Large-Scale Dynamic Predictive Regressions

Daniele Bianchi

University of Warwick
Warwick Business School

Ken McAlinn

University of Chicago
Booth School of Business

10th ECB Workshop on Forecasting Techniques

Introduction

Set the Stage

- The goal is to predict some (univariate) quantity y based on a set of predictors $\boldsymbol{x}$.
- A canonical approach is

$$y_t = \beta' \boldsymbol{x}_t + \epsilon_t, \qquad \epsilon_t \sim N(0, \nu)$$

- The dimension of $\boldsymbol{x}_t$ could be large, possibly "too large".
- The out-of-sample performance of standard techniques such as OLS, MLE, or Bayesian inference with uninformative priors, tends to deteriorate as the dimensionality of the data increases, i.e., curse of dimensionality.

Set the Stage

- The goal is to predict some (univariate) quantity y based on a set of predictors $\boldsymbol{x}$.
- A canonical approach is

$$y_t = \boldsymbol{\beta}'\boldsymbol{x}_t + \epsilon_t, \qquad \epsilon_t \sim N(0, \nu)$$

- The dimension of $\boldsymbol{x}_t$ could be large, possibly "too large".
- The out-of-sample performance of standard techniques such as OLS, MLE, or Bayesian inference with uninformative priors, tends to deteriorate as the dimensionality of the data increases, i.e., curse of dimensionality.

Set the Stage

- The goal is to predict some (univariate) quantity y based on a set of predictors $\boldsymbol{x}$.
- A canonical approach is

$$y_t = \beta' \boldsymbol{x}_t + \epsilon_t, \qquad \epsilon_t \sim N(0, \nu)$$

- The dimension of $\boldsymbol{x}_t$ could be large, possibly "too large".
- The out-of-sample performance of standard techniques such as OLS, MLE, or Bayesian inference with uninformative priors, tends to deteriorate as the dimensionality of the data increases, i.e., curse of dimensionality.

Set the Stage

- The goal is to predict some (univariate) quantity y based on a set of predictors $\boldsymbol{x}$.

- A canonical approach is

$$y_t = \boldsymbol{\beta}'\boldsymbol{x}_t + \epsilon_t, \qquad \epsilon_t \sim N(0, \nu)$$

- The dimension of $\boldsymbol{x}_t$ could be large, possibly "too large".

- The out-of-sample performance of standard techniques such as OLS, MLE, or Bayesian inference with uninformative priors, tends to deteriorate as the dimensionality of the data increases, i.e., curse of dimensionality.

Set the Stage (Cont'd)

- Two modeling frameworks:
  - **Regularization/Shrinkage**: selection of few variables with the highest explanatory power (e.g., LASSO, SSVS, etc.).
  - **Factor** modeling: few common components capture the statistical features of a large set of predictors (e.g., PCA, etc.).
- By lowering the model complexity both approaches introduce some form of bias, i.e., bias-variance tradeoff;
  - LASSO/Shrinkage: increasing shrinkage increases bias (cross-validation to sub-optimally solve bias-variance trade-off).
  - PCA/Factor models: fewer factors introduce bias (number of factors chosen, e.g., via BIC, to balance bias and variance).
- Both approaches have pros and cons, e.g. interpretability vs. forecasting performance/parsimony.

Set the Stage (Cont'd)

- Two modeling frameworks:
  - **Regularization/Shrinkage**: selection of few variables with the highest explanatory power (e.g., LASSO, SSVS, etc.).
  - **Factor** modeling: few common components capture the statistical features of a large set of predictors (e.g., PCA, etc.).
- By lowering the model complexity both approaches introduce some form of bias, i.e., bias-variance tradeoff;
  - LASSO/Shrinkage: increasing shrinkage increases bias (cross-validation to sub-optimally solve bias-variance trade-off).
  - PCA/Factor models: fewer factors introduce bias (number of factors chosen, e.g., via BIC, to balance bias and variance).
- Both approaches have pros and cons, e.g. interpretability vs. forecasting performance/parsimony.

Set the Stage (Cont'd)

- Two modeling frameworks:
  - **Regularization/Shrinkage**: selection of few variables with the highest explanatory power (e.g., LASSO, SSVS, etc.).
  - **Factor** modeling: few common components capture the statistical features of a large set of predictors (e.g., PCA, etc.).
- By lowering the model complexity both approaches introduce some form of bias, i.e., bias-variance tradeoff;
  - LASSO/Shrinkage: increasing shrinkage increases bias (cross-validation to sub-optimally solve bias-variance trade-off).
  - PCA/Factor models: fewer factors introduce bias (number of factors chosen, e.g., via BIC, to balance bias and variance).
- Both approaches have pros and cons, e.g. interpretability vs. forecasting performance/parsimony.

This Paper

- We partition $\boldsymbol{x}_t$ into smaller $j = 1 : J$ groups

$$y_t = \boldsymbol{\beta}_1' \boldsymbol{x}_{t,1} + ... + \boldsymbol{\beta}_j' \boldsymbol{x}_{t,j} + ... + \boldsymbol{\beta}'_J \boldsymbol{x}_{t,J} + \epsilon_t, \quad \epsilon_t \sim N(0, \nu)$$

- Two-step **Decouple-Recouple** dynamic predictive strategy.
  - Step 1: **Decouple** the regression into J smaller predictive regression models, i.e.,

$$y_t = \boldsymbol{\beta}_j' \boldsymbol{x}_{t,j} + \epsilon_{t,j}, \qquad \epsilon_{t,j} \sim N(0, \nu_j) \quad \longrightarrow \quad p(y_{t+k}|\mathcal{A}_j)$$

  - Step 2: **Recouple** each $p(y_{t+k}|\mathcal{A}_j)$ by sequentially learn the aggregate bias and cross-density latent interdependencies.
- **Idea:** The decoupling step maintain a **low variance** while the recoupling step sequentially **learn** and **adjust** for **the bias**.

This Paper

- We partition $\boldsymbol{x}_t$ into smaller $j = 1 : J$ groups

$$y_t = \boldsymbol{\beta}_1' \boldsymbol{x}_{t,1} + ... + \boldsymbol{\beta}_j' \boldsymbol{x}_{t,j} + ... + \boldsymbol{\beta}'_J \boldsymbol{x}_{t,J} + \epsilon_t, \quad \epsilon_t \sim N(0, \nu)$$

- Two-step **Decouple-Recouple** dynamic predictive strategy.
  - Step 1: **Decouple** the regression into J smaller predictive regression models, i.e.,

$$y_t = \boldsymbol{\beta}_j' \boldsymbol{x}_{t,j} + \epsilon_{t,j}, \qquad \epsilon_{t,j} \sim N(0, \nu_j) \quad \longrightarrow \quad p(y_{t+k}|\mathcal{A}_j)$$

  - Step 2: **Recouple** each $p(y_{t+k}|\mathcal{A}_j)$ by sequentially learn the aggregate bias and cross-density latent interdependencies.

