exceed three in a day. We additionally exclude Swiss bills.[9] For the UK specifically, we exclude gilts within the 7-business-day ex-dividend window prior to a coupon date. Ex-dividend means that if bonds are purchased within the ex-dividend period, the buyer does not receive the coupon payment. The accrued interest is thus negative in this case.[10] Also for the UK, we exclude gilt quotes observed during their partly-paid issuance periods.[11] Finally, we drop dates with fewer than five observed bonds and days reported as holidays by Bloomberg. We compute accrued interest and construct the corresponding cash-flow streams according to each market's conventions, including coupon schedules, holidays, and day-count conventions. We also collect foreign exchange spot and forward rates from Bloomberg.

Figure 5 displays the number of bonds used in our estimation on each date for all markets after applying the filters described above. The US market consistently exhibits high coverage, from around 120 bonds in the late 1980s to over 250 in recent years. In contrast, non-US markets have much thinner coverage throughout the sample, except for Japan, where the number of available bonds increases sharply around 2000 after starting from the fewest bonds in the sample. Canada, Germany, and the UK typically range between 30 and 70 bonds. Note that the periodic drops in the number of securities in the UK are due to the exclusion of ex-dividend bonds. Switzerland has the sparsest coverage of around 10 to 20. This disparity in data coverage motivates our transfer learning framework.

The panel of available bonds in a given market is highly uneven across maturities. Taking the Japan market as an example, Figure 6a shows the time-to-maturity distribution of outstanding bonds over time. The large white bands illustrate persistent maturity gaps that a model must interpolate or extrapolate across. Before 2000, there are large gaps with no observations at both the short and long ends of the maturity spectrum. The coverage improves over time but remains

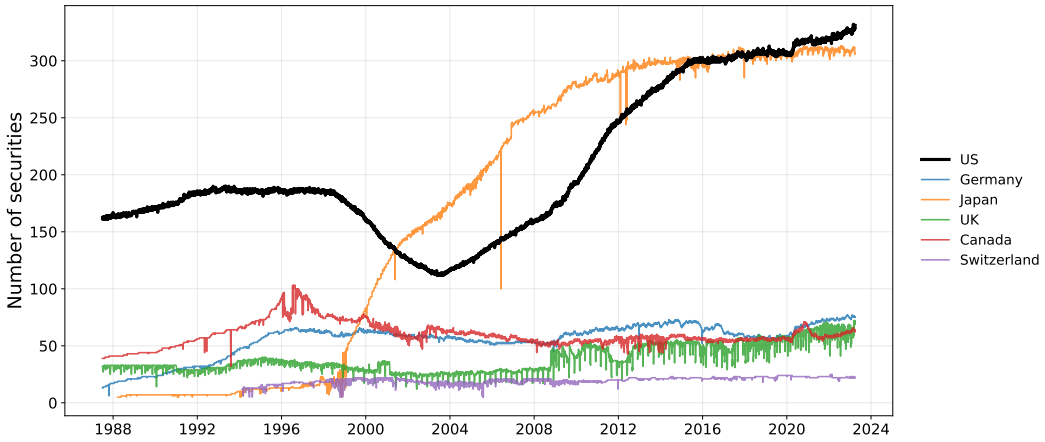
The Bank of England excludes zero-coupon bonds with residual maturities of 1 year or less. The European Central Bank estimates euro-area yield curves excluding bonds with residual maturities of less than 3 months.

[9]Swiss bills are nearly absent in the data, the Bloomberg source returns almost no quotes for them, so they contribute essentially nothing to the Swiss cross-section. This is consistent with market practice and related Swiss yield curve literature (Camenzind and Filipovic, 2023).

[10]They are subject to preferential tax treatment and past literature either excludes them or adjusts for the tax benefit.

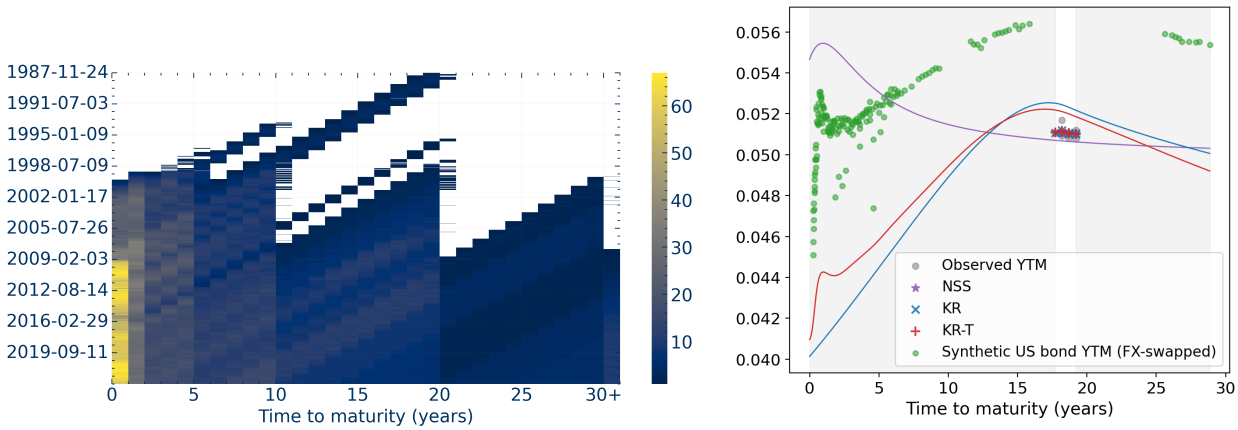
[11]Several gilts issued in 1992–94 were sold on a partly-paid basis. This means that the buyer settles the purchase price in installments over the weeks following issue, and until the final installment is paid the gilt is quoted at the paid-up amount only (roughly half of par). We exclude each affected gilt's quotes over its installment period only.

Figure 5: Number of Securities Used in Estimation by Country Over Time



The figure shows the number of securities that can be used in the yield curve estimation on each date for each country, after applying all data filters. The periodic drops in the number of securities in the UK are due to the exclusion of ex-dividend bonds.

Figure 6: Japan: observed maturities and an example date with limited data



(a) Observed maturities (rounded to years)

(b) Example date: 1989-07-11

The figure shows the observed bond maturities in Japan and the yield estimates on an example date. The left plot shows the observed government bond time to maturity distribution (rounded in years) for Japan. The large white bands illustrate persistent maturity gaps that a model must interpolate or extrapolate across. The right plot shows yield estimates and observed and fitted yields for Japan on the example date 1989-07-11, on which such a maturity gap exists. Shaded regions mark the extrapolation region. On this example date, the observed bond maturities are concentrated around 18 years; the fitted yields from the different methods (NSS, standalone KR, and KR-T) overlap at the observed maturities, while the estimated yield curves differ substantially away from them.

uneven, with some maturities having scarce observations until 2008. This pattern is common to all markets we study; Figure A.2 provides the full set of observed time-to-maturity distributions for the different markets.

Figure 6b illustrates the data availability for Japan on 1989-07-11. The few available bonds cluster around 18 years of maturity, so estimation at shorter and longer maturities becomes an extrapolation. For intuition, we also show the synthetic bonds obtained from the estimated FX

curve and the US yield curve. The methods we use to estimate the Japanese yield curve on this date differ substantially in the extrapolation region: our transfer learning approach KR-T pulls the KR estimate toward the synthetic bonds, which, as we show in subsequent sections, yields a more precise curve there. The figure also illustrates the weak CIP: the synthetic bonds generally have a level difference relative to the observed bonds, but our weak CIP penalty leverages only the information in the shape of the synthetic bond curve.

Figure A.1 shows the availability of foreign exchange spot and forward rates. The availability of FX forward maturities is below 5 years in the early period, but stable around 10 years for the past decade. This has important implications for the maturity range over which transfer learning can help. We begin the transfer learning analysis for each market according to the availability of FX forwards. For example, we start the analysis for Germany after 1998-12-14, when EUR/USD FX forward rates become available. Table A.2 summarizes the exact sample periods for the different estimations.

All our prices are end-of-day prices, which affects the interpretation of our yield curves estimated with transfer learning. There is a natural asynchronicity in the data because bond and FX markets close at different times across regions.[12] For bonds, the price corresponds to the last eligible contributed price from Bloomberg's composite (BGN) pricing source.[13] For FX spot and forward rates, the close is the New York close at 5:00 p.m. New York time. Thus, the transfer learning estimates combine local bond closes with US bond and FX observations available later in the New York trading day.[14] Our transfer learning curves should therefore be interpreted as daily end-of-day curves constructed from the latest available Bloomberg observations for each market.

3.2 Implementation

We compare the performance of standalone KR estimation (Filipovic, Pelger, and Ye, 2024), our novel KR-T estimation, and the widely used Nelson–Siegel–Svensson (NSS) model. We implement the NSS model on the same underlying dataset as our KR and KR-T methods following the procedure in Svensson (1994). The NSS model is the natural benchmark as it is used for the official yield curves in Switzerland and Germany[15], and its implementations, such as the publicly available yield curves of Gürkaynak, Sack, and Wright (2007), are a standard reference in academia and industry.

3.2.1 KR Implementation

We implement KR as in Filipovic, Pelger, and Ye (2024) by solving the KR problem (4). Our objective function minimizes duration-weighted pricing errors, (7), which correspond up to first

[12] We use Bloomberg PX_LAST field with BGN (Bloomberg Generic) for bonds and FX rates.

[13] These prices are consensus quotes rather than necessarily executed transaction prices. The end-of-day value is recorded according to Bloomberg's market-specific end-of-day cutoff time. This cutoff defines the "close" for recording purposes, although Bloomberg can continue to receive and incorporate quotes after this time if they meet quality criteria.

[14] For example, a German government bond quote reflects the local European bond market close, while the corresponding FX rate is measured at the New York close and US bond prices reflect the US bond market close.

[15] Appendix E provides an overview of the methods used by central banks in the different countries.

order to yield pricing errors. This is the same weighting as used, among others, by Gürkaynak, Sack, and Wright (2007). We use cross-validation to select the optimal tuning parameters. In order to have a meaningful scale for the smoothness parameter λ_a, we scale the smoothness penalty term by 2,000 (so that "$\lambda_a = 1$" means a smoothness penalty factor of 1/2000). Under this scaling, the duration-weighted pricing errors and the smoothness term have a comparable magnitude. The scaling is convenient, as it turns out that for all the countries it implies a close-to-optimal $\lambda = 1$.[16] The resulting estimate is $g_a^{\mathrm{KR}}(x)$ for each country a.

3.2.2 KR-T Implementation

Our KR-T framework operates in three steps:

(i) **Step 1 (independent KR estimation for each country)**:
We estimate the discount curve $g_a^{\mathrm{KR}}(x)$ with KR separately for each country a by solving the KR problem (4). These are the same estimates that we use for the KR standalone evaluation. Hence, we also use the optimal tuning parameters from the standalone estimation.

(ii) **Step 2 (FX forward curve estimation)**:
We estimate the FX forward curve with the KR method for each US/country pair. Specifically, we estimate the log-FX spread curve η_a^{KR} by solving (11) using observed FX spot and forwards and the CIP-baseline curve $\log(g_{\mathrm{US}}^{\mathrm{KR}}(x)/g_a^{\mathrm{KR}}(x))$. This yields the estimated FX forward curve F_a^{KR} in (12).

Based on cross-validation results, we set the regularization parameter for the FX curve estimation to $\lambda^{\mathrm{FX}} = 1$ across all currency pairs. This optimal choice balances smoothness against accuracy to the observed FX forward quotes and produces stable forward exchange rate curves across all markets and sample periods.

(iii) **Step 3 (KR-T estimation with synthetic bonds).**
We construct synthetic bonds $(\widetilde{P}_a, \widetilde{C}_a)$ via (8) using the FX spot S_a and estimated forwards $F_a^{\mathrm{KR}}(x_j)$. We then jointly re-estimate the target discount curve g_a and synthetic discount curve $\widetilde{g}_a$ by solving the KR-Transfer problem (9), selecting the transfer parameter θ via cross-validation.

The three-step structure ensures that the synthetic bonds are economically meaningful, as they are priced using actual FX market data rather than assumed parity, while the closed-form KR-Transfer solution (Theorem 2) keeps computation tractable.

3.2.3 NSS Implementation

The Nelson–Siegel–Svensson (NSS) model (Nelson and Siegel, 1987; Svensson, 1994) is the most widely used parametric method for estimating the discount curve. The NSS forward curve is

[16] Filipovic, Pelger, and Ye (2024) use a slightly different scale normalization of the smoothness term. The close-to-optimal λ is still very similar.

parametrized as

$$f_{\text{NSS}}(x) = \gamma_0 + \gamma_1 e^{-\frac{x}{\tau_1}} + \frac{x}{\tau_1}\gamma_2 e^{-\frac{x}{\tau_1}} + \frac{x}{\tau_2}\gamma_3 e^{-\frac{x}{\tau_2}} \tag{13}$$

for the parameters γ_0, γ_1, γ_2, γ_3 and $\tau_1, \tau_2 > 0$. This results in the discount curve $g_{\text{NSS}}(x) = \mathrm{e}^{-\int_0^x f_{\text{NSS}}(t)\,dt}$. Our implementation minimizes the sum of squared YTM errors. Because estimating this non-linear model can be numerically unstable, we use multiple numerical solvers to ensure convergence.

3.3 Evaluation Design

Our main results are based on out-of-sample (OOS) comparisons of the different models. Given the unevenness in the observed data, we conduct two sets of out-of-sample cross-validation: stratified cross-validation, which ensures that each fold is representative of the overall data set, and contiguous block cross-validation, which mimics real-world missing-data patterns where we usually need to interpolate over a gap. We use stratified cross-validation to study the effect of the hyperparameters λ, α, and θ. We divide the data into the following nine maturity buckets: 3M-1y, 1y-2y, 2y-3y, 3y-4y, 4y-5y, 5y-7y, 7y-10y, 10y-20y and $>$ 20y.[17] Given that the availability of bonds in some markets is limited, we split the bond data each day into 3 stratified folds. Each fold has roughly the same maturity distribution as the whole sample. We use two folds for estimation and the remaining fold for out-of-sample evaluation, and rotate the folds until each fold has been evaluated out-of-sample.

The contiguous 5-year block cross-validation provides an overall quantitative measure to assess a model in interpolating and extrapolating across these maturity gaps. This design mimics the real-world situation where yields need to be estimated at untraded maturities. Each fold withholds all bonds with time-to-maturity in a contiguous 5-year interval ($(0, 5]$, $(5, 10]$, or $(10, 15]$), estimates the curves on the remaining data, and calculates OOS YTM errors on the held-out block. Each bond is held out and evaluated exactly once; every maturity block is tested. The overall OOS YTM RMSE then provides a more appropriate measure of how well a method estimates yields at maturity gaps that frequently appear in different markets.

We use three root-mean-squared error (RMSE) metrics to compare performance: duration-weighted pricing error, relative pricing error, and yield error. In the following, $P^g_{i,t}$ and $Y^g_{i,t}$ denote the model implied prices and yields, where g is a placeholder for the estimated discount curve. The errors RMSE_t are reported for each day t within the sample and aggregated over the full sample,

$$\text{RMSE} = \frac{1}{T}\sum_{t=1}^{T} \text{RMSE}_t,$$

and given by either:

- **Duration weighted pricing RMSE** This approximates yield errors and corresponds to the

[17] As we filter out bonds with time to maturity less than 3 months, the shortest maturity bucket starts at 3M. The definition of maturity buckets follows closely with the one by the Treasury, Liu and Wu (2021) and Filipovic, Pelger, and Ye (2024).

objective of the optimization problems (4) and (9),

$$\text{RMSE}_t = \sqrt{\frac{1}{M_t}\sum_{i=1}^{M_t} \frac{1}{(D_{i,t}P_{i,t})^2}\left(P_{i,t} - P_{i,t}^g\right)^2}.$$

- **Relative pricing RMSE** This is the percentage pricing error of coupon bonds, also used in Filipovic, Pelger, and Ye (2024). It can be computed by scaling bond prices and cash flows so that prices are normalized to 100, and then calculating RMSE_t as the mean pricing difference at time t. Equivalently,

$$\text{RMSE}_t = \sqrt{\frac{1}{M_t}\sum_{i=1}^{M_t} \frac{1}{P_{i,t}^2}\left(P_{i,t} - P_{i,t}^g\right)^2}.$$

- **Yield RMSE** This measures the discrepancy between the observed yields of coupon bonds and the model-implied yields.

$$\text{RMSE}_t = \sqrt{\frac{1}{M_t}\sum_{i=1}^{M_t} \left(Y_{i,t} - Y_{i,t}^g\right)^2}$$

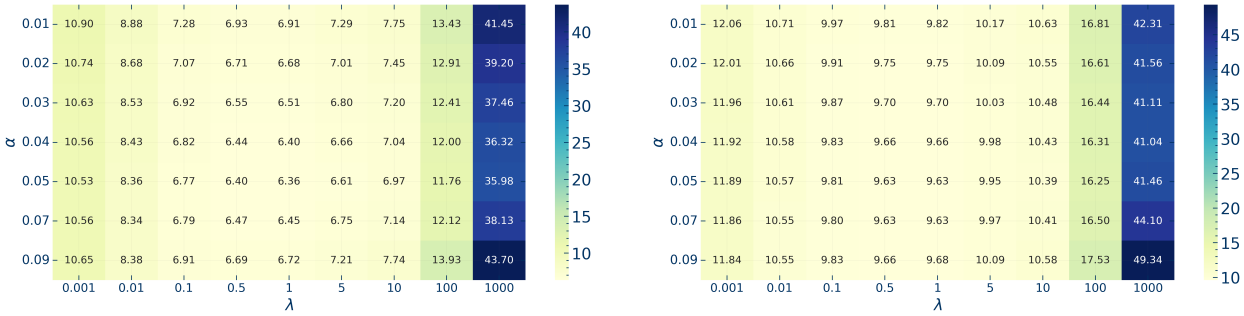
The out-of-sample RMSE of yields and discount bond prices is a meaningful way to evaluate how close the estimated discount curves are to the true discount curve. The observed bond prices represent the true, fundamental values plus noise, (1). A flexible model can obtain in-sample RMSE of zero by overfitting the noise. In fact, our KR estimator obtains a perfect in-sample fit for a vanishing smoothness parameter λ. However, we show that such an overfitted model has a large out-of-sample error, and hence does not provide a good representation of the true discount curve.

Table A.2 provides the exact sample periods for the different estimations. The metrics that are aggregated over time include all the dates for which they can be calculated. Specifically, for block cross-validation, we only include dates for which at least two non-empty 5-year maturity blocks (one for each train and test) exist in the target market. The 3-fold stratified CV can be applied on each day as we require at least 5 bonds for a date to be included in our sample.

4 Empirical Results: KR

Before analyzing the transfer learning framework, we establish that standalone kernel ridge regression (KR) uniformly dominates the Nelson–Siegel–Svensson (NSS) parametric benchmark across all markets. This confirms that the findings of Filipovic, Pelger, and Ye (2024) for US government bonds generalize to all major government bond markets in our sample. In sum, KR provides the best yield curves for international markets when countries are considered individually.

Figure 7: Cross-validation YTM RMSE for α and λ



(a) UK **(b)** US

The figure shows the out-of-sample yield RMSE in basis points (BPS) from stratified cross-validation on monthly data for different values of maturity weight α and smoothness parameter λ. Panel (a) shows the UK; panel (b) shows the US. The performance surface is remarkably flat across a wide range of λ values in both markets, confirming that $\lambda = 1$ and $\alpha = 0.05$ provide a universally robust configuration.

4.1 Parameter Selection

We start our analysis by selecting the optimal parameter values for the KR method, which is completely specified up to the maturity weight α and the smoothness parameter λ. We choose the optimal parameters in a data-driven way by minimizing the out-of-sample yield error using 3-fold stratified cross-validation on monthly data. This ensures that the selected parameters are robust across maturity buckets.

Figure 7 shows the out-of-sample yield RMSE from stratified cross-validation for different values of α and λ for the UK and the US. First, the choice of α has a negligible effect on the fitted yields if it is within a reasonable range. This is expected as α corresponds to the yield of an infinite-maturity discount bond as formalized in Filipovic, Pelger, and Ye (2024), and it mainly affects the extrapolation for maturities that are not observed. In contrast, the choice of the smoothness parameter λ has a large effect on the estimation. The optimal value for the US and the UK is attained for $\lambda = 1$. We conclude that an infinite-maturity yield of $\alpha = 0.05$ and smoothness parameter of $\lambda = 1$ provides a (close to) optimal model.

Figure A.4 collects the results for all markets. The baseline parameters $\lambda = 1$ and $\alpha = 0.05$ are optimal or near-optimal for every country, so they provide a universally robust configuration without per-market tuning.[18]

Our findings show that discount curves with small out-of-sample errors need to be sufficiently smooth. This suggests that the true discount curve itself has to be sufficiently smooth with respect to our smoothness measure. A perfect in-sample fit by setting $\lambda = 0$ results in large out-of-sample errors. On the other hand, inflexible curves with large λ also lead to large out-of-sample errors. This suggests that the true discount curve cannot be too simplistic either. Based on the cross-

[18] We have also studied the extension of the smoothness penalty that includes a tension parameter δ to penalize the first derivative of the discount curve as in Filipovic, Pelger, and Ye (2024). Their findings from the US market generalize to all other markets, and $\delta = 0$ is (close to) optimal. For the sake of brevity we only include the empirically relevant case of measuring smoothness with the second derivative. The results are available upon request.

Table 1: KR vs NSS: Stratified Cross-Validation Comparison

Country	Method	YTM RMSE (BPS)	Dur. Wt. Pricing RMSE (BPS)	Rel. Pricing RMSE (BPS)
US	KR	**9.63**	**9.61**	**26.02**
	NSS	11.01	11.00	40.48
Germany	KR	**4.57**	**4.56**	**25.72**
	NSS	6.83	6.82	53.52
Japan	KR	**3.53**	**3.54**	**36.20**
	NSS	6.69	6.71	75.00
UK	KR	**6.36**	**6.35**	**37.94**
	NSS	9.57	9.52	103.60
Canada	KR	**6.49**	**6.49**	**42.87**
	NSS	8.77	8.75	69.10
Switzerland	KR	**6.43**	**6.47**	**71.94**
	NSS	8.76	8.78	83.39

The table compares the out-of-sample performance of standalone KR ($\lambda = 1$, $\alpha = 0.05$) and NSS for each market using three-fold stratified cross-validation on monthly data. Columns are ordered by YTM RMSE, duration-weighted pricing RMSE, and relative pricing RMSE. All numbers are in basis points (BPS). KR uniformly dominates NSS across all markets.

validation, we define our baseline KR model as $\lambda = 1$, $\alpha = 0.05$, which we use in the following sections for our comparison study and extension to KR-T.

4.2 Comparison Study

The KR method uniformly dominates NSS in every metric and market. Table 1 summarizes the out-of-sample performance of KR versus NSS for each market using 3-fold stratified cross-validation on monthly data. The yield RMSE reductions range from 13% (US) to 47% (Japan), with reductions of 34% (UK), 33% (Germany), 26% (Canada), and 27% (Switzerland). These results confirm that the superior performance of KR over NSS documented by Filipovic, Pelger, and Ye (2024) for US bonds generalizes to all major government bond markets.

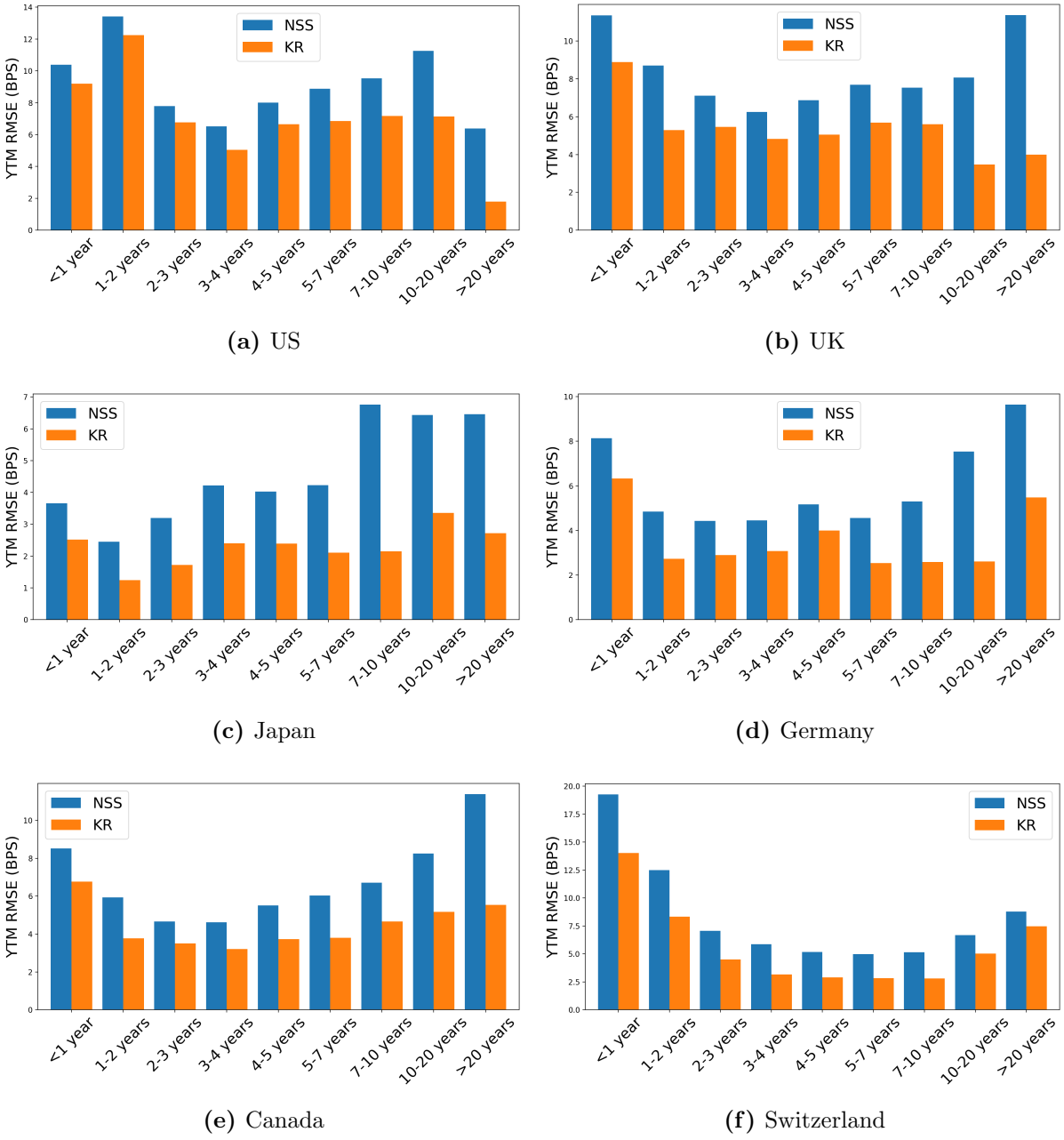
Importantly, KR's outperformance is larger outside the US: the RMSE improvement over NSS is about 1.4 basis points in the US, more than twice as large in Japan and the UK (about 3.2 basis points), and about 2.3 basis points in Germany, Canada, and Switzerland. The gains are thus more pronounced in countries with less data than the US.

As expected, the duration-weighted pricing error is essentially identical to the yield error. By construction, the relative pricing error puts larger emphasis on longer maturities as the same yield error has a larger effect for a longer maturity bond. Hence, different metrics put different weights on different parts of the maturity spectrum.

4.3 Comparison Across Maturities

Figure 8 confirms that KR dominates NSS across maturity buckets for all markets. These uniformly better results imply that the KR discount curves are closer to the ground truth than the NSS curves.

Figure 8: YTM RMSE by Maturity Buckets Across Markets



(a) US **(b)** UK

(c) Japan **(d)** Germany

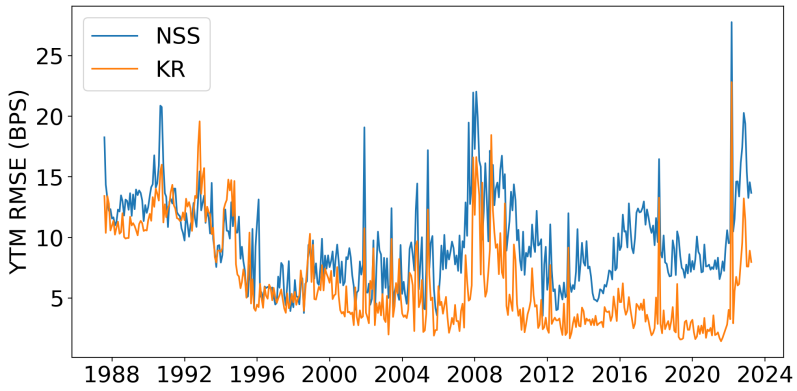
(e) Canada **(f)** Switzerland

This figure shows the out-of-sample YTM RMSE across maturity buckets for each standalone market yield estimation with the KR and NSS method. Each bar shows the error of the maturity bucket. Monthly data are used for cross-validation. Errors are reported in basis points (BPS).

Importantly, the bucket-level results show that the accuracy gains at certain maturities are even larger than the averaged metrics suggest. For example, for UK bonds with maturity beyond 20 years, KR improves on NSS by 65%. For other countries, the improvements are similarly substantial and unevenly distributed across maturities.

