Optimal Asset Allocation with Multivariate Bayesian Dynamic Linear Models

Jared D. Fisher[a]
with Davide Pettenuzzo[b], Carlos M. Carvalho[c]

[a]University of California, Berkeley
[b]Brandeis University
[c]University of Texas at Austin

2020 NBER-NSF Seminar on Bayesian Inference
in Econometrics and Statistics

Saturday, August 8, 2020

Introduction

- A recurring question in empirical asset pricing is whether or not asset returns can be predicted
  - Fama and Schwert (1977), Campbell and Shiller (1988), Lettau and Ludvigson (2001), Lewellen (2004)
- Large debate over in-sample vs. out-of-sample and over statistical vs. economic predictability
  - Ang and Bekaert (2007), Goyal and Welch (2008)
  - Thornton and Valente (2012)
- For a single risky asset, importance of features such as model/parameter instability and model/parameter uncertainty are by now well established
  - Stock returns: see e.g. Dangl and Halling (2012), Johannes et al (2014)
  - Bond returns: see e.g. Gargano et al (2017)
- **This paper: What is the role of these features when the focus is on multiple risky assets?**

Introduction

- A recurring question in empirical asset pricing is whether or not asset returns can be predicted
  - Fama and Schwert (1977), Campbell and Shiller (1988), Lettau and Ludvigson (2001), Lewellen (2004)
- Large debate over in-sample vs. out-of-sample and over statistical vs. economic predictability
  - Ang and Bekaert (2007), Goyal and Welch (2008)
  - Thornton and Valente (2012)
- For a single risky asset, importance of features such as model/parameter instability and model/parameter uncertainty are by now well established
  - Stock returns: see e.g. Dangl and Halling (2012), Johannes et al (2014)
  - Bond returns: see e.g. Gargano et al (2017)
- **This paper: What is the role of these features when the focus is on multiple risky assets?**

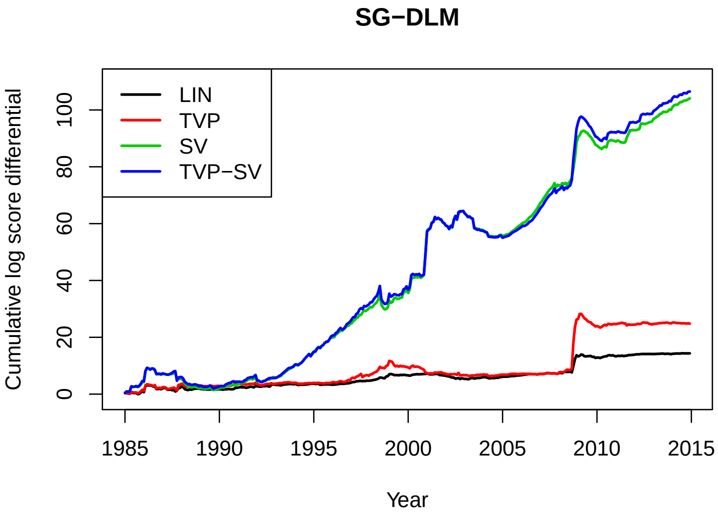
Introduction

- A recurring question in empirical asset pricing is whether or not asset returns can be predicted
  - Fama and Schwert (1977), Campbell and Shiller (1988), Lettau and Ludvigson (2001), Lewellen (2004)
- Large debate over in-sample vs. out-of-sample and over statistical vs. economic predictability
  - Ang and Bekaert (2007), Goyal and Welch (2008)
  - Thornton and Valente (2012)
- For a single risky asset, importance of features such as model/parameter instability and model/parameter uncertainty are by now well established
  - Stock returns: see e.g. Dangl and Halling (2012), Johannes et al (2014)
  - Bond returns: see e.g. Gargano et al (2017)
- **This paper: What is the role of these features when the focus is on multiple risky assets?**

Introduction

- A recurring question in empirical asset pricing is whether or not asset returns can be predicted
  - Fama and Schwert (1977), Campbell and Shiller (1988), Lettau and Ludvigson (2001), Lewellen (2004)
- Large debate over in-sample vs. out-of-sample and over statistical vs. economic predictability
  - Ang and Bekaert (2007), Goyal and Welch (2008)
  - Thornton and Valente (2012)
- For a single risky asset, importance of features such as model/parameter instability and model/parameter uncertainty are by now well established
  - Stock returns: see e.g. Dangl and Halling (2012), Johannes et al (2014)
  - Bond returns: see e.g. Gargano et al (2017)
- **This paper: What is the role of these features when the focus is on multiple risky assets?**

Cumulative sum of MVALS



Contributions of this paper

1. We introduce a flexible yet computationally simple framework to study the role of these features when forecasting the returns of multiple risky assets
    - Build on the Bayesian Dynamic Linear Models of West and Harrison (1997) and Gruber and West (2016)
    - We allow for asset-specific predictors, time-varying parameters and multivariate stochastic volatility, and derive closed-form expressions for all the moments of the posterior distributions and predictive densities (no need for MCMC!)
2. We estimate, rather than assume, the degree of time variation in the model parameters and asset variances/covariances
3. We evaluate the performance of our approach by jointly modeling monthly stock and bond returns over the 1985–2014 period, and uncover important gains from using our methods, both in statistical and economic terms

Contributions of this paper

1. We introduce a flexible yet computationally simple framework to study the role of these features when forecasting the returns of multiple risky assets
   - Build on the Bayesian Dynamic Linear Models of West and Harrison (1997) and Gruber and West (2016)
   - We allow for asset-specific predictors, time-varying parameters and multivariate stochastic volatility, and derive closed-form expressions for all the moments of the posterior distributions and predictive densities (no need for MCMC!)
2. We estimate, rather than assume, the degree of time variation in the model parameters and asset variances/covariances
3. We evaluate the performance of our approach by jointly modeling monthly stock and bond returns over the 1985–2014 period, and uncover important gains from using our methods, both in statistical and economic terms