- **Idea:** The decoupling step maintain a **low variance** while the recoupling step sequentially **learn** and **adjust** for **the bias**.

This Paper

- We partition $\boldsymbol{x}_t$ into smaller $j = 1 : J$ groups

$$y_t = \boldsymbol{\beta}_1' \boldsymbol{x}_{t,1} + ... + \boldsymbol{\beta}_j' \boldsymbol{x}_{t,j} + ... + \boldsymbol{\beta}'_J \boldsymbol{x}_{t,J} + \epsilon_t, \quad \epsilon_t \sim N(0, \nu)$$

- Two-step **Decouple-Recouple** dynamic predictive strategy.
  - Step 1: **Decouple** the regression into J smaller predictive regression models, i.e.,

$$y_t = \boldsymbol{\beta}_j' \boldsymbol{x}_{t,j} + \epsilon_{t,j}, \qquad \epsilon_{t,j} \sim N(0, \nu_j) \quad \longrightarrow \quad p(y_{t+k}|\mathcal{A}_j)$$

  - Step 2: **Recouple** each $p(y_{t+k}|\mathcal{A}_j)$ by sequentially learn the aggregate bias and cross-density latent interdependencies.

- **Idea:** The decoupling step maintain a **low variance** while the recoupling step sequentially **learn** and **adjust** for **the bias**.

This Paper (Cont'd)

Empirical Setting:

- Test the decouple-recouple strategy (DRS) on two datasets:
  - Macroeconomic data, i.e., forecasting inflation in U.S.
  - Financial data, i.e. forecasting industry-specific stock returns.
- The $j = 1, ..., J$, sub-models are based on economic rationale, although *the procedure is more general*.
- We compare our DRS against: dynamic BMA, linear pooling with equal weights, LASSO-type regularization, dynamic PCA, historical mean, and the predictive densities $\forall j = 1, ..., J$.

Main result:

- DRS scheme outperforms all competing strategies, both in statistical and economic terms + keeps some interpretability.

This Paper (Cont'd)

Empirical Setting:

- Test the decouple-recouple strategy (DRS) on two datasets:
  - Macroeconomic data, i.e., forecasting inflation in U.S.
  - Financial data, i.e. forecasting industry-specific stock returns.
- The $j = 1, ..., J$, sub-models are based on economic rationale, although *the procedure is more general*.
- We compare our DRS against: dynamic BMA, linear pooling with equal weights, LASSO-type regularization, dynamic PCA, historical mean, and the predictive densities $\forall j = 1, ..., J$.

Main result:

- DRS scheme outperforms all competing strategies, both in statistical and economic terms + keeps some interpretability.

This Paper (Cont'd)

Empirical Setting:

- Test the decouple-recouple strategy (DRS) on two datasets:
  - Macroeconomic data, i.e., forecasting inflation in U.S.
  - Financial data, i.e. forecasting industry-specific stock returns.
- The $j = 1, ..., J$, sub-models are based on economic rationale, although *the procedure is more general*.
- We compare our DRS against: dynamic BMA, linear pooling with equal weights, LASSO-type regularization, dynamic PCA, historical mean, and the predictive densities $\forall j = 1, ..., J$.

Main result:

- DRS scheme outperforms all competing strategies, both in statistical and economic terms + keeps some interpretability.

This Paper (Cont'd)

Empirical Setting:

- Test the decouple-recouple strategy (DRS) on two datasets:
  - Macroeconomic data, i.e., forecasting inflation in U.S.
  - Financial data, i.e. forecasting industry-specific stock returns.
- The $j = 1, \ldots, J$, sub-models are based on economic rationale, although *the procedure is more general*.
- We compare our DRS against: dynamic BMA, linear pooling with equal weights, LASSO-type regularization, dynamic PCA, historical mean, and the predictive densities $\forall j = 1, \ldots, J$.

Main result:

- DRS scheme outperforms all competing strategies, both in statistical and economic terms + keeps some interpretability.

Some Reference Literature

A non-exhaustive list:

1. Pooling Agents' Opinions and Bayesian Predictive Synthesis:
(e.g., West and Crosse (1992), West (1992), McAlinn and West (2018), McAlinn, Aastveit, Nakajima and West (2018), etc.)

2. Decouple-recouple in multi-variate time series analysis:
(e.g., Gruber and West (2017), Chen et al. (2017), etc.)

3. Dense modeling, regularization and sparsity:
(e.g., De Mol, Giannone and Reichlin (2008), Diebold and Shin (2017), Giannone, Lenza, and Primiceri (2017), George and Ročková (2016), Ročková (2018), etc.)

4. Forecasts combination and calibration:
(e.g., Timmermann (2004), Gneiting, Balabdaoui, and Raftery (2007), Amisano and Geweke (2011), Billio, Casarin, Ravazzolo and van Dijk (2013), Casarin, Grassi, Ravazzolo and van Dijk (2016))

Some Reference Literature

A non-exhaustive list:

1. Pooling Agents' Opinions and Bayesian Predictive Synthesis: (e.g., West and Crosse (1992), West (1992), McAlinn and West (2018), McAlinn, Aastveit, Nakajima and West (2018), etc.)
2. Decouple-recouple in multi-variate time series analysis: (e.g., Gruber and West (2017), Chen et al. (2017), etc.)
3. Dense modeling, regularization and sparsity: (e.g., De Mol, Giannone and Reichlin (2008), Diebold and Shin (2017), Giannone, Lenza, and Primiceri (2017), George and Ročková (2016), Ročková (2018), etc.)
4. Forecasts combination and calibration: (e.g., Timmermann (2004), Gneiting, Balabdaoui, and Raftery (2007), Amisano and Geweke (2011), Billio, Casarin, Ravazzolo and van Dijk (2013), Casarin, Grassi, Ravazzolo and van Dijk (2016))

Some Reference Literature

A non-exhaustive list:

1. Pooling Agents' Opinions and Bayesian Predictive Synthesis:
   (e.g., West and Crosse (1992), West (1992), McAlinn and West (2018), McAlinn, Aastveit, Nakajima and West (2018), etc.)

2. Decouple-recouple in multi-variate time series analysis:
   (e.g., Gruber and West (2017), Chen et al. (2017), etc.)

3. Dense modeling, regularization and sparsity:
   (e.g., De Mol, Giannone and Reichlin (2008), Diebold and Shin (2017), Giannone, Lenza, and Primiceri (2017), George and Ročková (2016), Ročková (2018), etc.)

