Shrinking the Term Structure

Filipovic, Pelger and Ye (2024)

Gustavo Amarante

Summary

- Develop a conditional factor model for the term structure, which unifies **non-parametric curve estimation** with **cross-sectional asset pricing**
- Factors are **investable portfolios** and estimated with **cross-sectional ridge regressions**
- Empirically, **four factors** explain the discount bond **excess return** curve and **term structure premium**
- The **fourth factor,** capturing complex and state-dependent shapes of the term structure premium, is a **hedge for bad economic times** and pays off during recessions

Introduction

- Term structure factor modeling is the **joint** problem of estimating the discount bond return curve and cross-sectional factor modeling
- A precise **return curve** is needed to **discover all relevant pricing factors**
- By **formally** showing the connection between **curve fitting** and **factor modeling**, we can explain the structure of the cross-sectional factors and why they arise.
- The trade off is between **return fitting** and an economically motivated **smoothness** reward of the discount bond return curve

- $d_t(x)$ denotes the price of a discount bond with face value of 1 and maturity x at date t
  - Not observable
- We have a sparse set of coupon bonds
  - TS modeling becomes a curve estimation problem that requires techniques to fill in the gaps to generate estimates for all maturituties
  - Let C_t denote the matrix of cashflows of coupon bonds at time t. Element $C_{t,ij}$ denotes the cash flow of coupon bond i at x_j time units after t

The LOOP connects the observed coupon bond prices with the unobserved discount bonds, and we can interpret a coupon bond as a portfolio of dicount bonds

$$P_{t,i} = \sum_{j=1}^{N} C_{t,ij} d_t\left(x_j\right) \qquad P_{t,i} = \sum_{j=1}^{N} C_{t-1,ij+k_t} d_t\left(x_j\right)$$

The gross return of the coupon bonds from $t-1$ to t

$$\frac{P_{t,i} + C_{t,ik_t}}{P_{t-1,i}} = \frac{\sum_{j=0}^{N} C_{t-1,ij+k_t} d_t\left(x_j\right)}{P_{t-1,i}} = \sum_{j=0}^{N} Z_{t-1,ij} \frac{d_t\left(x_j\right)}{d_{t-1}\left(x_j + \Delta_t\right)}$$

where $Z_{t-1,ij} = \frac{C_{t-1,ij+k_t} d_{t-1}\left(x_j + \Delta_t\right)}{P_{t-1,i}}$ denotes the normalized discounted cash flows of bond i.

One-day Risk-free bond returns are given by $\frac{1}{d_{t-1}(\Delta_t)}$. So we represent the term structure of **discount** bond excess returns as a function

$$r_t\left(x\right)=\frac{d_t\left(x\right)}{d_{t-1}\left(x+\Delta_t\right)}-\frac{1}{d_{t-1}\left(\Delta_t\right)}$$

So excess returns of coupon bonds are

$$R_{t,i}^{bond}=\sum_{j=1}^{N}Z_{t-1,ij}r_t\left(x_j\right)$$

In matrix notation

$$R_t^{bond}=Z_{t-1}R_t \tag{1}$$

where R_t^{bond} is $M_t\times 1$, Z_{t-1} is $M_t\times N$ and R_t is $N\times 1$.

- TS factor modeling is the **joint** problem of discount **bond return curve** estimation and **cross-sectional factor** modeling.
- Standard approach for modeling TS factors **neglects the relationship between the factors and bond returns**, and takes a panel of pre-selected discount bond returns as given.
- This is problematic for at least two reasons:
  1. All the **potential problems** of the discount curve estimation **carry over** to the cross-sectional factor model
     - Too simplistic curve estimation, for example with the parametric Nelson–Siegel–Svensson model, neglects relevant complex shapes in the curve, and implies an erroneous and simplistic factor structure for the cross-section
  2. Cross-sections of discount bonds are often **selected arbitrarily**, which may overweight some maturities and thus not reflect the properties of the actually traded coupon bonds

A factor model corresponds to a set of $n \leq N$ basis functions that approximately span the discount bond excess return curve

$$R_t \approx \underbrace{\beta}_{N\times n} \underbrace{F_t}_{n\times 1} \tag{2}$$

Where the columns of β capture the basis functions, or the "loadings", and varies with the cross-section of maturities, not over time. F_t holds the same factors for all maturities, which vary over time.