Contributions of this paper

1. We introduce a flexible yet computationally simple framework to study the role of these features when forecasting the returns of multiple risky assets
    - Build on the Bayesian Dynamic Linear Models of West and Harrison (1997) and Gruber and West (2016)
    - We allow for asset-specific predictors, time-varying parameters and multivariate stochastic volatility, and derive closed-form expressions for all the moments of the posterior distributions and predictive densities (no need for MCMC!)
2. We estimate, rather than assume, the degree of time variation in the model parameters and asset variances/covariances
3. We evaluate the performance of our approach by jointly modeling monthly stock and bond returns over the 1985–2014 period, and uncover important gains from using our methods, both in statistical and economic terms

Contributions of this paper

1. We introduce a flexible yet computationally simple framework to study the role of these features when forecasting the returns of multiple risky assets
    - Build on the Bayesian Dynamic Linear Models of West and Harrison (1997) and Gruber and West (2016)
    - We allow for asset-specific predictors, time-varying parameters and multivariate stochastic volatility, and derive closed-form expressions for all the moments of the posterior distributions and predictive densities (no need for MCMC!)
2. We estimate, rather than assume, the degree of time variation in the model parameters and asset variances/covariances
3. We evaluate the performance of our approach by jointly modeling monthly stock and bond returns over the 1985–2014 period, and uncover important gains from using our methods, both in statistical and economic terms

Wishart DLM (West and Harrison, 1997)

- Begin with:

$$\boldsymbol{r}_t = \boldsymbol{B}_t' \boldsymbol{x}_{t-1} + \boldsymbol{v}_t \qquad \boldsymbol{v}_t | \boldsymbol{\Sigma}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_t)$$

$\boldsymbol{r}_t$: $q \times 1$ vector of log excess returns ($t = 1, ..., T$)
$\boldsymbol{x}_{t-1}$: $p \times 1$ vector of lagged predictor variables

- Allow for TVP...

$$vec(\boldsymbol{B}_t) = vec(\boldsymbol{B}_{t-1}) + \boldsymbol{\omega}_t \qquad \boldsymbol{\omega}_t | \boldsymbol{\Sigma}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_t \otimes \boldsymbol{W}_t)$$

- ... and multivariate SV

$$\boldsymbol{\Sigma}_t = \begin{bmatrix} \sigma_{1,t}^2 & \sigma_{12,t} & \cdots & \sigma_{1q,t} \\ \sigma_{12,t} & \sigma_{2,t}^2 & \cdots & \sigma_{2q,t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1q,t} & \sigma_{2q,t} & \cdots & \sigma_{q,t}^2 \end{bmatrix}$$

Wishart DLM (West and Harrison, 1997)

- Begin with:

$$\boldsymbol{r}_t = \boldsymbol{B}_t' \boldsymbol{x}_{t-1} + \boldsymbol{v}_t \qquad\qquad \boldsymbol{v}_t | \boldsymbol{\Sigma}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_t)$$

$\boldsymbol{r}_t$: $q \times 1$ vector of log excess returns $(t = 1, ..., T)$
$\boldsymbol{x}_{t-1}$: $p \times 1$ vector of lagged predictor variables

- Allow for TVP...

$$vec(\boldsymbol{B}_t) = vec(\boldsymbol{B}_{t-1}) + \boldsymbol{\omega}_t \qquad\qquad \boldsymbol{\omega}_t | \boldsymbol{\Sigma}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_t \otimes \boldsymbol{W}_t)$$

- ... and multivariate SV

$$\Sigma_t = \begin{bmatrix} \sigma^2_{1,t} & \sigma_{12,t} & \cdots & \sigma_{1q,t} \\ \sigma_{12,t} & \sigma^2_{2,t} & \cdots & \sigma_{2q,t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1q,t} & \sigma_{2q,t} & \cdots & \sigma^2_{q,t} \end{bmatrix}$$

Wishart DLM (West and Harrison, 1997)

- Begin with:

$$\boldsymbol{r}_t = \boldsymbol{B}_t' \boldsymbol{x}_{t-1} + \boldsymbol{v}_t \qquad \boldsymbol{v}_t | \boldsymbol{\Sigma}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_t)$$

$\boldsymbol{r}_t$: $q \times 1$ vector of log excess returns ($t = 1, ..., T$)
$\boldsymbol{x}_{t-1}$: $p \times 1$ vector of lagged predictor variables

- Allow for TVP...

$$vec(\boldsymbol{B}_t) = vec(\boldsymbol{B}_{t-1}) + \boldsymbol{\omega}_t \qquad \boldsymbol{\omega}_t | \boldsymbol{\Sigma}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_t \otimes \boldsymbol{W}_t)$$

- ... and multivariate SV

$$\boldsymbol{\Sigma}_t = \begin{bmatrix} \sigma_{1,t}^2 & \sigma_{12,t} & \dots & \sigma_{1q,t} \\ \sigma_{12,t} & \sigma_{2,t}^2 & \dots & \sigma_{2q,t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1q,t} & \sigma_{2q,t} & \dots & \sigma_{q,t}^2 \end{bmatrix}$$

Discounting with the Wishart DLM

- At any point in time t, posteriors are given by

$$vec(\boldsymbol{B}_t)|\boldsymbol{\Sigma}_t, \mathcal{D}_{t-1} \sim \mathcal{N}(vec(\boldsymbol{M}_{t-1}), \boldsymbol{\Sigma}_t \otimes \hat{\boldsymbol{C}}_t)$$
$$\boldsymbol{\Sigma}_t|\, \mathcal{D}_{t-1} \sim \mathcal{IW}(\hat{n}_t, \boldsymbol{S}_{t-1})$$

with

$$\hat{\boldsymbol{C}}_t = \frac{1}{\delta_\beta}\boldsymbol{C}_{t-1}, \qquad \hat{n}_t = \delta_v n_{t-1}$$