The maturity buckets with the largest improvements vary by country. For the US, the largest

Figure 9: OOS YTM RMSE over time: UK



Out-of-sample YTM RMSE for KR and NSS in the UK over time. Monthly data are used for stratified cross-validation. Errors are in basis points (BPS). KR consistently outperforms NSS over time and across all maturity buckets.

error reductions relative to NSS are at long maturities, whereas Japan, the UK, and Switzerland also show large improvements at short maturities. Data availability differs substantially across countries: Japan, the UK, and Switzerland have few short-maturity bonds in the earlier part of the sample, and a robust method like KR provides better estimates in such cases. This lack of short-maturity observations becomes particularly relevant under transfer learning, where we can exploit FX prices and observed US bonds at short maturities. In summary, the improvements at different maturities also depend on bond availability, which is heterogeneous across countries.

4.4 Comparison Over Time

The relative comparison results are robust over time. Figure 9 shows the monthly yield errors for the UK, and Figure A.5 collects the corresponding results for the other countries. Overall, KR has smaller yield errors in almost all time periods.

In some periods, KR's outperformance is particularly large. During the 2008–2009 financial crisis and the COVID-19 period, for example, the gap between KR and NSS widens, and for the UK, KR reduces yield errors by around 10 basis points relative to NSS. The aggregate numbers can therefore understate how much better KR is than NSS.

We can also replicate the other main results of Filipovic, Pelger, and Ye (2024) and show that they generalize to the other countries, including the robustness, stability, and forward-rate results.

5 Empirical Results: KR-Transfer

Having established KR as the dominant single-market method, we now analyze the transfer learning framework. We begin by selecting the transfer parameter θ.

5.1 Transfer Parameter Selection

As the KR estimation is robust to a range of α and λ values, we use the common parameters $\alpha = 0.05$ and $\lambda = 1$ for all markets in KR-T. We then select the transfer parameter θ that controls the strength of information sharing. Our objective is to use information from data-rich markets to improve estimation in markets with sparse observations, particularly at maturities where local data are limited or missing. To assess the impact of different θ values, we use the contiguous 5-year block cross-validation, which creates a realistic test environment for evaluating the model's ability to fill maturity gaps. We conduct out-of-sample cross-validation on country pairs for a range of transfer parameters $\theta \in \{0, 0.1, 1, 10, 100\}$. We estimate the FX curves with $\lambda^{\text{FX}} = 1$, which is close to the optimal choice based on cross-validation.[19] As for the discount curve smoothness parameter, we scale the spread and FX penalty terms by 2,000.

Figure 10a shows the out-of-sample yield errors for each market and value of θ under block cross-validation. The optimal transfer parameter θ is 1 for Japan and 10 for Canada, the UK, Germany, and Switzerland; recall that $\theta = 0$ corresponds to standalone KR. Allowing information sharing across markets cuts the OOS YTM RMSE by about 50% for Canada, 45% for Germany, and more than 40% for the UK and Switzerland.

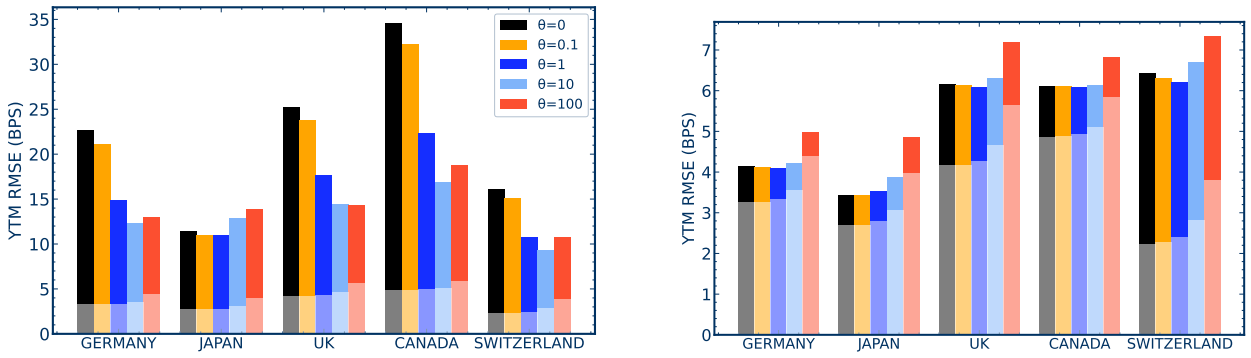
To visualize the benefit of transfer learning in filling maturity gaps, Figure 11 shows the yield estimates for Canada on 2004-06-30, where bonds with time to maturity below five years are masked from estimation. Standalone KR ($\theta = 0$) fails to capture the shape of the short-end curve in the absence of local data, while the transfer learning approach (KR-T with $\theta = 10$) successfully extrapolates the short-end yields using cross-market information from the US market.

To confirm the robustness of our findings, we also perform 3-fold stratified cross-validation for KR-T. Recall that stratified cross-validation randomly partitions bonds across folds while preserving the overall maturity distribution. Because each fold is a miniature of the full panel, this design tests whether KR-T is robust to price noise. Figure 10b shows that KR-T remains competitive in this setting, where bonds of different maturities are generally present in both the training and test folds. Using the θ selected from the block cross-validation, the extra cross-market information does not distort the fit when local data are present.

In both Figures 10a and 10b, the overlapping transparent bars represent in-sample performance for each θ when all observed bonds are used without masking. The gap between the transparent bars (in-sample, all bonds) and the solid bars (out-of-sample, masked bonds) is much more pronounced under block cross-validation. This indicates that when the observed data contain a maturity gap, interpolation and extrapolation become much harder, and naive cross-validation underestimates the true out-of-sample error. Given its competitive performance under stratified cross-validation and its superior performance under the 5-year block cross-validation, KR-T is the more robust choice when real-world maturity gaps inevitably arise.

[19] The FX curve smoothness parameter λ^{FX} can also be selected jointly with θ to optimize the yield error for KR-T. We find that $\lambda^{\text{FX}} = 1$ is also close to the optimal value for this different objective and that the results are generally robust to the choice of λ^{FX}.

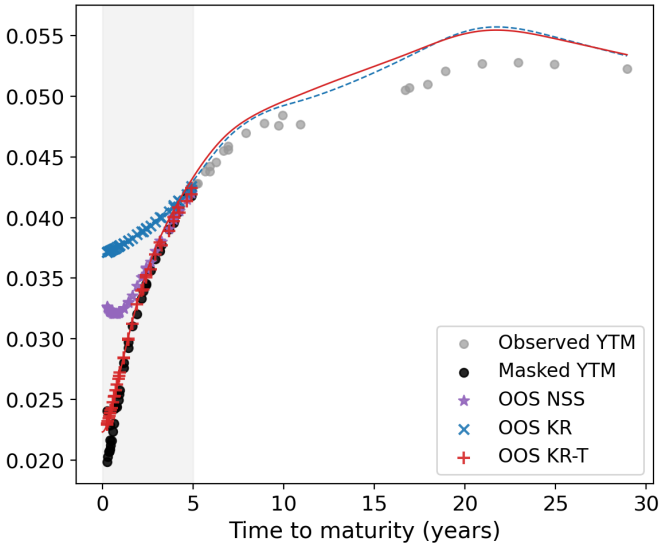
Figure 10: Cross-Validation: OOS YTM RMSE



(a) Block Cross-Validation **(b)** Stratified Cross-Validation

The figure shows the out-of-sample YTM RMSE from cross-validation of the transfer learning method KR-T for values of the transfer parameter $\theta \in \{0, 0.1, 1, 10, 100\}$, where $\theta = 0$ corresponds to the standalone KR method. Each pair of overlapping bars at a given country and θ contrasts the in-sample benchmark (transparent) with the out-of-sample YTM RMSE (solid). The left panel shows the 5-year block cross-validation; the right panel shows the 3-fold stratified cross-validation. Under block cross-validation, the optimal transfer parameter is $\theta = 1$ for Japan and $\theta = 10$ for Canada, the UK, Germany, and Switzerland. Information sharing cuts the OOS YTM RMSE by about 50% for Canada, 45% for Germany, and more than 40% for the UK and Switzerland. The right panel confirms that the OOS performance is also robust under the standard stratified cross-validation.

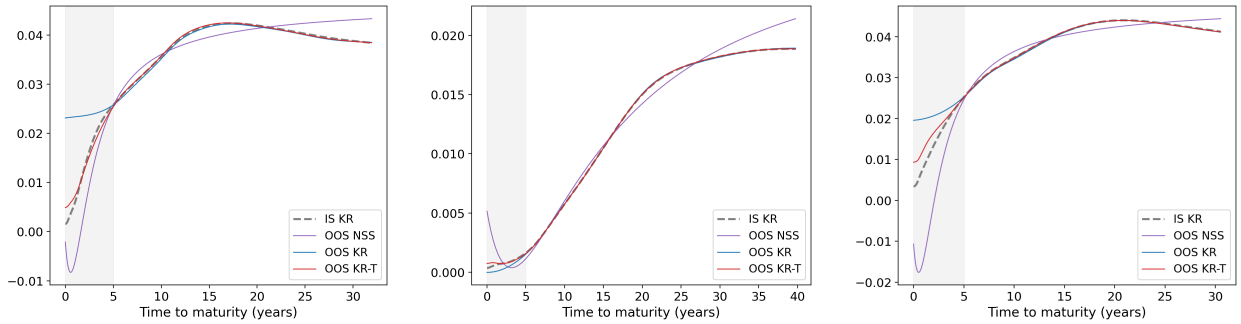
Figure 11: Yield estimates with masked short maturities: Canada, 2004-06-30



Illustrative example of transfer learning between the US and Canada. Bonds with 0–5 year maturities are masked from estimation (light grey band). Short-end yields are successfully extrapolated using cross-market information (KR-T with $\theta = 10$), while the standalone KR method fails to capture the shape of the short-end curve in the absence of local data.

In the following, we set the baseline parameters $\lambda = 1$, $\theta = 10$, and $\alpha = 0.05$ for KR-T across all countries with the exception of $\theta = 1$ for Japan; $\lambda = 1$ and $\alpha = 0.05$ for KR; and $\lambda^{\text{FX}} = 1$ for FX curve estimation.

Figure 12: Yield curve estimates with extrapolation for 0–5y



(a) Canada, 2009-06-30 **(b)** Japan, 2014-06-30 **(c)** Germany, 2009-12-31

This figure shows the extrapolated yield estimates on three example dates in different markets. For each example date, bonds with 0–5y time to maturity are excluded from estimation (light grey band). We illustrate the effect of transfer learning by comparing our KR-T method with the KR method and NSS method. KR-T uses transfer parameter $\theta = 10$ for all markets except Japan, where $\theta = 1$, with cross-market information from the US market. The in-sample (IS) KR curve, which corresponds to the KR estimates when no bonds are masked, is added as a proxy for the true yield curve.

5.2 Comparison Study

KR-T provides the best discount curves when data are missing in larger blocks and matches KR when local data are sufficient. Both KR-based estimators substantially outperform the Nelson–Siegel–Svensson benchmark used by many central banks.

Table 2 reports the out-of-sample results under both contiguous 5-year block cross-validation and 3-fold stratified cross-validation, and Figure A.8 visualizes the block CV comparisons. We evaluate each market with quarterly data up to Q1 2023, with start dates determined by data availability (UK from 1987, Canada from 1988, Japan from 1992, Switzerland from 1994, Germany from 1998); Table A.2 reports the exact periods. Across all five target markets and all three error metrics, KR-T uniformly dominates both benchmarks under block cross-validation. Relative to NSS, KR-T reduces yield RMSE by 71–86%, with the largest improvements in Germany (86%) and Switzerland (80%). Relative to standalone KR, KR-T reduces yield RMSE by 4–51%, with the most pronounced gains in Canada (51%) and Germany (46%), precisely the markets with the most severe maturity gaps. For Japan, the reduction versus standalone KR is more modest (4%), reflecting the extreme sparsity of the early sample period (Figure 6a), which limits meaningful cross-validation. Importantly, even in this challenging case, transfer learning does not hurt performance.

Figure 12 shows the extrapolated yield curve on three example days when no local quotes of bonds with maturity within 0–5 years are available for estimation. We simulate this scenario by masking bonds with maturity below five years on these dates. The figure shows that KR-T (red line) provides more stable and reliable short-end extrapolation than both NSS (purple line) and standalone KR (blue line). Without short-end local information, both KR and NSS deviate from the true yield curve (the in-sample KR curve), exhibiting too much or too little curvature, whereas KR-T recovers the short-end shape as if the short-maturity local bonds were present.

Table 2: Combined Cross-Validation Summary Results

		5-yr Block CV			Stratified CV		
Market	Method	YTM	Dur. Wt.	Rel. Pricing	YTM	Dur. Wt.	Rel. Pricing
GERMANY	KR	22.62	22.57	44.42	4.13	4.13	17.90
	KR-T	**12.30**	**12.29**	**39.43**	4.21	4.21	19.18
	NSS	87.05	87.76	133.38	6.26	6.25	52.24
JAPAN	KR	11.40	11.48	84.26	3.42	3.44	34.42
	KR-T	**10.93**	**11.00**	**82.86**	3.52	3.53	35.60
	NSS	43.14	43.26	146.93	6.69	6.71	75.10
UK	KR	25.23	25.15	52.86	6.15	6.15	36.70
	KR-T	**14.45**	**14.42**	**48.27**	6.29	6.29	37.67
	NSS	51.09	51.27	125.64	9.06	9.02	97.18
CANADA	KR	34.61	34.49	68.67	6.10	6.10	39.65
	KR-T	**16.84**	**16.80**	**57.40**	6.14	6.14	40.96
	NSS	57.99	58.26	99.81	8.21	8.19	63.17
SWITZERLAND	KR	16.14	16.12	82.15	6.42	6.46	71.25
	KR-T	**9.29**	**9.33**	**80.82**	6.70	6.75	74.91
	NSS	45.62	46.00	121.51	9.40	9.41	85.37

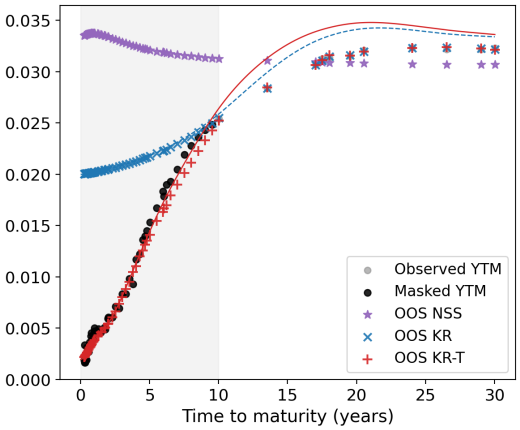
The table shows aggregate evaluation metrics (all in BPS) calculated out-of-sample using both 5-year block cross-validation (left) and 3-fold stratified cross-validation (right) for KR, KR-T, and NSS in each market. KR-T uses transfer learning from the US market with transfer parameter $\theta = 10$ for all markets except Japan, where $\theta = 1$. Quarterly data are used. The Germany sample is restricted to the period with available EUR/USD FX data. Under block CV, KR-T uniformly dominates both benchmarks. Under stratified CV, where each fold preserves the maturity distribution, KR-T remains competitive with standalone KR (within about 0.3 BPS), confirming that transfer learning does not distort local fitting when local data are sufficient. Bold indicates the best method for each market and metric under block CV.

The effect holds and is more pronounced when a larger maturity block is unobserved. Figure 13 shows the yield estimates for Germany on the example date 2010-06-30, when bonds with maturity within 0–10 years are masked from estimation. In the extrapolation region (0–10y), only the KR-T curve reflects the true (in-sample) shape; the other two methods fail to capture it because they lack local data. Figures A.16 to A.20 collect more examples for all countries, which illustrate that the same findings carry over to other dates and countries.

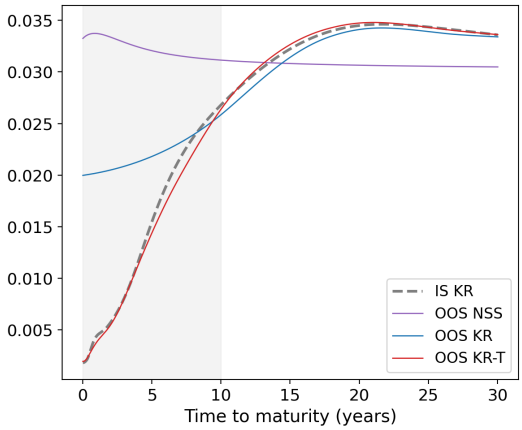
These results highlight the benefits of transfer learning in yield curve estimation to address data sparsity issues commonly encountered in smaller or less liquid bond markets. By learning from the US market, which typically has richer data across maturities, our KR-T estimation framework provides robust and reliable extrapolation of yields in scenarios with limited local maturity coverage.

The right panel of Table 2 shows the results under stratified cross-validation. Here, where each fold preserves the maturity distribution of the full sample, KR-T remains competitive with standalone KR across all markets (within about 0.3 basis points of yield RMSE), while both KR and KR-T substantially outperform NSS (reductions of roughly 40%). The key finding is that transfer learning does not compromise local fitting accuracy when local data are sufficient, confirming that

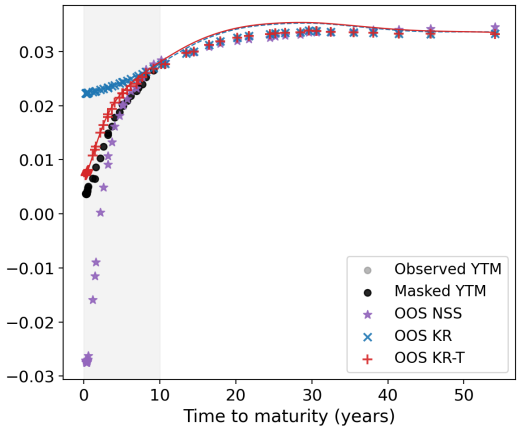
Figure 13: Yield curve estimates with extrapolation for 0–10y



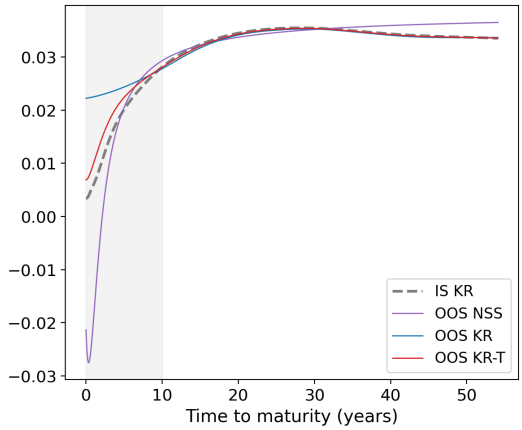
(a) Fitted vs observed YTM (Germany, 2010-06-30)



(b) Yield estimates (Germany, 2010-06-30)



(c) Fitted vs observed YTM (UK, 2014-06-30)



(d) Yield estimates (UK, 2014-06-30)

This figure shows out-of-sample fitted and observed YTM (left) and yield estimates (right) with all bonds of 0–10y maturity masked from estimation (light grey band). Panels (a)–(b) show Germany on 2010-06-30; panels (c)–(d) show the UK on 2014-06-30. KR-T uses transfer parameter $\theta = 10$, with cross-market information from the US market. Out-of-sample NSS and KR estimates that also extrapolate the 0–10y region are included for comparison. The in-sample (IS) KR curve is estimated using all available bonds.

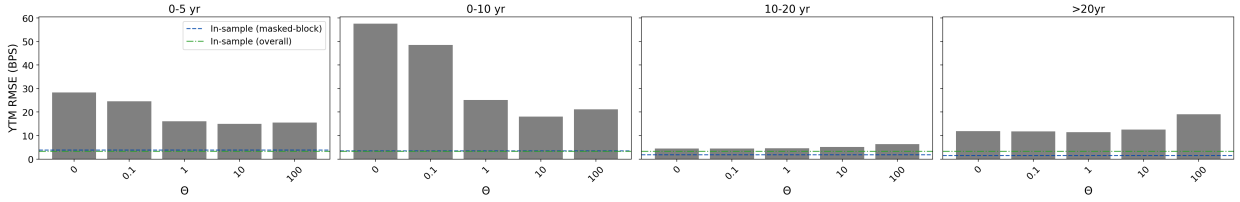
the framework "does no harm" in data-rich regimes while delivering substantial improvements when maturity gaps are present.

5.3 When and Where Does Transfer Learning Help?

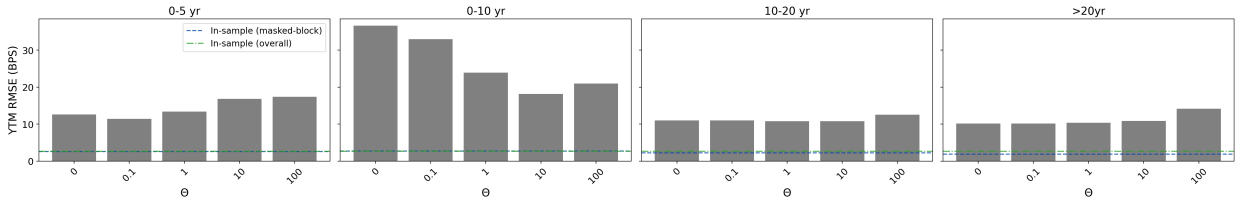
A central question is at which maturities, and under what data conditions, cross-market information improves yield curve estimation. We address this through a masked-maturity block analysis that isolates the contribution of transfer learning by maturity segment.

For each country pair (target market paired with the US), we withhold all bonds whose time-to-maturity falls within a specific maturity window (short: 0–5y or 0–10y; medium: 10–20y; or

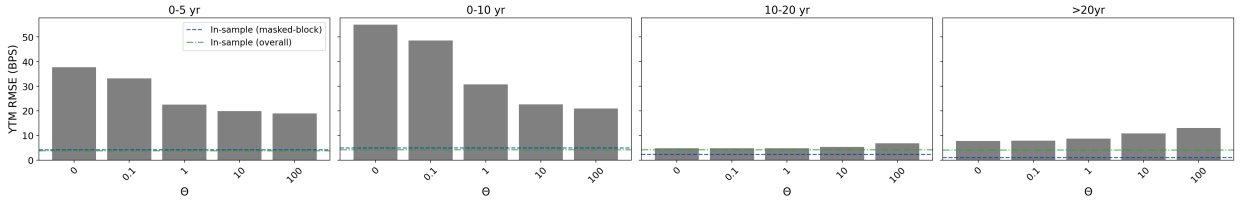
Figure 14: Out-of-sample YTM RMSE by masked maturity block



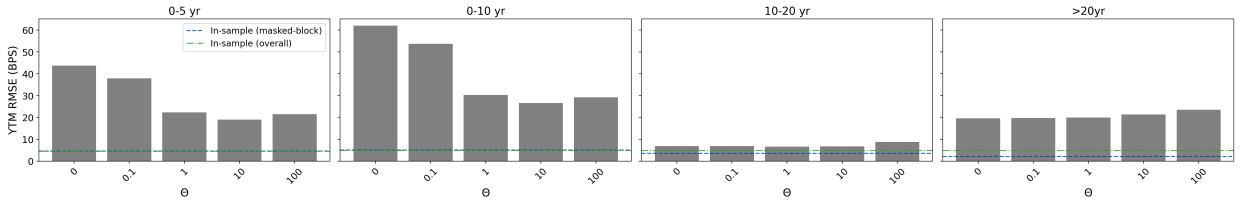
(a) Germany



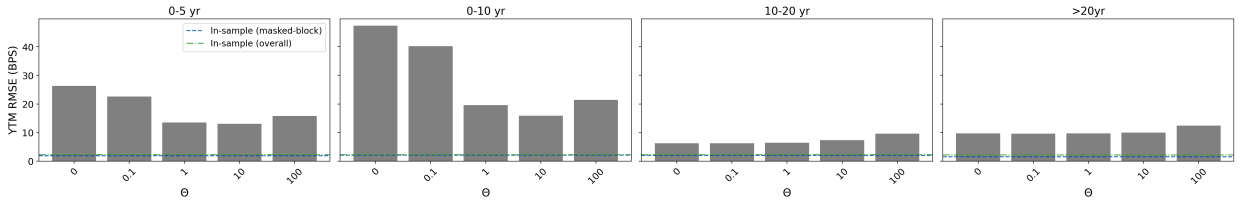
(b) Japan



(c) UK



(d) Canada



(e) Switzerland

The figure shows out-of-sample YTM RMSE across $\theta \in \{0, 0.1, 1, 10, 100\}$ for short, medium, and long maturity blocks when jointly estimating each country's yield curve with the US. For each country, all bonds with time-to-maturity in the indicated window are withheld from estimation. The overall in-sample line is the YTM RMSE evaluated over all bonds when the yield curve is independently estimated with all observed bonds. Transfer learning gains are largest for the short-end blocks, where FX forwards are directly observed, grow with the size of the masked block, and vanish at medium and long maturities.

long: >20y), estimate the joint curves on the remaining bonds, and evaluate out-of-sample errors on the withheld instruments. This creates a controlled experiment that directly measures the

model's ability to extrapolate into maturity gaps using cross-market information. The two short-end windows are nested by design. Masking 0–10y removes twice as much of the maturity spectrum as masking 0–5y and pushes the nearest observed bond further away from the withheld maturities, turning a moderate extrapolation problem into a severe one. Comparing the two windows therefore reveals how the value of transfer learning grows as the local information deteriorates. This missing-data structure is also the empirically relevant one: as documented in Section 3, short-maturity bonds are missing for extended periods in several markets, with contiguous gaps that extend from the short end well beyond five years (Figures 6a and A.2). The masked blocks thus replicate the estimation problem that arises in the actual data rather than an artificial one.

Figure 14 presents the results. The pattern is striking: the gains from transfer learning are determined by the availability of observed FX forward rates, which are the sole channel through which US bond information becomes genuinely informative about the target currency. When short maturities are masked, transfer learning produces large error reductions; these are precisely the maturities where FX forward rates are most abundantly observed and most liquid. Even without any local short-maturity bonds, the synthetic bonds carry real price information because their cash flows are converted using observed FX forwards. The US bond data, combined with observed exchange rates at these maturities, provides a direct economic signal about what discount factors should be in the target currency. Moreover, the gains grow with the severity of the gap: for the larger 0–10y block, the errors of the standalone methods increase sharply relative to the 0–5y block, while KR-T contains the deterioration, so the improvements from transfer learning are largest exactly where extrapolation is hardest. For Germany, for example, KR-T at the baseline $\theta = 10$ cuts the out-of-sample RMSE by roughly half for the 0–5y mask and by roughly two-thirds for the 0–10y mask.

For medium maturities (10–20y), transfer learning provides little additional benefit. This is consistent with the economic channel: FX forward availability thins out beyond 10 years, so synthetic bonds carry little genuinely new information there. It also reflects that the masked 10–20y block poses an interpolation rather than an extrapolation problem: local bonds on both sides of the gap already anchor the curve, out-of-sample errors are small for all methods, and there is little room for improvement. At long maturities (>20y), where FX forwards are generally not observed, the exchange rate link between the two markets is missing. Synthetic bond cash flows at these maturities are converted using extrapolated rather than observed FX rates, so the resulting synthetic bonds do not contain genuinely new information from the US market; they largely reflect the extrapolation assumptions embedded in the FX curve estimation. Consequently, transfer learning offers no improvement over standalone estimation at long horizons. This maturity-dependent pattern provides strong evidence that the benefits of our framework are driven by economic fundamentals, namely the observed FX forward rates that establish the cross-currency link, rather than by a statistical regularization effect.

5.4 Persistence and Time Variation

We demonstrate that the transfer learning advantage is robust over time. Figure 15 shows the time series of in-sample and out-of-sample yield RMSE under block cross-validation, comparing KR and KR-T with $\theta = 10$ (and $\theta = 1$ for Japan). Three patterns stand out. First, KR-T consistently outperforms standalone KR across the full sample. Second, on some days the gains are substantial, so average results may understate the potential gains; the largest accuracy gains of KR-T relative to KR occur around the 2008–2009 financial crisis. Third, the in-sample estimates give essentially the same curves for KR and KR-T, confirming again that for an interpolation task with sufficient observations, transfer learning does not hurt.

Figures A.11 to A.14 show the out-of-sample yield RMSE for different masked blocks and values of θ over time, giving a more refined picture of how the choice of θ affects results in specific periods. Consider Germany during 2008–2009. For 0–5 year block masking, KR-T with $\theta = 10$ improves the yield RMSE from over 100 to around 25 basis points relative to KR. The larger value $\theta = 100$, however, degrades accuracy at longer maturities, whereas $\theta = 10$ closely matches KR there. This confirms that $\theta = 10$ strikes the right balance for transfer learning.

Lastly, we study the time-variation in the choice of an optimal θ. Cross-validation allows us to select an optimal θ on each date separately. Table A.3 reports the percentage of dates on which each value of θ achieves the lowest OOS error. For all markets except Japan, higher values ($\theta \geq 10$) are optimal on the majority of dates, indicating that strong cross-market information sharing is beneficial. The choice $\theta = 10$, which is optimal for the aggregate RMSE, is also among the most frequently optimal values for Germany, the UK, Canada, and Switzerland. Figure A.15 shows that the optimal θ does vary over time, suggesting that the informativeness of cross-market data can change with market conditions. Table A.4 shows that a time-varying θ (selecting the best θ on each date) can lead to modest improvements compared to a fixed θ.