- Basis functions in this model will be derived from a kernel that is based on economic first principles.
- Curve estimation will provide factors F_t that are portfolios of the traded assets. Indeed, if the term structure can be spanned by a sparse set of traded factors, then investors only need to trade in a few factor portfolios to replicate the full term structure.This is in contrast to most curve estimators.

Using equations (1) and (2)

$$R_t^{bond} \approx \underbrace{Z_{t-1}\beta}_{\beta_{t-1}^{bond}} F_t$$

The loadings of the coupon bonds β_{t-1}^{bond} are linear functions of their normalized discounted cash flows and therefore time-varying.

- Any asset pricing model for the TS requires an estimate of the discount bond return curve
- It is common to first estimate the discount curves d_{t-1} and d_t with parametric or non-parametric methods, and then derive the implied discount bond excess return curve.
- In contrast, they directly estimate the curve r_t that can best explain the observed coupon bond returns.

Assuming there is an observational error, or deviation from fundamental price

$$R_t^{bond} = Z_{t-1} R_t + \epsilon_t$$

The naive estimator would be to minimize the mean squared return errors of the coupon bonds for each day

$$\min_{r_t} \left\| R_t^{bond} - Z_{t-1} R_t \right\|_2^2$$

Shrinking Solution

Discount Bond Excess Return Curve Estimation

- Number of observed return M_t is substantially smaller than the number of cash flow dates N
  - On a typical day, M_t is around 300 (observable bond returns) and N is around 3650 (non-observable discount bond returns)
- Impose **regularization** to reduce the dimensionality of the parameter space
- Choose **class of functions** for the discount bond excess return curve
- These choices will be based in the principle of **smoothness**
  - Smooth discount bond excess curves imply that bonds with similar cash flows and maturities have similar returns.
  - Large sudden changes along the return curve imply trading strategies with **extreme payoffs**, which are likely infeasible out-of-sample

The authors regularize the problem by the selecting the optimal level of smoothness, measured by the weighted average of the second derivative

$$\|h\|_{\mathcal{H}_\alpha}^2 = \int_0^\infty h''(x)^2 \, e^{\alpha x} dx$$

The function $e^{\alpha x}$ with maturity weight $\alpha > 0$ allows the smoothness to be maturity dependent. $\mathcal{H}_\alpha$ is the space of all twice weakly differentiable functions.

The fundamental problem of estimating the discount bond excess return curve is a trade-off between fitting the returns of observed assets and rewarding its smoothness

$$\min_{r_t \in \mathcal{H}_\alpha} \left\| R_t^{bond} - Z_{t-1} R_t \right\|_2^2 + \lambda \left\| r_t \right\|_{\mathcal{H}_\alpha}^2$$

R_t is r_t evaluated at the cash flow dates. Parameters α and λ are selected via cross-validation.

Authors provide a closed form solution

$$\hat{R}_t = \beta \hat{F}_t \qquad \beta = VS^{\frac{1}{2}} \qquad K = VSV'$$

where K is the evaluated kernel from the previous problem. $\hat{F}_t$ is the solution to

$$\min_{F_t \in \mathbb{R}^n} \left\| R_t^{bond} - \beta_{t-1}^{bond} F_t \right\|_2^2 + \lambda \left\| F_t \right\|_2^2 \qquad \beta_{t-1}^{bond} = Z_{t-1}\beta$$

which is

$$\hat{F}_t = \omega_{t-1} R_t^{bond} \qquad \omega_{t-1} = \left(\left(\beta_{t-1}^{bond} \right)' \beta_{t-1}^{bond} + \lambda M_t I_N \right)^{-1} \left(\beta_{t-1}^{bond} \right)'$$