Key: $\delta_\beta \in (0,1]$ and $\delta_v \in (0,1]$ represent discount factors

- The predictive density for $\boldsymbol{r}_t$ is given by

$$\boldsymbol{r}_t|\delta_\beta, \delta_v, \mathcal{D}_{t-1} \sim \mathcal{T}_{\hat{n}_t}\left(\boldsymbol{M}'_{t-1}\boldsymbol{x}_{t-1}, \quad \boldsymbol{S}_{t-1}(1+\boldsymbol{x}'_{t-1}\hat{\boldsymbol{C}}_t\boldsymbol{x}_{t-1})\right).$$

where $\mathcal{T}_{\hat{n}_t}$ denotes a Student's t-distribution with $\hat{n}_t$ degrees of freedom.

Discounting with the Wishart DLM

- At any point in time t, posteriors are given by

$$\begin{gathered} vec(\boldsymbol{B}_t)|\boldsymbol{\Sigma}_t, \mathcal{D}_{t-1} \sim \mathcal{N}(vec(\boldsymbol{M}_{t-1}), \boldsymbol{\Sigma}_t \otimes \hat{\boldsymbol{C}}_t) \\ \boldsymbol{\Sigma}_t | \, \mathcal{D}_{t-1} \sim \mathcal{IW}(\hat{n}_t, \boldsymbol{S}_{t-1}) \end{gathered}$$

  with

$$\hat{\boldsymbol{C}}_t = \frac{1}{\delta_\beta} \boldsymbol{C}_{t-1}, \qquad \hat{n}_t = \delta_v n_{t-1}$$

  Key: $\delta_\beta \in (0, 1]$ and $\delta_v \in (0, 1]$ represent discount factors

- The predictive density for $\boldsymbol{r}_t$ is given by

$$\boldsymbol{r}_t | \delta_\beta, \delta_v, \mathcal{D}_{t-1} \sim \mathcal{T}_{\hat{n}_t} \left(\boldsymbol{M}'_{t-1} \boldsymbol{x}_{t-1}, \quad \boldsymbol{S}_{t-1}(1 + \boldsymbol{x}'_{t-1} \hat{\boldsymbol{C}}_t \boldsymbol{x}_{t-1}) \right).$$

  where $\mathcal{T}_{\hat{n}_t}$ denotes a Student's t-distribution with $\hat{n}_t$ degrees of freedom.

Limitations of the Wishart DLM

- The Wishart DLM is a simple and computationally very fast model, but presents three key drawbacks:
  1. It forces the same predictors $\boldsymbol{x}_{t-1}$ to each risky asset
  2. It imposes a single discount factor for all the elements of $\boldsymbol{B}_t$
  3. It features a single discount factor for the whole covariance matrix $\boldsymbol{\Sigma}_t$

Limitations of the Wishart DLM

- The Wishart DLM is a simple and computationally very fast model, but presents three key drawbacks:

  1. It forces the same predictors $\boldsymbol{x}_{t-1}$ to each risky asset
  2. It imposes a single discount factor for all the elements of $\boldsymbol{B}_t$
  3. It features a single discount factor for the whole covariance matrix $\boldsymbol{\Sigma}_t$

Limitations of the Wishart DLM

- The Wishart DLM is a simple and computationally very fast model, but presents three key drawbacks:

  1. It forces the same predictors $\boldsymbol{x}_{t-1}$ to each risky asset
  2. It imposes a single discount factor for all the elements of $\boldsymbol{B}_t$
  3. It features a single discount factor for the whole covariance matrix Σ_t

Limitations of the Wishart DLM

- The Wishart DLM is a simple and computationally very fast model, but presents three key drawbacks:

  1. It forces the same predictors $\boldsymbol{x}_{t-1}$ to each risky asset
  2. It imposes a single discount factor for all the elements of $\boldsymbol{B}_t$
  3. It features a single discount factor for the whole covariance matrix $\boldsymbol{\Sigma}_t$

Simultaneous Graphical DLM (Gruber and West, 2016)

- Write the joint dynamic system for the q excess returns $\boldsymbol{r}_t$ as follows:

$$\boldsymbol{r}_t = \begin{pmatrix} \boldsymbol{x}'_{1,t-1}\boldsymbol{\beta}_{1t} \\ \vdots \\ \boldsymbol{x}'_{q,t-1}\boldsymbol{\beta}_{qt} \end{pmatrix} + \begin{pmatrix} \boldsymbol{r}'_{-1,t}\gamma_{1t} \\ \vdots \\ \boldsymbol{r}'_{-q,t}\gamma_{qt} \end{pmatrix} + \boldsymbol{\nu}_t \quad \boldsymbol{\nu}_t | \boldsymbol{\Omega}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_t)$$

  $\boldsymbol{x}_{j,t-1}$ $(j = 1, ..., q)$: $p_j \times 1$ vector of asset j's lagged predictors
  $\boldsymbol{r}_{-j,t}$: contemporaneous returns on all assets other than j
  $\boldsymbol{\Omega}_t = diag\left(\sigma^2_{1t}, ..., \sigma^2_{qt}\right)$

- Key: Assuming that this system is fully-recursive leads to important computational savings. That is, for the generic j-th equation write

$$r_{jt} = \boldsymbol{x}'_{j,t-1}\boldsymbol{\beta}_{jt} + \boldsymbol{r}'_{<j,t}\gamma_{<j,t} + \nu_{jt} \qquad \nu_{jt} \sim N(0, \sigma^2_{jt})$$

  where $\gamma_{<j,t}$ is the $(j-1) \times 1$ vector of coefficients associated with the contemporaneous excess returns $\boldsymbol{r}_{<j,t}$.