4. Forecasts combination and calibration:
   (e.g., Timmermann (2004), Gneiting, Balabdaoui, and Raftery (2007), Amisano and Geweke (2011), Billio, Casarin, Ravazzolo and van Dijk (2013), Casarin, Grassi, Ravazzolo and van Dijk (2016))

Some Reference Literature

A non-exhaustive list:

1. Pooling Agents' Opinions and Bayesian Predictive Synthesis:
   (e.g., West and Crosse (1992), West (1992), McAlinn and West (2018), McAlinn, Aastveit, Nakajima and West (2018), etc.)

2. Decouple-recouple in multi-variate time series analysis:
   (e.g., Gruber and West (2017), Chen et al. (2017), etc.)

3. Dense modeling, regularization and sparsity:
   (e.g., De Mol, Giannone and Reichlin (2008), Diebold and Shin (2017), Giannone, Lenza, and Primiceri (2017), George and Ročková (2016), Ročková (2018), etc.)

4. Forecasts combination and calibration:
   (e.g., Timmermann (2004), Gneiting, Balabdaoui, and Raftery (2007), Amisano and Geweke (2011), Billio, Casarin, Ravazzolo and van Dijk (2013), Casarin, Grassi, Ravazzolo and van Dijk (2016))

Modeling Framework

Background

- The de-coupling step provides "prior" information in the form of a set of predictive densities $\forall j = 1, \ldots, J$ groups.

$$\mathcal{H} = (h_1(\cdot), \ldots, h_J(\cdot)) \qquad \text{with} \qquad h_j(x_j) = p(y|\mathcal{A}_j)$$

- The re-coupled predictive density $p(y|\mathcal{H})$ effectively represents a posterior distribution.
- West and Crosse (1992) and West (1992) show that multiple prior information can be "synthesized" as

$$p(y|\mathcal{H}) = \int \alpha(y|\boldsymbol{x}) h(\boldsymbol{x}) d\boldsymbol{x} \qquad \text{where} \qquad h(\boldsymbol{x}) = \prod_{j=1}^{J} h_j(x_j)$$

where $\alpha(y|\boldsymbol{x})$ is a conditional density function that "aggregates" the group-specific predictive densities.

Background

- The de-coupling step provides “prior” information in the form of a set of predictive densities $\forall j = 1, \ldots, J$ groups.

$$\mathcal{H} = \left(h_1\left(\cdot\right), \ldots, h_J\left(\cdot\right)\right) \qquad \text{with} \qquad h_j\left(x_j\right) = p\left(y|\mathcal{A}_j\right)$$

- The re-coupled predictive density $p\left(y|\mathcal{H}\right)$ effectively represents a posterior distribution.
- West and Crosse (1992) and West (1992) show that multiple prior information can be “synthesized” as

$$p(y|\mathcal{H}) = \int \alpha(y|\boldsymbol{x})h(\boldsymbol{x})d\boldsymbol{x} \qquad \text{where} \qquad h(\boldsymbol{x}) = \prod_{j=1}^{J} h_j\left(x_j\right)$$

where $\alpha(y|\boldsymbol{x})$ is a conditional density function that “aggregates” the group-specific predictive densities.

Background

- The de-coupling step provides “prior” information in the form of a set of predictive densities $\forall j = 1, \ldots, J$ groups.

$$\mathcal{H} = \left(h_1\left(\cdot\right), \ldots, h_J\left(\cdot\right)\right) \qquad \text{with} \qquad h_j\left(x_j\right) = p\left(y|\mathcal{A}_j\right)$$

- The re-coupled predictive density $p\left(y|\mathcal{H}\right)$ effectively represents a posterior distribution.
- West and Crosse (1992) and West (1992) show that multiple prior information can be “synthesized” as

$$p(y|\mathcal{H}) = \int \alpha(y|\boldsymbol{x})h(\boldsymbol{x})d\boldsymbol{x} \qquad \text{where} \qquad h(\boldsymbol{x}) = \prod_{j=1}^{J} h_j\left(x_j\right)$$

where $\alpha(y|\boldsymbol{x})$ is a conditional density function that “aggregates” the group-specific predictive densities.

Background (Cont'd)

- We follow McAlinn and West (2018) and McAlinn et al. (2018) and assume a stochastic linear synthesis function

$$\alpha_t(y_t|\boldsymbol{x}_t,\ldots) = N(y_t|\boldsymbol{F}_t'\boldsymbol{\theta}_t, v_t) \quad \text{with} \quad \boldsymbol{F}_t = (1, h_t(\boldsymbol{x}_t)')'$$

- Time variation in the synthesis coefficients $\boldsymbol{\theta}_t$ is specified as

$$y_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim N(0, v_t),$$
$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(0, v_t\boldsymbol{W}_t),$$

where the dynamics of $\boldsymbol{W}_t$ is imposed by a discount factor.

Background (Cont'd)

- We follow McAlinn and West (2018) and McAlinn et al. (2018) and assume a stochastic linear synthesis function

$$\alpha_t(y_t|\boldsymbol{x}_t, \ldots) = N(y_t|\boldsymbol{F}_t'\boldsymbol{\theta}_t, v_t) \quad \text{with} \quad \boldsymbol{F}_t = (1, h_t(\boldsymbol{x}_t)')'$$

- Time variation in the synthesis coefficients $\boldsymbol{\theta}_t$ is specified as

$$\begin{aligned} y_t &= \boldsymbol{F}_t'\boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim N(0, v_t), \\ \boldsymbol{\theta}_t &= \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(0, v_t\boldsymbol{W}_t), \end{aligned}$$

where the dynamics of $\boldsymbol{W}_t$ is imposed by a discount factor.

Background (Cont'd)

Estimation strategy: Two-components block Gibbs sampler.

Block 1: Conditional on $h_t(\boldsymbol{x}_t)$, the parameters $\boldsymbol{\theta}_t, v_t, \boldsymbol{W}_t$ are sampled by using conjugate updates, e.g., FFBS.

Block 2: Conditional on $\theta_t, v_t, \boldsymbol{W}_t$ we update the each $j = 1, \ldots, J$ group predictions as

$$p(\boldsymbol{x}_t | y_t, \mathcal{H}_t, \ldots) \propto N\left(y_t | \mathbf{F}_t' \theta_t, v_t\right) \prod_{j=1:J} h_{tj}(x_{tj}) dx_{tj}$$

e.g., with $h_{tj}(x_{tj})$ Student-t density, we use a scale mixture of normals which allows for conjugate updating.