Linearity of the solution implies tradeability, allowing to replicate and hedge any default free fixed income security. This version is called the **"Full KR"** model

To make sure the solution has low rank

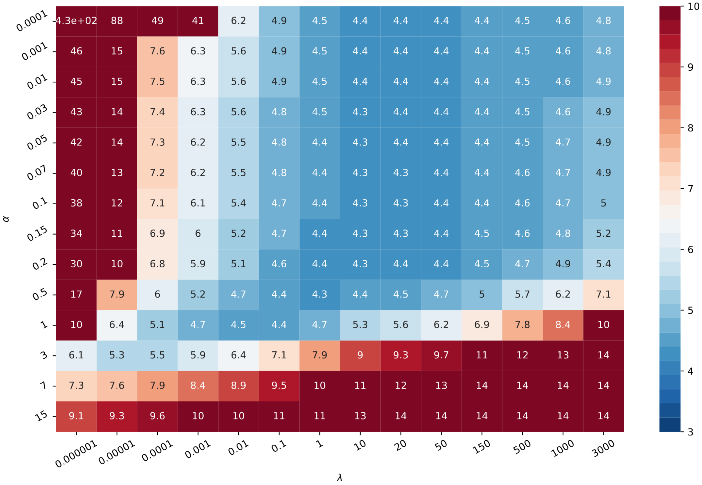
$$\min_{F_t \in \mathbb{R}^n} \left\| R_t^{bond} - \beta_{t-1}^{bond} F_t \right\|_2^2 + \lambda \left\| F_t \right\|_2^2 + \lambda_1 \left\| F_t \right\|_1$$

There is a one-to-one mapping between λ_1 and the number of selected factors n. This version is called "KR-n" model.

- CUSIP-level coupon-bearing Treasury bond data from the CRSP Treasuries Time Series
- Main sample are daily observations of securities maturing within 10 years
- The sample period is from June 1961 to December 2020
- Filter out issues that trade at a premium due to their specialness (tax benefits, option-like features, etc) or liquidity.

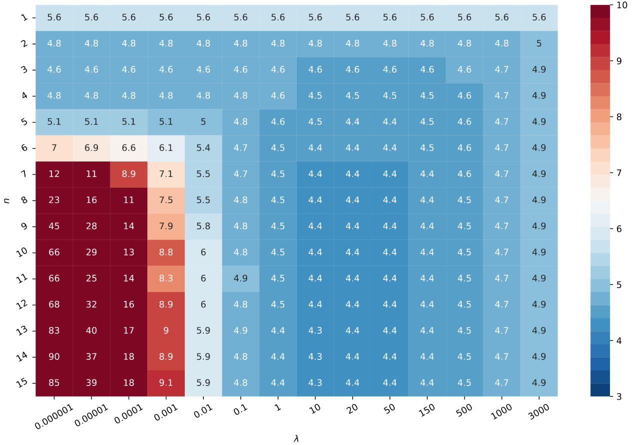
- Selecting the optimal parameter values for λ and α in the KR model
  - Error metric is the error in excess returns of observed Treasury bonds, which can be interpreted as a hedging error
  - **Cross-sectional leave-one-out cross-validation**: in a cross section, each security is left out only once, and the remaining securities form a training set, on which we fit the model. Out-of-sample performance is then evaluated on the single security being left out, and results are averaged across all securities in the cross-section to get out-of-sample performance

Figure 1: Cross-Validation Excess Return RMSE for λ and α



This figure shows the cross-validation excess return RMSE of the full KR model in basis points (BPS) as a function of the smoothness parameter λ and the maturity weight α. We apply cross-sectional leave-one-out cross-validation, this means on each day, each each security is left out only once for out-of-sample evaluation and the model is estimated on the remaining securities. The out-of-sample results are averaged across all securities. We use the last day of each quarter from June 1961 to December 2020 to speed up the calculation.

