Bayesian Dynamic Linear Models for Strategic Asset Allocation

Jared Fisher
Carlos Carvalho, University of Texas
Davide Pettenuzzo, Brandeis University

THE UNIVERSITY OF TEXAS AT AUSTIN
McCOMBS SCHOOL OF BUSINESS

August 31, 2018

Purpose

- To improve the forecasting model of the joint distribution of asset returns, particularly for assets of different classes

How should an investor optimally create a portfolio?

- Establish beliefs about future performance of assets $(\hat{\mu}_t, \hat{\Sigma}_t)$ using all current data:
    - Asset excess returns $\boldsymbol{r}_1, ..., \boldsymbol{r}_{t-1}$, where $\boldsymbol{r}_t = (r_{1t}, ..., r_{mt})'$
    - Predictive covariates $\boldsymbol{x}_1, ..., \boldsymbol{x}_{t-1}$, where $\boldsymbol{x}_t = (x_{1t}, ..., x_{pt})'$
- Choose how much to invest based on these beliefs

Literature on Predicting Risk Premia

$$r_t = \boldsymbol{x}'_{t-1}\boldsymbol{\beta} + \nu_t, \quad var(\nu_t) = \sigma^2$$

Literature on Predicting Risk Premia

$$r_t = \boldsymbol{x}'_{t-1}\boldsymbol{\beta} + \nu_t, \quad var(\nu_t) = \sigma^2$$

- Predictors

Literature on Predicting Risk Premia

$$r_t = \boldsymbol{x}'_{t-1}\boldsymbol{\beta} + \nu_t, \quad var(\nu_t) = \sigma^2$$

- Predictors
- Model uncertainty

Literature on Predicting Risk Premia

$$r_t = \boldsymbol{x}'_{t-1}\boldsymbol{\beta} + \nu_t, \quad var(\nu_t) = \sigma^2$$

- Predictors
- Model uncertainty
- Time-variation: rolling windows, structural breaks, dynamic linear models, stochastic volatility

Literature on Predicting Multi-Class Risk Premia

- Vector Auto-Regression (VAR)
- Other computationally intensive models needing MCMC
- Features found in univariate literature not all incorporated

How should covariates inform excess returns?

- We look at two methods
  - Assets share a vector of covariates:

$$
\begin{aligned}
r_{jt} &= \boldsymbol{x}'_{t-1}\boldsymbol{\beta}_{jt} + \nu_{jt} \\
\Rightarrow \boldsymbol{r}'_t &= \boldsymbol{x}'_{t-1}B_t + \boldsymbol{\nu}_t
\end{aligned}
$$

  - Include contemporaneous returns of other assets:

$$
\begin{aligned}
r_{jt} &= \boldsymbol{x}'_{j,t-1}\boldsymbol{\beta}_{jt} + \boldsymbol{r}'_{-j,t}\gamma_{jt} + \nu_{jt} \\
\Rightarrow \boldsymbol{r}_t &= \boldsymbol{\mu}_t + \Gamma_t \boldsymbol{r}_t + \boldsymbol{\nu}_t \\
\boldsymbol{r}_t &= (I - \Gamma_t)^{-1}(\boldsymbol{\mu}_t + \boldsymbol{\nu}_t)
\end{aligned}
$$

where $\boldsymbol{\mu}_t = (\boldsymbol{x}'_{1,t-1}\boldsymbol{\beta}_{1t}, ..., \boldsymbol{x}'_{m,t-1}\boldsymbol{\beta}_{mt})'$ and $j = 1, ..., m$

Wishart DLM - West and Harrison (1997)

Model

$$r_{jt} = \boldsymbol{x}'_{t-1}\boldsymbol{\beta}_{jt} + \nu_{jt}, \qquad \boldsymbol{\nu}_t|\Sigma_t \sim N(\mathbf{0}, \Sigma_t)$$
$$B_t = B_{t-1} + \Omega_t, \qquad \Omega_t|\Sigma_t \sim N(\mathbf{0}, W_t, \Sigma_t)$$

Prior

$$(B_0, \Sigma_0|D_0) \sim NW^{-1}_{n_0}(M_0, C_0, S_0)$$

Discount

$$n_t = \delta_v n_{t-1} + 1 \qquad \delta_v \in (0, 1]$$
$$W_t = \left(\frac{1}{\delta_\beta} - 1\right) C_{t-1} \qquad \delta_\beta \in (0, 1]$$

Simultaneous Graphical DLM - Gruber and West (2016)