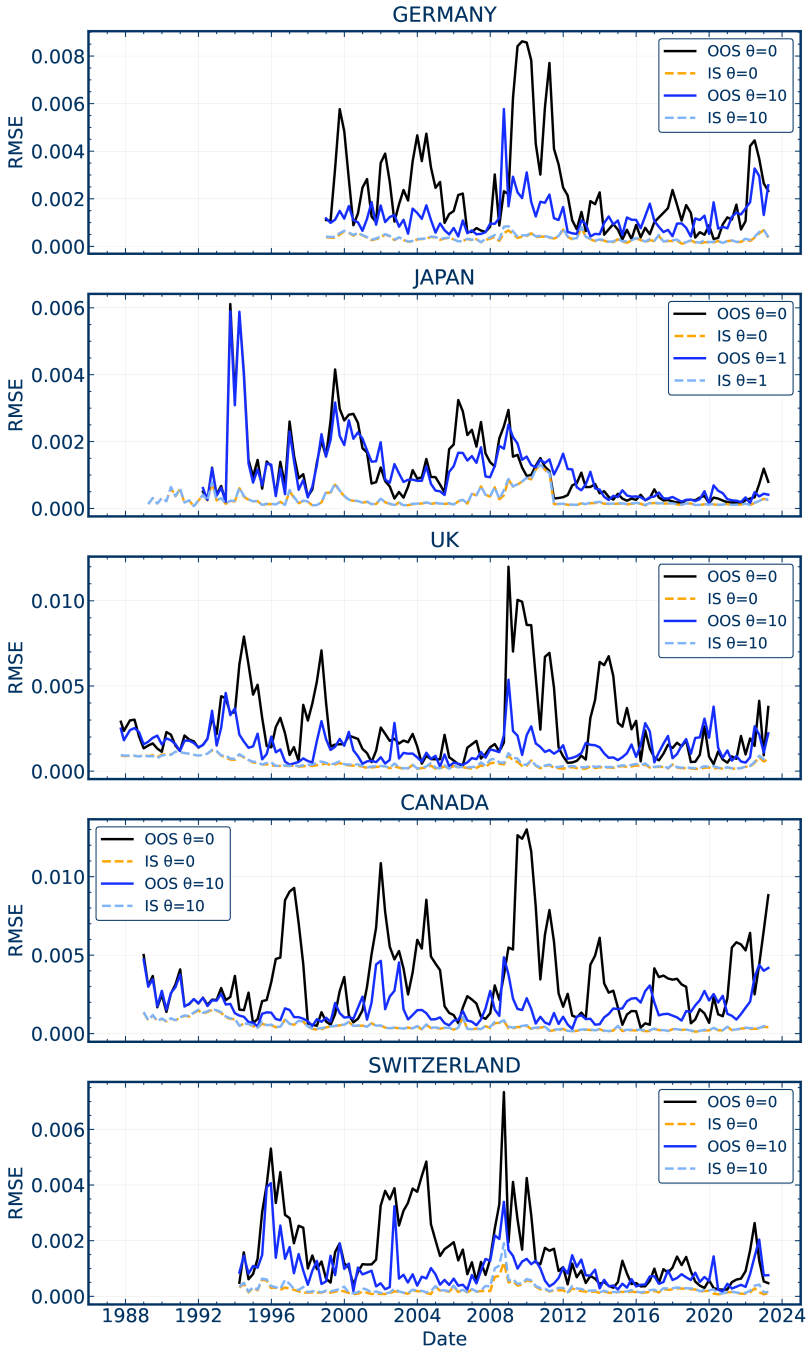
5.5 Rate Estimates over Time

We now show that KR-T provides the most reliable and stable interest-rate dynamics. Cross-sectional RMSE does not capture the stability of the time series, so we provide an additional perspective by inspecting the estimated rates over time.

Figure 16 illustrates the effect on the estimated yield curves themselves for Japan, showing the 30-day yield and 10-year forward rate estimates over time. The 30-day (one-month) yield is an important input for academic research and a crucial economic indicator. For long maturities, we use forward rates rather than yields, because long-maturity yields are averages of forward rates, so effects at the long end are partly averaged out in the yields.

Figure 16a shows that the KR-T estimates for the 30-day yield are smooth and stable throughout the sample, while both standalone KR and NSS exhibit substantial fluctuations in the pre-2000 period when short-maturity bonds were scarce. The NSS estimates are particularly problematic, producing economically implausible negative one-month yields around 1994–2000.

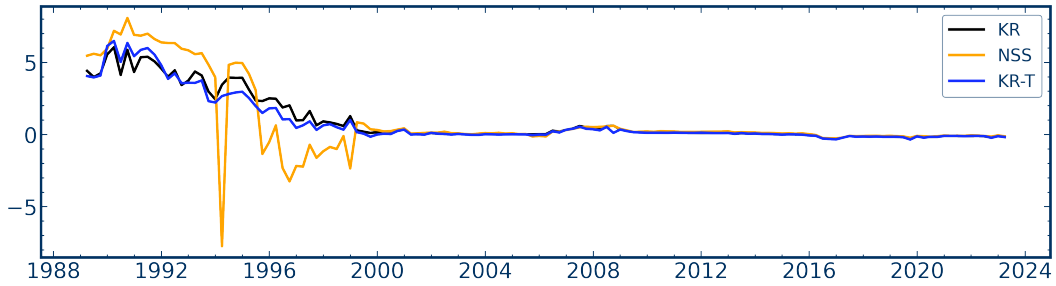
Figure 15: Block Cross-Validation: OOS YTM RMSE over time



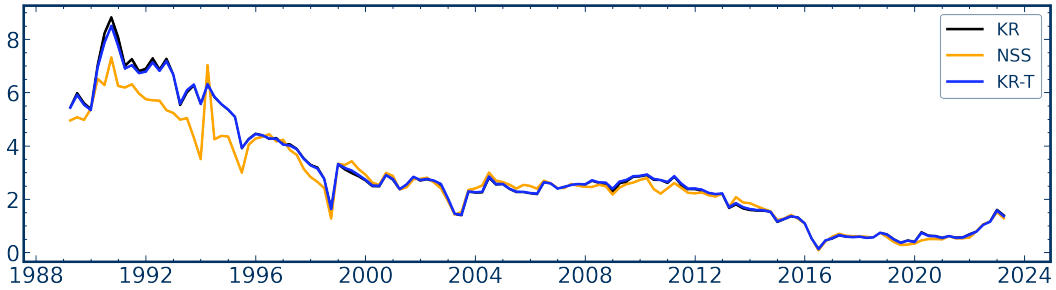
Time series of out-of-sample YTM RMSE under contiguous 5-year block cross-validation, for each country estimated jointly with the US. The solid lines show the out-of-sample RMSE for standalone KR ($\theta = 0$) and KR-T ($\theta = 10$; $\theta = 1$ for Japan); the dashed lines show the corresponding in-sample RMSE. KR-T consistently outperforms standalone KR across the full sample, with the largest improvements during periods of data sparsity.

Figure 16b shows the 10-year forward rate time-series. At the beginning of the sample before 1995, we see instabilities for KR and NSS. This is exactly the time period when no bonds with maturity less than 10 years are observed, and hence standalone estimators have to extrapolate.

Figure 16: Short and long maturity rate estimates over time



(a) 30-day yield estimates



(b) 10-year forward rate estimates

The figure shows the time series of short and long maturity rate estimates for different methods over time for Japan. Subplot (a) shows the annualized 30-day yield estimates; subplot (b) shows the 10-year forward rate for a one-year horizon, that is, the time-t forward rate for a one-year investment from $t+10$ to $t+11$. The transfer parameter for Japan is $\theta = 1$. Quarterly data are used.

KR-T benefits from the information in US Treasuries and FX markets. This confirms that transfer learning produces smoother, more stable rate estimates with the largest improvements during periods of data sparsity.

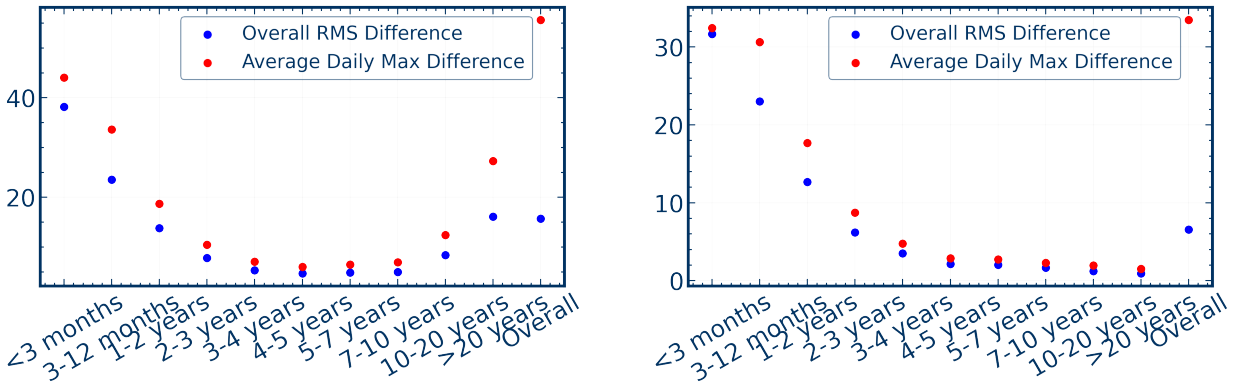
Figures A.21 and A.22 show the corresponding results for the other countries, which exhibit the same pattern. The NSS estimates for the 30-day yield are problematic for all countries, with implausible negative rate spikes for the UK, Germany, and Switzerland. The KR and KR-T time series are generally close, but the KR-T time series are more stable. The findings are similar for the 10-year forward rate time series.

5.6 Yield Curve Comparison

To characterize where transfer learning alters the yield curve, we compare the RMS and maximum absolute yield curve differences across maturity buckets.[20] Figure 17 shows the results for the UK. Differences between NSS and KR-T are large (RMS differences of up to about 40 basis points) and concentrated at the shortest maturities, reflecting the parametric limitations of NSS. Differences

[20]The RMS difference for yield curves for NSS is estimated as $\frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(Y_{i,t}^{\text{NSS}}-Y_{i,t}^{\text{KR-T}}\right)^2}$, and analogously for other estimators. The average maximum yield difference for NSS is calculated as $\frac{1}{T}\sum_{t=1}^{T}\max_{i\in\{1,\ldots,M\}}\left|Y_{i,t}^{\text{NSS}}-Y_{i,t}^{\text{KR-T}}\right|$, and analogously for other estimators. We calculate these metrics also for the subset of maturities within a maturity bucket.

Figure 17: UK: yield RMS vs max differences



(a) NSS vs KR-T

(b) KR vs KR-T

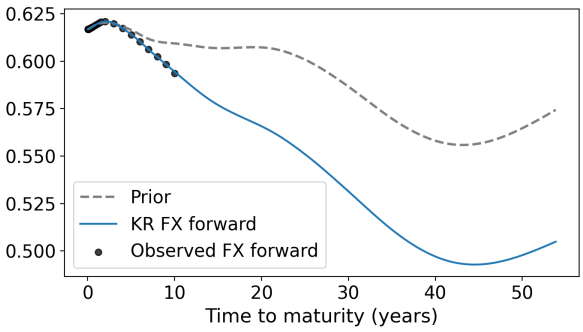
This figure compares yield curve differences in the UK between the transfer learning method KR-T and the standalone KR and NSS methods. Each panel shows the RMS difference (blue) and the average maximum absolute difference (red) across maturity buckets. The left panel compares NSS with KR-T; the right panel compares standalone KR with KR-T. Quarterly data are used. KR-T uses transfer parameter $\theta = 10$.

between standalone KR and KR-T are also concentrated at the short end: they reach about 30 basis points below one year, where local bonds are scarcest and FX-implied information is most valuable, and shrink to a few basis points beyond two years. This is exactly the footprint of the transfer channel: KR-T introduces targeted adjustments at the maturities where cross-market information is most informative while preserving local fitting accuracy elsewhere.

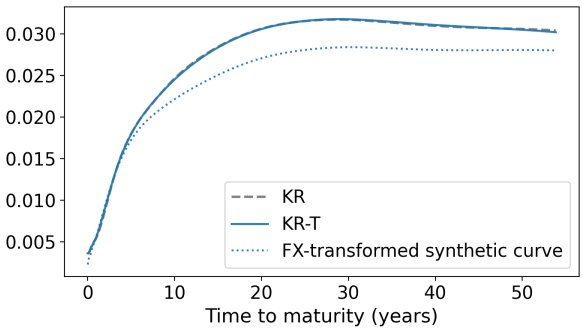
Figures A.23–A.25 show the yield curve differences for the other countries, with generally the same findings. The NSS curves differ substantially from the KR-T estimates, by up to about 80 basis points for Japan, which is economically large. KR and KR-T remain close, with RMS differences below 5 basis points for Germany.

To illustrate how the fitted curves differ across methods on a typical date, Figure A.26 presents the yield curves and fitted YTM for all five target markets on 2013-12-30, a common date where all markets have ample bond coverage. On this representative date, the KR and KR-T yield curves are nearly indistinguishable across the maturity spectrum in most markets, consistent with the design of our framework: when local data are sufficient, KR-T preserves the standalone KR fit. The NSS curves, by contrast, exhibit visible deviations, particularly at short maturities where the parametric form struggles to capture the curvature of the term structure, and at long maturities where the exponential basis functions impose a rigid asymptotic shape. In Japan and Switzerland, where the term structure is relatively flat, all three methods produce similar curves, but the NSS fit to individual bond YTMs shows larger residuals. In the UK and Canada, the differences between NSS and the kernel-based methods are more pronounced at the short end, reflecting the greater flexibility of the RKHS approach. These examples confirm that the aggregate statistical comparisons reported above are not driven by a few outlier dates but reflect systematic differences in curve-fitting methodology.

Figure 18: FX forward curve estimation and implied yield comparison



(a) FX forward curve: UK, 2014-09-30



(b) KR, KR-T, and synthetic yield: UK, 2014-09-30

Panel (a) shows the estimated forward exchange rate curve for GBP/USD on 2014-09-30, fitted using KR with a CIP-implied prior derived from independently estimated discount curves. The observed FX forward quotes (dots) anchor the estimate at short and intermediate maturities, while the KR-estimated curve (solid line) provides smooth extrapolation beyond the observed range. Panel (b) shows how this FX information enters the yield curve estimation: the synthetic FX-implied curve (dotted line) converts the US yield curve into GBP using the estimated FX forwards. At maturities where FX forwards are observed, the synthetic curve carries genuinely new cross-market information; beyond the FX forward horizon, it converges to the CIP-implied prior. The KR-T estimate (solid line) blends this synthetic information with local bond data.

5.7 FX Curve Estimation

The FX forward curve estimation is a key intermediate step in the KR-T framework. Figure 18a illustrates the estimated FX forward curve for the UK on a representative example day. It shows the observed FX forward quotes, the CIP-implied prior derived from the independently estimated discount curves, and the KR-estimated FX forward curve. The KR-based approach produces smooth forward FX curves that incorporate both the CIP-implied prior and observed forward rates, yielding economically plausible extrapolation beyond the observed range.

Figure 18b compares the KR, KR-T, and CIP-implied synthetic yield curves for the UK on the same example day. The synthetic FX-transformed curve, obtained by converting the US yield curve into the target currency with the KR-estimated FX forwards, shows how FX market information enters the estimation. At maturities where FX forwards are observed, the synthetic curve carries genuinely new information; beyond the FX forward horizon, the synthetic curve converges to the CIP-implied prior.

Figures A.6 and A.7 show that this finding generalizes to other dates and countries. Figure A.6 shows how the FX forward curve is estimated from observed quotes using the CIP-implied prior, and Figure A.7 shows how the resulting synthetic yield curve compares to the standalone KR and KR-T estimates.

The economic foundation of our transfer learning framework rests on a powerful observation: FX forward rates are the only source of genuinely new information that links yield curves across currencies. Under CIP, the forward exchange rate is determined by the ratio of discount curves across two currencies, meaning that observing an FX forward rate at maturity x is economically

equivalent to observing a constraint on the ratio $g_{\text{US}}(x)/g_a(x)$. Our framework accommodates CIP deviations through the FX basis spread estimation, and the weak CIP condition ensures that only curvature dynamics must be shared, while level and slope differences remain unconstrained. This mechanism produces a sharp empirical prediction, confirmed in Figure 14, namely that gains concentrate at maturities where FX forwards are observed.

5.8 Yield Curve Factor Structure

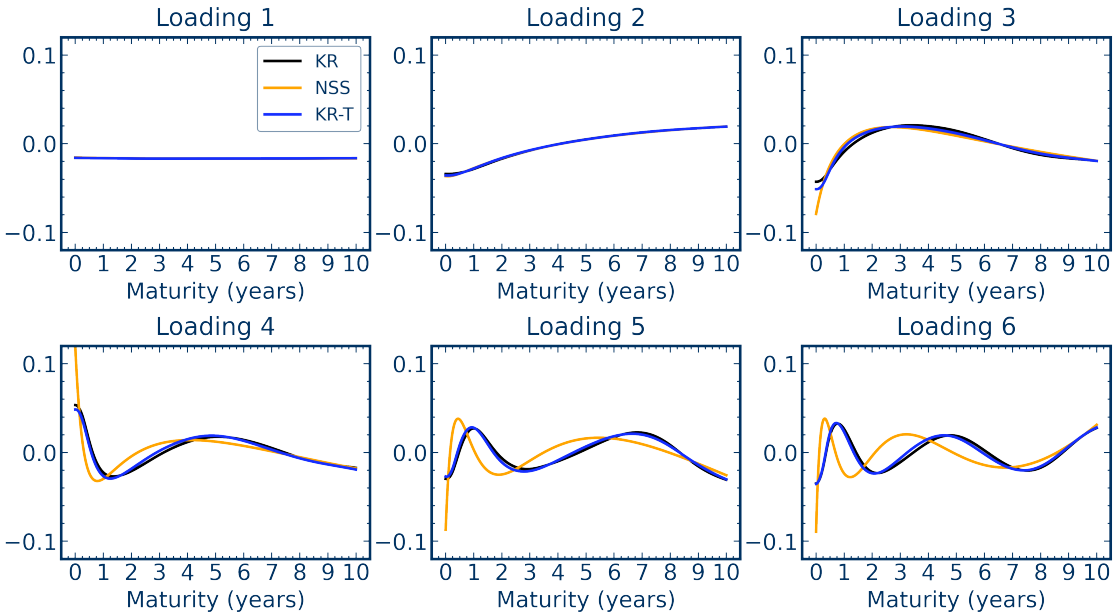
What is the underlying factor structure in international bond markets? To examine this question, we apply PCA to the panel of yield curves estimated from each method.[21] Figure 19 shows the first six eigenvectors for Canada, and Figures A.29–A.32 show the corresponding loadings for the other countries. The first two PCs (level and slope) are nearly identical across all methods, confirming that these dominant factors are robustly identified regardless of the estimation approach. KR and KR-T remain closely aligned through all six PCs, confirming that transfer learning preserves the data-driven factor structure. NSS diverges from the KR methods starting at the third PC, exhibiting stronger short-end sensitivity and more abrupt oscillations. This reflects the parametric constraints of the Nelson–Siegel functional form, which cannot flexibly capture higher-order curvature patterns. For Germany and Japan, subtle differences between standalone KR and KR-T emerge in the fourth to sixth PCs: KR-T produces smoother oscillations in Germany and oscillatory patterns in Japan that align more closely with those observed in other developed markets, reflecting the stabilizing influence of cross-market information on higher-order yield curve dynamics.

These findings connect directly to the results of Filipović, Pelger, and Ye (2025), who show that smoothness is a fundamental principle of the term structure of returns for US Treasuries. The optimal basis functions that span the discount bond return curve are the smoothest basis functions in decreasing order of smoothness, which explains why PCA applied to yield curves recovers the familiar level, slope, and curvature shapes. Our international results confirm that this smoothness principle extends across all six major bond markets: the KR and KR-T loadings exhibit the same ordered smooth structure documented by Filipović, Pelger, and Ye (2025) for the US. Importantly, Filipović, Pelger, and Ye (2025) identify a fourth "complexity" factor that captures complex shapes of the term structure premium and carries a large risk premium. This factor requires precise yield curve estimation to be detected and is missed by overly restrictive parametric methods like NSS. Our finding that NSS diverges from the kernel-based methods starting at the third PC is consistent with this: the Nelson–Siegel functional form cannot flexibly capture the higher-order curvature patterns that correspond to economically important risk factors. Figure 21 shows that NSS exhibits the same distortions systematically across countries. By providing more precise yield curve estimates across international markets, our KR and KR-T framework enables the discovery and study of these higher-order factors globally.

Figure 20 shows the KR-T loadings for all five target markets, confirming the universality of this

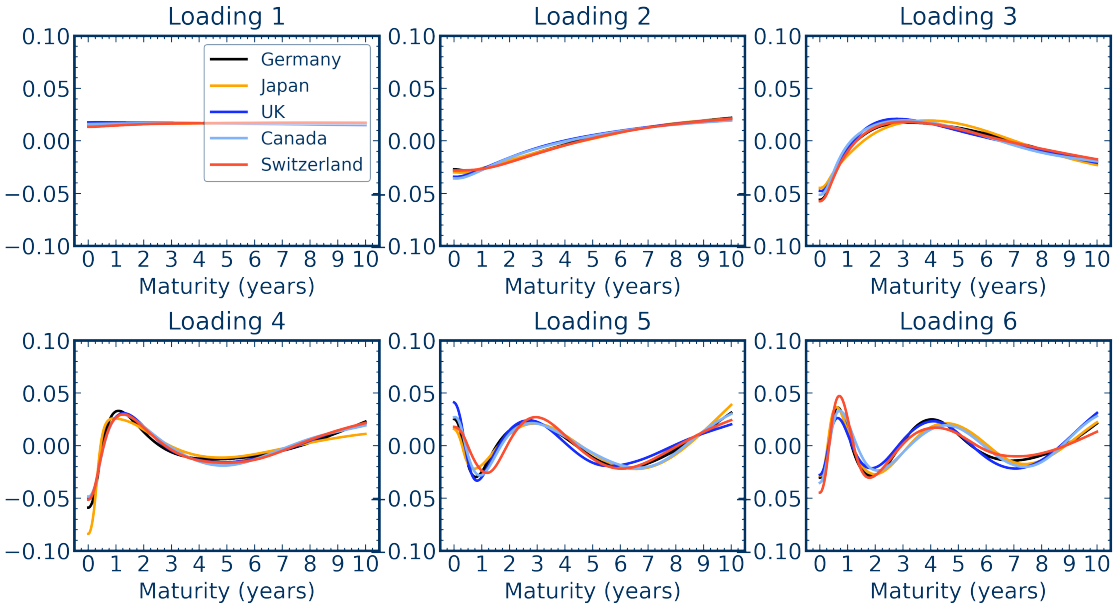
[21] In the following, we restrict the analysis to maturities of up to 10 years, for which the panel of estimated yield curves is available in all markets without excessive extrapolation.

Figure 19: Yield curve loadings for Canada



The figure shows the first six loadings estimated from the panel of Canadian yield curves for the KR, KR-T, and NSS methods. The panel consists of the estimated yields with maturities up to 10 years for the sample from 1990-01-02 to 2023-03-31. Quarterly data are used.

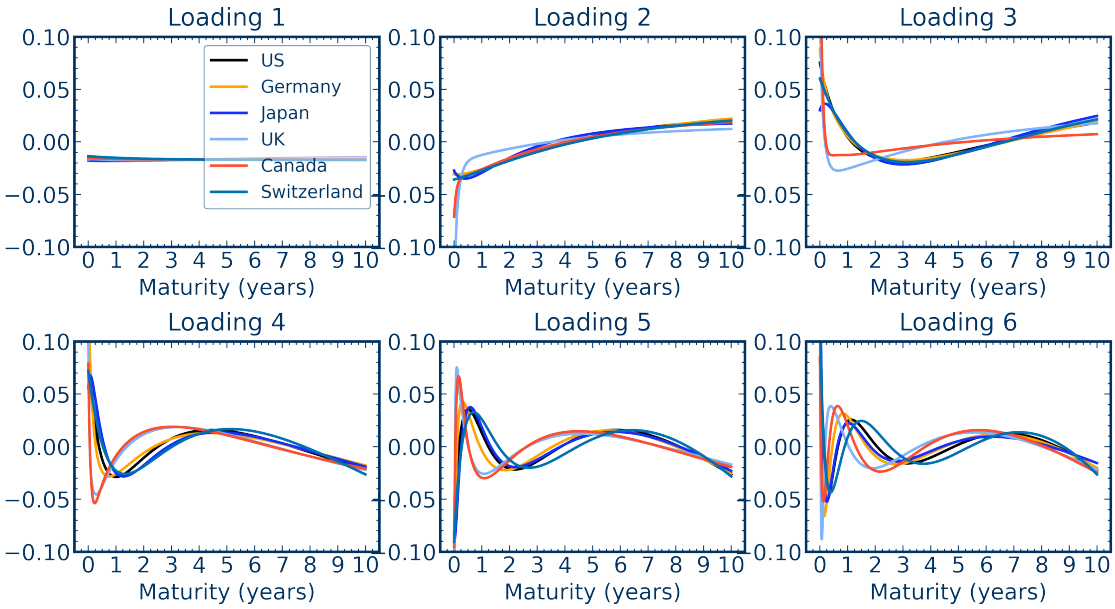
Figure 20: Yield curve loadings for all markets (KR-T)



The figure shows the first six loadings estimated from the panel of yield curves for the KR-T method across Germany, Japan, the UK, Canada, and Switzerland. The panel consists of the estimated yields with maturities up to 10 years for the sample from 1990-01-02 to 2023-03-31. Quarterly data are used.

factor structure across markets. Importantly, these PCA loadings are identical to the leading basis functions of the KR method, as explained in Filipović, Pelger, and Ye (2025): the eigenvectors

Figure 21: Yield curve loadings for all markets (NSS)



This figure shows the first six loadings estimated from the panel of yield curves for the NSS method across all markets. The panel consists of the estimated yields with maturities up to 10 years for the sample from 1990-01-02 to 2023-03-31. Daily data are used.

associated with the dominant eigenvalues of the kernel are the most important basis functions for KR and have the same shape as the PCs extracted from a panel. Hence, beyond the factor perspective, these results help explain which basis functions drive the superior fit of KR. While the same loadings, or basis functions, span the different bond markets, the coefficients on them generally differ: the factors representing different markets have different time-series dynamics but capture the same cross-sectional relationships between bonds. The results confirm that KR and KR-T estimate the same fundamental relationships, but KR-T provides a smoother estimate in data-poor environments.

6 Economic Implications

Yield curves are not merely statistical objects; they are fundamental inputs for economic analysis and financial decision-making. Forward rates are derivatives of the yield curve and are therefore inherently more sensitive to estimation noise than yields themselves. Small errors in the discount curve amplify into large forward rate distortions, particularly at maturities where bond data are sparse. This matters because forward rates are central inputs for monetary policy analysis, pension liability valuation, and fixed-income portfolio management.

All the policy implications that Filipovic, Pelger, and Ye (2024) and Filipović, Pelger, and Ye (2025) identify for the US market carry over to, and are amplified in, international markets. The error reductions of KR and KR-T relative to the incumbent NSS benchmark are several times larger abroad than in the US bond market, so adopting KR or KR-T for Germany, the UK, Japan,

Canada, and Switzerland should bring large benefits.

The policy consequences of KR-T relative to KR are more subtle. KR-T substantially improves the extrapolation at short maturities. This matters because such bonds are sparser in international markets than in the US. In addition, the KR-T framework produces more stable forward rate estimates and more economically plausible yield curve dynamics across all markets. The time series of 30-day yields and 10-year forward rates consistently show that transfer learning produces smoother, more plausible rate estimates, with the largest improvements during periods of data sparsity. The factor structure analysis confirms that KR-T preserves the dominant level and slope factors while producing more stable higher-order loadings, which are crucial for asset pricing and investment management.

The improvements documented in this paper have concrete implications for the practitioners and policymakers who rely on yield curves daily. For *central banks*, our framework offers a path to more reliable yield curve estimates for monetary policy analysis, particularly for economies with thin government bond markets. The conventional approach at many central banks, fitting a Nelson–Siegel or Svensson model to local bond data, produces large extrapolation errors at maturities where bonds are scarce. Adopting our transfer learning approach would improve the accuracy of term premium estimates, forward rate signals used for inflation expectations, and the short-rate estimates used in monetary policy rules. For *government debt management offices*, accurate yield curves at untraded maturities are essential for pricing new bond issues. Our method provides more reliable reference rates at maturities where no local bond currently exists, potentially reducing mispricing and improving the efficiency of government debt auctions. For *pension funds and insurance companies*, long-horizon discount rates determine the present value of liabilities. The stakes are easily quantified: the value of a liability moves approximately with its duration times the discount rate error, so a 10 basis point yield error misvalues a liability with a 20-year duration by roughly 2%. The yield error reductions we document are therefore economically large for long-horizon investors. In addition, the more stable and economically plausible yield curve estimates produced by our framework, particularly the smoother forward rates shown in Figure 16, would improve liability valuation and reduce artificial volatility in funded status ratios. For *global asset managers*, the availability of consistent, high-quality yield curves across countries is essential for cross-border portfolio construction, relative value analysis, and risk management.

7 Conclusion

Government bond yield curves are essential inputs for monetary policy, debt management, and financial market functioning, yet most countries outside the US face a persistent data challenge: too few bonds, too many maturity gaps, and too much price noise. This paper provides the most accurate and robust international yield curves, validated through comprehensive out-of-sample comparisons against the parametric benchmarks used by central banks. Our estimation framework combines the kernel ridge regression (KR) method with a novel transfer learning extension (KR-T)

that exploits economic linkages between international bond markets.

Our results establish a clear hierarchy of methods. First, the KR method of Filipovic, Pelger, and Ye (2024) generalizes from the US to all major international government bond markets, uniformly dominating the Nelson–Siegel–Svensson parametric benchmark used by many central banks. This makes KR the best available single-market method for yield curve estimation worldwide. Second, when additional information is available through foreign exchange markets, our KR-T framework further improves upon KR by constructing synthetic bonds via FX spot and forward rates and estimating discount curves jointly under a weak CIP condition. The KR-T estimation is formulated as vector-valued kernel ridge regression with a curvature-based spread penalty: rather than requiring yield equalization across currencies, we penalize only curvature differences in the cross-currency spread, leaving level and slope unconstrained. A single parameter θ controls the degree of information transfer, and the resulting optimization admits a closed-form solution.