Figure 2: Cross-Validation Excess Return RMSE for λ and n

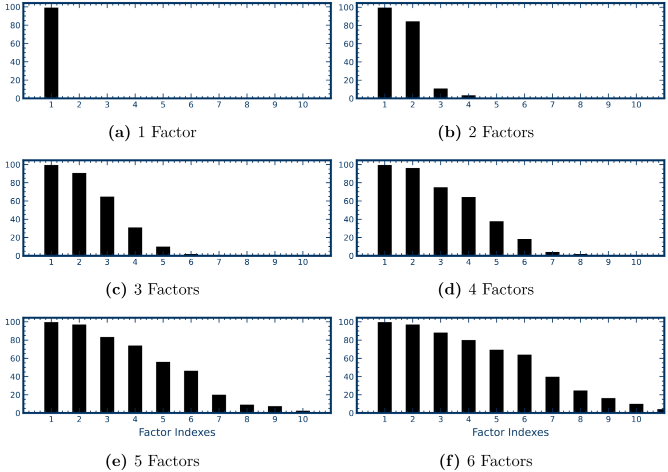


This figure shows the cross-validation excess return RMSE of the KR-n factor model in basis points (BPS) as a function of the smoothness parameter λ and the sparse number of factors n. Factors are added by the order implied by the kernel decomposition, that is, in decreasing order of the eigenvalues of the kernel matrix $\boldsymbol{K}$. We apply cross-sectional leave-one-out cross-validation, this means on each day, each each security is left out only once for out-of-sample evaluation and the model is estimated on the remaining securities. The out-of-sample results are averaged across all securities. We use the last day of each quarter from June 1961 to December 2020 to speed up the calculation.

To evaluate the importance of each factor:

- Factors are selected by the data and the LASSO, and not by order of eigenvalues of the kernel matrix
- Run LASSO on the last day of every quarter in the sample period, and aggregate results to calculate the fraction of times each factor has non-zero coefficient.

Figure 3: Frequency (%) of factors selected by LASSO



(a) 1 Factor

(b) 2 Factors

(c) 3 Factors

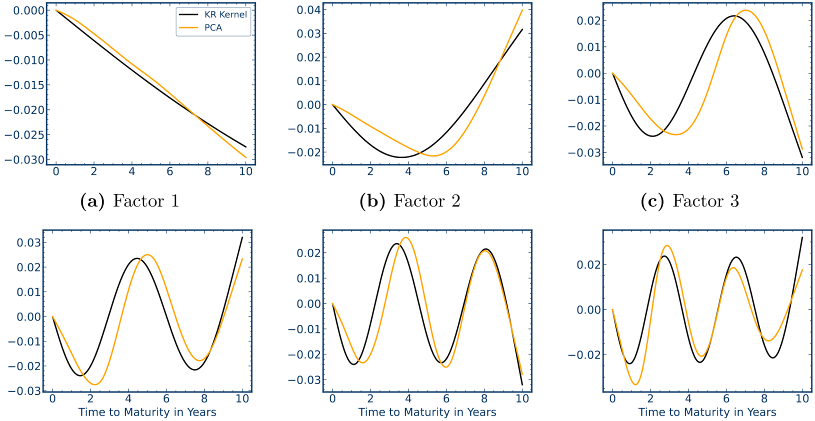
(d) 4 Factors

(e) 5 Factors

(f) 6 Factors

This figure shows fraction of times the i-th factor (horizontal axis) is selected by LASSO when at most n factors can have non-zero coefficients for n ranging from 1 to 6. We run LASSO separately on the last day of every quarter in the sample period, and aggregate results to calculate the fraction of times each factor has non-zero coefficient. We use the baseline model with parameters $\alpha = 0.05$ and $\lambda = 10$. The sample is daily data from June 1961 to December 2020.