Model

$$r_{jt} = \boldsymbol{x}'_{j,t-1}\boldsymbol{\beta}_{jt} + \boldsymbol{r}'_{-j,t}\boldsymbol{\gamma}_{jt} + \nu_{jt}, \qquad \nu_{jt} \sim N(\mathbf{0}, \lambda_{jt}^{-1})$$

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

Prior

$$(\boldsymbol{\theta}_{jt}, \boldsymbol{\lambda}_{j0} | D_0) \sim NG(\boldsymbol{m}_{j0}, C_{j0}, n_{j0}, s_{j0})$$

Discount

$$n_{jt} = \delta_v n_{j,t-1} + 1 \qquad \delta_v \in (0,1]$$

$$W_{jt} = \left(\frac{1}{\delta} - 1\right)^{***} C_{j,t-1} \qquad \delta_\beta, \delta_\gamma \in (0,1]$$

Specific Research Questions

1. Should we put the extra care into modeling assets' correlation?
2. Is there unincorporated information in X?
   - Does predictability exist/are markets efficient? Should X contain predictors or not?
3. Does it matter how information is incorporated?
   - Should the coefficient parameters B_t and/or volatility Σ_t vary over time, or not $(\delta_\beta = \delta_\gamma = \delta_v = 1)$?

Model Uncertainty

- We don't know the ideal X, δ_β, δ_γ, δ_v *a priori*.
- Solution: average and share strength across models.
  - Predictability: model for each pair of the stock predictors and bond predictors
  - Time-variation: 3 values of each δ are considered, equally spaced in the range $\delta_\beta, \delta_\gamma \in [0.98, 1.0]$, $\delta_v \in [0.95, 1.0]$.
- Weight the model average based on past performance, via score, akin to Bayesian model averaging

Model Evaluation: Forward-looking OOS Measures

- Economic: annualized certainty equivalent returns (CER)
  - Power utility investor, leverage constrained to invest within $[-2, 3]$ in any single asset.
  - $CER_\tau = \left(\left[(1-\gamma) \frac{1}{\tau} \sum_{t=1}^{\tau} U_t \right]^{\frac{12}{1-\gamma}} - 1 \right) * 100$
  - $U(wealth) = \frac{1}{1-\gamma} (wealth)^{1-\gamma}$
- Statistical fit measure: average log score (ALS)
  - $ALS_\tau = \frac{1}{\tau} \sum_{t=1}^{\tau} log \left(N[\boldsymbol{r}_t | \hat{\boldsymbol{\mu}}_t, \hat{\Sigma}_t] \right)$

Data Description

- Assets
  - US Treasurey Bonds with 5 and 10 year maturities
  - Large-, Mid-, and Small-cap stock portfolios
  - Risk-free T-Bill rate
- Bond Predictors
  - Cochrane and Piazzesi (2005)'s linear combination of forward rates
  - Fama and Bliss (1987)'s forward spread
  - Ludvigson and Ng (2009)'s macro economic factor
- Stock Predictors, from Welch and Goyal (2008)
  - Dividend Yield
  - Dividend/Payout ratio
  - Stock Variance

Methodology

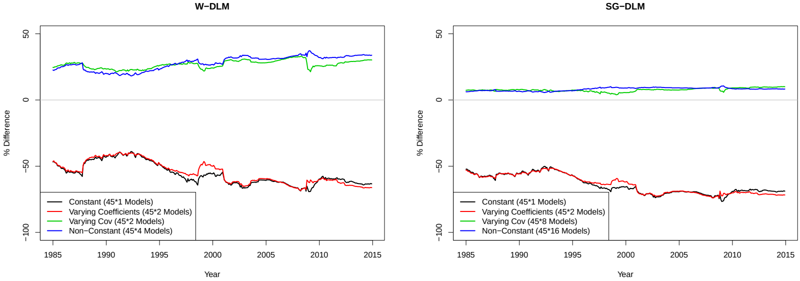
- The dataset contains monthly observations, spanning 1962 - 2014
    - Prior created on 1962-1971
    - Models evaluated on 1985-2014

Models

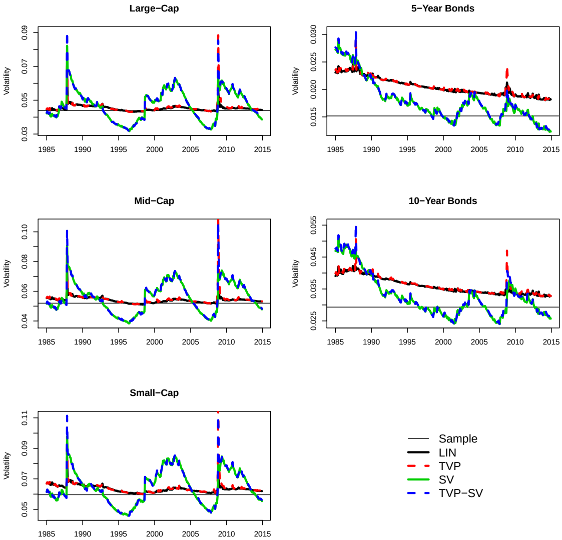
- Different Models
  - SG-DLM
  - W-DLM
- Different Feature Sets