Key: *The predictive densities for each $j = 1, ..., J$ are updated by explicitly considering latent interdependencies and aggregate bias.*

Background (Cont'd)

Estimation strategy: Two-components block Gibbs sampler.

Block 1: Conditional on $h_t\,(\boldsymbol{x}_t)$, the parameters $\boldsymbol{\theta}_t, v_t, \boldsymbol{W}_t$ are sampled by using conjugate updates, e.g., FFBS.

Block 2: Conditional on $\boldsymbol{\theta}_t, v_t, \boldsymbol{W}_t$ we update the each $j = 1, \ldots, J$ group predictions as

$$p(\boldsymbol{x}_t | y_t, \mathcal{H}_t, \ldots) \propto N\left(y_t | \mathbf{F}_t' \boldsymbol{\theta}_t, v_t\right) \prod_{j=1:J} h_{tj}(x_{tj}) dx_{tj}$$

e.g., with $h_{tj}(x_{tj})$ Student-t density, we use a scale mixture of normals which allows for conjugate updating.

Key: *The predictive densities for each $j = 1, ..., J$ are updated by explicitly considering latent interdependencies and aggregate bias.*

Background (Cont'd)

Estimation strategy: Two-components block Gibbs sampler.

Block 1: Conditional on $h_t\,(\boldsymbol{x}_t)$, the parameters $\boldsymbol{\theta}_t, v_t, \boldsymbol{W}_t$ are sampled by using conjugate updates, e.g., FFBS.

Block 2: Conditional on $\boldsymbol{\theta}_t, v_t, \boldsymbol{W}_t$ we update the each $j = 1, \ldots, J$ group predictions as

$$p(\boldsymbol{x}_t | y_t, \mathcal{H}_t, \ldots) \propto N\left(y_t | \mathbf{F}_t' \boldsymbol{\theta}_t, v_t\right) \prod_{j=1:J} h_{tj}(x_{tj}) dx_{tj}$$

e.g., with $h_{tj}(x_{tj})$ Student-t density, we use a scale mixture of normals which allows for conjugate updating.

Key: <u>*The predictive densities for each $j = 1, ..., J$ are updated by explicitly considering latent interdependencies and aggregate bias.*</u>

Empirical Study

Empirical Study: Setting

- Two applications: (1) Macro data, i.e., forecasting inflation, and (2) Financial data, i.e., forecasting industry stock returns.
- We compare our strategy against:
  - Bayesian Model Averaging (BMA) of $p(y|\mathcal{A}_j)$.
  - Factor model, k recursively chosen by IC.
  - LASSO, leave-one-out cross-validation.
  - Equal-weighted linear pooling of predictive densities.
  - Historical average (for the finance application).
  - Predictions from single subsets $p(y|\mathcal{A}_j)$, $j = 1, ..., J$.

Empirical Study: Setting

- Two applications: (1) Macro data, i.e., forecasting inflation, and (2) Financial data, i.e., forecasting industry stock returns.
- We compare our strategy against:
  - Bayesian Model Averaging (BMA) of $p(y|\mathcal{A}_j)$.
  - Factor model, k recursively chosen by IC.
  - LASSO, leave-one-out cross-validation.
  - Equal-weighted linear pooling of predictive densities.
  - Historical average (for the finance application).
  - Predictions from single subsets $p(y|\mathcal{A}_j)$, $j = 1, ..., J$.

Empirical Study: Setting

- Two applications: (1) Macro data, i.e., forecasting inflation, and (2) Financial data, i.e., forecasting industry stock returns.
- We compare our strategy against:
  - Bayesian Model Averaging (BMA) of $p(y|\mathcal{A}_j)$.
  - Factor model, k recursively chosen by IC.
  - LASSO, leave-one-out cross-validation.
  - Equal-weighted linear pooling of predictive densities.
  - Historical average (for the finance application).
  - Predictions from single subsets $p(y|\mathcal{A}_j)$, $j = 1, ..., J$.

Empirical Study: Setting

- Two applications: (1) Macro data, i.e., forecasting inflation, and (2) Financial data, i.e., forecasting industry stock returns.
- We compare our strategy against:
  - Bayesian Model Averaging (BMA) of $p\left(y|\mathcal{A}_j\right)$.
  - Factor model, k recursively chosen by IC.
  - LASSO, leave-one-out cross-validation.
  - Equal-weighted linear pooling of predictive densities.
  - Historical average (for the finance application).
  - Predictions from single subsets $p\left(y|\mathcal{A}_j\right)$, $j = 1, ..., J$.

Empirical Study: Setting

- Two applications: (1) Macro data, i.e., forecasting inflation, and (2) Financial data, i.e., forecasting industry stock returns.
- We compare our strategy against:
  - Bayesian Model Averaging (BMA) of $p(y|\mathcal{A}_j)$.
  - Factor model, k recursively chosen by IC.
  - LASSO, leave-one-out cross-validation.
  - Equal-weighted linear pooling of predictive densities.
  - Historical average (for the finance application).
  - Predictions from single subsets $p(y|\mathcal{A}_j)$, $j = 1, ..., J$.

Empirical Study: Setting

- Two applications: (1) Macro data, i.e., forecasting inflation, and (2) Financial data, i.e., forecasting industry stock returns.
- We compare our strategy against:
  - Bayesian Model Averaging (BMA) of $p(y|\mathcal{A}_j)$.
  - Factor model, k recursively chosen by IC.
  - LASSO, leave-one-out cross-validation.
  - Equal-weighted linear pooling of predictive densities.
  - Historical average (for the finance application).
  - Predictions from single subsets $p(y|\mathcal{A}_j)$, $j = 1, ..., J$.

Empirical Study: Setting

- Two applications: (1) Macro data, i.e., forecasting inflation, and (2) Financial data, i.e., forecasting industry stock returns.
- We compare our strategy against:
  - Bayesian Model Averaging (BMA) of $p\left(y|\mathcal{A}_j\right)$.
  - Factor model, k recursively chosen by IC.
  - LASSO, leave-one-out cross-validation.
  - Equal-weighted linear pooling of predictive densities.
  - Historical average (for the finance application).
  - Predictions from single subsets $p\left(y|\mathcal{A}_j\right)$, $j = 1, ..., J$.

Empirical Study: Setting (Cont'd)

- For the decouple step $p\left(y|\mathcal{A}_j\right)$ we use a dynamic linear model for each subgroup, $j = 1{:}J$,

$$
\begin{aligned}
y_t &= \boldsymbol{\beta}_{tj}' \boldsymbol{x}_{tj} + \epsilon_{tj}, \quad \epsilon_{tj} \sim N(0, \nu_{tj}), \\
\boldsymbol{\beta}_{tj} &= \boldsymbol{\beta}_{t-1,j} + \boldsymbol{u}_{tj}, \quad \boldsymbol{u}_{tj} \sim N(0, \nu_{tj} \boldsymbol{U}_{tj}),
\end{aligned}
$$

where both ν_{tj} and $\boldsymbol{U}_{tj}$ have a discount-like dynamics.