Figure 4: Cross-sectional factor loadings of KR and PCA factors on discount bond excess returns



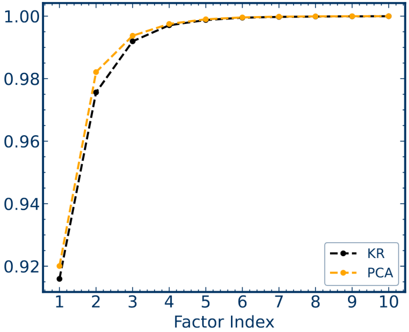
(a) Factor 1 **(b)** Factor 2 **(c)** Factor 3

(d) Factor 4 **(e)** Factor 5 **(f)** Factor 6

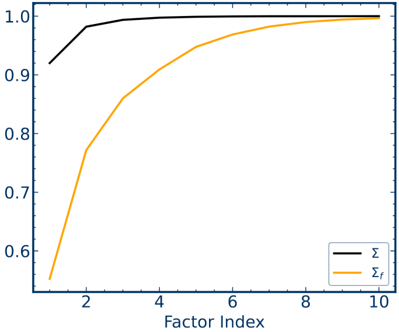
This figure shows the normalized loadings β of the first 6 KR factors and the loadings of the first 6 PCA factors on a panel of discount bond excess returns. The norm of the loadings is normalized to one. The loadings correspond to portfolio weights on discount bond excess returns to construct the KR and PCA factors. The KR loadings β equal the eigenvectors of the KR kernel matrix $\boldsymbol{K}$. The PCA loadings are the eigenvectors of the sample covariance matrix from the panel of discount bond excess returns estimated with the full KR model. The sample is daily data from June 1961 to December 2020.

- The most widely used test assets are panels of discount bond excess returns or yields
  - Common perception is that 2 to 3 factors are sufficient to explain essentially all the variation in discount bond returns
  - This perspective is misleading. The returns of discount bonds have a mechanical overlapping dependency structure which inflates the eigenvalues
  - Forward returns are a more meaningful set of test assets for measuring dependencies in the cross-section

Figure 8: Explained variation and eigenvalue structure for different number of factors



(a) Explained variation in discount bond excess returns for KR and PCA factors

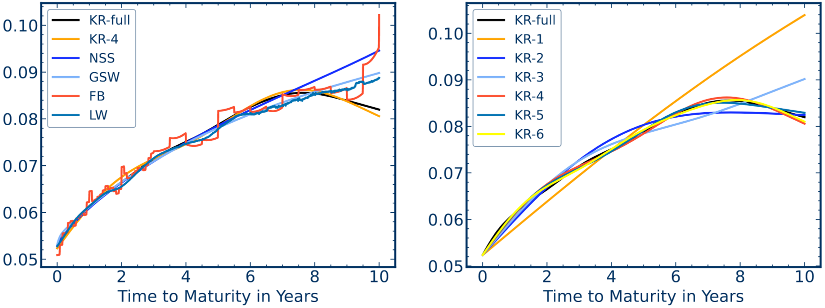


(b) Normalized cumulative sum of eigenvalues of Σ and Σ_f

This figure shows the explained variation and eigenvalue structure for different number of factors. The left subplot (a) displays the cumulative fraction of variation in the panel of discount bond excess returns of all daily maturities estimated by the full KR model, which is explained by the first 10 KR factors and the first 10 PCA factors. The right plot shows the normalized cumulative sum of eigenvalues of Σ and Σ_f. Σ is the covariance matrix of the full panel of discount bond excess returns with all daily maturities estimated with the full KR model. Σ_f is the covariance matrix of forward returns, which is the cross-sectional first difference of discount bond excess returns. Each eigenvalue is normalized by the sum of all eigenvalues. The sample is daily observations from June 1961 to December 2020 for all daily maturities.