Simultaneous Graphical DLM (Gruber and West, 2016)

- Write the joint dynamic system for the q excess returns $\boldsymbol{r}_t$ as follows:

$$\boldsymbol{r}_t = \begin{pmatrix} \boldsymbol{x}'_{1,t-1}\boldsymbol{\beta}_{1t} \\ \vdots \\ \boldsymbol{x}'_{q,t-1}\boldsymbol{\beta}_{qt} \end{pmatrix} + \begin{pmatrix} \boldsymbol{r}'_{-1,t}\gamma_{1t} \\ \vdots \\ \boldsymbol{r}'_{-q,t}\gamma_{qt} \end{pmatrix} + \boldsymbol{\nu}_t \qquad \boldsymbol{\nu}_t | \, \boldsymbol{\Omega}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}_t)$$

  $\boldsymbol{x}_{j,t-1}$ $(j = 1, ..., q)$: $p_j \times 1$ vector of asset j's lagged predictors
  $\boldsymbol{r}_{-j,t}$: contemporaneous returns on all assets other than j
  $\boldsymbol{\Omega}_t = diag\left(\sigma^2_{1t}, ..., \sigma^2_{qt}\right)$

- Key: Assuming that this system is fully-recursive leads to important computational savings. That is, for the generic j-th equation write

$$r_{jt} = \boldsymbol{x}'_{j,t-1}\boldsymbol{\beta}_{jt} + \boldsymbol{r}'_{<j,t}\gamma_{<j,t} + \nu_{jt} \qquad\qquad \nu_{jt} \sim N(0, \sigma^2_{jt})$$

  where $\gamma_{<j,t}$ is the $(j-1) \times 1$ vector of coefficients associated with the contemporaneous excess returns $\boldsymbol{r}_{<j,t}$.

Discounting in the SG-DLM

- As for the $\boldsymbol{\beta}_{jt}$ and $\boldsymbol{\gamma}_{<j,t}$ coefficients, write

$$\begin{pmatrix} \boldsymbol{\beta}_{jt} \\ \boldsymbol{\gamma}_{<j,t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_{j,t-1} \\ \boldsymbol{\gamma}_{<j,t-1} \end{pmatrix} + \boldsymbol{\omega}_{jt} \qquad \boldsymbol{\omega}_{jt} \sim N(\mathbf{0}, \boldsymbol{W}_{jt}).$$

- At any point in time t, posteriors are given by

$$\begin{pmatrix} \beta_{jt} \\ \gamma_{<j,t} \end{pmatrix}, \sigma^2_{jt} \Bigg| \mathcal{D}_{t-1} \sim \mathcal{NG}(\boldsymbol{m}_{j,t-1}, \hat{\boldsymbol{C}}_{jt}, \hat{n}_{jt}, s_{j,t-1}).$$

with

$$\hat{\boldsymbol{C}}_{j,t} = \begin{bmatrix} \boldsymbol{C}_{\beta\beta j,t-1}/\delta_{\beta j} & \boldsymbol{C}_{\beta\gamma j,t-1} \\ \boldsymbol{C}_{\gamma\beta j,t-1} & \boldsymbol{C}_{\gamma\gamma j,t-1}/\delta_{\gamma j} \end{bmatrix}, \qquad \hat{n}_{jt} = \delta_{vj} n_{j,t-1}$$

- Key: $\delta_{\beta j} \in (0,1]$, $\delta_{\gamma j} \in (0,1]$, and $\delta_{vj} \in (0,1]$ denote asset-specific discount factors

- As with the W-DLM, the predictive density of $\boldsymbol{r}_t$ is readily available and in closed-form

Discounting in the SG-DLM

- As for the $\boldsymbol{\beta}_{jt}$ and $\boldsymbol{\gamma}_{<j,t}$ coefficients, write

$$\begin{pmatrix} \boldsymbol{\beta}_{jt} \\ \boldsymbol{\gamma}_{<j,t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_{j,t-1} \\ \boldsymbol{\gamma}_{<j,t-1} \end{pmatrix} + \boldsymbol{\omega}_{jt} \qquad \boldsymbol{\omega}_{jt} \sim N(\mathbf{0}, \boldsymbol{W}_{jt}).$$

- At any point in time t, posteriors are given by

$$\begin{pmatrix} \boldsymbol{\beta}_{jt} \\ \boldsymbol{\gamma}_{<j,t} \end{pmatrix}, \sigma^2_{jt} \Bigg| \mathcal{D}_{t-1} \sim \mathcal{NG}(\boldsymbol{m}_{j,t-1}, \hat{\boldsymbol{C}}_{jt}, \hat{n}_{jt}, s_{j,t-1}).$$

with

$$\hat{\boldsymbol{C}}_{j,t} = \begin{bmatrix} \boldsymbol{C}_{\beta\beta j,t-1}/\delta_{\beta j} & \boldsymbol{C}_{\beta\gamma j,t-1} \\ \boldsymbol{C}_{\gamma\beta j,t-1} & \boldsymbol{C}_{\gamma\gamma j,t-1}/\delta_{\gamma j} \end{bmatrix}, \qquad \hat{n}_{jt} = \delta_{vj} n_{j,t-1}$$

- Key: $\delta_{\beta j} \in (0,1]$, $\delta_{\gamma j} \in (0,1]$, and $\delta_{vj} \in (0,1]$ denote asset-specific discount factors

- As with the W-DLM, the predictive density of $\boldsymbol{r}_t$ is readily available and in closed-form

Discounting in the SG-DLM

- As for the $\boldsymbol{\beta}_{jt}$ and $\gamma_{<j,t}$ coefficients, write

$$\begin{pmatrix} \boldsymbol{\beta}_{jt} \\ \gamma_{<j,t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_{j,t-1} \\ \gamma_{<j,t-1} \end{pmatrix} + \boldsymbol{\omega}_{jt} \qquad \boldsymbol{\omega}_{jt} \sim N(\mathbf{0}, \boldsymbol{W}_{jt}).$$