Our empirical analysis covers six major economies with data spanning several decades. The results are clear and consistent. Under contiguous block cross-validation that mimics real-world maturity gaps, KR-T reduces yield RMSE by up to 51% relative to standalone KR and by 71–86% relative to NSS. The improvements are most pronounced at short and intermediate maturities where FX forward rates are directly observed, exactly where the exchange rate link carries genuine cross-market information. We identify FX forward rate availability as the sole economic channel driving transfer learning gains. At longer maturities beyond the FX forward horizon, synthetic bonds rely on extrapolated exchange rates rather than observed prices, and accordingly the marginal contribution of transfer learning diminishes. For maturities with sufficient local data coverage, KR-T is essentially identical to KR, confirming that the framework preserves local accuracy. Beyond point estimates, the framework produces more stable forward rate estimates, more economically plausible short-rate dynamics, and yield curve loadings that more accurately capture the underlying factor structure.

More broadly, our approach demonstrates a general principle for combining economic structure with machine learning: when related entities share underlying economic linkages but differ in data availability, domain-specific constraints, such as our weak CIP condition, can guide the transfer of information from data-rich to data-poor environments. This principle extends beyond yield curves to any setting where economic theory provides structure that can discipline statistical learning across related but heterogeneous data sources.

References

Adrian, T., R. K. Crump, and E. Moench (2013): "Pricing the Term Structure with Linear Regressions," *Journal of Financial Economics*, 110(1), 110–138.

Anderson, N., and J. Sleath (2001): "New estimates of the UK real and nominal yield curves," *Bank of England Working Paper*, 126.

Bianchi, D., M. Büchner, and A. Tamoni (2021): "Bond Risk Premiums with Machine Learning," *The Review of Financial Studies*, 34(2), 1046–1089.

Bolder, D. J., and S. Gusba (2002): "Exponentials, polynomials, and Fourier series: More yield curve modelling at the Bank of Canada," Working Paper 2002-29, Bank of Canada.

Bolder, D. J., G. Johnson, and A. Metzler (2004): "An empirical analysis of the Canadian term structure of zero-coupon interest rates," Working Paper 2004-48, Bank of Canada.

Brezis, H. (2011): *Functional analysis, Sobolev spaces and partial differential equations*, Universitext. Springer, New York.

Camenzind, N., and D. Filipovic (2023): "Stripping the Swiss Discount Curve using Kernel Ridge Regression," Swiss finance institute research paper no. 23-97, Swiss Finance Institute, European Actuarial Journal, forthcoming.

——— (2025): "Transfer Learning Across Fixed-Income Product Classes," Swiss finance institute research paper no. 25-50, Swiss Finance Institute.

Caruana, R. (1997): "Multitask Learning," *Machine Learning*, 28, 41–75.

Chen, L., M. Pelger, and J. Zhu (2024): "Deep Learning in Asset Pricing," *Management Science*, 70(2), 714–750.

Christensen, J. H. E., F. X. Diebold, and G. D. Rudebusch (2011): "The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models," *Journal of Econometrics*, 164(1), 4–20.

Diebold, F. X., C. Li, and V. Z. Yue (2008): "Global Yield Curve Dynamics and Interactions: A Dynamic Nelson-Siegel Approach," *Journal of Econometrics*, 146(2), 351–363.

Du, W., A. Tepper, and A. Verdelhan (2018): "Deviations from Covered Interest Rate Parity," *The Journal of Finance*, 73(3), 915–957.

Fama, E. F., and R. R. Bliss (1987): "The Information in Long-Maturity Forward Rates," *The American Economic Review*, 77(4), 680–692.

Filipovic, D., M. Pelger, and Y. Ye (2024): "Stripping the Discount Curve - a Robust Machine Learning Approach," *Management Science*, Accepted for publication.

Filipović, D., M. Pelger, and Y. Ye (2025): "Shrinking the Term Structure," *Working Paper*.

Gu, S., B. T. Kelly, and D. Xiu (2020): "Empirical Asset Pricing via Machine Learning," *The Review of Financial Studies*, 33(5), 2223–2273.

Gürkaynak, R. S., B. Sack, and J. H. Wright (2007): "The U.S. Treasury yield curve: 1961 to the present," *Journal of Monetary Economics*, 54(8), 2291–2304.

Hellum, O., L. H. Pedersen, and A. Rønn-Nielsen (2024): "How Global is Predictability? The Power of Financial Transfer Learning," *Working Paper*.

Ioannides, M. (2003): "A comparison of yield curve estimation techniques using UK data," *Journal of Banking & Finance*, 27, 1–26.

Joslin, S., K. J. Singleton, and H. Zhu (2011): "A New Perspective on Gaussian Dynamic Term Structure Models," *The Review of Financial Studies*, 24(3), 926–970.

Jotikasthira, C., A. Le, and C. T. Lundblad (2015): "Why Do Term Structures in Different Currencies Comove?," *Journal of Financial Economics*, 115(1), 58–83.

Liu, Y., and J. C. Wu (2021): "Reconstructing the yield curve," *Journal of Financial Economics*, 142(3), 1395–1425.

Monetary, and E. Department (2005): "Zero-coupon yield curves: technical documentation," *BIS Papers*, 25.

Nelson, C. R., and A. F. Siegel (1987): "Parsimonious Modeling of Yield Curves," *Journal of Business*, 60(4), 473–489.

Pan, S. J., and Q. Yang (2010): "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359.

Paulsen, V. I., and M. Raghupathi (2016): *An introduction to the theory of reproducing kernel Hilbert spaces*, vol. 152 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

Rasmussen, C. E., and C. K. I. Williams (2006): *Gaussian Processes for Machine Learning*. MIT Press.

Schich, S. T. (1997): "Estimating the German term structure," *Deutsche Bundesbank, Economic Research Group, Discussion paper*, no 4.

Svensson, L. E. (1994): "Estimating and Interpreting Forward Interest Rates: Sweden 1992 - 1994," Discussion paper, National Bureau of Economic Research, Cambridge, MA.

Weiss, K., T. M. Khoshgoftaar, and D. Wang (2016): "A Survey of Transfer Learning," *Journal of Big Data*, 3(1), 1–40.

Appendices

Appendix Contents

A Theory **45**

A.1 Reproducing Kernel Hilbert Spaces and KR Problem 45

A.2 KR-Transfer: Proof of Theorem 2 46

A.3 Notation 47

B Empirical Results **48**

B.1 Data 48

C Empirical Results for KR **52**

C.1 KR Parameter Selection 52

C.2 KR Estimation Comparison 53

D Empirical Results for KR-T **55**

D.1 FX Forward Curve Examples 55

D.2 Overall Comparison 57

D.3 Additional Cross-Validation Results 59

D.3.1 KR-T Parameter Selection on Finer Grid 59

D.3.2 Alternative Block Cross-Validation 59

D.3.3 Time-Varying Transfer Learning 64

D.3.4 Example dates for masking maturity blocks 66

D.4 Time Series of Yields and Forward Rates 72

D.5 Yield Curve Difference 74

D.6 Robustness to Short-Maturity Bonds 78

D.7 Basis Functions 83

E Official Government Yield Curve Sources **86**

A Theory

This appendix collects the theoretical background for the kernel ridge regression (KR) and transfer KR (KR-Transfer) problems stated in Section 2. The results for the scalar KR problem are developed in Filipovic, Pelger, and Ye (2024). We extend the framework to the vector-valued setting for transfer learning. We include self-contained statements in the notation of our paper for ease of reference.

A.1 Reproducing Kernel Hilbert Spaces and KR Problem

A fundamental notion in statistical machine learning is that of a reproducing kernel Hilbert space. Here we briefly recap its definition. For more background and applications we refer to Rasmussen and Williams (2006); Paulsen and Raghupathi (2016). Let E be a set; in our case, $E = [0, \infty)$. A reproducing kernel Hilbert space (RKHS) $\mathcal{H}$ is a Hilbert space of functions $h : E \to \mathbb{R}$ such that the point evaluations $h \mapsto h(x)$ are bounded linear functionals for all $x \in E$. By the Riesz representation theorem, for any $x \in E$, there exists a function $k_x \in \mathcal{H}$ such that the scalar product $\langle h, k_x \rangle_{\mathcal{H}} = h(x)$. The function $k : E \times E \to \mathbb{R}$ induced by $k(x, y) = \langle k_x, k_y \rangle_{\mathcal{H}} = k_y(x)$ is called the reproducing kernel of $\mathcal{H}$. It has the property that for any finite selection of points $x_1, \dots, x_n \in E$ the $n \times n$ matrix with elements $k(x_i, x_j)$ is symmetric and positive semi-definite. Thanks to the powerful property called the representer theorem, many kernel-based machine learning problems boil down to finite-dimensional standard convex optimization. Our results are a variant thereof.

Specifically, we fix a maturity weight $\alpha > 0$, and define the weighted Sobolev space using the curvature metric (3),

$$\mathcal{H} \coloneqq \Big\{ h : [0, \infty) \to \mathbb{R} \;\Big|\; h \text{ twice weakly differentiable},\ h(0) = 0,\ \lim_{x \to \infty} h'(x) = 0,\ \|h\|_\alpha < \infty \Big\},$$

equipped with the inner product

$$\langle h_1, h_2 \rangle_\alpha \coloneqq \int_0^\infty h_1''(x)\, h_2''(x)\, \mathrm{e}^{\alpha x}\, dx \qquad \text{and norm} \qquad \|h\|_\alpha^2 = \langle h, h \rangle_\alpha.$$

Here, a function h is twice weakly differentiable if it is continuously differentiable with absolutely continuous first derivative $h'(x) = h'(0) + \int_0^x h''(s)\, ds$ for some locally integrable h''. As shown in (Filipovic, Pelger, and Ye, 2024, Appendix A.3–A.4), $\mathcal{H}$ is a separable RKHS with reproducing kernel $k(x, y)$ given by (5).

The affine space $\mathcal{G} = 1 + \mathcal{H}$ inherits the metric: every curve $g \in \mathcal{G}$ is twice weakly differentiable and admits a unique representation $g = 1 + h$ with $h \in \mathcal{H}$, and $\|g\|_\alpha = \|h\|_\alpha < \infty$.

Theorem 1 is a restatement of (Filipovic, Pelger, and Ye, 2024, Theorem 2) stating that the KR problem (4) admits a unique solution of the form (6). The proof follows the standard representer theorem argument: the solution lies in the finite-dimensional subspace $1 + \operatorname{span}\{k(\cdot, x_1), \dots, k(\cdot, x_N)\}$, and the coefficients β are found in closed form by solving a linear system. We make these arguments more explicit in the proof of Theorem 2 below.

A.2 KR-Transfer: Proof of Theorem 2

We write $g_a = 1 + h_1$ and $\widetilde{g}_a = 1 + h_2$, for $h_1, h_2 \in \mathcal{H}$, and define the tuple $\boldsymbol{h} = [h_1, h_2]^\top$, which is an element in the product space $\mathcal{H} \times \mathcal{H}$. The sum of the regularization terms in (9) can then be written as

$$\mathcal{R}(\boldsymbol{h}) \coloneqq \langle \boldsymbol{h}^\top, Q\boldsymbol{h} \rangle_\alpha$$

for the inverse of the transfer matrix

$$Q \coloneqq B_a^{-1} = \begin{bmatrix} \lambda_a + \theta & -\theta \\ -\theta & \widetilde{\lambda}_a + \theta \end{bmatrix}.$$

As Q is positive definite, $\langle \boldsymbol{f}^\top, Q\boldsymbol{h} \rangle_\alpha$ defines a scalar product between tuples $\boldsymbol{f}, \boldsymbol{h}$ that renders the product $\mathcal{H} \times \mathcal{H}$ a separable Hilbert space, so that $\mathcal{R}(\boldsymbol{h})$ is the square of the norm of $\boldsymbol{h}$.

The sum of the error terms in (9) can be written as

$$\mathcal{E}(\boldsymbol{h}) \coloneqq (\boldsymbol{P} - \boldsymbol{C}\mathbf{1} - \boldsymbol{C}S\boldsymbol{h})^\top \boldsymbol{J}^{-1} (\boldsymbol{P} - \boldsymbol{C}\mathbf{1} - \boldsymbol{C}S\boldsymbol{h})$$

for the bounded linear sampling operator $S : \mathcal{H} \times \mathcal{H} \to \mathbb{R}^{2N}$ defined as $S\boldsymbol{h} \coloneqq \mathrm{vec}(\boldsymbol{h}^\top(\boldsymbol{x}))$.

Minimizing (9) over $g_a, \widetilde{g}_a \in \mathcal{G}$ is thus equivalent to minimizing $\mathcal{E}(\boldsymbol{h}) + \mathcal{R}(\boldsymbol{h})$ over $\boldsymbol{h} \in \mathcal{H} \times \mathcal{H}$. This objective function is continuous, strictly convex, and coercive in $\boldsymbol{h}$ in the Hilbert space $\mathcal{H} \times \mathcal{H}$, which asserts existence of a unique solution (Brezis, 2011, Corollary 3.23). Moreover, as $\boldsymbol{h}$ enters $\mathcal{E}(\boldsymbol{h})$ only in the form $\boldsymbol{C}S\boldsymbol{h}$, the minimizer must lie in the orthogonal complement of the null space of $\boldsymbol{C}S$. Indeed, adding to $\boldsymbol{h}$ any non-zero element $\boldsymbol{f}$ in the null space of $\boldsymbol{C}S$ does not affect the error, $\mathcal{E}(\boldsymbol{h} + \boldsymbol{f}) = \mathcal{E}(\boldsymbol{h})$, but strictly increases the regularization component, $\mathcal{R}(\boldsymbol{h} + \boldsymbol{f}) > \mathcal{R}(\boldsymbol{h})$. In fact, this is the representer theorem, which formally states that the minimizer is of the form

$$\boldsymbol{h} = S^* \boldsymbol{C}^\top q \quad \text{for some } q \in \mathbb{R}^{M_a + M_{\mathrm{US}}}, \tag{14}$$

for the adjoint $S^* : \mathbb{R}^{2N} \to \mathcal{H} \times \mathcal{H}$, which is given by

$$S^* v = \sum_{j=1}^{N} k(\cdot, x_j) B_a W_j \tag{15}$$

where W_j is the jth column of the matrix $W = [W_1, \ldots, W_N] \in \mathbb{R}^{2 \times N}$ with $\mathrm{vec}(W^\top) = v$. Indeed, elementary calculations show

$$\langle S\boldsymbol{h}, v \rangle_{\mathbb{R}^{2N}} = \sum_{j=1}^{N} \boldsymbol{h}^\top(x_j) W_j = \sum_{j=1}^{N} \langle \boldsymbol{h}^\top, k(\cdot, x_j) W_j \rangle_\alpha = \langle \boldsymbol{h}^\top, Q \sum_{j=1}^{N} k(\cdot, x_j) B_a W_j \rangle_\alpha,$$

which proves (15). Plugging (14) with (15) into the objective function $\mathcal{E}(\boldsymbol{h}) + \mathcal{R}(\boldsymbol{h})$ and solving the first-order conditions, as in the proof of the scalar case (Filipovic, Pelger, and Ye, 2024, Theorem A.1), then proves Theorem 2.

A.3 Notation

Table A.1 summarizes the main notation of the paper.

Table A.1: Notation Summary

Symbol	Description
Markets and data	
$\mathcal{A}$	set of markets: US (source), Germany, Japan, UK, Canada, and Switzerland
a	target market, $a \in \mathcal{A} \setminus \{\text{US}\}$
M_a	number of quoted bonds in market a on a given day
$\boldsymbol{x} = [x_1, \ldots, x_N]^\top$	common cash-flow dates
P_a, C_a	bond prices and cash-flow matrix in market a (local currency)
$\omega_{a,i}$	duration-based pricing error weight of bond i, equation (7)
S_a, $F_a(x)$	FX spot and time-x forward price of one US dollar in currency a
$\tau_{a,1} < \cdots < \tau_{a,m_a}$	observed FX forward maturities
$\widetilde{P}_a$, $\widetilde{C}_a$	prices and cash flows of FX-adjusted synthetic bonds, equation (8)
Curves and estimators	
$g_a(x)$	discount curve of market a
$\widetilde{g}_a(x)$	synthetic discount curve that prices the synthetic bonds
$\eta_a(x)$	log-FX spread curve capturing CIP deviations, $\eta_a = \log g_a - \log \widetilde{g}_a$
g_a^{KR}, $g_a^{\text{KR}-\text{T}}$	KR and KR-Transfer estimates of the discount curve
$F_a^{\text{KR}}(x)$	estimated FX forward curve, equation (12)
Function spaces and kernels	
$\|f\|_\alpha$	weighted curvature metric, equation (3)
$\mathcal{H}$	RKHS of twice weakly differentiable functions with $h(0) = 0$, Appendix A.1
$\mathcal{G} = 1 + \mathcal{H}$	affine space of candidate discount curves
$k(x, y)$, K	reproducing kernel, equation (5), and kernel matrix $K_{ij} = k(x_i, x_j)$
B_a	2×2 transfer matrix of Theorem 2
Hyperparameters (baseline values)	
α	maturity weight in the curvature metric ($\alpha = 0.05$)
λ_a, $\widetilde{\lambda}_a$	smoothness parameters of the target and synthetic curves ($\lambda = 1$)
θ	transfer parameter of the weak CIP spread penalty ($\theta = 10$; $\theta = 1$ for Japan)
λ_a^{FX}	smoothness parameter of the log-FX spread curve ($\lambda^{\text{FX}} = 1$)

The table summarizes the main notation of the paper. The baseline hyperparameter values are selected by cross-validation in Sections 4 and 5.

B Empirical Results

This appendix presents supplementary empirical findings that complement the main results.

B.1 Data

This subsection documents the data availability, filters, and FX forward maturity coverage. Figure A.1 shows FX forward maturity availability, which constrains the effective range of synthetic bond construction. Figure A.2 displays time-to-maturity heatmaps for all markets, highlighting persistent maturity gaps. Figure A.3 confirms that our outlier filter removes very few observations (at most 3 per date).

Figure A.2 shows the time-to-maturity distribution of outstanding bonds for different markets. The large white bands illustrate persistent maturity gaps that a model must interpolate or extrapolate across. This observation highlights the fact that the panel of available bonds in a given market is highly uneven across maturities.

Figure A.3 shows the counts of bonds flagged as yield to maturity (YTM) outliers according to the cross-sectional rule. An observation is marked as an outlier if its YTM deviates from the cross-sectional median by more than 0.1 on that date.

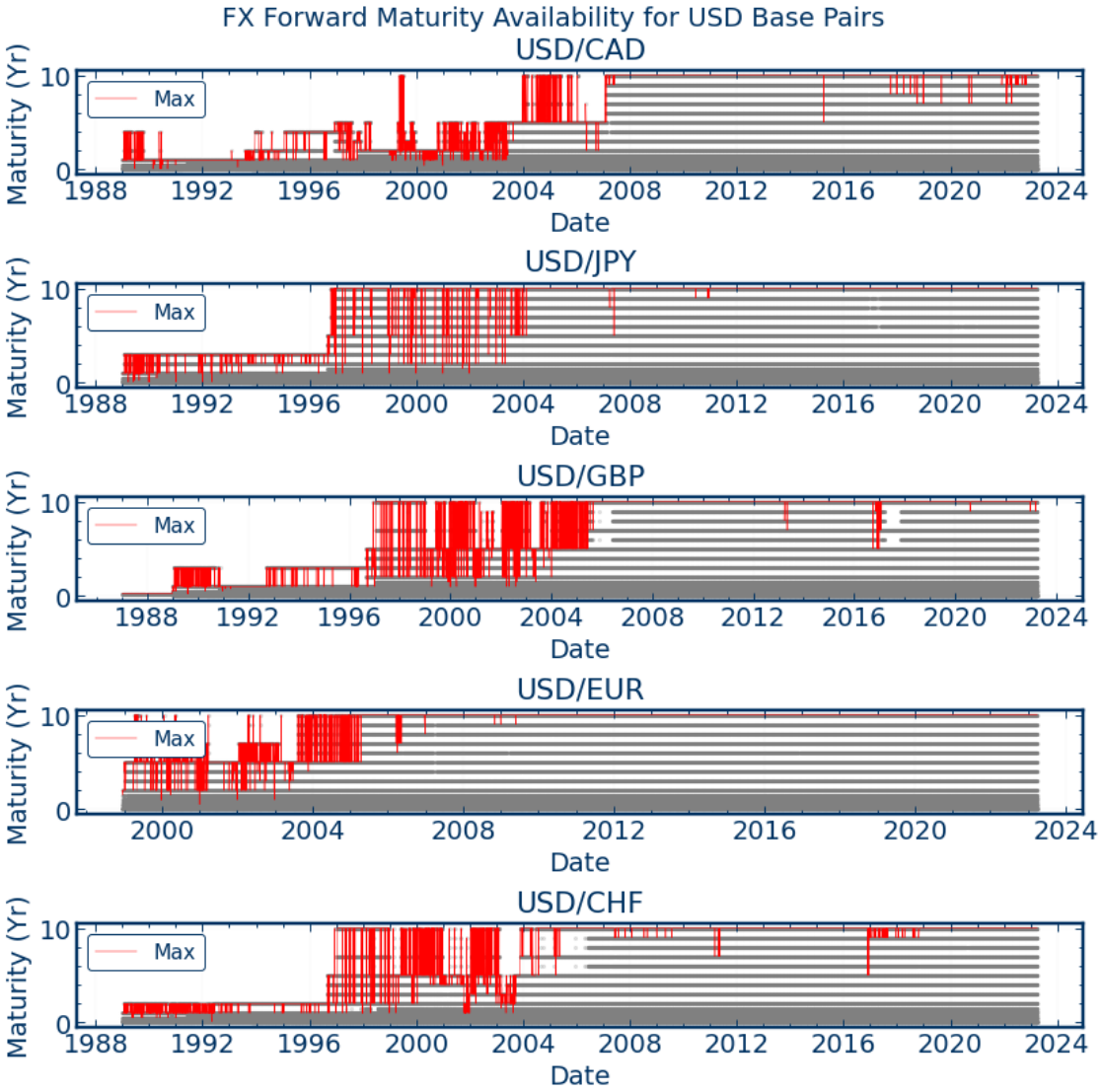
Table A.2 reports the sample periods used in the standalone and transfer-learning estimation.

Table A.2: Sample Periods for Estimation

Market	Standalone KR/NSS (monthly)	Transfer Learning (quarterly)
US	1961-06-30 to 2023-03-31	Reference market
Germany	1987-07-31 to 2023-03-31	1998-12-30 to 2023-03-31
Japan	1987-12-28 to 2023-03-31	1988-12-30 (stratified CV) / 1992-03-31 (block CV) to 2023-03-31
UK	1987-07-31 to 2023-03-31	1987-09-30 to 2023-03-31
Canada	1987-07-31 to 2023-03-31	1988-12-30 to 2023-03-31
Switzerland	1994-02-28 to 2023-03-31	1994-03-31 to 2023-03-31

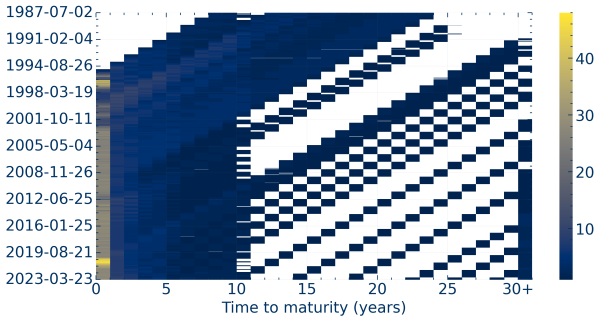
The table shows the standalone KR and NSS and transfer-learning evaluation sample periods. The standalone KR and NSS samples are determined separately for each local bond market and therefore require only that the target market have sufficient local bond observations. Transfer-learning samples require the intersection of the US reference market, the target local market, and the corresponding FX spot and forward data. This is the sample that we use for the comparison between KR-T, KR, and NSS. The US is used as the reference market and is therefore not reported for transfer learning results. For Japan, the stratified cross-validation sample begins earlier than the 5-year block cross-validation sample because the nature of the block CV design requires the target market bonds to span at least two non-empty 5-year maturity blocks on a given date (so that the within-block test set has a non-empty complementary training set). Germany's standalone sample starts in 1987 and includes the pre-euro Deutsche Mark-era Bund data, while the transfer learning sample starts at the end of 1998 when EUR/USD FX forwards first become available.

Figure A.1: FX Forward Maturity Availability (USD Base Pairs)

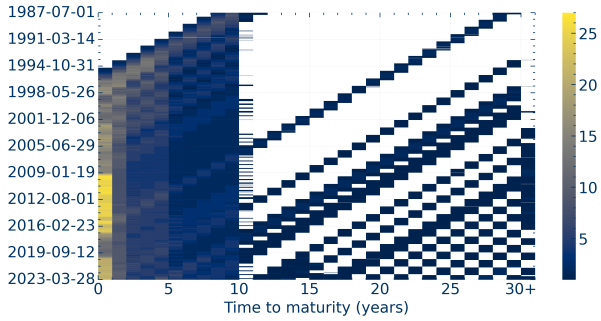


The figure shows availability of FX forward rates for USD base pairs across different currencies. Each dot represents the presence of an FX forward rate at a specific maturity on a given day. The red line traces the maximum available maturity over time.

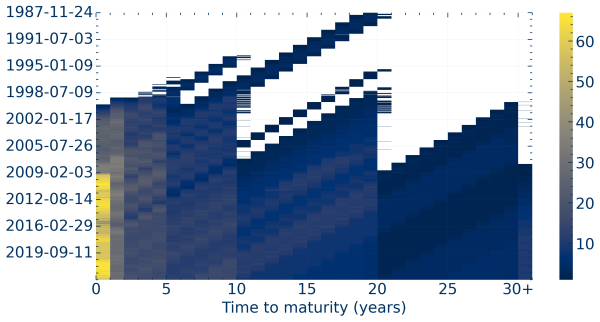
Figure A.2: Observed maturities (rounded in years)



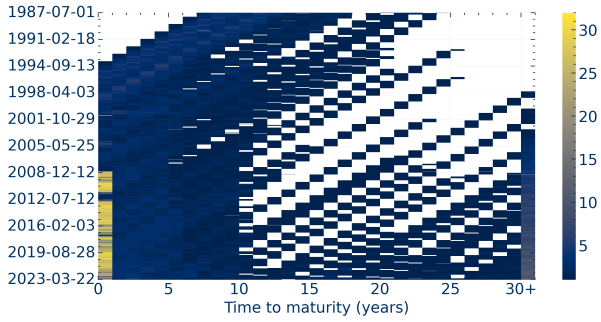
(a) Canada



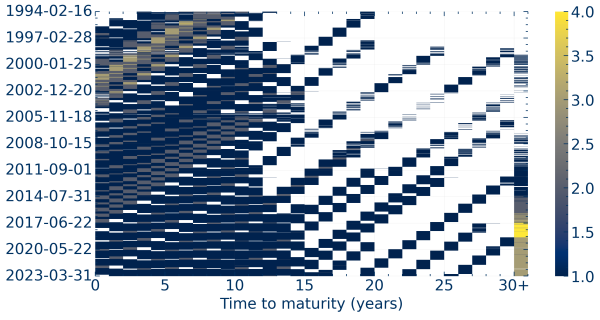
(b) Germany



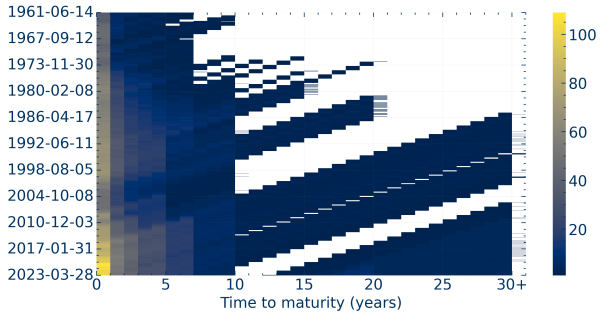
(c) Japan



(d) UK

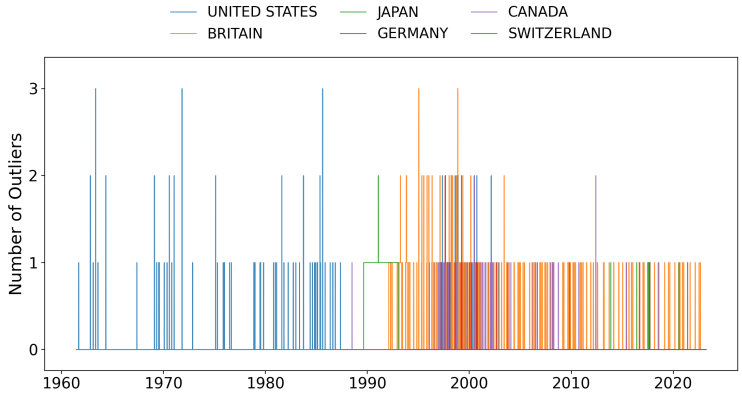


(e) Switzerland



(f) US

Figure A.3: YTM outlier counts over time by market



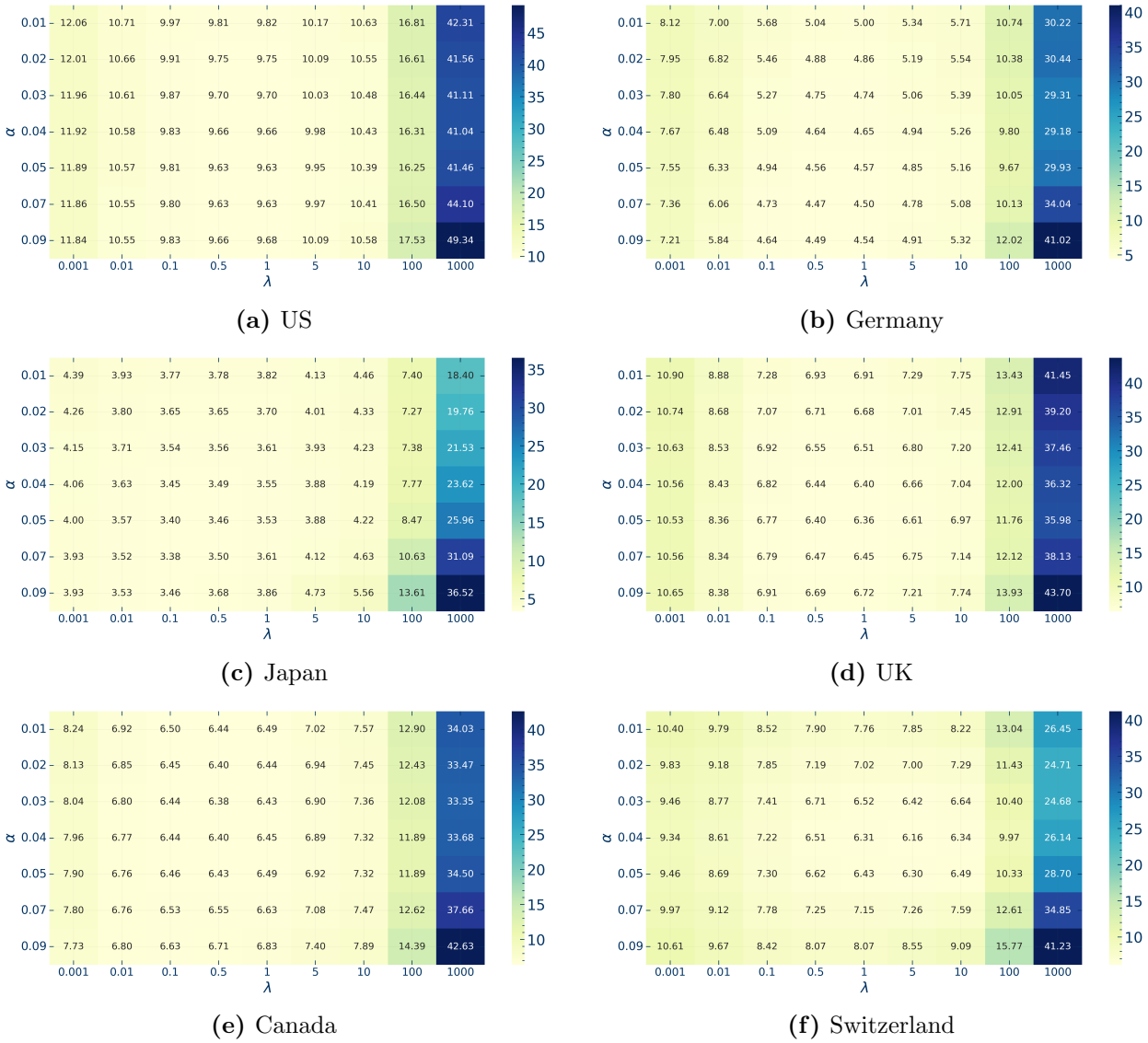
This figure shows the counts of bonds flagged as YTM outliers over time in each market. An observation is marked as an outlier if its YTM deviates from the cross-sectional median by more than 0.1 on that date.