- Density forecasts are compared based on the cumulative Log Predictive Density Ratios (LPDR) that is,

$$
\mathrm{LPDR}_t(k) = \sum_{i=1}^{t} \log\{p(y_{i+k}|y_{1:i}, \mathcal{M}_s)/p(y_{i+k}|y_{1:i}, \mathcal{M}_0)\},
$$

where $\mathcal{M}_s$ denotes the competing strategy.

- N.B: For the finance application we also compare based on realized Certainty Equivalent Returns (CER)s.

Empirical Study: Setting (Cont'd)

- For the decouple step $p\,(y|\mathcal{A}_j)$ we use a dynamic linear model for each subgroup, $j = 1{:}J$,

$$
\begin{aligned}
y_t &= \boldsymbol{\beta}'_{tj}\boldsymbol{x}_{tj} + \epsilon_{tj}, \quad \epsilon_{tj} \sim N(0, \nu_{tj}), \\
\boldsymbol{\beta}_{tj} &= \boldsymbol{\beta}_{t-1,j} + \boldsymbol{u}_{tj}, \quad \boldsymbol{u}_{tj} \sim N(0, \nu_{tj}\boldsymbol{U}_{tj}),
\end{aligned}
$$

  where both ν_{tj} and $\boldsymbol{U}_{tj}$ have a discount-like dynamics.

- Density forecasts are compared based on the cumulative Log Predictive Density Ratios (LPDR) that is,

$$
\mathrm{LPDR}_t(k) = \sum_{i=1}^{t} \log\{p(y_{i+k}|y_{1:i}, \mathcal{M}_s)/p(y_{i+k}|y_{1:i}, \mathcal{M}_0)\},
$$

  where $\mathcal{M}_s$ denotes the competing strategy.

- N.B: For the finance application we also compare based on realized Certainty Equivalent Returns (CER)s.

Empirical Study: Setting (Cont'd)

- For the decouple step $p\left(y|\mathcal{A}_j\right)$ we use a dynamic linear model for each subgroup, $j = 1{:}J$,

$$y_t = \boldsymbol{\beta}'_{tj}\boldsymbol{x}_{tj} + \epsilon_{tj}, \quad \epsilon_{tj} \sim N(0, \nu_{tj}),$$
$$\boldsymbol{\beta}_{tj} = \boldsymbol{\beta}_{t-1,j} + \boldsymbol{u}_{tj}, \quad \boldsymbol{u}_{tj} \sim N(0, \nu_{tj}\boldsymbol{U}_{tj}),$$

  where both ν_{tj} and $\boldsymbol{U}_{tj}$ have a discount-like dynamics.

- Density forecasts are compared based on the cumulative Log Predictive Density Ratios (LPDR) that is,

$$\mathrm{LPDR}_t(k) = \sum_{i=1}^{t} \log\{p(y_{i+k}|y_{1:i}, \mathcal{M}_s)/p(y_{i+k}|y_{1:i}, \mathcal{M}_0)\},$$

  where $\mathcal{M}_s$ denotes the competing strategy.

- N.B: For the finance application we also compare based on realized Certainty Equivalent Returns (CER)s.

Empirical Study 1: Forecasting Inflation

Objective: Monthly forecast of the annual US inflation rate (1- and 3-month ahead) over the period 1986:01-2015:12.

- $N = 128$ macro variables classified in $J = 9$ categories:
  - Output and Income, Labor Market, Consumption, Inventories, Money and Credit, Interest and Exchange Rates, Prices, and Stock Market.
- Decoupled regressions are estimated in parallel over 1986:01-1993:06 as a training period.
- From 1993:07-2015:12 both the decouple models and the re-coupling strategies are run.
- Forecast of inflation from 1993:07-2000:12 are discarded as DRS training period.

Empirical Study 1: Forecasting Inflation

Objective: Monthly forecast of the annual US inflation rate (1- and 3-month ahead) over the period 1986:01-2015:12.

- $N = 128$ macro variables classified in $J = 9$ categories:
  - Output and Income, Labor Market, Consumption, Inventories, Money and Credit, Interest and Exchange Rates, Prices, and Stock Market.
- Decoupled regressions are estimated in parallel over 1986:01-1993:06 as a training period.
- From 1993:07-2015:12 both the decouple models and the re-coupling strategies are run.
- Forecast of inflation from 1993:07-2000:12 are discarded as DRS training period.

Empirical Study 1: Forecasting Inflation

Objective: Monthly forecast of the annual US inflation rate (1- and 3-month ahead) over the period 1986:01-2015:12.

- $N = 128$ macro variables classified in $J = 9$ categories:
  - Output and Income, Labor Market, Consumption, Inventories, Money and Credit, Interest and Exchange Rates, Prices, and Stock Market.
- Decoupled regressions are estimated in parallel over 1986:01-1993:06 as a training period.
- From 1993:07-2015:12 both the decouple models and the re-coupling strategies are run.
- Forecast of inflation from 1993:07-2000:12 are discarded as DRS training period.

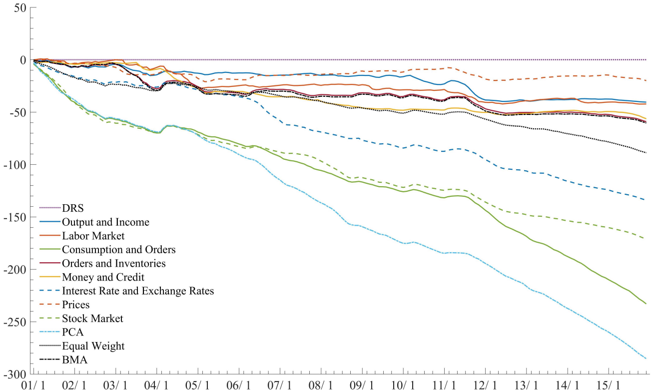
Empirical Study 1: Forecasting Inflation

Objective: Monthly forecast of the annual US inflation rate (1- and 3-month ahead) over the period 1986:01-2015:12.

- $N = 128$ macro variables classified in $J = 9$ categories:
  - Output and Income, Labor Market, Consumption, Inventories, Money and Credit, Interest and Exchange Rates, Prices, and Stock Market.
- Decoupled regressions are estimated in parallel over 1986:01-1993:06 as a training period.
- From 1993:07-2015:12 both the decouple models and the re-coupling strategies are run.
- Forecast of inflation from 1993:07-2000:12 are discarded as DRS training period.