Name	Single Asset Formula	
Prevailing Mean	$r_{jt} = \beta_{j0} + \nu_{jt},$	$\nu_t \sim N(0, \sigma^2)$
LIN	$r_{jt} = \beta_{j0} + \beta_{j1} x_{t-1} + \nu_{jt},$	$\nu_t \sim N(0, \sigma^2)$
TVP	$r_{jt} = \beta_{j0t} + \beta_{j1t} x_{t-1} + \nu_{jt},$	$\nu_t \sim N(0, \sigma^2)$
SV	$r_{jt} = \beta_{j0} + \beta_{j1} x_{t-1} + \nu_{jt},$	$\nu_t \sim N(0, \sigma_t^2)$
TVP-SV	$r_{jt} = \beta_{j0t} + \beta_{j1t} x_{t-1} + \nu_{jt},$	$\nu_t \sim N(0, \sigma_t^2)$

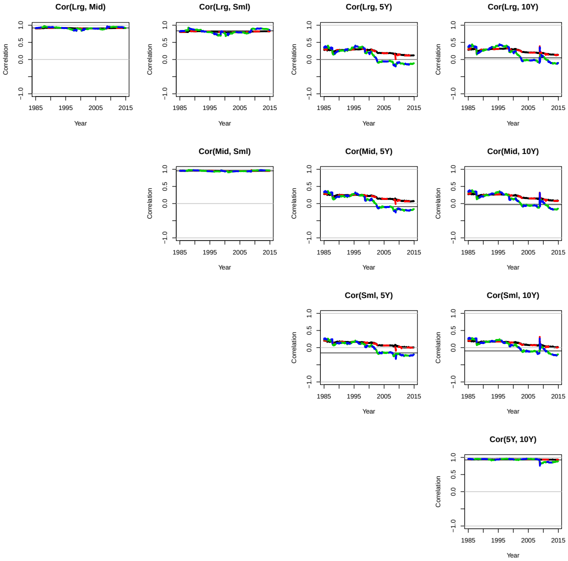
Time series of score-based weights by feature set



Time series of predicted volatilities for SG-DLM models

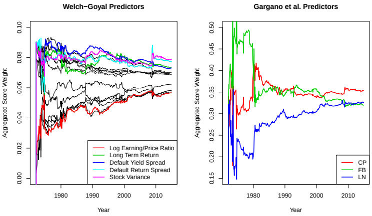


Time series of predicted correlations for SG-DLM models

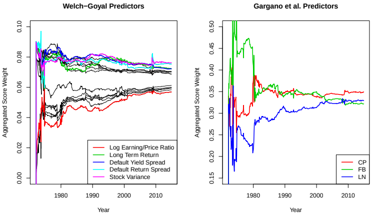


Time-series of score-based weights by predictor

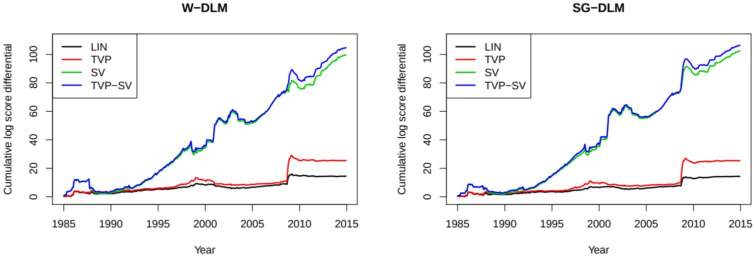
Panel A: W-DLM



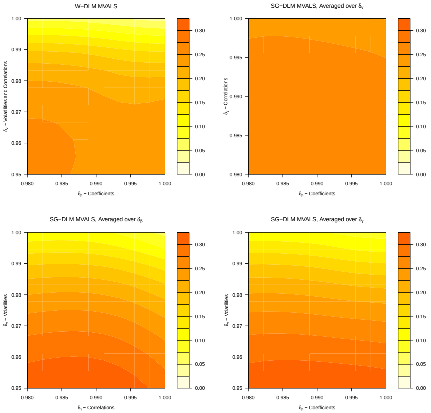
Panel B: SG-DLM



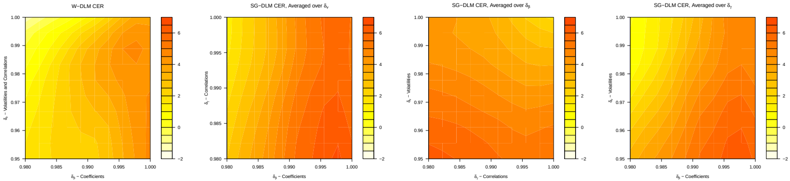
Cumulative sum of the multivariate log score differentials for W-DLM and SG-DLM models