C Empirical Results for KR

C.1 KR Parameter Selection

The key takeaway from the parameter selection analysis is that the KR method is remarkably robust to the choice of hyperparameters. Figure A.4 shows the out-of-sample yield RMSE from stratified cross-validation on monthly data for different values of α and λ for each market. The performance is robust for λ within the range $[0.5, 5]$ across all markets. This robustness is practically important: the method does not require careful per-market tuning.

Figure A.4: Cross-validation YTM RMSE for α and λ

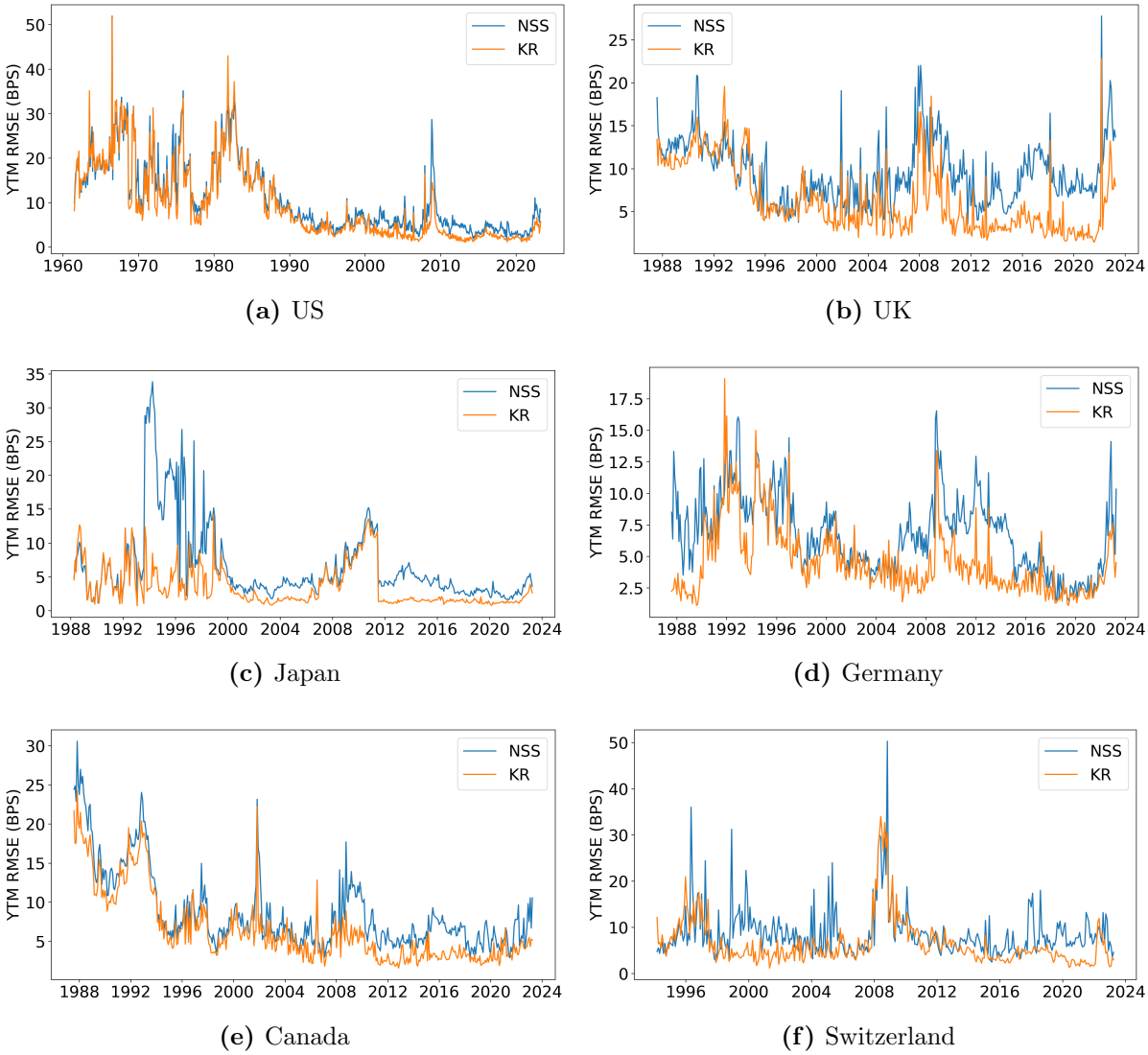


The figure shows the out-of-sample yield RMSE in basis points (BPS) from stratified cross-validation on monthly data for different values of maturity weight α and smoothness parameter λ for each market.

C.2 KR Estimation Comparison

This subsection confirms that, even without transfer learning, standalone KR substantially outperforms the widely used NSS benchmark across all markets, robustly over time and across maturity buckets. Figure 9 shows the out-of-sample YTM RMSE over time and across maturity buckets for standalone KR and NSS in the UK. Full results over time for each market are in Figure A.5, and full results by maturity bucket are in Figure 8. KR consistently outperforms NSS over time and across maturity buckets in all markets.

Figure A.5: YTM RMSE Over Time Across Markets



(a) US **(b)** UK

(c) Japan **(d)** Germany

(e) Canada **(f)** Switzerland

This figure shows the out-of-sample YTM RMSE for KR and NSS methods across the sample period for standalone market yield estimation. Monthly data are used for cross-validation. Errors are reported in basis points (BPS).

D Empirical Results for KR-T

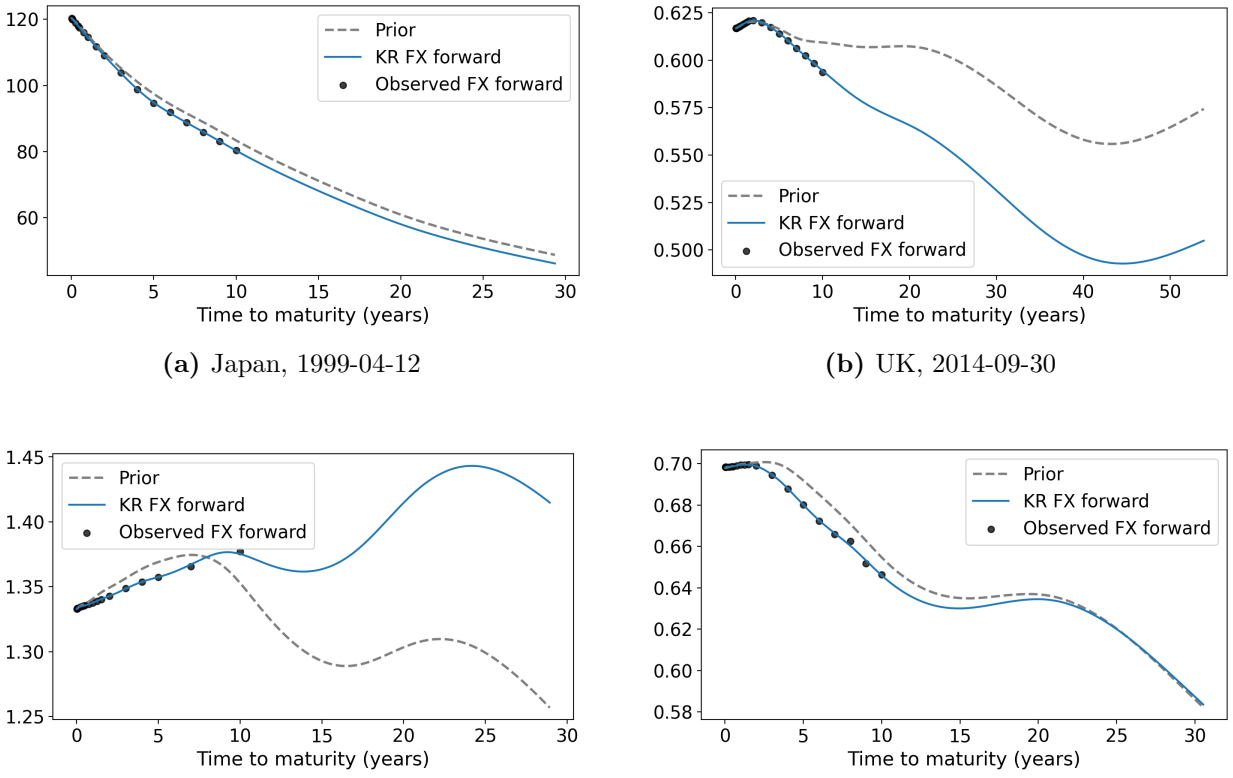
D.1 FX Forward Curve Examples

This subsection illustrates the FX forward curve estimation on representative dates. The key takeaway is that the KR-based approach produces a smooth forward FX curve that incorporates both the CIP-implied prior (from estimated discount curves) and observed forward rates.

Figure A.6 displays the estimated forward exchange rate curves for four markets on representative dates. Each panel shows the observed FX forward quotes, the CIP-implied prior derived from independently estimated discount curves, and the KR-estimated forward curve.

Figure A.7 compares the KR, KR-T, and CIP-implied synthetic yield curves on selected example dates across four markets. The synthetic FX-transformed curve is the yield curve implied by converting the US yield curve into the target currency using the KR-estimated FX forward curve.

Figure A.6: Forward exchange rate curve



(a) Japan, 1999-04-12

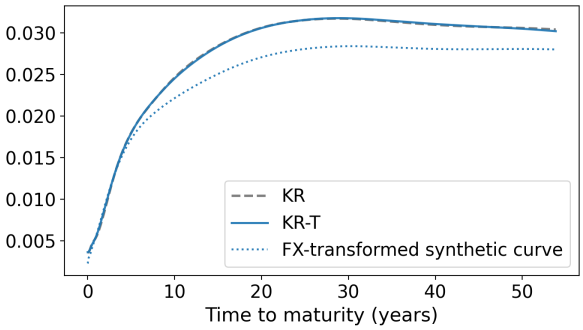
(b) UK, 2014-09-30

(c) Canada, 2004-06-30

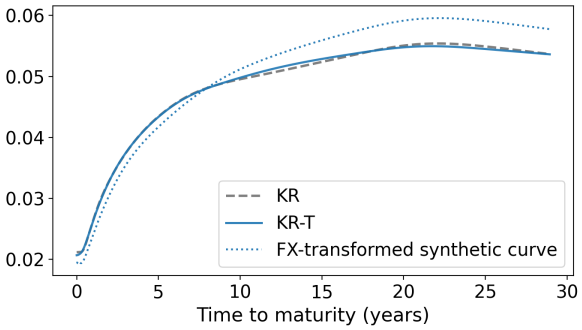
(d) Germany, 2009-12-31

This figure shows the estimated forward exchange rate curves on example dates, estimated using KR with a CIP-implied prior. Forward exchange rates are against USD. The prior is estimated from independently estimated KR discount curves for the US and each target market. The FX curve is estimated with parameters $\alpha = 0.05$ and $\lambda^{\text{FX}} = 1$, with the smoothness term scaled by 2,000.

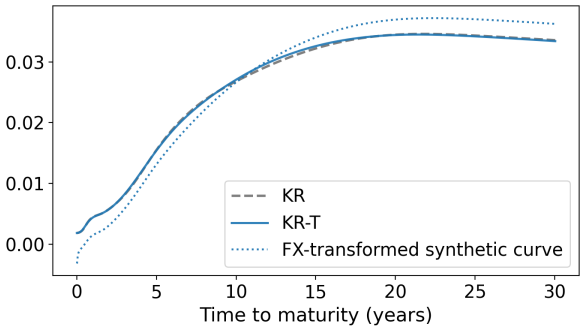
Figure A.7: KR, KR-T, and synthetic FX-implied yield curve comparison



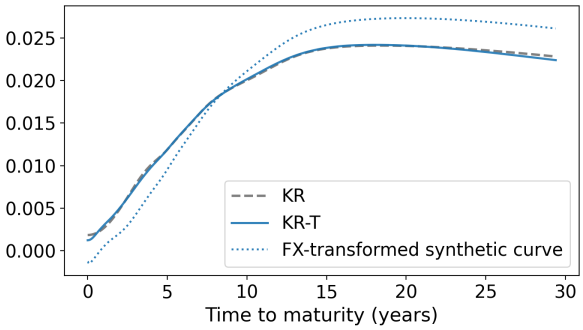
(a) UK, 2014-09-30



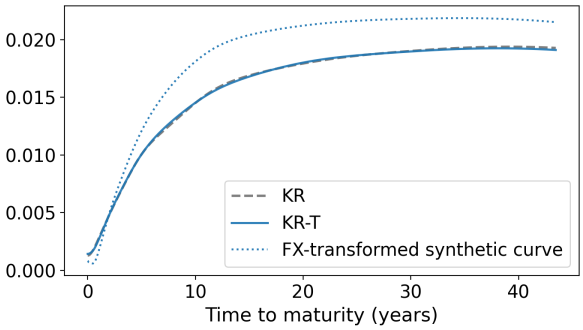
(b) Canada, 2004-06-30



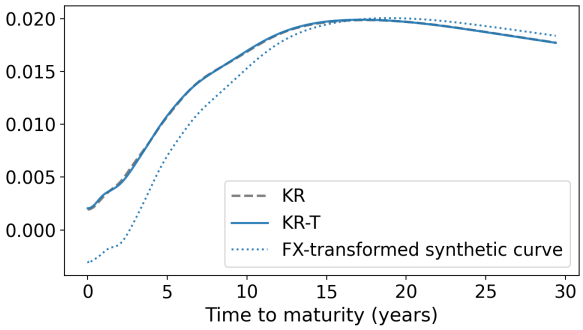
(c) Germany, 2010-06-30



(d) Switzerland, 2009-12-31



(e) Canada, 2021-06-30



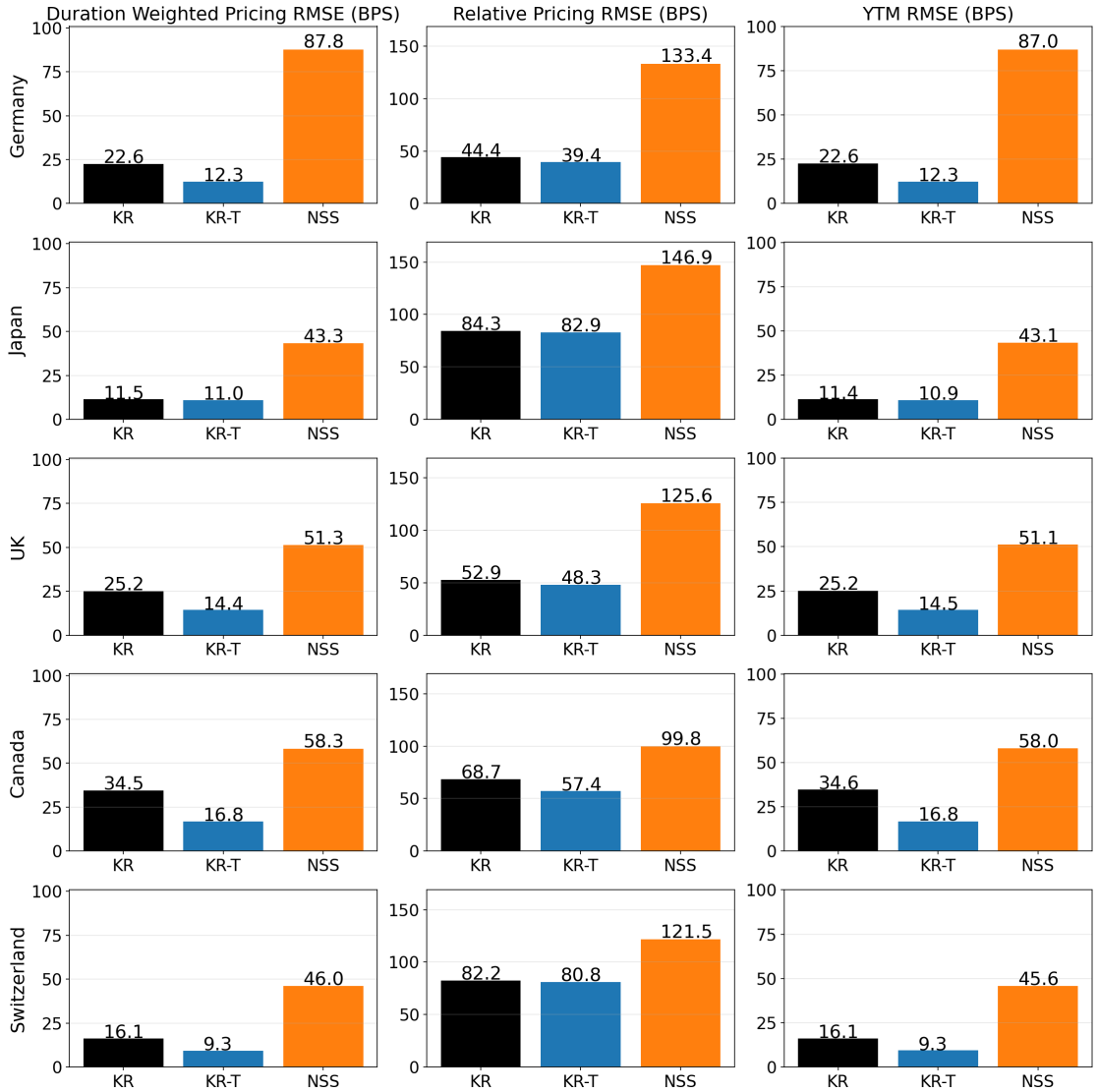
(f) Switzerland, 2010-12-31

This figure compares the standalone KR, transfer learning KR-T ($\theta = 10$), and synthetic FX-implied yield curves on selected example dates. The synthetic FX-implied curve is the target-market yield curve implied by converting the US KR discount curve into the target currency using the KR-estimated FX forward curve via CIP. The FX curve is estimated with parameters $\alpha = 0.05$ and $\lambda^{\text{FX}} = 1$, with the smoothness term scaled by 2,000. Forward exchange rates are against USD.

D.2 Overall Comparison

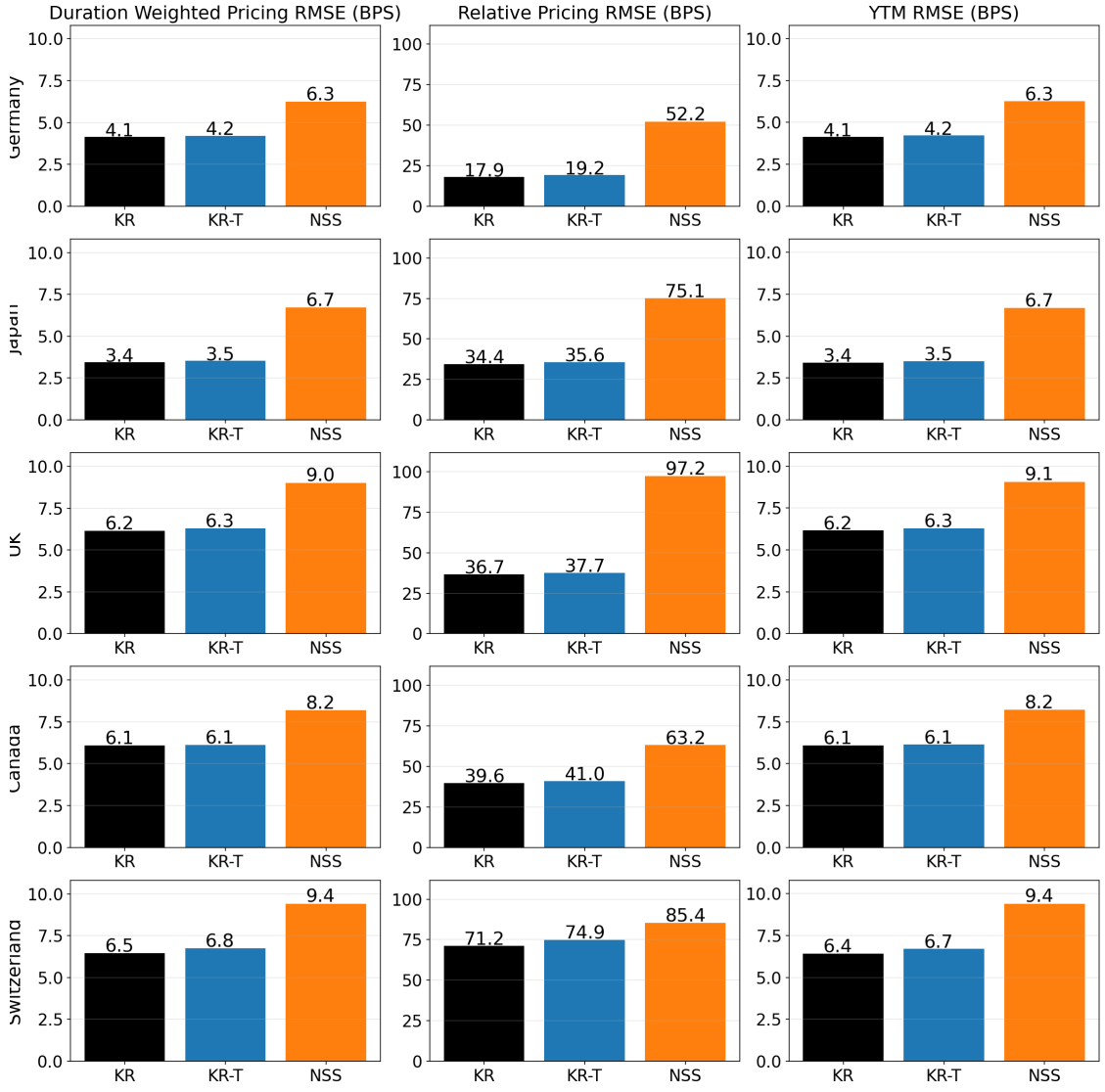
This subsection presents the overall comparison across all markets in a unified format. We evaluate each market with quarterly data up to Q1 2023, with start dates determined by data availability (Table A.2); dates with fewer than two bonds in either the target market or the US are skipped. The bar plots confirm that KR-T dominates under block cross-validation (Figure A.8) while remaining competitive under stratified cross-validation (Figure A.9): transfer learning helps precisely when it is needed and does no harm when it is not.

Figure A.8: 5-yr Block Cross-Validation Summary Bar Plot



The figure shows out-of-sample metrics under 5-year block cross-validation for the KR, KR-T, and NSS methods. KR-T uses transfer learning from the US market with transfer parameter $\theta = 10$ for all markets except Japan, where $\theta = 1$. We use quarterly data. The Germany sample is restricted to the period with available EUR/USD FX data. Columns correspond to duration-weighted pricing RMSE, relative pricing RMSE, and YTM RMSE. Numbers are in basis points (BPS).

Figure A.9: 3-Fold Stratified Cross-Validation Summary Bar Plot



The figure shows out-of-sample metrics under 3-fold stratified cross-validation for the KR, KR-T, and NSS methods. KR-T uses transfer learning from the US market with transfer parameter $\theta = 10$ for all markets except Japan, where $\theta = 1$. We use quarterly data. The Germany sample is restricted to the period with available EUR/USD FX data. Columns correspond to duration-weighted pricing RMSE, relative pricing RMSE, and YTM RMSE. Numbers are in basis points (BPS).

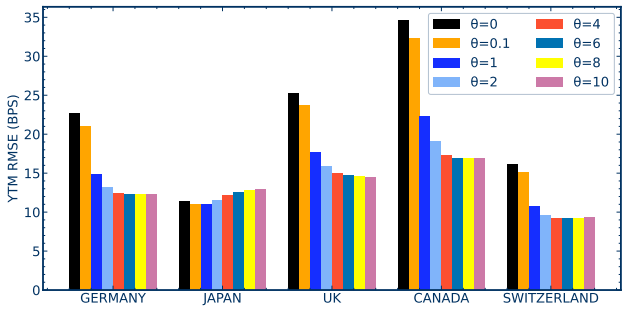
D.3 Additional Cross-Validation Results

This subsection provides detailed block cross-validation results, including time-varying performance and the stability of the optimal transfer parameter θ.

D.3.1 KR-T Parameter Selection on Finer Grid

Figure A.10 shows the block cross-validation results on a finer grid of the transfer parameter θ. The key finding is that the optimal θ varies across markets but the performance surface is relatively flat for $\theta \in [1, 100]$, confirming that the transfer learning gains are robust to the specific choice of this tuning parameter.

Figure A.10: Block Cross-Validation: OOS YTM RMSE (finer grid)



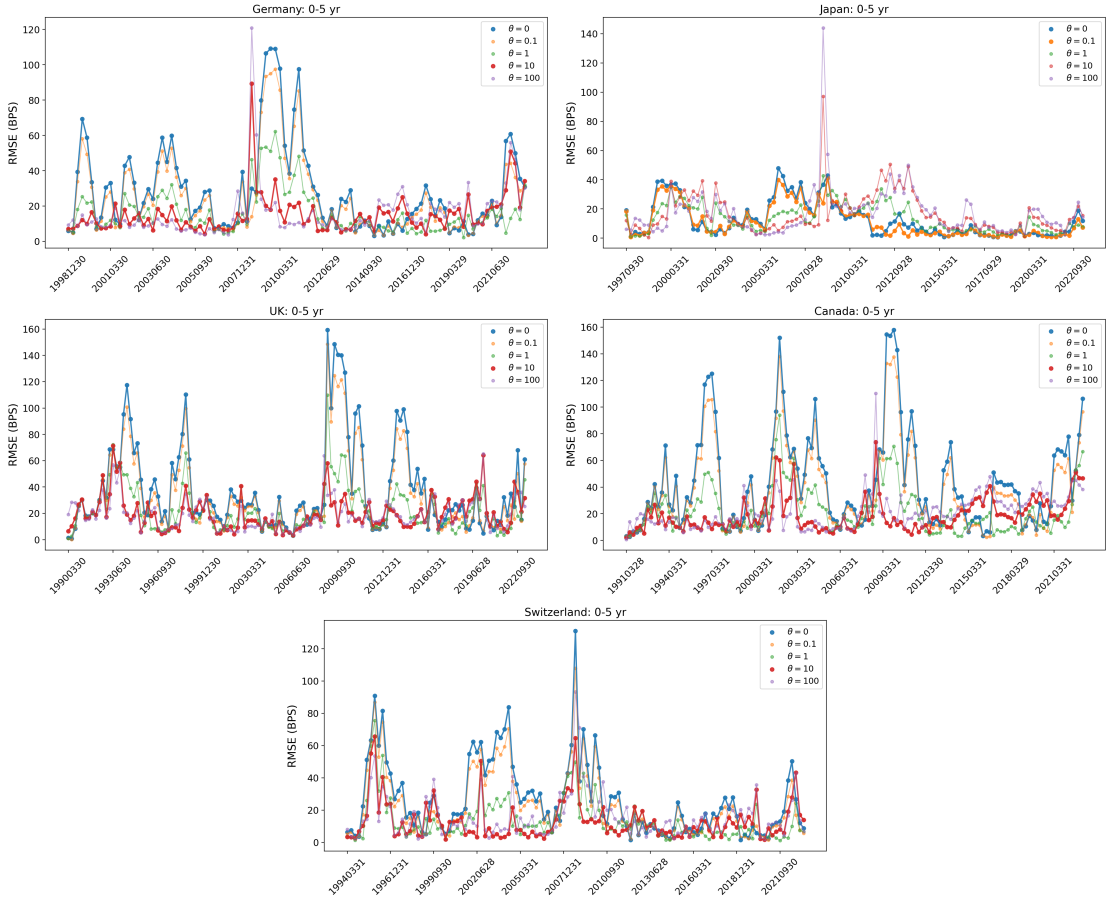
The figure shows the OOS YTM RMSE for 5-year block cross-validation on a finer grid of the transfer parameter θ.