- At any point in time t, posteriors are given by

$$\begin{pmatrix} \boldsymbol{\beta}_{jt} \\ \gamma_{<j,t} \end{pmatrix}, \sigma_{jt}^2 \bigg| \mathcal{D}_{t-1} \sim \mathcal{NG}(\boldsymbol{m}_{j,t-1}, \hat{\boldsymbol{C}}_{jt}, \hat{n}_{jt}, s_{j,t-1}).$$

with

$$\hat{\boldsymbol{C}}_{j,t} = \begin{bmatrix} \boldsymbol{C}_{\beta\beta j,t-1}/\delta_{\beta j} & \boldsymbol{C}_{\beta\gamma j,t-1} \\ \boldsymbol{C}_{\gamma\beta j,t-1} & \boldsymbol{C}_{\gamma\gamma j,t-1}/\delta_{\gamma j} \end{bmatrix}, \qquad \hat{n}_{jt} = \delta_{vj} n_{j,t-1}$$

- Key: $\delta_{\beta j} \in (0,1]$, $\delta_{\gamma j} \in (0,1]$, and $\delta_{vj} \in (0,1]$ denote asset-specific discount factors

- As with the W-DLM, the predictive density of $\boldsymbol{r}_t$ is readily available and in closed-form

Discounting in the SG-DLM

- As for the $\boldsymbol{\beta}_{jt}$ and $\gamma_{<j,t}$ coefficients, write

$$\begin{pmatrix} \boldsymbol{\beta}_{jt} \\ \gamma_{<j,t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_{j,t-1} \\ \gamma_{<j,t-1} \end{pmatrix} + \boldsymbol{\omega}_{jt} \qquad\qquad \boldsymbol{\omega}_{jt} \sim N(\mathbf{0}, \boldsymbol{W}_{jt}).$$

- At any point in time t, posteriors are given by

$$\begin{pmatrix} \boldsymbol{\beta}_{jt} \\ \gamma_{<j,t} \end{pmatrix}, \sigma^2_{jt} \Bigg| \mathcal{D}_{t-1} \sim \mathcal{NG}(\boldsymbol{m}_{j,t-1}, \hat{\boldsymbol{C}}_{jt}, \hat{n}_{jt}, s_{j,t-1}).$$

  with

$$\hat{\boldsymbol{C}}_{j,t} = \begin{bmatrix} \boldsymbol{C}_{\beta\beta j,t-1}/\delta_{\beta j} & \boldsymbol{C}_{\beta\gamma j,t-1} \\ \boldsymbol{C}_{\gamma\beta j,t-1} & \boldsymbol{C}_{\gamma\gamma j,t-1}/\delta_{\gamma j} \end{bmatrix}, \qquad \hat{n}_{jt} = \delta_{vj} n_{j,t-1}$$

- Key: $\delta_{\beta j} \in (0,1]$, $\delta_{\gamma j} \in (0,1]$, and $\delta_{vj} \in (0,1]$ denote asset-specific discount factors
- As with the W-DLM, the predictive density of $\boldsymbol{r}_t$ is readily available and in closed-form

Model averaging

- Both W-DLM and SG-DLM models require the user to make a number of choices ex-ante, e.g.:
  - What's the right set of predictors to include in the model, asset by asset?
  - What's the optimal degree of time variation for parameters, variances, and covariances?
- Instead of relying on *ad-hoc* choices, we opt for the following:
  1. We estimate a different version of each W-DLM and SG-DLM model for each combination of predictor variables and discount factors
  2. We combine the predictive densities of all such models, favoring the model permutations that are most supported by the data
     (we rely on the multivariate log score for that)

Model averaging

- Both W-DLM and SG-DLM models require the user to make a number of choices ex-ante, e.g.:
  - What's the right set of predictors to include in the model, asset by asset?
  - What's the optimal degree of time variation for parameters, variances, and covariances?
- Instead of relying on *ad-hoc* choices, we opt for the following:
  1. We estimate a different version of each W-DLM and SG-DLM model for each combination of predictor variables and discount factors
  2. We combine the predictive densities of all such models, favoring the model permutations that are most supported by the data
     (we rely on the multivariate log score for that)

Model averaging

- Both W-DLM and SG-DLM models require the user to make a number of choices ex-ante, e.g.:
  - What's the right set of predictors to include in the model, asset by asset?
  - What's the optimal degree of time variation for parameters, variances, and covariances?
- Instead of relying on *ad-hoc* choices, we opt for the following:
  1. We estimate a different version of each W-DLM and SG-DLM model for each combination of predictor variables and discount factors
  2. We combine the predictive densities of all such models, favoring the model permutations that are most supported by the data
     (we rely on the multivariate log score for that)

Data Description and Model Setup

- Each DLM will predict the excess returns on
  - US Treasury Bonds with 5 and 10 year maturities
  - Large-, Mid-, and Small-cap stock portfolios

- Each DLM will have both a stock predictor and a bond predictor
  - Stock Predictors, 15 total from Welch and Goyal (2008), e.g.
    - Dividend Yield
    - Term and Default spreads
    - Stock Variance
  - Bond Predictors
    - Cochrane and Piazzesi (2005)'s factor
    - Fama and Bliss (1987)'s forward spread
    - Ludvigson and Ng (2009)'s macro factor

- Dataset spans Jan 1962 – Dec 2014. First ten years are used to initialize filters, and last 30 years (Jan 1985 – Dec 2014) are used for evaluation.