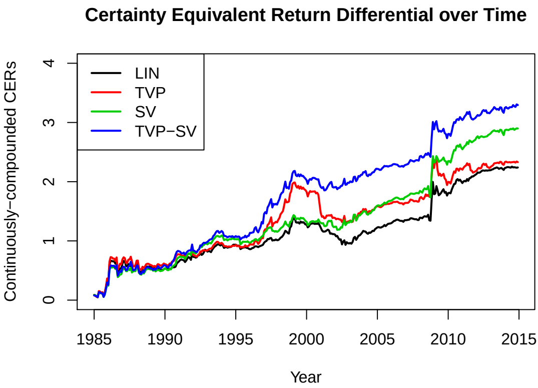
Heat map of multivariate average log scores for different discount factors



Heat map of certainty equivalent returns from discount factor combinations



Continuously compounded CER differentials



Weighted Mean-squared forecast errors

	W-DLM		SG-DLM	
	Equal	Score	Equal	Score
LIN	**0.989**	0.991	**0.989**	0.991
TVP	0.996	0.999	0.996	0.998
SV	**0.989**	0.990	0.989	0.990
TVP-SV	0.996	0.997	0.996	0.997

$$WMSFE_i = \frac{\sum_{\tau=\underline{t}}^{T} \boldsymbol{e}'_{i\tau} \left[\widehat{Cov(\boldsymbol{r}_t)}\right]^{-1} \boldsymbol{e}_{i\tau}}{\sum_{\tau=\underline{t}}^{T} \boldsymbol{e}'_{bcmk,\tau} \left[\widehat{Cov(\boldsymbol{r}_t)}\right]^{-1} \boldsymbol{e}_{bcmk,\tau}}$$

Multivariate Average log score differentials

	W-DLM		SG-DLM	
	Equal	Score	Equal	Score
LIN	0.043	0.040	0.042	0.040
TVP	0.069	0.070	0.069	0.070
SV	0.265	0.277	0.257	0.285
TVP-SV	0.280	**0.291**	0.270	**0.296**

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

Annualized certainty equivalent returns

	W-DLM		SG-DLM	
	Equal	Score	Equal	Score
LIN	4.608	3.804	4.526	3.814
TVP	2.220	1.808	2.442	2.100
SV	**5.612**	5.293	5.865	**5.870**
TVP-SV	4.941	4.583	5.459	5.657

$$CER_i = 100 \times \left(\left[\frac{1}{T - \underline{t} + 1} \sum_{t=\underline{t}}^{T} \widehat{W}_{it}^{1-A} \right]^{\frac{12}{1-A}} - 1 \right)$$

Conclusions

1. Should we model correlation? Yes, depending on utility function.
2. Is their information to aid predictability? Yes, if incorporated appropriately.
3. Should time-variation be incorporated? Yes, especially stochastic volatility.

Questions, Comments?

Thank you!

jared.fisher@utexas.edu

Myopic, leverage-constrained investor with power utility

- Power utility of wealth (CRRA): $\frac{(w_t)^{1-\gamma}}{1-\gamma}$
- Risk aversion $\gamma = 5$
- Investor is limited/restricted to placing [-2,3] times personal wealth in any given asset
- Following Cambell and Viceira (2002)'s example, we derive the investor's expected utility maximization
- For vector of portfolio weights on risky assets $\boldsymbol{w}_t$,

$$
\begin{aligned}
arg \max_{\boldsymbol{w}_t} \quad & \boldsymbol{w}_t' \left(\hat{\boldsymbol{\mu}}_t + \frac{1}{2} diag(\hat{\Sigma}_t) \right) - \frac{\gamma}{2} \boldsymbol{w_t'} \hat{\Sigma}_t \boldsymbol{w_t} && (1) \\
s.t. \quad & \boldsymbol{w}_t' \mathbf{1} \in [-2, 3] && (2) \\
& \boldsymbol{w}_{jt} \in [-2, 3], \ \forall j && (3)
\end{aligned}
$$