D.3.2 Alternative Block Cross-Validation

This subsection presents alternative block cross-validation results that mask maturity blocks of varying length, representing the data gaps commonly observed in each market. To evaluate the ability of transfer learning to extrapolate or interpolate yields when no bonds within a maturity window are observed, we implement the following masked-maturity block analysis. Using the US as the reference country, for each country pair we withhold all bonds whose time-to-maturity falls in a short, medium, or long maturity bucket and apply the joint learning to the remaining bonds. Because the bonds in the masked block are never seen during estimation, the experiment provides a pseudo-out-of-sample evaluation that mimics the data gaps observed in government bond markets. We compute the YTM errors on these masked bonds as a measure of the model's extrapolation and interpolation ability in such situations. We use this analysis as a data-driven way to choose the hyperparameter θ, and the results suggest that jointly estimating curves provides robustness relative to single-market estimation.

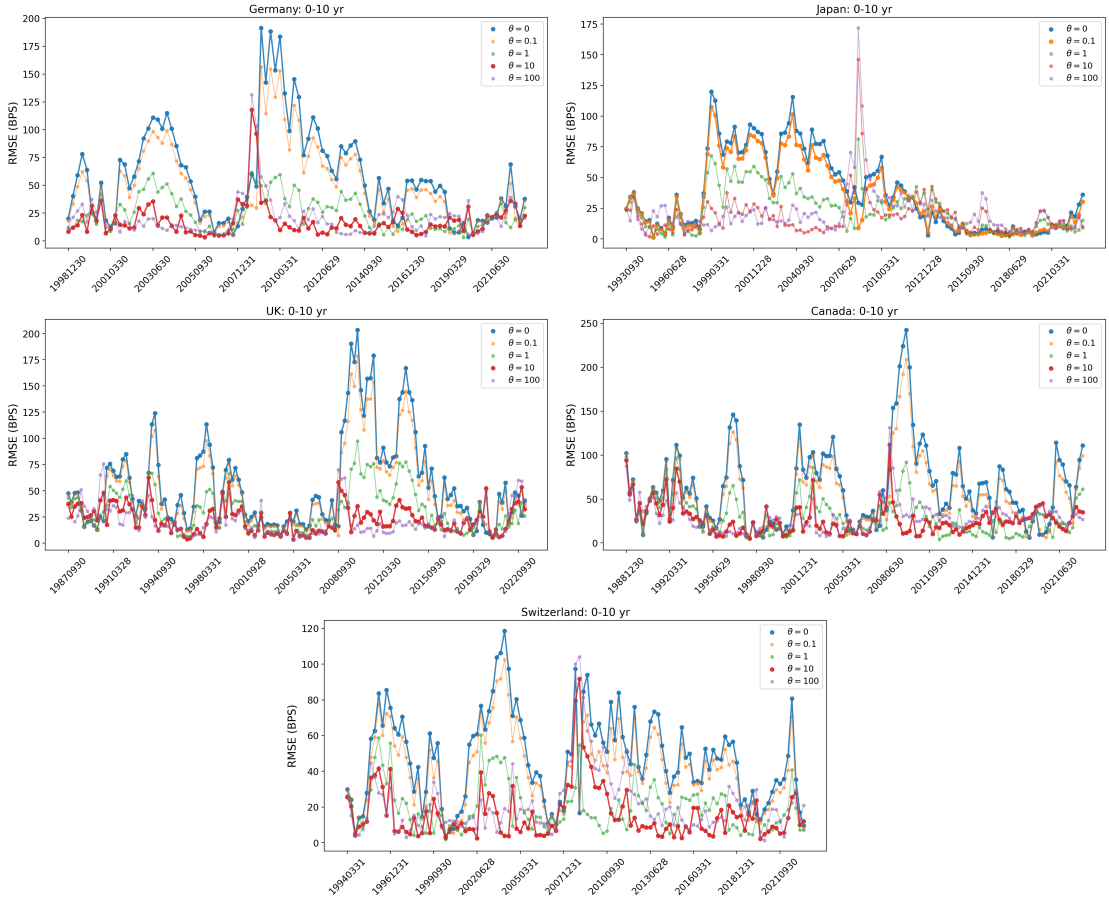
The OOS RMSE reduction from KR-T is substantial when short maturities are masked. Even though we do not directly observe short-maturity bonds in the masked market, the FX forward rates observed there let the swapped US bonds reflect effective prices at short maturities. At long maturities, where FX forwards are not observed, the swapped US bonds carry no extra information for computing the YTM in the masked market.

Figure A.11: Out-of-sample YTM RMSE over time for the masked 0-5 year block



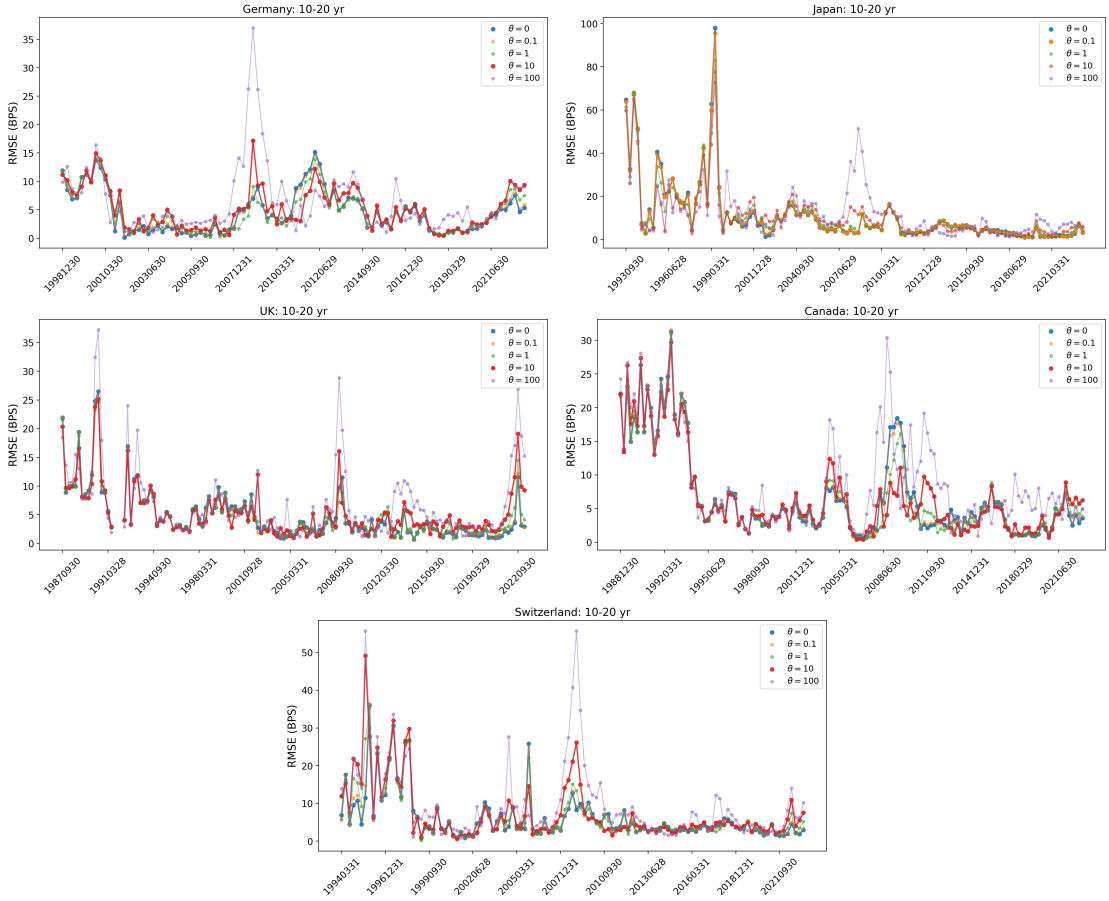
The figure shows the OOS YTM RMSE over time for the 0–5 year maturity block when each country's yield curve is estimated jointly with the US. The overall in-sample line is the YTM RMSE evaluated over all bonds when the yield curve is estimated independently with all observed bonds; the in-sample masked-block YTM RMSE is evaluated only on the masked bonds in the respective maturity block.

Figure A.12: Out-of-sample YTM RMSE over time for the masked 0-10 year block



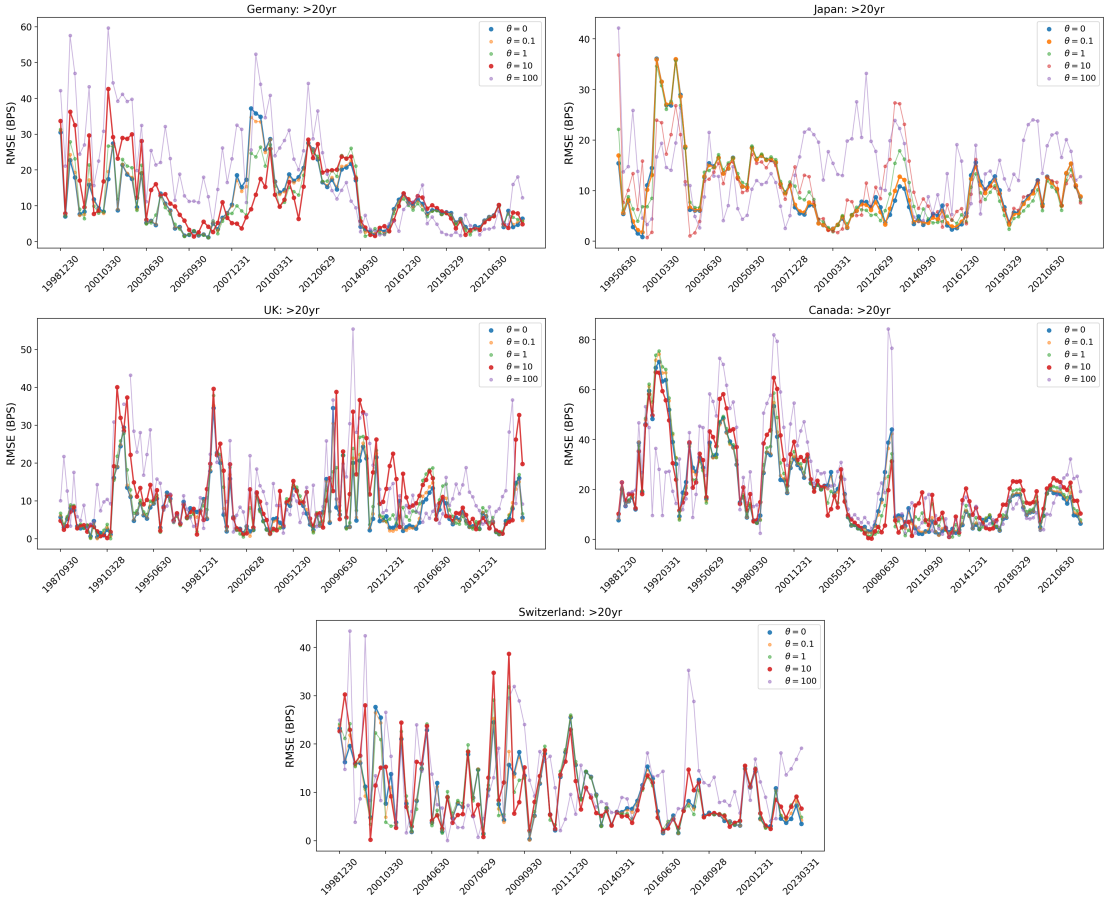
The figure shows the OOS YTM RMSE over time for the 0–10 year maturity block when each country's yield curve is estimated jointly with the US. The overall in-sample line is the YTM RMSE evaluated over all bonds when the yield curve is estimated independently with all observed bonds; the in-sample masked-block YTM RMSE is evaluated only on the masked bonds in the respective maturity block.

Figure A.13: Out-of-sample YTM RMSE over time for the masked medium block



The figure shows the OOS YTM RMSE over time for the medium (10–20 year) maturity block when each country's yield curve is estimated jointly with the US. The overall in-sample line is the YTM RMSE evaluated over all bonds when the yield curve is estimated independently with all observed bonds; the in-sample masked-block YTM RMSE is evaluated only on the masked bonds in the respective maturity block.

Figure A.14: Out-of-sample YTM RMSE over time for the masked long block

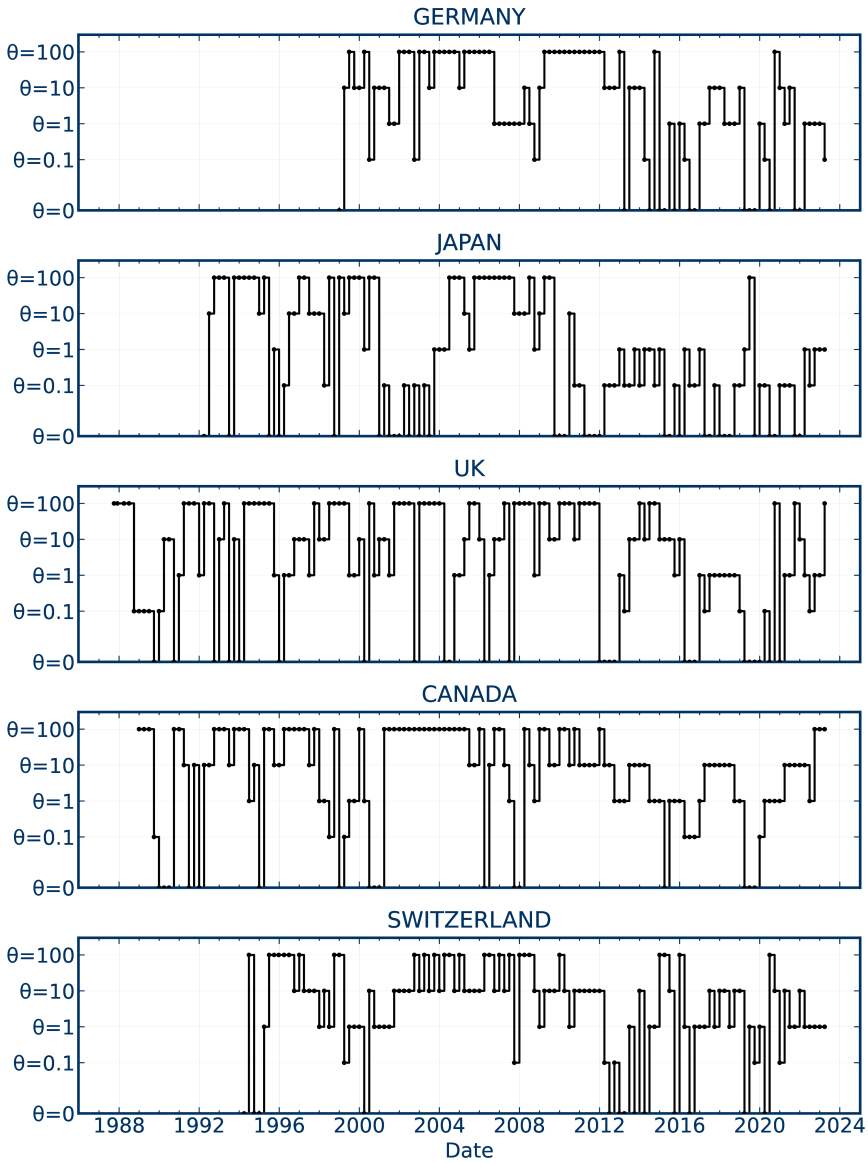


The figure shows the OOS YTM RMSE over time for the long (>20 year) maturity block when each country's yield curve is estimated jointly with the US. The overall in-sample line is the YTM RMSE evaluated over all bonds when the yield curve is estimated independently with all observed bonds; the in-sample masked-block YTM RMSE is evaluated only on the masked bonds in the respective maturity block.

D.3.3 Time-Varying Transfer Learning

Figure A.15 shows the best transfer parameter θ over time for each market. Table A.3 reports the percentage of dates on which each θ is best.

Figure A.15: Best transfer parameter θ over time



This figure shows for each market over time, the value of θ that minimizes the out-of-sample YTM RMSE under contiguous 5-year block cross-validation.

Table A.3: Percentage of dates each θ is best

	$\theta = 0$	$\theta = 0.1$	$\theta = 1$	$\theta = 10$	$\theta = 100$
GERMANY	14.3	7.1	22.4	22.4	33.7
JAPAN	25.6	21.6	15.2	11.2	26.4
UK	17.5	7.0	19.6	18.9	37.1
CANADA	12.3	5.1	18.1	27.5	37.0
SWITZERLAND	11.1	5.1	25.6	35.0	23.1

This table shows the percentage of dates (in %) on which each θ yields the lowest out-of-sample YTM RMSE under the 5-year block cross-validation. $\theta = 0$ corresponds to standalone KR estimation, and higher θ values indicate stronger cross-market learning.

Table A.4: OOS RMSE comparison between time-varying and fixed θ

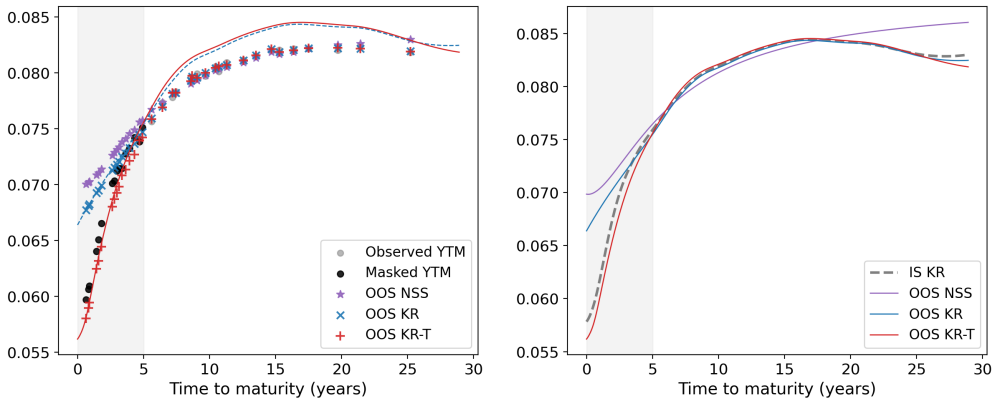
	Dynamic θ (BPS)	Fixed θ (BPS)	Diff (BPS)
GERMANY	8.24	12.30	4.06
JAPAN	7.94	10.93	2.99
UK	10.97	14.45	3.48
CANADA	12.15	16.84	4.69
SWITZERLAND	7.54	9.29	1.75

The table shows the OOS YTM RMSE comparison for 5-yr block cross-validation between using time-varying optimal θ and fixed θ chosen from overall block cross-validation results. The fixed θ is the baseline value, $\theta = 10$ for all markets except $\theta = 1$ for Japan. All numbers are in basis points (BPS).

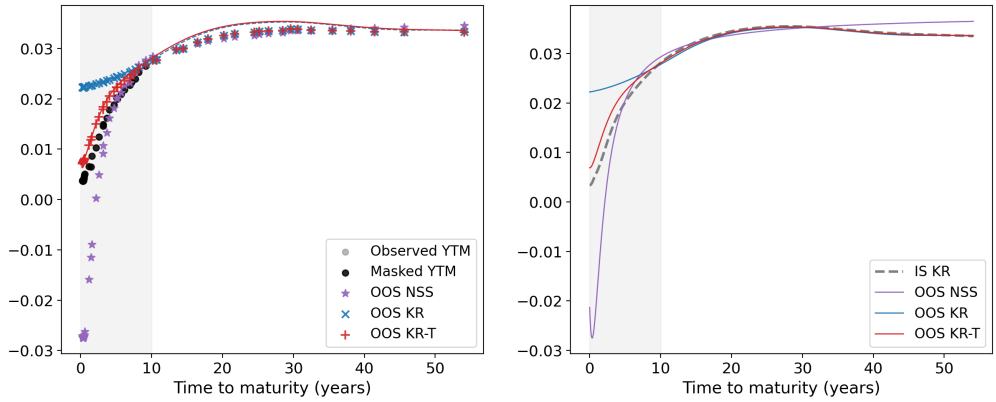
D.3.4 Example dates for masking maturity blocks

The following figures collect fitted yield curves and YTM fits on selected example dates under maturity masking. On individual dates, KR-T consistently recovers plausible curve shapes in the masked region, while both standalone KR and NSS produce unreliable extrapolations. This visual evidence complements the aggregate RMSE results in the main text.

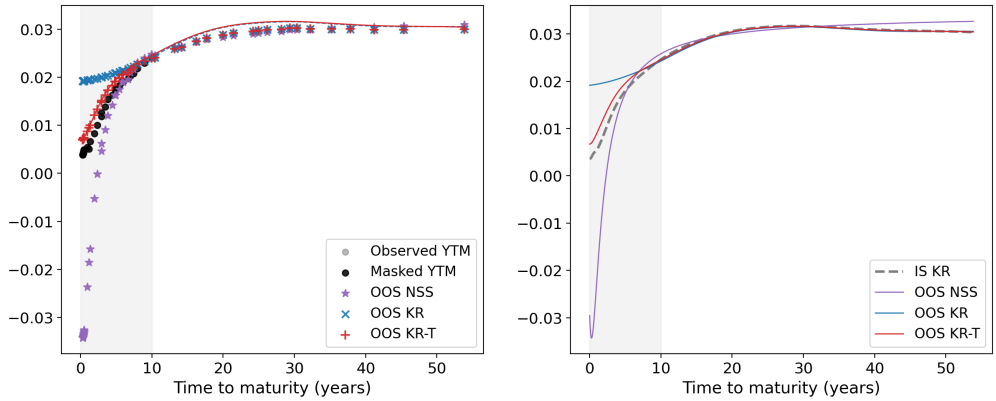
Figure A.16: UK example dates



(a) 1996-03-29, masked 0–5y



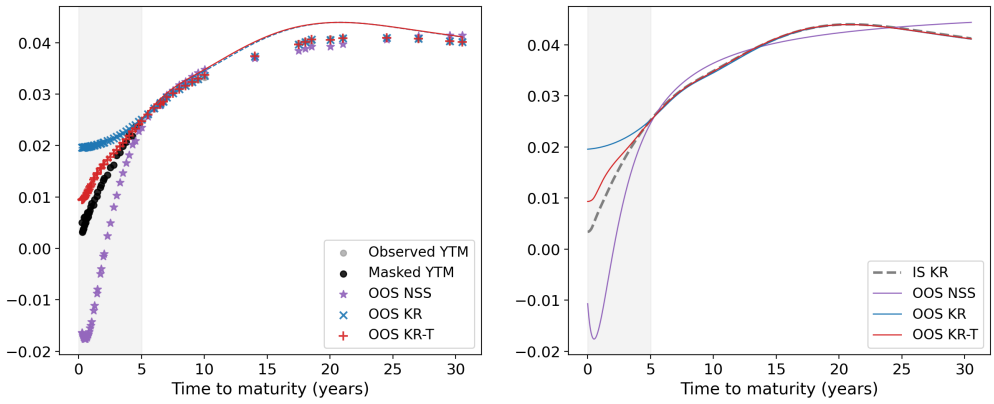
(b) 2014-06-30, masked 0–10y



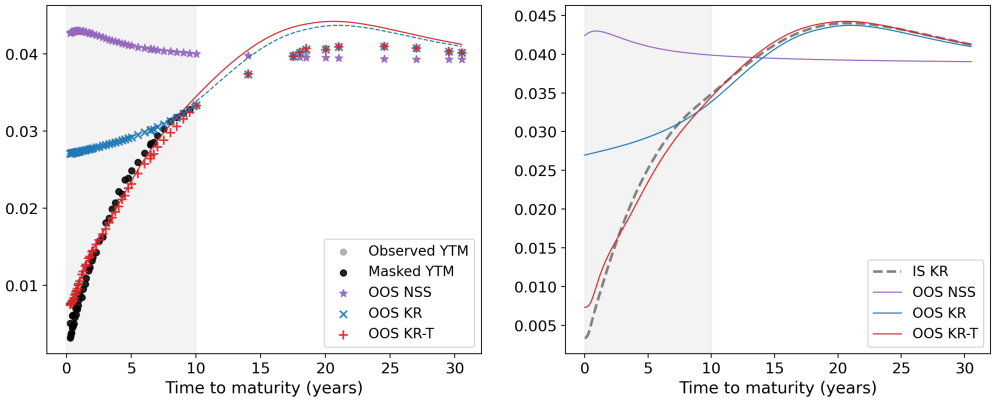
(c) 2014-09-30, masked 0–10y

This figure collects out-of-sample fitted and observed YTM (left) and yield estimates (right) for the UK on selected example dates. The light grey shaded region represents the masked maturities; bonds in this region are excluded from the estimation. KR-T uses transfer parameter $\theta = 10$. NSS is added for comparison. The in-sample (IS) KR curve is estimated using all available bonds.

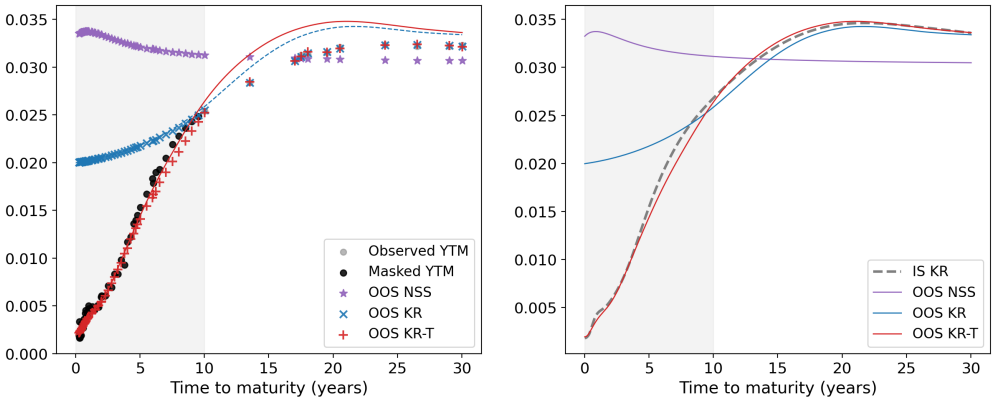
Figure A.17: Germany example dates



(a) 2009-12-31, masked 0–5y



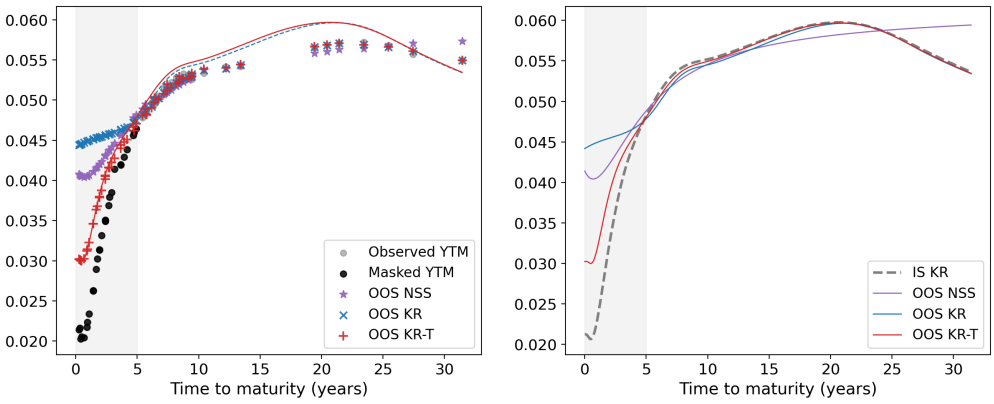
(b) 2009-12-31, masked 0–10y



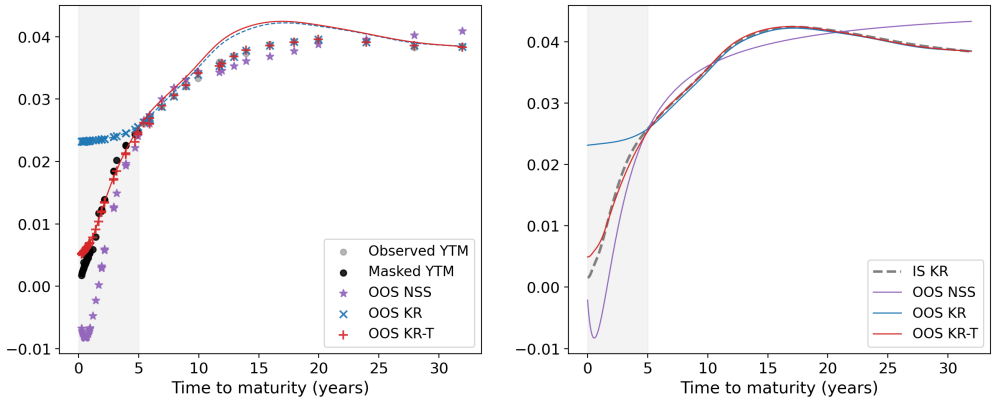
(c) 2010-06-30, masked 0–10y

This figure collects out-of-sample fitted and observed YTM (left) and yield estimates (right) for Germany on selected example dates. The light grey shaded region represents the masked maturities; bonds in this region are excluded from the estimation. KR-T uses transfer parameter $\theta = 10$. NSS is added for comparison. The in-sample (IS) KR curve is estimated using all available bonds.