- Estimates from different methodologies share a similar shape with a premium increasing in maturity
- However, the benchmark estimates show systematic biases and instabilities
  - Returns of FB estimates result in unstable timeseries, which is reflected by the kinks and spikes in the average returns. This makes them also problematic as an input for portfolio optimization.
  - Less precise methods GSW, NSS, and LW **systematically overestimate the average returns for 6 to 10 years** of maturity. The instability in the price estimates of FB and LW also results in more volatile estimated return time-series.

Figure 10: Expected returns of discount bonds for different estimators



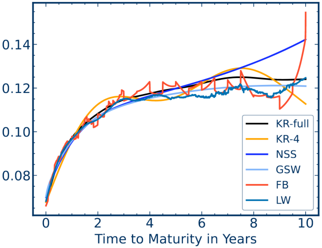
(a) Different estimators

(b) Number of KR factors

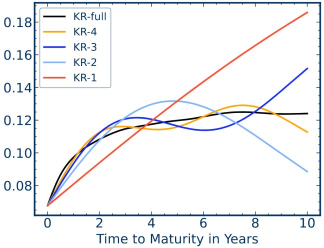
This figure shows expected one-day discount bond returns for different estimators. The left subplot shows the expected discount bond returns implied by the full KR, four-factor KR, GSW, NSS, FB, and LW estimates. The KR estimator directly estimates the term structure of returns, while GSW, NSS, FB and LW estimate yields for discount bonds. The discount bond returns are averaged over time from November 1971 to December 2013, as the 10-year rate is unavailable in the FB data set until November 1971. The right subplot shows the average discount bond returns given by KR models for different number of factors. We consider the full KR model and one to six KR factors. For better comparison we use the time period as in the left plot. Figure A.4 in the appendix shows average returns for daily data from June 1961 to December 2020 with the same results. The average one-business day returns are annualized.

- During times of booms, the shape of the term structure takes a simpler form.
  - A simple slope factor captures most of the information about the TS premium for up to 6 years of maturity
- The term structure is substantially more complex during recessions.
  - Using models with only the first three factors result in strongly biased estimates of the term structure premium
  - Including the fourth factor is key for capturing the same shape as with the full KR model
- Conditioning on level of yield spreads leads to similar results
  - Times of high yield spreads coincide with simpler shapes of the term structure premium