- As for the discount factors, we consider $\delta_{\beta j}, \delta_{\gamma j} \in \{0.98, 0.99, 1.0\}$, and $\delta_{vj} \in \{0.95, 0.975, 1.0\} \rightarrow$ 405 W-DLMs and 1215 SG-DLMs

Data Description and Model Setup

- Each DLM will predict the excess returns on
  - US Treasury Bonds with 5 and 10 year maturities
  - Large-, Mid-, and Small-cap stock portfolios

- Each DLM will have both a stock predictor and a bond predictor
  - Stock Predictors, 15 total from Welch and Goyal (2008), e.g.
    - Dividend Yield
    - Term and Default spreads
    - Stock Variance
  - Bond Predictors
    - Cochrane and Piazzesi (2005)'s factor
    - Fama and Bliss (1987)'s forward spread
    - Ludvigson and Ng (2009)'s macro factor

- Dataset spans Jan 1962 – Dec 2014. First ten years are used to initialize filters, and last 30 years (Jan 1985 – Dec 2014) are used for evaluation.

- As for the discount factors, we consider $\delta_{\beta j}, \delta_{\gamma j} \in \{0.98, 0.99, 1.0\}$, and $\delta_{vj} \in \{0.95, 0.975, 1.0\} \rightarrow$ 405 W-DLMs and 1215 SG-DLMs

Data Description and Model Setup

- Each DLM will predict the excess returns on
  - US Treasury Bonds with 5 and 10 year maturities
  - Large-, Mid-, and Small-cap stock portfolios
- Each DLM will have both a stock predictor and a bond predictor
  - Stock Predictors, 15 total from Welch and Goyal (2008), e.g.
    - ★ Dividend Yield
    - ★ Term and Default spreads
    - ★ Stock Variance
  - Bond Predictors
    - ★ Cochrane and Piazzesi (2005)'s factor
    - ★ Fama and Bliss (1987)'s forward spread
    - ★ Ludvigson and Ng (2009)'s macro factor
- Dataset spans Jan 1962 – Dec 2014. First ten years are used to initialize filters, and last 30 years (Jan 1985 – Dec 2014) are used for evaluation.
- As for the discount factors, we consider $\delta_{\beta j}$, $\delta_{\gamma j} \in \{0.98, 0.99, 1.0\}$, and $\delta_{vj} \in \{0.95, 0.975, 1.0\} \rightarrow$ 405 W-DLMs and 1215 SG-DLMs

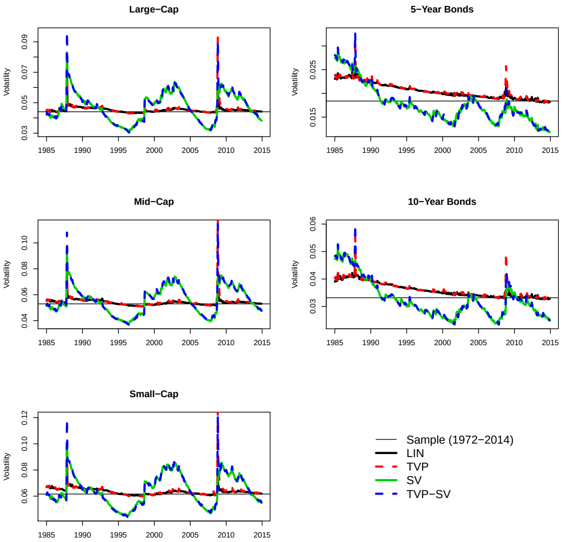
Data Description and Model Setup

- Each DLM will predict the excess returns on
  - US Treasury Bonds with 5 and 10 year maturities
  - Large-, Mid-, and Small-cap stock portfolios
- Each DLM will have both a stock predictor and a bond predictor
  - Stock Predictors, 15 total from Welch and Goyal (2008), e.g.
    - Dividend Yield
    - Term and Default spreads
    - Stock Variance
  - Bond Predictors
    - Cochrane and Piazzesi (2005)'s factor
    - Fama and Bliss (1987)'s forward spread
    - Ludvigson and Ng (2009)'s macro factor
- Dataset spans Jan 1962 – Dec 2014. First ten years are used to initialize filters, and last 30 years (Jan 1985 – Dec 2014) are used for evaluation.
- As for the discount factors, we consider $\delta_{\beta j}$, $\delta_{\gamma j} \in \{0.98, 0.99, 1.0\}$, and $\delta_{vj} \in \{0.95, 0.975, 1.0\} \rightarrow$ 405 W-DLMs and 1215 SG-DLMs

Data Description and Model Setup

- Each DLM will predict the excess returns on
  - US Treasury Bonds with 5 and 10 year maturities
  - Large-, Mid-, and Small-cap stock portfolios
- Each DLM will have both a stock predictor and a bond predictor
  - Stock Predictors, 15 total from Welch and Goyal (2008), e.g.
    - Dividend Yield
    - Term and Default spreads
    - Stock Variance
  - Bond Predictors
    - Cochrane and Piazzesi (2005)'s factor
    - Fama and Bliss (1987)'s forward spread
    - Ludvigson and Ng (2009)'s macro factor
- Dataset spans Jan 1962 – Dec 2014. First ten years are used to initialize filters, and last 30 years (Jan 1985 – Dec 2014) are used for evaluation.
- As for the discount factors, we consider $\delta_{\beta j}$, $\delta_{\gamma j} \in \{0.98, 0.99, 1.0\}$, and $\delta_{vj} \in \{0.95, 0.975, 1.0\} \rightarrow$ 405 W-DLMs and 1215 SG-DLMs

Time series of predicted volatilities (SG-DLMs)



Out of sample evaluation

- Multivariate average log score differentials (MVALS) between model i and the benchmark,

$$MVALS_i = \frac{1}{T - \underline{t} + 1} \sum_{\tau=\underline{t}}^{T} \left(\ln\left(S_{i,\tau}\right) - \ln\left(S_{bcmk,\tau}\right) \right)$$

$S_{i,\tau}$ ($S_{bcmk,\tau}$): model i's (benchmark's) multivariate log score at time τ