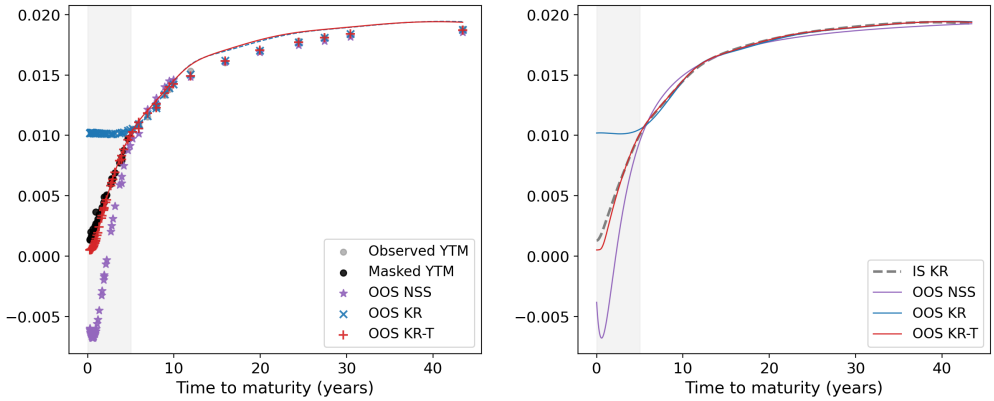
Figure A.18: Canada example dates



(a) 2001-12-31, masked 0–5y



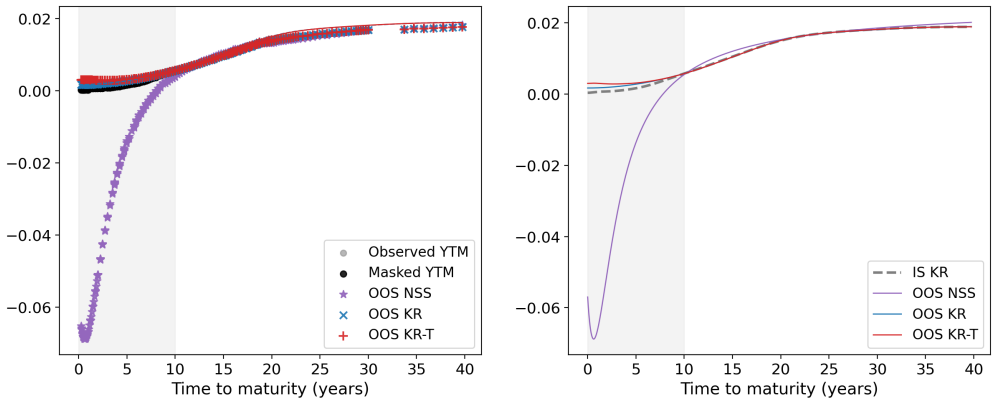
(b) 2009-06-30, masked 0–5y



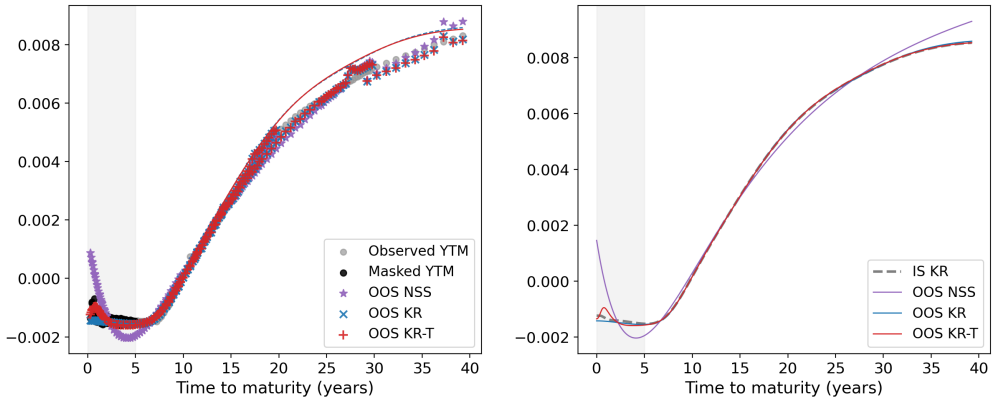
(c) 2021-06-30, masked 0–5y

This figure collects out-of-sample fitted and observed YTM (left) and yield estimates (right) for Canada on selected example dates. The light grey shaded region represents the masked maturities; bonds in this region are excluded from the estimation. KR-T uses transfer parameter $\theta = 10$. NSS is added for comparison. The in-sample (IS) KR curve is estimated using all available bonds.

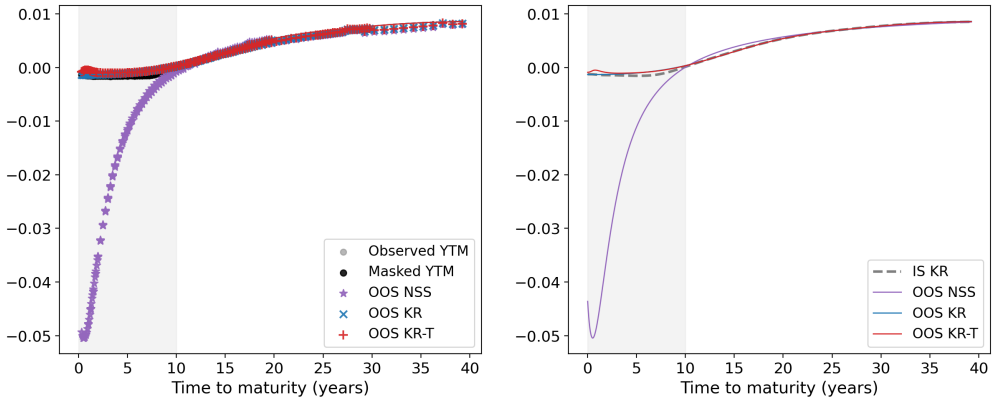
Figure A.19: Japan example dates



(a) 2014-06-30, masked 0–10y



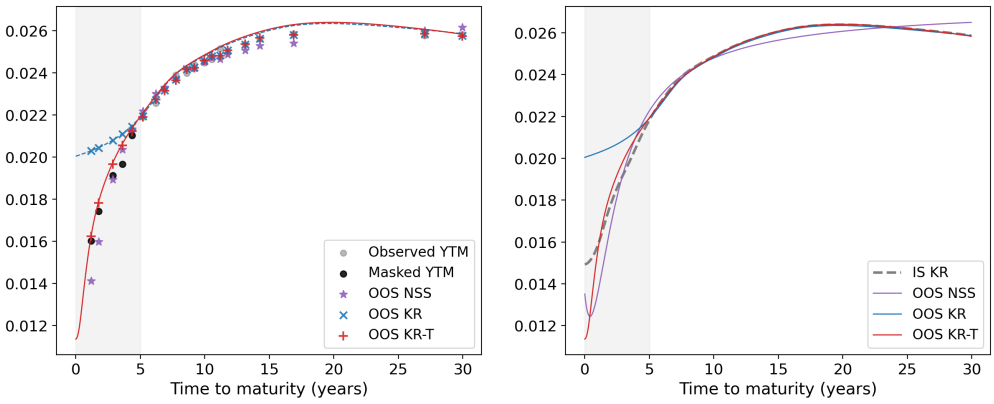
(b) 2018-12-31, masked 0–5y



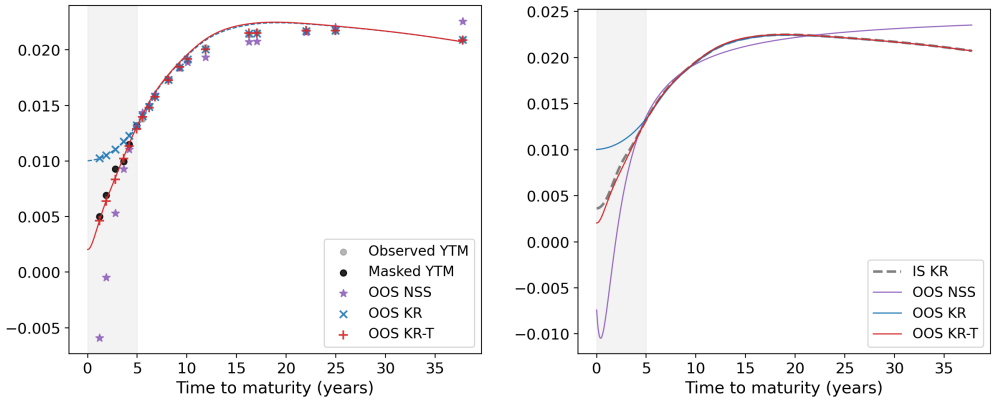
(c) 2018-12-31, masked 0–10y

This figure collects out-of-sample fitted and observed YTM (left) and yield estimates (right) for Japan on selected example dates. The light grey shaded region represents the masked maturities; bonds in this region are excluded from the estimation. KR-T uses transfer parameter $\theta = 1$. NSS is added for comparison. The in-sample (IS) KR curve is estimated using all available bonds.

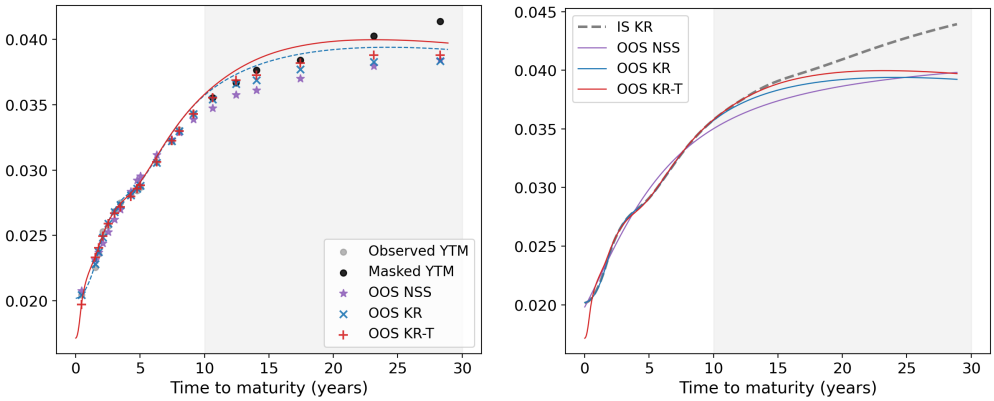
Figure A.20: Switzerland example dates



(a) 2006-03-31, masked 0–5y



(b) 2011-03-31, masked 0–5y



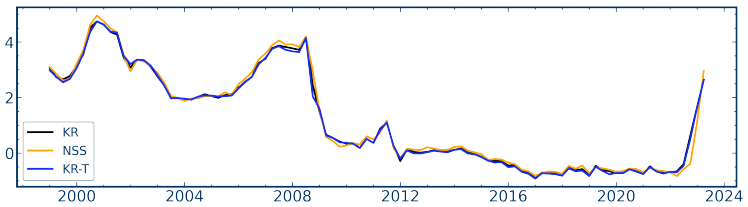
(c) 1999-12-30, masked 10–30y

This figure collects out-of-sample fitted and observed YTM (left) and yield estimates (right) for Switzerland on selected example dates. The light grey shaded region represents the masked maturities; bonds in this region are excluded from the estimation. KR-T uses transfer parameter $\theta = 10$. NSS is added for comparison. The in-sample (IS) KR curve is estimated using all available bonds.

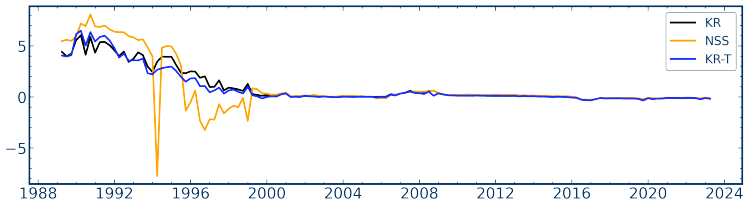
D.4 Time Series of Yields and Forward Rates

This subsection extends the time-series analysis of Section 6 to all markets. Transfer learning produces smoother and more stable short-rate and forward-rate estimates across all countries, with the most pronounced improvements in the early sample periods when bond data were sparsest.

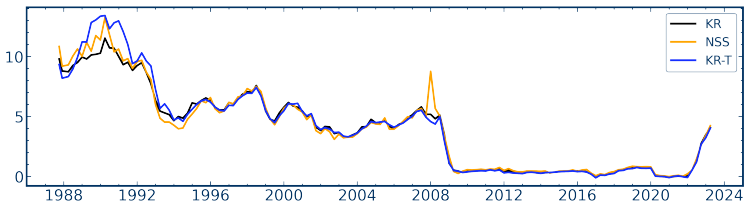
Figure A.21: 30-day yield estimates



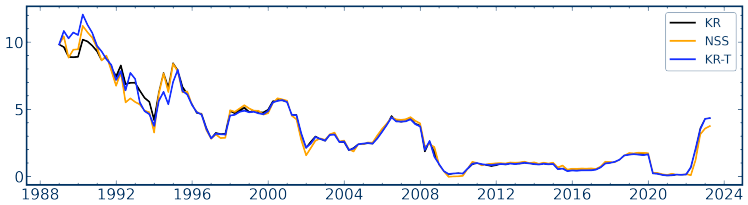
(a) Germany



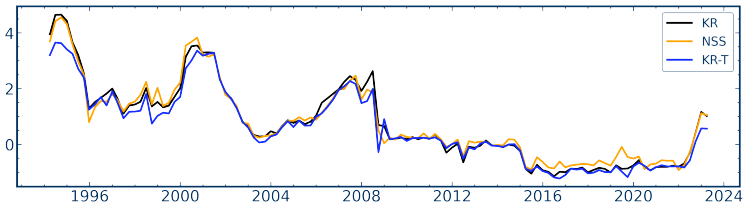
(b) Japan



(c) UK



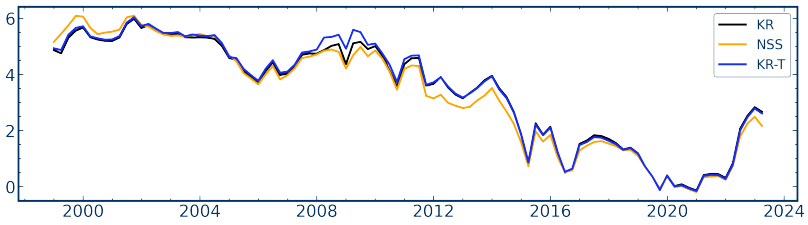
(d) Canada



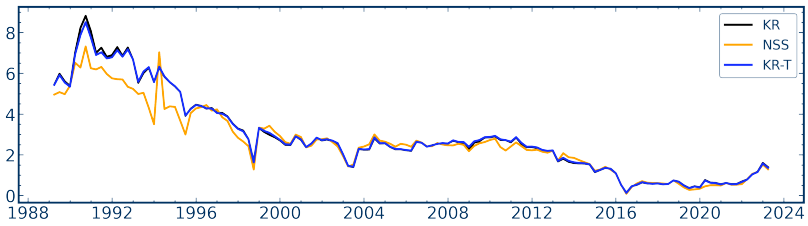
(e) Switzerland

The figure shows the time series of 30-day yield estimates from the different methods. Each panel corresponds to a different market. The transfer parameter is $\theta = 10$ for all markets except Japan, where $\theta = 1$. Quarterly data are used.

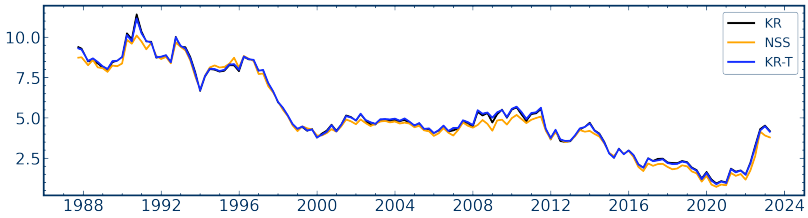
Figure A.22: 10-year forward rates



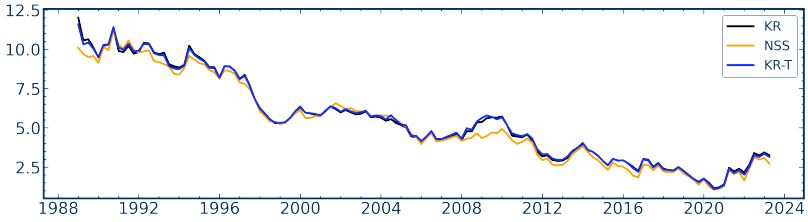
(a) Germany



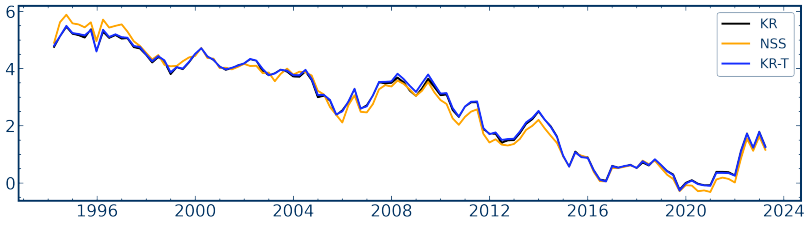
(b) Japan



(c) UK



(d) Canada



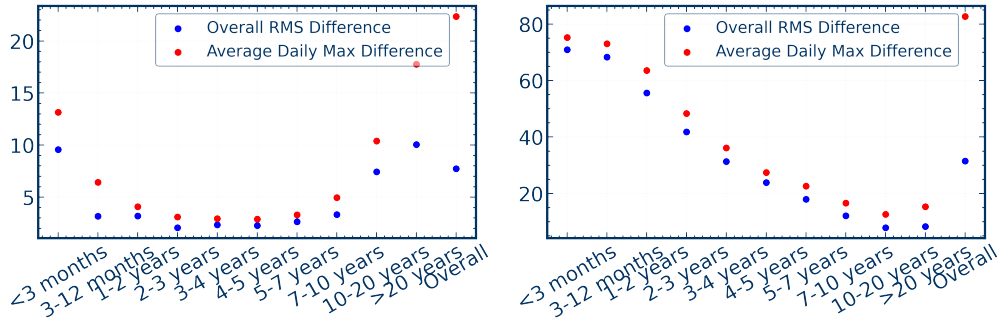
(e) Switzerland

The figure shows the time series of 10-year forward rates implied by the estimated yield curves from the KR, KR-T, and NSS methods. Each panel corresponds to one market. The transfer parameter is $\theta = 10$ for all markets except Japan, where $\theta = 1$. Quarterly data are used.

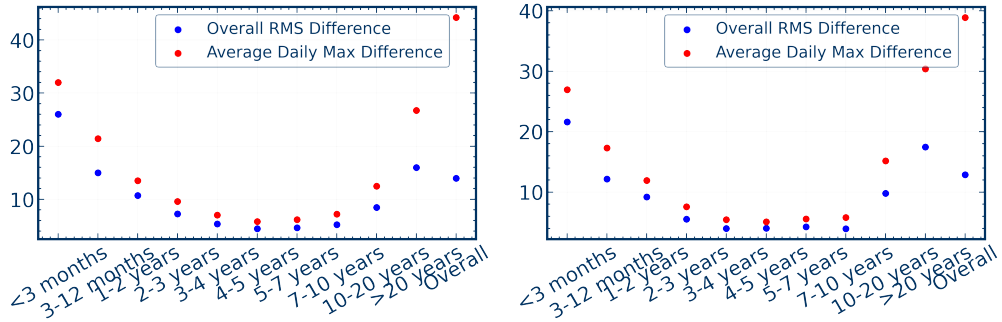
D.5 Yield Curve Difference

This subsection presents the full set of yield curve comparisons across all markets. The findings are consistent: differences between NSS and KR-T are large (up to about 80 basis points for Japan) and concentrated at short maturities, reflecting the parametric limitations of NSS, while differences between standalone KR and KR-T are concentrated at the very short end and substantially smaller beyond it, confirming that transfer learning introduces only targeted adjustments where cross-market information is most informative.

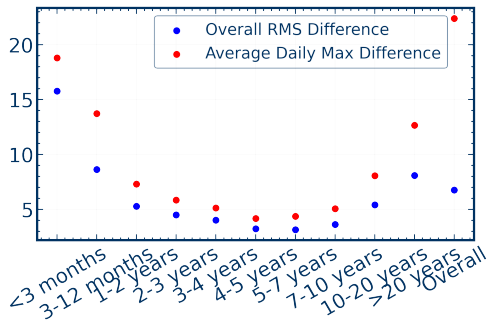
Figure A.23: KR and NSS yield RMS vs max difference



(a) Germany and Japan



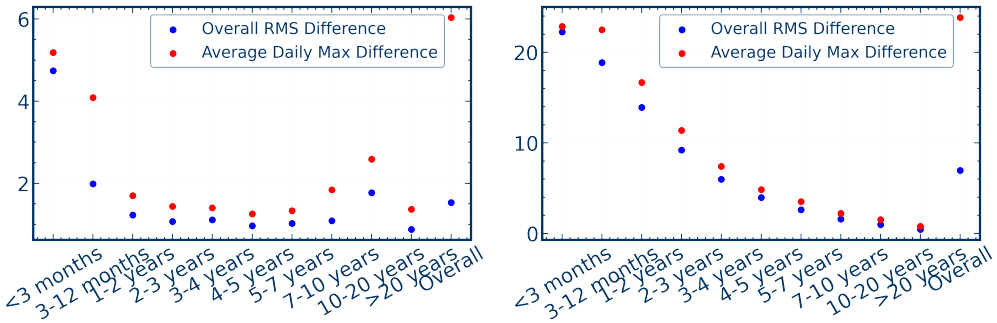
(b) UK and Canada



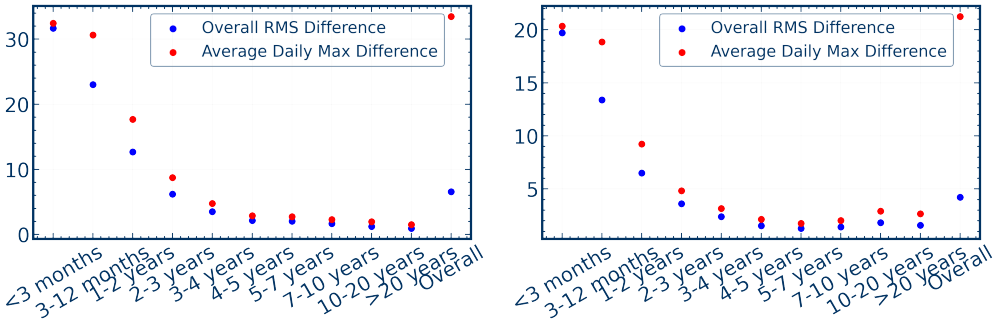
(c) Switzerland

This figure compares yield curve differences between the KR and NSS methods. Each panel shows the RMS difference and the average maximum absolute difference for the corresponding market. Quarterly data are used.

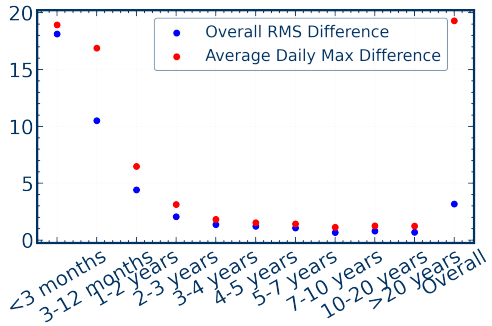
Figure A.24: KR and KR-T yield RMS vs max difference



(a) Germany and Japan



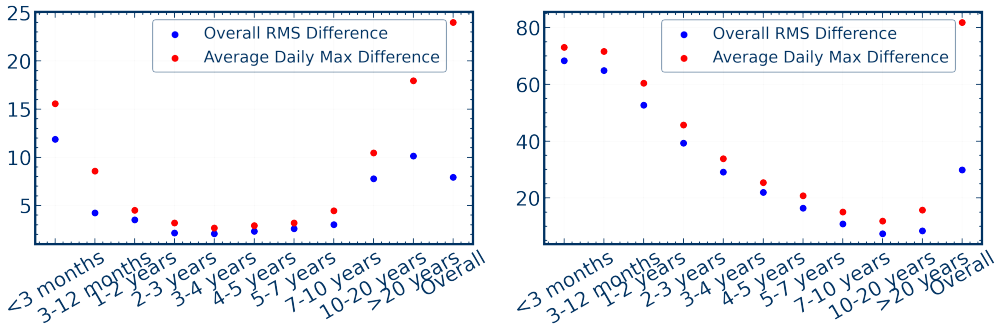
(b) UK and Canada



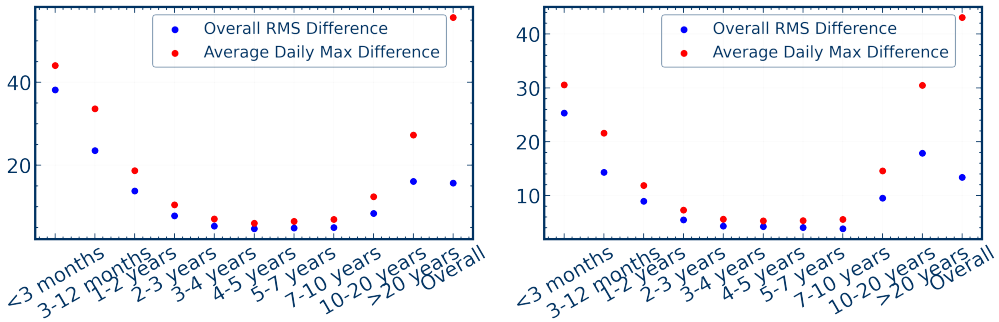
(c) Switzerland

This figure compares yield curve differences between the KR-T and KR methods. Each panel shows the RMS difference and the average maximum absolute difference for the corresponding market. Quarterly data are used. The transfer parameter is $\theta = 10$ for all markets except Japan, where $\theta = 1$.

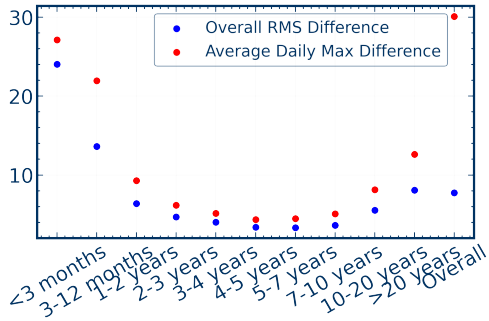
Figure A.25: NSS and KR-T yield RMS vs max difference



(a) Germany and Japan



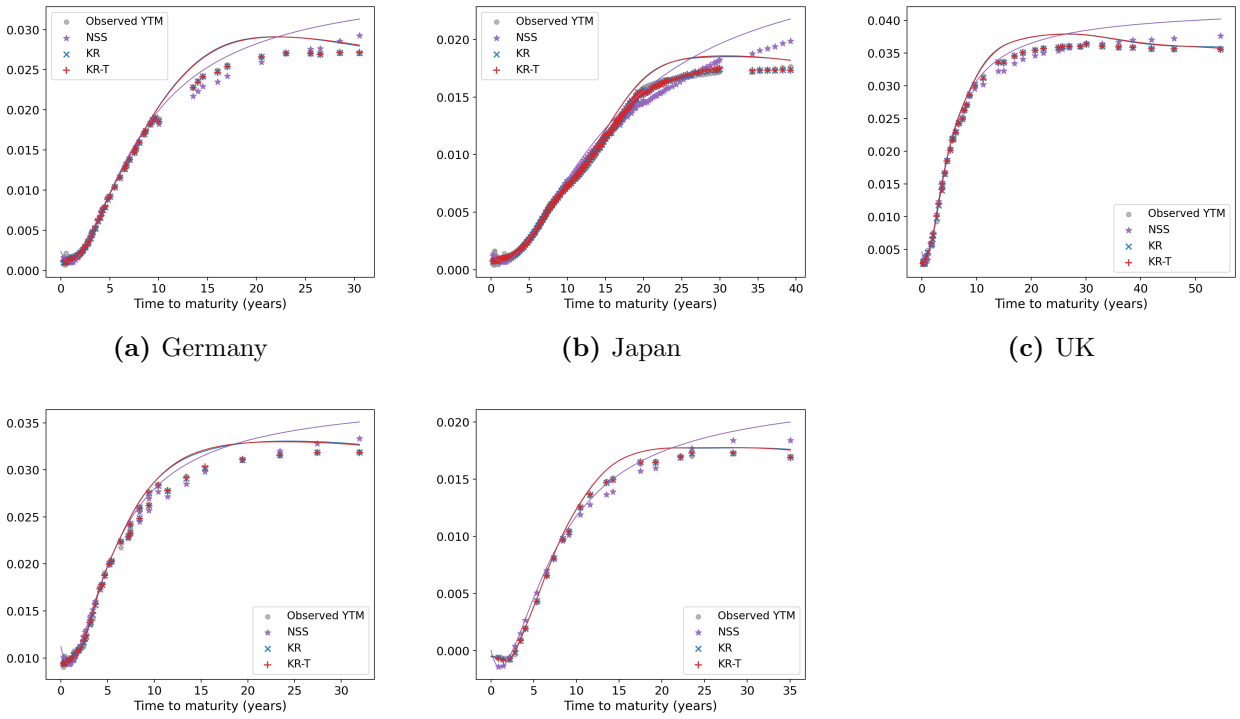
(b) UK and Canada



(c) Switzerland

This figure compares yield curve differences between the KR-T and NSS methods. Each panel shows the RMS difference and the average maximum absolute difference for the corresponding market. Quarterly data are used. The transfer parameter is $\theta = 10$ for all markets except Japan, where $\theta = 1$.