Panel A: Recession

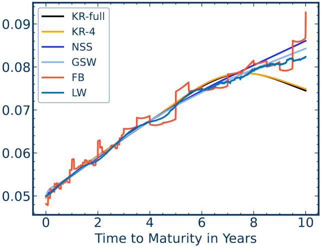


(a) Recession: Different estimators

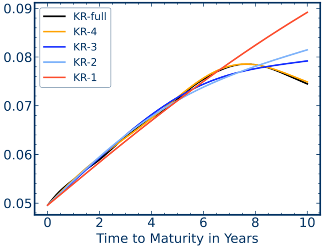


(b) Recession: Number of factors

Panel B: Boom



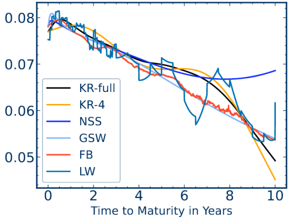
(c) Boom: Different estimators



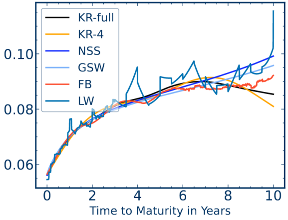
(d) Boom: Number of factors

Figure 13: Average discount bond returns conditioned on yield spreads

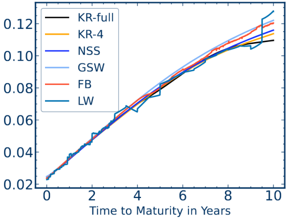
PANEL A: Different estimators



(a) low yield spread

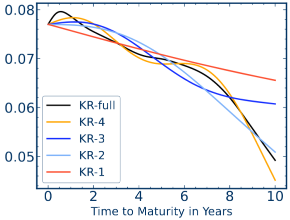


(b) medium yield spread

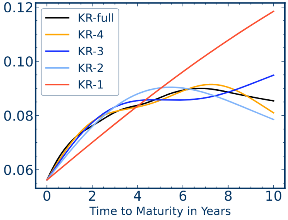


(c) high yield spread

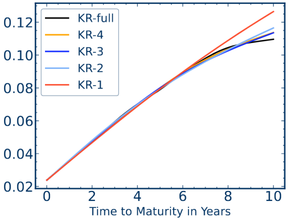
PANEL B: Number of factors



(d) low yield spread



(e) medium yield spread



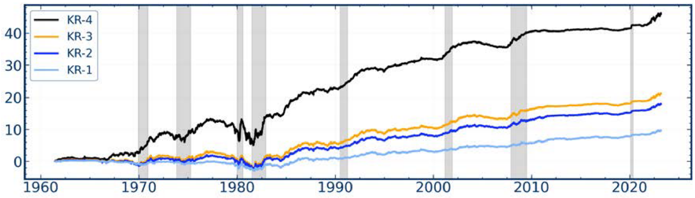
(f) high yield spread

- **Bad times require more complex curvature patterns** and alternative methods are unstable and/or omit flexible patterns
- The complex **fourth term structure factor is a hedge for bad economic times** and pays off during recessions

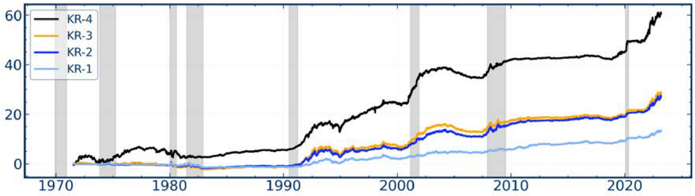
Performance of portfolios built from the factors lead to 3 insights:

1. Investment patterns for the in-sample and out-of-sample analysis are qualitatively very similar, confirming that our **results are stable and do not suffer from overfitting**
2. The largest increase in cumulative returns is due to **including the fourth factor**.
3. A **large part of this increase in performance** from including the fourth factor **occurs during recession** periods. This motivates the economic label for the fourth KR factor as a hedging factor for bad times.

Figure 14: Cumulative excess returns of implied SDF



(a) In-sample cumulative excess returns



(b) Out-of-sample cumulative excess returns

The figure shows the in-sample and out-of-sample cumulative excess returns of mean-variance efficient portfolios formed by 1 to 4 KR factors. The in-sample mean-variance efficient portfolio weights are estimated using the full panel of daily data from June 1961 to December 2020. The out-of-sample mean-variance efficient portfolio weights are estimated on a 10-year rolling window and are re-balanced monthly. Shaded areas indicate recession periods.

Conclusions

- Introduce a conditional factor model for the term structure, which **unifies non-parametric curve estimation with cross-sectional asset pricing**.
  - Learns the return curve directly from observed returns of Treasury bonds
  - This curve lies in a reproducing kernel Hilbert space that optimally trades off smoothness against return fitting.
- **Low dimensional factor model** arises.
  - First four estimated factors are investable portfolios , which replicate the full term structure and are sufficient to hedge against interest rate changes.
  - Factors are essentially the same as PCA factors, but do not rely on time series to be estimated and have an economic motivation behind them
- Changes in **market complexity** affect the term structure premium.
  - The "complex" fourth term structure factor is a hedge for bad economic times and pays off during recessions.

… and opinions

- **Positives**
  - Computable everyday, without having to pre-select maturities. Just use all the bonds available.
  - No need for a full time series. Fit happens by the trade off between cross sectional returns and smoothness
- **Negatives**
  - Very complex methodology that in the end gets similar results and interpretations as PCA
  - Focused on the new methodology, no new empirical insight.