- Certainty equivalent returns (CERs) – in percentage annualized terms:

$$CER_i = 100 \times \left(\left[\frac{1}{T - \underline{t} + 1} \sum_{t=\underline{t}}^{T} \exp\left(r_{p,it}\right)^{1-A} \right]^{\frac{12}{1-A}} - 1 \right)$$

$r_{p,it}$: the realized portfolio return at time t, as implied by model i
A: relative risk aversion ($A = 5$)

Out of sample evaluation

- Multivariate average log score differentials (MVALS) between model i and the benchmark,

$$MVALS_i = \frac{1}{T - \underline{t} + 1} \sum_{\tau=\underline{t}}^{T} \left(\ln \left(S_{i,\tau} \right) - \ln \left(S_{bcmk,\tau} \right) \right)$$

$S_{i,\tau}$ ($S_{bcmk,\tau}$): model i's (benchmark's) multivariate log score at time τ

- Certainty equivalent returns (CERs) – in percentage annualized terms:

$$CER_i = 100 \times \left(\left[\frac{1}{T - \underline{t} + 1} \sum_{t=\underline{t}}^{T} \exp \left(r_{p,it} \right)^{1-A} \right]^{\frac{12}{1-A}} - 1 \right)$$

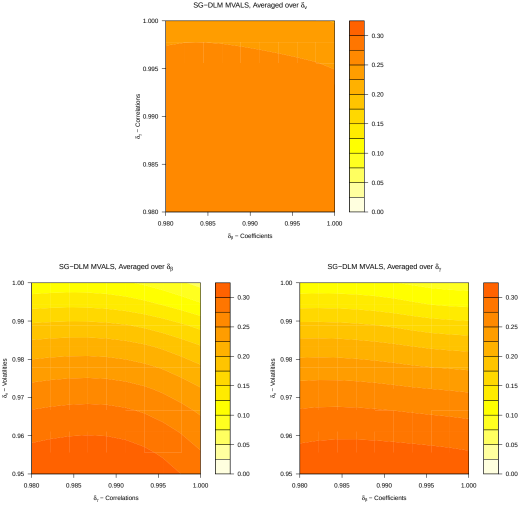
$r_{p,it}$: the realized portfolio return at time t, as implied by model i
A: relative risk aversion ($A = 5$)

Multivariate average log score differentials (MVALS)

	W-DLM		SG-DLM	
	Equal	Score	Equal	Score
LIN	0.043	0.040	0.042	0.040
TVP	0.066	0.069	0.066	0.069
SV	0.266	0.255	0.271	0.289
TVP-SV	0.265	**0.266**	0.279	**0.296**

$$MVALS_i = \frac{1}{T - \underline{t} + 1} \sum_{\tau=\underline{t}}^{T} \left(\ln\left(S_{i,\tau}\right) - \ln\left(S_{bcmk,\tau}\right) \right)$$

Heat maps of multivariate average log scores for different discount factors (SG-DLMs)

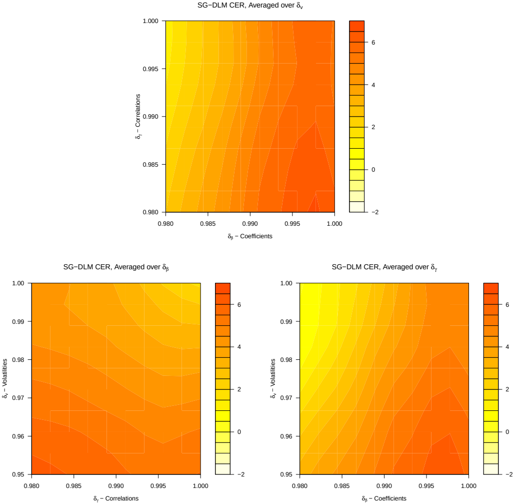


Annualized certainty equivalent returns

	W-DLM		SG-DLM	
	Equal	Score	Equal	Score
LIN	4.608	3.804	4.526	3.814
TVP	0.361	0.262	0.513	0.460
SV	**5.429**	4.623	**5.944**	5.892
TVP-SV	2.940	2.628	3.772	3.947

$$CER_i = 100\times \left[\left(\left[\frac{1}{T-\underline{t}+1} \sum_{t=\underline{t}}^{T} \exp\left(r_{p,it}\right)^{1-A} \right]^{\frac{12}{1-A}} - 1 \right) - \left(\left[\frac{1}{T-\underline{t}+1} \sum_{t=\underline{t}}^{T} \exp\left(r_{p,bench,t}\right)^{1-A} \right]^{\frac{12}{1-A}} - 1 \right) \right]$$

Heat map of certainty equivalent returns for different discount factors (SG-DLMs)



Conclusions

- We build on the W-DLM and the SG-DLM models and introduce a flexible, yet computationally efficient, approach to model and forecast multiple asset returns
  - We integrate a number of useful features into the predictive system: (i) model and parameter uncertainty, (ii) time-varying parameters, (iii) stochastic volatility, (iv) time varying covariances
- We combine these DLMs with a fully automated data-based model-averaging procedure to objectively determine the optimal set of features to include in the model
- We employ our approach to jointly forecast monthly stock and bond excess returns over the 1985–2014 period and find that:
  - The W-DLMs and SG-DLMs with stochastic volatility and time-varying correlations bring the largest gains in terms of statistical predictability
  - SG-DLM models with stochastic volatility and time-varying correlations performed best overall, with gains of over 500 basis points (annualized).

Conclusions

- We build on the W-DLM and the SG-DLM models and introduce a flexible, yet computationally efficient, approach to model and forecast multiple asset returns
  - We integrate a number of useful features into the predictive system: (i) model and parameter uncertainty, (ii) time-varying parameters, (iii) stochastic volatility, (iv) time varying covariances
- We combine these DLMs with a fully automated data-based model-averaging procedure to objectively determine the optimal set of features to include in the model
- We employ our approach to jointly forecast monthly stock and bond excess returns over the 1985–2014 period and find that:
  - The W-DLMs and SG-DLMs with stochastic volatility and time-varying correlations bring the largest gains in terms of statistical predictability
  - SG-DLM models with stochastic volatility and time-varying correlations performed best overall, with gains of over 500 basis points (annualized).