Figure A.26: Illustrative yield curves and fitted YTM on 2013-12-30



(a) Germany **(b)** Japan **(c)** UK

(d) Canada **(e)** Switzerland

The figure shows the fitted yield curves and YTM fits for all markets on 2013-12-30, a common date across samples. The panels display fitted YTM by method and the full yield curve. Blue, red, and purple lines correspond respectively to KR, KR-T, and NSS fits. KR-T uses transfer parameter $\theta = 10$ for all markets except Japan, where $\theta = 1$.

Table A.5: KR vs NSS excluding less than one-year securities from evaluation

	KR YTM RMSE (BPS)		NSS YTM RMSE (BPS)	
Country	All bonds	> 1y only	All bonds	> 1y only
US	**9.63**	**10.34**	11.01	11.92
Germany	**4.57**	**3.65**	6.83	6.34
Japan	**3.53**	**3.49**	6.69	6.72
UK	**6.36**	**5.67**	9.57	9.33
Canada	**6.49**	**5.64**	8.77	8.24
Switzerland	**6.43**	**5.57**	8.76	7.60

The table shows out-of-sample YTM RMSE (BPS) from 3-fold stratified cross-validation on monthly data, for standalone KR and NSS. KR uses the common parameters $\alpha = 0.05$, $\lambda = 1$. "All bonds" evaluates the error over the full estimation data set (time-to-maturity > 90 days); "> 1y only" excludes bonds with time-to-maturity in [90d, 365d] from the error metric while estimating the curves from all bonds. KR dominates NSS under either evaluation.

D.6 Robustness to Short-Maturity Bonds

Our results in the main text use a 90-day minimum time-to-maturity filter (Section 3). Some studies and central banks instead apply a more conservative one-year filter because very short maturity securities can be noisier and may trade at a liquidity discount. This section shows that this data filter choice does not affect our results: all conclusions are unchanged when the analysis is restricted to maturities above one year. We proceed in two steps. First, we re-evaluate the out-of-sample errors of Tables 1 and 2 on the subset of securities with more than one year to maturity, so that the short securities inform the fit but are excluded from the error metric. Second, we measure how much the standalone KR curve itself moves when the securities with maturities between 90 days and one year are removed from the estimation.

Table A.5 repeats the KR and NSS comparison with securities of less than one year to maturity excluded from the error metric. KR continues to dominate NSS in every market: the OOS YTM RMSE reduction relative to NSS ranges from 13% (US) to 48% (Japan), compared with 13–47% on the full evaluation set. Excluding the short end lowers the absolute errors in five of six markets and raises them only slightly for the US.

Table A.6 repeats the KR-T comparison. Under block cross-validation, KR-T still uniformly dominates both benchmarks, cutting the OOS YTM RMSE by 6–44% relative to KR and by 65–87% relative to NSS (compared with 4–51% and 71–86% in the main text). The reduction relative to KR is somewhat smaller because the securities with maturities between 90 days and one year lie at the short end, where transfer learning is most valuable, but the gains remain substantial. Under stratified cross-validation, KR-T stays within about 0.3 BPS of KR in every market. Hence, none of our conclusions depend on the securities with less than one year to maturity.

Finally, we estimate the standalone KR yield curve with the one-year filter instead of the 90-day filter and compare the two fitted curves directly. Table A.7 reports the RMS difference over time between the two curves at each maturity of one year and above. We do not report maturities below

Table A.6: KR-T comparison excluding less than one-year securities from evaluation

	Block CV			Stratified CV		
Country	KR	KR-T	NSS	KR	KR-T	NSS
Germany	14.52	**8.10**	62.04	**2.59**	2.84	5.51
Japan	10.35	**9.77**	32.01	**3.39**	3.48	6.73
UK	16.34	**10.80**	30.93	**5.52**	5.64	8.85
Canada	21.16	**11.79**	36.65	**5.19**	5.34	7.58
Switzerland	12.95	**8.01**	35.97	**5.62**	5.75	8.09

The table shows out-of-sample YTM RMSE (BPS) under 5-year block and 3-fold stratified cross-validation, with bonds in [90d, 365d] excluded from the error metric (curves are still estimated from all bonds > 90 days). KR-T uses $\theta = 10$ for all markets except Japan ($\theta = 1$). KR-T dominates under block CV, where maturity gaps are present, and matches KR under stratified CV, where local data are dense.

one year: there the one-year-filter fit contains no bonds and must extrapolate, so a comparison would mechanically reflect the additional information in the short-maturity bonds rather than the stability of the fit. For maturities above one year, where both fits are informed by bonds, the two curves are essentially indistinguishable: the RMS difference is below 1 BPS in five of six markets and below 3 BPS for the US, whose short end is the most densely populated. At the one-year boundary itself, where the one-year-filter fit is anchored only by its shortest bonds, differences are naturally larger but remain contained; the out-of-sample comparisons above, which include these maturities, confirm that they do not affect any conclusions. Figures A.27 and A.28 confirm this visually: the one- and ten-year yield time series under the two filters lie on top of each other. In sum, including bonds with maturities between 90 days and one year leaves the curve unchanged where it is already identified and adds information only where the coarser filter would otherwise rely on extrapolation.

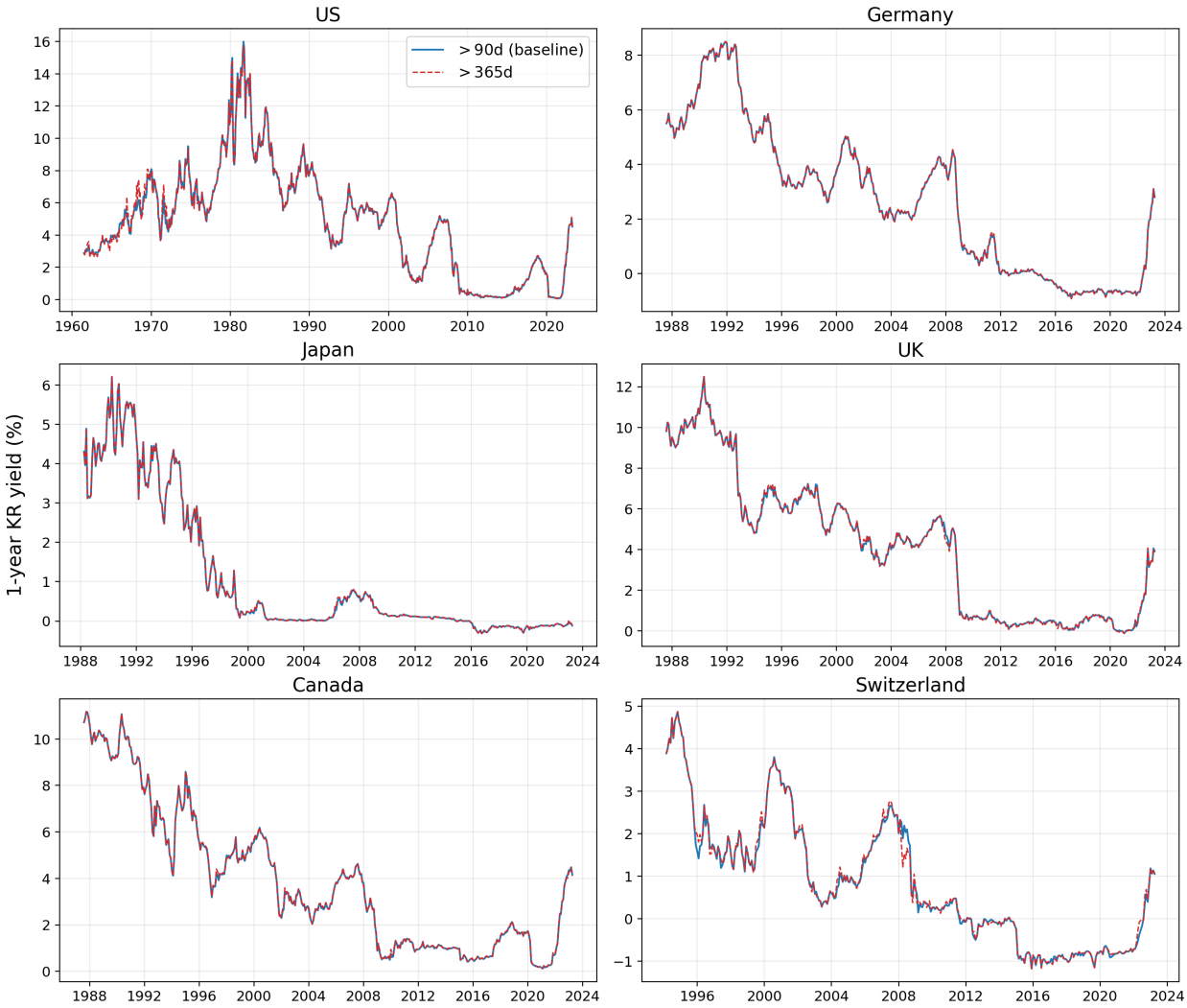
Taken together, these results show that our conclusions are robust to the inclusion of short-maturity securities. The KR/NSS and KR-T/KR comparisons continue to hold, and the estimated curves are unchanged wherever they are data-constrained. At the same time, the short securities stabilize the short end. Because our method is robust to their inclusion and they add information precisely where it is otherwise absent, we adopt the more inclusive 90-day maturity filter.

Table A.7: Standalone KR curve difference with vs without sub-one-year bonds, by maturity

Maturity	US	Germany	Japan	UK	Canada	Switzerland
1Y*	16.73	2.27	1.31	6.56	3.98	10.35
2Y	2.98	0.25	0.10	0.81	0.51	0.77
3Y	1.36	0.12	0.05	0.46	0.21	0.46
5Y	0.77	0.09	0.05	0.21	0.14	0.09
7Y	0.90	0.13	0.06	0.24	0.16	0.06
10Y	1.58	0.24	0.05	0.18	0.19	0.05
15Y	1.04	0.21	0.05	0.16	0.34	0.06
20Y	0.46	0.26	0.08	0.15	0.34	0.04
30Y	—	0.24	0.08	0.26	0.36	0.06

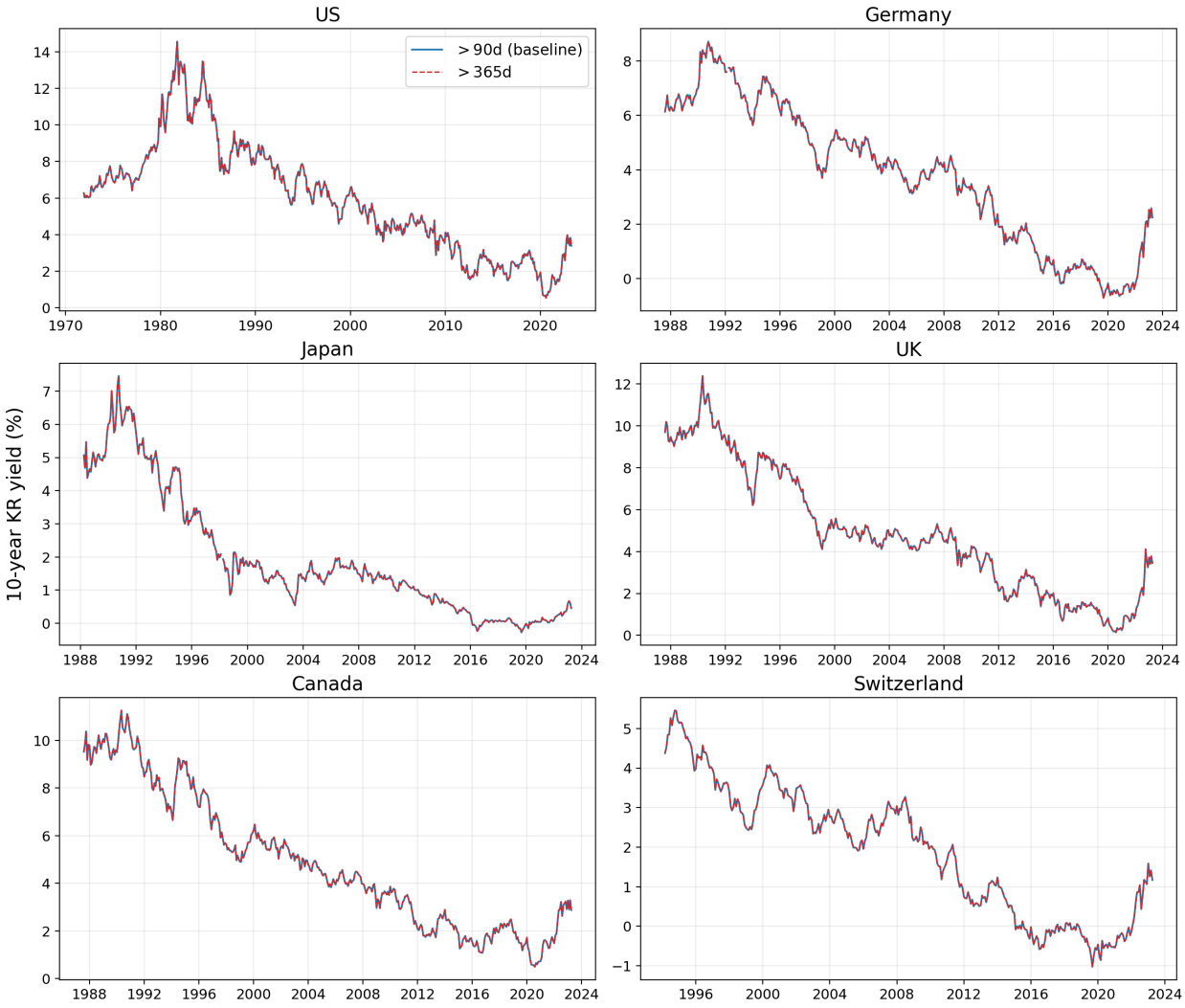
The table reports how much the estimated standalone KR curve changes between fitting with all bonds with time-to-maturity > 90 days ($y^{>90}$) and with only those > 365 days ($y^{>365}$). Each entry is the RMS across sample dates of the difference, in BPS, at the indicated maturity m ($\alpha = 0.05$, $\lambda = 1$), where t indexes monthly end dates: $\text{RMS}(m) = 10^4 \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(y_t^{>90}(m) - y_t^{>365}(m)\right)^2}$. Maturities below one year are not reported: there the > 365-day fit contains no bonds and must extrapolate, so the comparison would reflect the additional short-maturity information rather than the stability of the fit. *The one-year maturity is the boundary of the > 365-day fit's bond support, where that fit is anchored only by its shortest bonds. The US 30Y entry is omitted because only three sampled date carried a US bond that long.

Figure A.27: 1-year KR yield over time: > 90d filter vs > 365d filter



The figure compares the standalone KR fitted 1-year zero-coupon yield over time for each market, estimated from all bonds with time-to-maturity > 90 days (solid) versus only those > 365 days (dashed), on monthly end dates ($\alpha = 0.05$, $\lambda = 1$). Although one year sits at the boundary of the > 365d filter's bond support and the dashed curve extrapolates just below its shortest bond, the two curves track each other closely.

Figure A.28: 10-year KR yield over time: $>$ 90d filter vs $>$ 365d filter

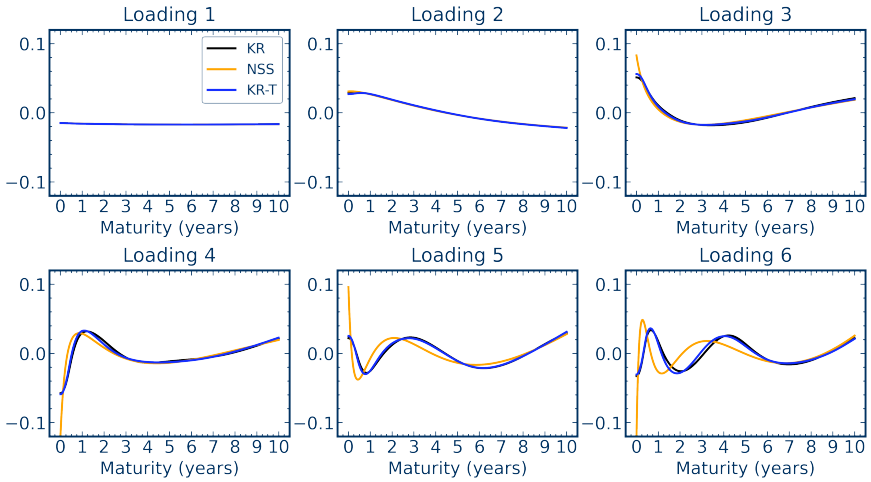


The figure compares the standalone KR fitted 10-year zero-coupon yield over time for each market, estimated from all bonds with time-to-maturity $>$ 90 days (solid) versus only those $>$ 365 days (dashed), on monthly end dates ($\alpha = 0.05$, $\lambda = 1$). At ten years, the two curves are visually identical.

D.7 Basis Functions

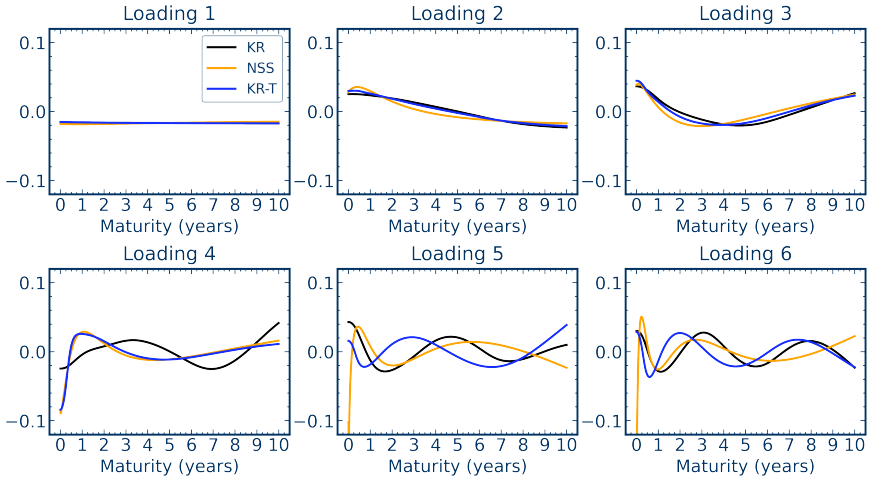
This subsection extends the principal component analysis from the main text to all markets and methods. The key takeaway is that KR and KR-T produce nearly identical loadings for the first two PCs (level and slope) across all markets. Differences emerge in higher-order PCs (third to sixth), where NSS tends to exhibit more abrupt short-end swings and KR-T displays smoother oscillations than standalone KR, reflecting the stabilizing effect of cross-market information.

Figure A.29: Yield curve loadings for Germany



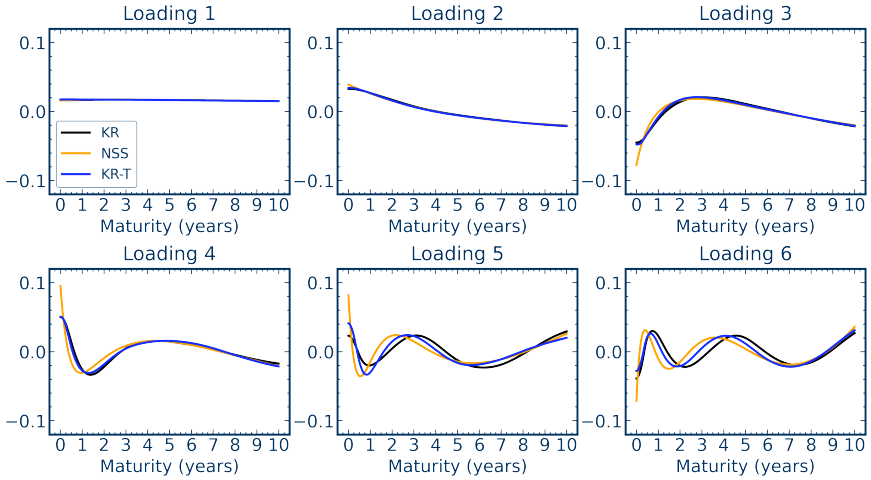
The figure shows the first six loadings estimated from the panel of Germany yield curves for the KR, KR-T, and NSS methods. The panel consists of the estimated yields with maturities up to 10 years for the sample from 1990-01-02 to 2023-03-31. Quarterly data are used.

Figure A.30: Yield curve loadings for Japan



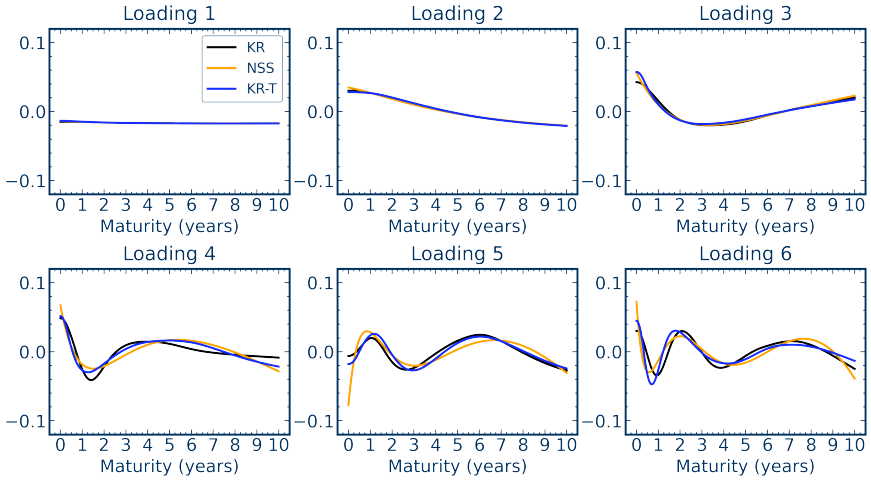
The figure shows the first six loadings estimated from the panel of Japan yield curves for the KR, KR-T, and NSS methods. The panel consists of the estimated yields with maturities up to 10 years for the sample from 1990-01-02 to 2023-03-31. Quarterly data are used.

Figure A.31: Yield curve loadings for the UK



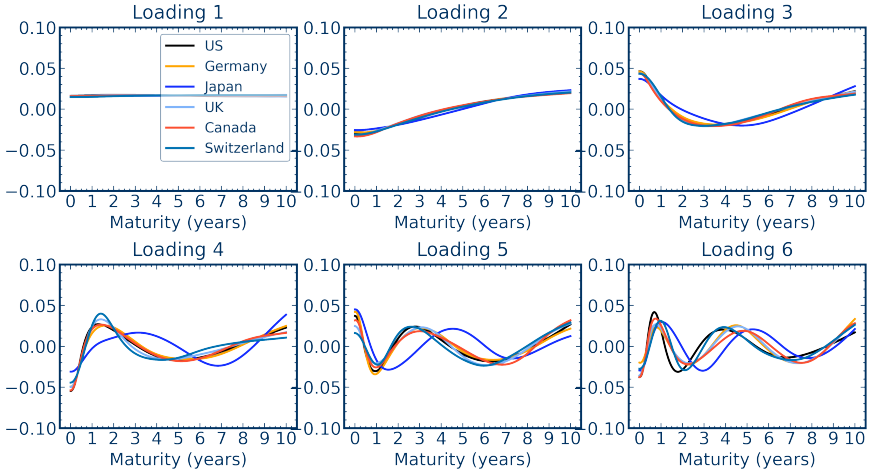
The figure shows the first six loadings estimated from the panel of UK yield curves for the KR, KR-T, and NSS methods. The panel consists of the estimated yields with maturities up to 10 years for the sample from 1990-01-02 to 2023-03-31. Quarterly data are used.

Figure A.32: Yield curve loadings for Switzerland



The figure shows the first six loadings estimated from the panel of Switzerland yield curves for the KR, KR-T, and NSS methods. The panel consists of the estimated yields with maturities up to 10 years for the sample from 1990-01-02 to 2023-03-31. Quarterly data are used.

Figure A.33: KR yield curve loadings for all markets



The figure shows the first six loadings estimated from the panel of yield curves for the KR method across the US, Germany, Japan, the UK, Canada, and Switzerland. The panel consists of the estimated yields with maturities up to 10 years for the sample from 1990-01-02 to 2023-03-31. Daily data are used.

E Official Government Yield Curve Sources

We summarize the publicly available official zero-coupon yield curves (or, where unavailable, constant-maturity yields) published by the relevant central bank or government agency for each country in our study. Table A.8 summarizes the data sources, estimation methodology, available maturities, maturity grid spacing, and sample period for each market.

Table A.8: Official Government Yield Curve Sources

Country	Source	Curve Type	Estimation Method	Maturities	Grid	From
UK	Bank of England[1]	Zero-coupon spot rates	VRP spline (Anderson and Sleath, 2001)	0.5–40 yr	0.5 yr	1979-01-02
Canada	Bank of Canada[2]	Zero-coupon spot rates	Merrill Lynch Exponential spline	0.25–30 yr	0.25 yr	1986-01-02
Germany	Deutsche Bundesbank[3]	Zero-coupon spot rates	Svensson	1–30 yr	1 yr	1997-08-07
Switzerland	Swiss National Bank[4]	Zero-coupon spot rates	Svensson	1–30 yr	Mixed[a]	1988-01-04
Japan	Ministry of Finance[5]	Constant-maturity YTMs	Yield curve interpolation[b]	1–40 yr	Mixed[c]	1974-09-24

The table shows the summary of official yield curve data sources. All curves except Japan provide zero-coupon spot rates estimated from coupon bond prices. Japan publishes constant-maturity yields to maturity of coupon-bearing bonds, which would differ from zero-coupon spot rates. For a comprehensive survey of central bank yield curve estimation methods, see Monetary and Department (2005). [a]Switzerland reports at 1, 2, 3, 4, 5, 10, 15, 20, and 30 years (1-year spacing up to 5 yr, then 5- and 10-year jumps). [b]Japan's constant-maturity rates are computed from a yield curve formed by interpolating prevailing market prices of selected JGBs, using "Reference Statistical Prices" provided by the Japan Securities Dealers Association (JSDA); see methodology at `https://www.mof.go.jp/english/policy/jgbs/reference/interest_rate/outline-e.pdf`. [c]Japan reports at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, and 40 years (1-year spacing up to 10 yr, then 5- and 10-year jumps).

For the UK, the Bank of England publishes daily nominal and real zero-coupon yield curves estimated using the variable roughness penalty (VRP) method of Anderson and Sleath (2001), a cubic spline technique fitted to conventional gilt prices and short-term GC repo rates. The downloaded data are available at a half-year grid from 0.5 to 40 years, starting from 2 January 1979. For Canada, the Bank of Canada publishes zero-coupon yield curves constructed from Government of Canada bonds and treasury bills using a Merrill Lynch exponential spline framework (Bolder, Johnson, and Metzler, 2004), reported at quarterly maturity intervals from 0.25 to 30 years, available from 2 January 1986. For Germany, the Deutsche Bundesbank estimates daily zero-coupon yields from federal securities (Bunds, Bobls, and Schätze) using the Svensson model (Svensson, 1994), as documented in Schich (1997). The downloaded data provide yields at integer maturities from 1 to 30 years on a 1-year grid, beginning 7 August 1997. For Switzerland, the Swiss National Bank (SNB) publishes spot interest rates on Swiss Confederation bonds, also estimated with the

[1]Bank of England yield curves: `https://www.bankofengland.co.uk/statistics/yield-curves`; methodology described in Anderson and Sleath (2001).

[2]Bank of Canada zero-coupon yield curves: `https://www.bankofcanada.ca/rates/interest-rates/bond-yield-curves/`; methodology described in Bolder, Johnson, and Metzler (2004).

[3]Deutsche Bundesbank term structure of interest rates: `https://www.bundesbank.de/en/statistics/money-and-capital-markets/interest-rates-and-yields/term-structure-of-interest-rates`; methodology described in Schich (1997).

[4]Swiss National Bank spot interest rates on Confederation bonds: `https://data.snb.ch/en/topics/ziredev/cube/rendoblid`.

[5]Japan Ministry of Finance interest rate data: `https://www.mof.go.jp/english/policy/jgbs/reference/interest_rate/index.htm`; methodology Q&A: `https://www.mof.go.jp/english/policy/jgbs/reference/interest_rate/qa.htm`.

Svensson model (Svensson, 1994), at maturities of 1, 2, 3, 4, 5, 10, 15, 20, and 30 years, available from 4 January 1988. For Japan, the Ministry of Finance publishes constant-maturity interest rates calculated as the semiannual compound yield on a constant maturity basis, derived from a yield curve formed from prevailing market prices of selected JGBs. The underlying prices are the "Reference Statistical Prices" provided by the Japan Securities Dealers Association (JSDA). These are coupon-bond yields to maturity rather than true zero-coupon spot rates. The available maturities are 1–10 (annually), 15, 20, 25, 30, and 40 years, with data available from 24 September 1974.