Empirical Study 1: Forecasting Inflation

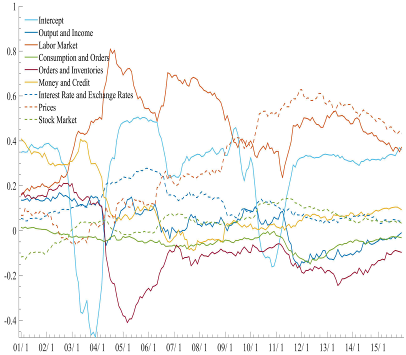
Figure: Out-of-sample log predictive density ratio



This figure shows the dynamics of the out-of-sample Log Predictive Density Ratio (LPDR) obtained for each competing specification. The sample period is 01:2001-12:2015, monthly. The objective function is the **one-step** ahead density forecast of annual inflation.

Empirical Study 1: Forecasting Inflation

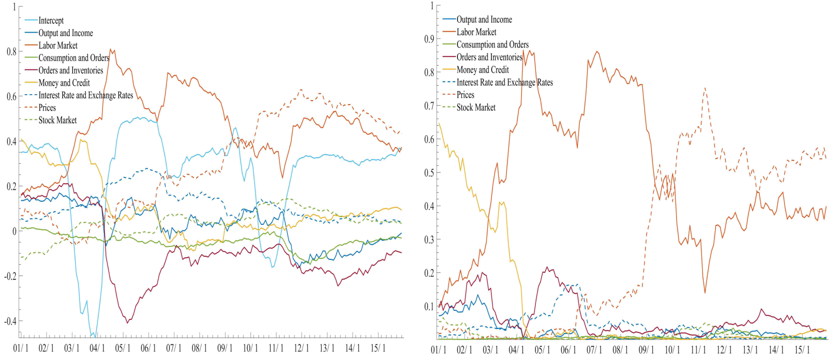
Figure: Posterior means of rescaled interdependencies.



This figure shows the latent interdependencies across groups of predictive densities used in the recoupling step. Left panel the posterior mean estimates and right panel the rescaled coefficients such that they are bounded between zero and one, and sum to one.

Empirical Study 1: Forecasting Inflation

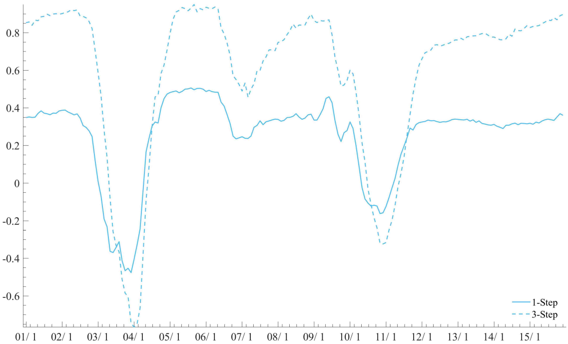
Figure: Posterior means of rescaled interdependencies.



This figure shows the latent interdependencies across groups of predictive densities used in the recoupling step. Left panel the posterior mean estimates and right panel the rescaled coefficients such that they are bounded between zero and one, and sum to one.

Empirical Study 1: Forecasting Inflation

Figure: Out-of-Sample Dynamic Predictive Bias



This figure shows the dynamics of the out-of-sample predictive bias obtained as the time-varying intercept from the recoupling step of the DRS strategy. The sample period is 01:2001-12:2015, monthly. The objective function is the one-step ahead density forecast of annual inflation.

Empirical Study 2: Forecasting Industry Stock Returns

Objective: Monthly forecast of the annual stock returns across industries over the period 1970:01-2015:12.

- $N = 92$ industry-specific and market-level predictors classified in $J = 10$ categories:
  - Value, Profitability, Capitalization, Financial Soundness, Solvency, Liquidity, Efficiency, Other, Agg. Financials, Macro.
- Industry aggregation based on the 4-digit SIC codes of firms.
- Decoupled regressions are estimated in parallel over 1970:01-1992:09 as a training period.
- From 1992:10-2015:12 both the decouple models and the re-coupling strategies are run.
- Forecast from 1992:10-2000:12 are discarded as DRS training period.

Empirical Study 2: Forecasting Industry Stock Returns

Objective: Monthly forecast of the annual stock returns across industries over the period 1970:01-2015:12.

- $N = 92$ industry-specific and market-level predictors classified in $J = 10$ categories:
  - Value, Profitability, Capitalization, Financial Soundness, Solvency, Liquidity, Efficiency, Other, Agg. Financials, Macro.
- Industry aggregation based on the 4-digit SIC codes of firms.
- Decoupled regressions are estimated in parallel over 1970:01-1992:09 as a training period.
- From 1992:10-2015:12 both the decouple models and the re-coupling strategies are run.
- Forecast from 1992:10-2000:12 are discarded as DRS training period.

Empirical Study 2: Forecasting Industry Stock Returns

Objective: Monthly forecast of the annual stock returns across industries over the period 1970:01-2015:12.

- $N = 92$ industry-specific and market-level predictors classified in $J = 10$ categories:
  - Value, Profitability, Capitalization, Financial Soundness, Solvency, Liquidity, Efficiency, Other, Agg. Financials, Macro.
- Industry aggregation based on the 4-digit SIC codes of firms.
- Decoupled regressions are estimated in parallel over 1970:01-1992:09 as a training period.
- From 1992:10-2015:12 both the decouple models and the re-coupling strategies are run.
- Forecast from 1992:10-2000:12 are discarded as DRS training period.

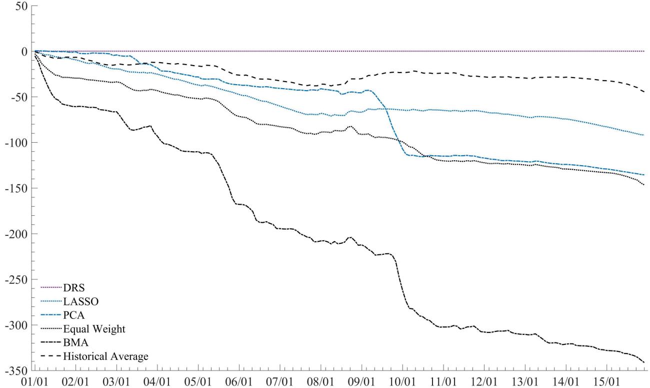
Empirical Study 2: Forecasting Industry Stock Returns

Objective: Monthly forecast of the annual stock returns across industries over the period 1970:01-2015:12.