Conclusions

- We build on the W-DLM and the SG-DLM models and introduce a flexible, yet computationally efficient, approach to model and forecast multiple asset returns
  - We integrate a number of useful features into the predictive system: (i) model and parameter uncertainty, (ii) time-varying parameters, (iii) stochastic volatility, (iv) time varying covariances
- We combine these DLMs with a fully automated data-based model-averaging procedure to objectively determine the optimal set of features to include in the model
- We employ our approach to jointly forecast monthly stock and bond excess returns over the 1985–2014 period and find that:
  - The W-DLMs and SG-DLMs with stochastic volatility and time-varying correlations bring the largest gains in terms of statistical predictability
  - SG-DLM models with stochastic volatility and time-varying correlations performed best overall, with gains of over 500 basis points (annualized).

Conclusions

- We build on the W-DLM and the SG-DLM models and introduce a flexible, yet computationally efficient, approach to model and forecast multiple asset returns
    - We integrate a number of useful features into the predictive system: (i) model and parameter uncertainty, (ii) time-varying parameters, (iii) stochastic volatility, (iv) time varying covariances
- We combine these DLMs with a fully automated data-based model-averaging procedure to objectively determine the optimal set of features to include in the model
- We employ our approach to jointly forecast monthly stock and bond excess returns over the 1985–2014 period and find that:
    - The W-DLMs and SG-DLMs with stochastic volatility and time-varying correlations bring the largest gains in terms of statistical predictability
    - SG-DLM models with stochastic volatility and time-varying correlations performed best overall, with gains of over 500 basis points (annualized).

Conclusions

- We build on the W-DLM and the SG-DLM models and introduce a flexible, yet computationally efficient, approach to model and forecast multiple asset returns
  - We integrate a number of useful features into the predictive system: (i) model and parameter uncertainty, (ii) time-varying parameters, (iii) stochastic volatility, (iv) time varying covariances
- We combine these DLMs with a fully automated data-based model-averaging procedure to objectively determine the optimal set of features to include in the model
- We employ our approach to jointly forecast monthly stock and bond excess returns over the 1985–2014 period and find that:
  - The W-DLMs and SG-DLMs with stochastic volatility and time-varying correlations bring the largest gains in terms of statistical predictability
  - SG-DLM models with stochastic volatility and time-varying correlations performed best overall, with gains of over 500 basis points (annualized).

Thank you! Questions?

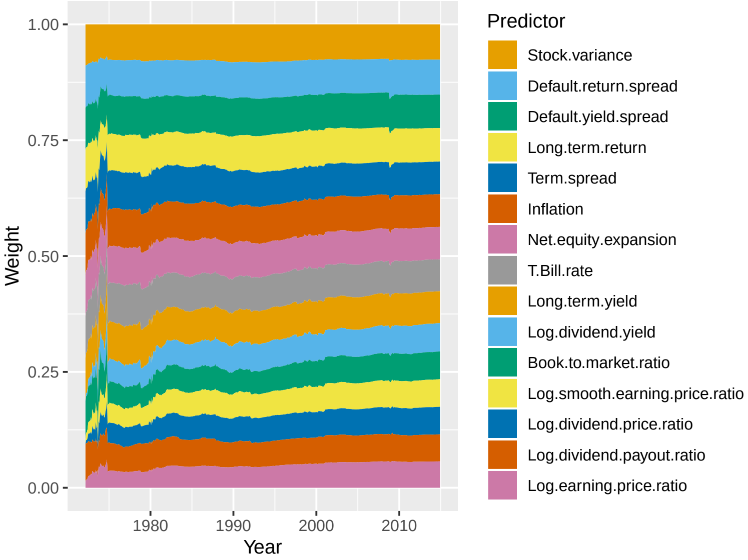
jared.fisher@berkeley.edu

Recently published:

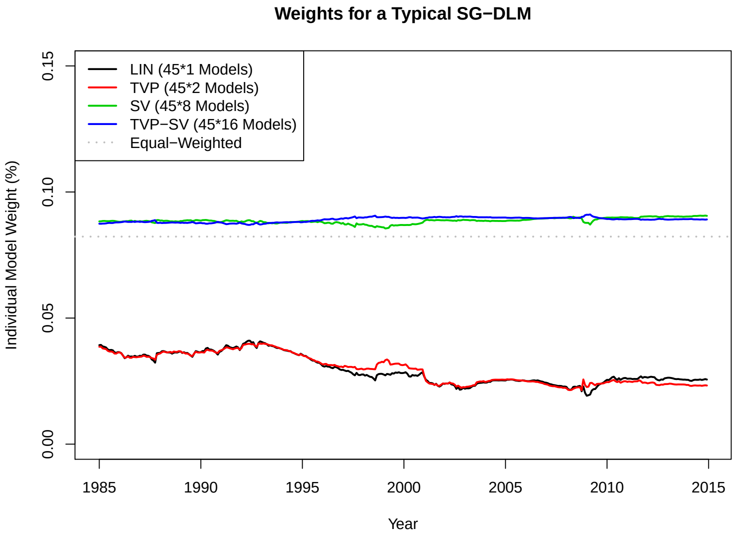
Fisher, J. D., Pettenuzzo, D.; Carvalho, C. M. "Optimal asset allocation with multivariate Bayesian dynamic linear models." Annals of Applied Statistics. 14 (2020), no. 1, 299–338.

Additional slides

Time series of score-based weights by predictor, SG-DLM



Time series of score-based weights by feature set



Continuously-compounded annualized CER differentials