- $N = 92$ industry-specific and market-level predictors classified in $J = 10$ categories:
  - Value, Profitability, Capitalization, Financial Soundness, Solvency, Liquidity, Efficiency, Other, Agg. Financials, Macro.
- Industry aggregation based on the 4-digit SIC codes of firms.
- Decoupled regressions are estimated in parallel over 1970:01-1992:09 as a training period.
- From 1992:10-2015:12 both the decouple models and the re-coupling strategies are run.
- Forecast from 1992:10-2000:12 are discarded as DRS training period.

Empirical Study 2: Forecasting Industry Stock Returns

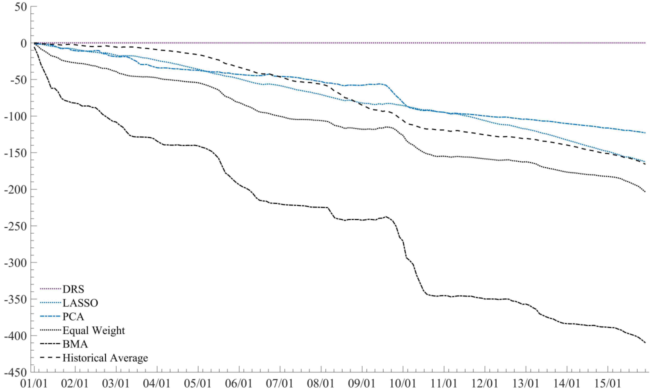
Figure: Out-of-sample LPDR for Consumer Durables



This figure shows the dynamics of the out-of-sample Log Predictive Density Ratio (LPDR) obtained from the competing specifications in addition to the historical average of the stock returns (HA). The sample period is 01:1970-12:2015, monthly.

Empirical Study 2: Forecasting Industry Stock Returns

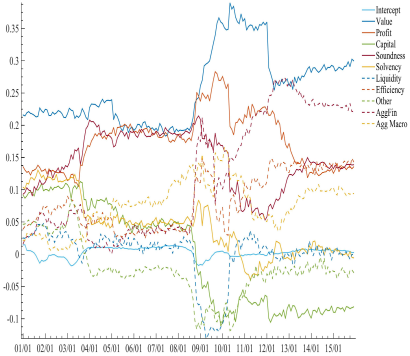
Figure: Out-of-sample LPDR for Telcm



This figure shows the dynamics of the out-of-sample Log Predictive Density Ratio (LPDR) obtained from the competing specifications in addition to the historical average of the stock returns (HA). The sample period is 01:1970-12:2015, monthly.

Empirical Study 2: Forecasting Industry Stock Returns

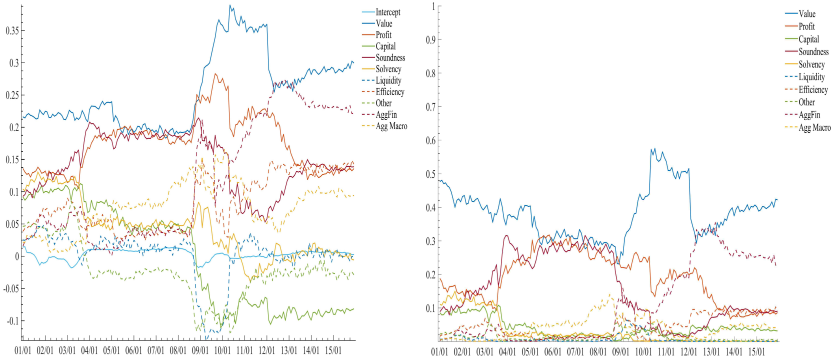
Figure: Posterior means of latent interdependencies for Manufacturing



This figure shows the one-step ahead latent interdependencies across groups of predictive densities– measured through the predictive coefficients– used in the recoupling step. The sample period is 01:1970-12:2015, monthly.

Empirical Study 2: Forecasting Industry Stock Returns

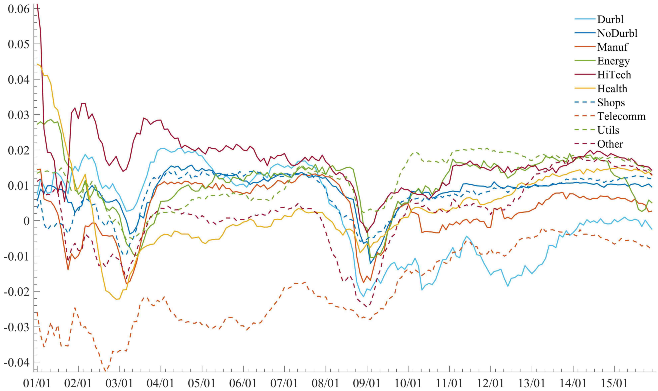
Figure: Posterior means of latent interdependencies for Manufacturing



This figure shows the one-step ahead latent interdependencies across groups of predictive densities– measured through the predictive coefficients– used in the recoupling step. The sample period is 01:1970-12:2015, monthly.

Empirical Study 2: Forecasting Industry Stock Returns

Figure: Out-of-Sample Dynamic Predictive Bias



This figure shows the dynamics of the out-of-sample predictive bias obtained as the time-varying intercept from the recoupling step of the DRS strategy. The objective function is the one-step ahead density forecast of stock excess returns across different industries. Industry classification is based on 4-digit SIC codes.

Empirical Study 2: Economic Significance

Question: Is there any economic insight/gain in using our decouple-recouple predictive strategy?

- Representative investor who wants to allocate wealth between stocks and a risk-less asset.
- The investor has a **power utility** and **moderate risk aversion**, i.e., cares about the whole distribution of returns.
- Two scenarios:
  - Short sales are allowed (unconstrained investor).
  - Short sales are not allowed, i.e., $w_{it} \geq 0$ (constrained investor).
- Performance is evaluated based on Certainty Equivalent Returns (CERs).

Empirical Study 2: Economic Significance

Question: Is there any economic insight/gain in using our decouple-recouple predictive strategy?

- Representative investor who wants to allocate wealth between stocks and a risk-less asset.
- The investor has a **power utility** and **moderate risk aversion**, i.e., cares about the whole distribution of returns.
- Two scenarios:
  - Short sales are allowed (unconstrained investor).
  - Short sales are not allowed, i.e., $w_{it} \geq 0$ (constrained investor).
- Performance is evaluated based on Certainty Equivalent Returns (CERs).

Empirical Study 2: Economic Significance

Question: Is there any economic insight/gain in using our decouple-recouple predictive strategy?

- Representative investor who wants to allocate wealth between stocks and a risk-less asset.
- The investor has a **power utility** and **moderate risk aversion**, i.e., cares about the whole distribution of returns.
- Two scenarios:
  - Short sales are allowed (unconstrained investor).
  - Short sales are not allowed, i.e., $w_{it} \geq 0$ (constrained investor).
- Performance is evaluated based on Certainty Equivalent Returns (CERs).

Empirical Study 2: Economic Significance

Question: Is there any economic insight/gain in using our decouple-recouple predictive strategy?

- Representative investor who wants to allocate wealth between stocks and a risk-less asset.
- The investor has a **power utility** and **moderate risk aversion**, i.e., cares about the whole distribution of returns.
- Two scenarios:
  - Short sales are allowed (unconstrained investor).
  - Short sales are not allowed, i.e., $w_{it} \geq 0$ (constrained investor).
- Performance is evaluated based on Certainty Equivalent Returns (CERs).

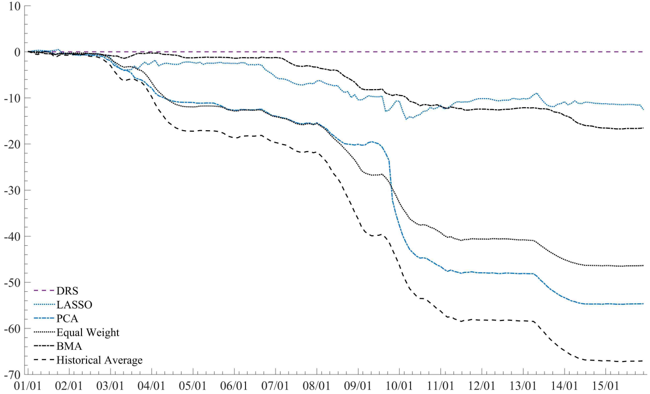
Empirical Study 2: Economic Significance

Question: Is there any economic insight/gain in using our decouple-recouple predictive strategy?

- Representative investor who wants to allocate wealth between stocks and a risk-less asset.
- The investor has a **power utility** and **moderate risk aversion**, i.e., cares about the whole distribution of returns.
- Two scenarios:
  - Short sales are allowed (unconstrained investor).
  - Short sales are not allowed, i.e., $w_{it} \geq 0$ (constrained investor).
- Performance is evaluated based on Certainty Equivalent Returns (CERs).

Empirical Study 2: Economic Significance of DRS

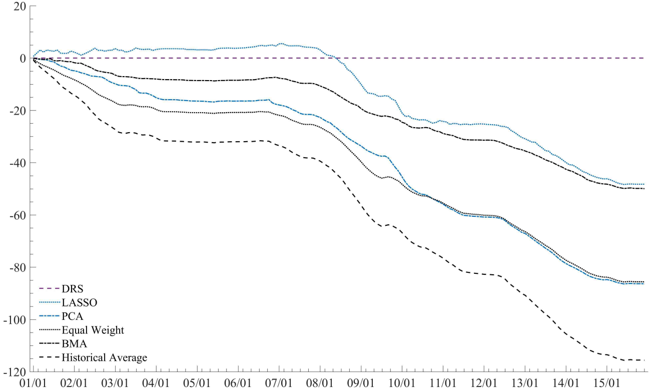
Figure: OOS cumulative CER for an **unconstrained** investor for Durables



This figure shows the dynamics of the out-of-sample Cumulative Certainty Equivalent Return (CER) for an **unconstrained** investor obtained for each of the competing model combination/shrinkage schemes. The sample period is 01:1970-12:2015, monthly.

Empirical Study 2: Economic Significance of DRS

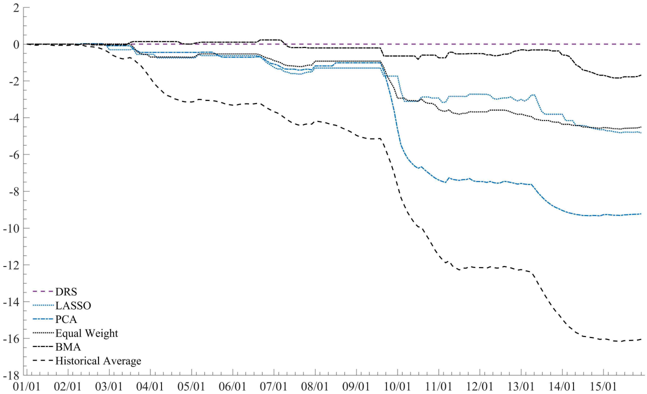
Figure: OOS cumulative CER for an **unconstrained** investor for Telcm



This figure shows the dynamics of the out-of-sample Cumulative Certainty Equivalent Return (CER) for an **unconstrained** investor obtained for each of the competing model combination/shrinkage schemes. The sample period is 01:1970-12:2015, monthly.

Empirical Study 2: Economic Significance of DRS

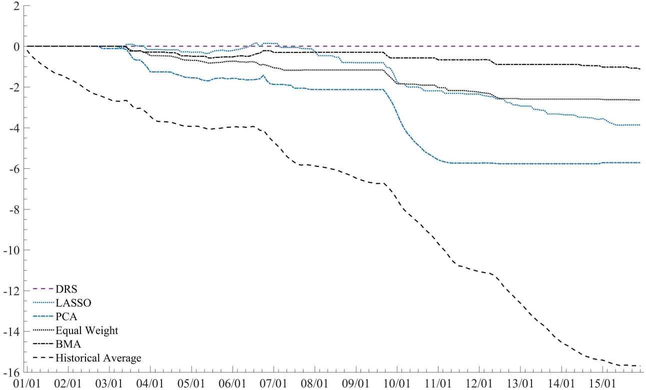
Figure: OOS cumulative CER for a **constrained** investor for Durables



This figure shows the dynamics of the out-of-sample Cumulative Certainty Equivalent Return (CER) for an **constrained** investor obtained for each of the competing model combination/shrinkage schemes. The sample period is 01:1970-12:2015, monthly.

Empirical Study 2: Economic Significance of DRS

Figure: OOS cumulative CER for a **constrained** investor for Telcm



This figure shows the dynamics of the out-of-sample Cumulative Certainty Equivalent Return (CER) for an **constrained** investor obtained for each of the competing model combination/shrinkage schemes. The sample period is 01:1970-12:2015, monthly.

Conclusions

In this paper:

- We propose an alternative decouple-recouple strategy (DRS) for predictive regressions with a (relatively) large number of covariates.
- We calibrate and implement the proposed methodology on both a macroeconomic and a finance application.
- We compare forecasts from DRS against a set of standard benchmark strategies.
- DRS emerges as the best for forecasting both the U.S. inflation rate as well as total excess returns across different industries in the U.S market.