Discussion of

Real-time Portfolio Choice Implications of Asset Pricing Models

by Francisco Barillas and Jay Shanken

Daniele Bianchi

University of Warwick

2019 FMA Consortium on Factor Investing

Why should we care?

A decision maker $\mathcal{D}$ wants to model asset returns y based on a set of risk factors $\boldsymbol{z}$. A canonical approach is

$$y_t = \boldsymbol{\beta}'\boldsymbol{z}_t + \epsilon_t, \qquad \epsilon_t \sim \pi(0, \nu_t)$$

Why should we care?

A decision maker $\mathcal{D}$ wants to model asset returns y based on a set of risk factors $\boldsymbol{z}$. A canonical approach is

$$y_t = \boldsymbol{\beta}' \boldsymbol{z}_t + \epsilon_t, \qquad \epsilon_t \sim \pi(0, \nu_t)$$

Data-rich environments:

- Many candidates for $\boldsymbol{z}$.
- i.e., "zoo" of firm's characteristics and risk factors.

Why should we care?

A decision maker $\mathcal{D}$ wants to model asset returns y based on a set of risk factors $\boldsymbol{z}$. A canonical approach is

$$y_t = \boldsymbol{\beta}' \boldsymbol{z}_t + \epsilon_t, \qquad \epsilon_t \sim \pi(0, \nu_t)$$

Data-rich environments:

- Many candidates for $\boldsymbol{z}$.
- i.e., "zoo" of firm's characteristics and risk factors.

Limitations of single-model approaches:

- Irrelevant/redundant factors.
- Misspecification bias.
- Structural breaks.

Why should we care?

A decision maker $\mathcal{D}$ wants to model asset returns y based on a set of risk factors $\boldsymbol{z}$. A canonical approach is

$$y_t = \boldsymbol{\beta}' \boldsymbol{z}_t + \epsilon_t, \qquad \epsilon_t \sim \pi(0, \nu_t)$$

Data-rich environments:

- Many candidates for $\boldsymbol{z}$.
- i.e., "zoo" of firm's characteristics and risk factors.

Limitations of single-model approaches:

- Irrelevant/redundant factors.
- Misspecification bias.
- Structural breaks.

Model averaging might help (Timmermann 2006):

- Variance reduction/regularization.

Why should we care?

A decision maker $\mathcal{D}$ wants to model asset returns y based on a set of risk factors $\boldsymbol{z}$. A canonical approach is

$$y_t = \boldsymbol{\beta}' \boldsymbol{z}_t + \epsilon_t, \qquad \epsilon_t \sim \pi(0, \nu_t)$$

Data-rich environments:

- Many candidates for $\boldsymbol{z}$.
- i.e., "zoo" of firm's characteristics and risk factors.

Limitations of single-model approaches:

- Irrelevant/redundant factors.
- Misspecification bias.
- Structural breaks.

Model averaging might help (Timmermann 2006):

- Variance reduction/regularization.
- Robustify estimates.

Why should we care?

A decision maker $\mathcal{D}$ wants to model asset returns y based on a set of risk factors $\boldsymbol{z}$. A canonical approach is

$$y_t = \boldsymbol{\beta}' \boldsymbol{z}_t + \epsilon_t, \qquad \epsilon_t \sim \pi(0, \nu_t)$$

Data-rich environments:

- Many candidates for $\boldsymbol{z}$.
- i.e., "zoo" of firm's characteristics and risk factors.

Limitations of single-model approaches:

- Irrelevant/redundant factors.
- Misspecification bias.
- Structural breaks.

Model averaging might help (Timmermann 2006):

- Variance reduction/regularization.
- Robustify estimates.
- Exploit $\neq$ adaptivity to breaks.

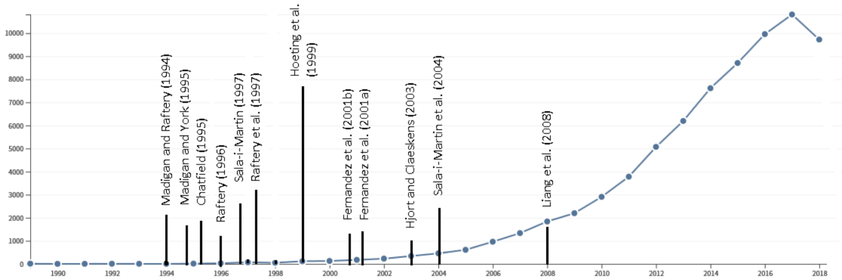
Why should we care?

> *'The main conclusion is that the composite set of forecasts can yield lower mean squared error than the original forecasts.'*
>
> Bates and Granger (1969)

Why should we care?

'The main conclusion is that the composite set of forecasts can yield lower mean squared error than the original forecasts.'

Bates and Granger (1969)



Total number of citations to papers with topic "model averaging" over years 1989-2018 in economics, finance, econometrics, statistics, etc. Google scholar returns over 52,500 papers in a search for "model averaging" and over 39,000 papers when searching for BMA.

This Paper

One of the problems in BMA is to figure out the "weight" ω_i, of each model $\mathcal{M}_i, i = 1, ..., N$, i.e.,

$$\omega_i \propto p\left(y_{t+1}|\mathcal{M}_i\right) p\left(\mathcal{M}_i\right)$$

Two ingredients:

- Marginal likelihood $p\left(y_{t+1}|\mathcal{M}_i\right)$.
- Prior distribution $p\left(\mathcal{M}_i\right)$.

This Paper

One of the problems in BMA is to figure out the "weight" ω_i, of each model $\mathcal{M}_i, i = 1, ..., N$, i.e.,

$$\omega_i \propto p\left(y_{t+1}|\mathcal{M}_i\right) p\left(\mathcal{M}_i\right)$$

Two ingredients:

- Marginal likelihood $p\left(y_{t+1}|\mathcal{M}_i\right)$.
- Prior distribution $p\left(\mathcal{M}_i\right)$.

The authors build on Barillas and Shanken (2018) to calculate ω_i consistent with asset pricing restrictions/implications.

This Paper

One of the problems in BMA is to figure out the "weight" ω_i, of each model $\mathcal{M}_i, i = 1, ..., N$, i.e.,

$$\omega_i \propto p\left(y_{t+1}|\mathcal{M}_i\right) p\left(\mathcal{M}_i\right)$$

Two ingredients:

- Marginal likelihood $p\left(y_{t+1}|\mathcal{M}_i\right)$.
- Prior distribution $p\left(\mathcal{M}_i\right)$.

The authors build on Barillas and Shanken (2018) to calculate ω_i consistent with asset pricing restrictions/implications.

This paper:

Investigates the implications for factor-based portfolio allocations when ω_i's are used for model averaging.

My Comments:

Comment 1: BMA implementation.

Comment 2: Real-time portfolio choice implementation.

Comment 3: Optimal portfolios formation.

Comment 1: Bayesian Model Averaging

Remind: BMA is "optimal" only in an $\mathcal{M}$-closed setting, i.e., when the *true* model $\mathcal{M}^T$ is in the set of models considered, i.e., $\mathcal{M}^T \in \mathcal{M}$

Comment 1: Bayesian Model Averaging

Remind: BMA is "optimal" only in an $\mathcal{M}$-closed setting, i.e., when the *true* model $\mathcal{M}^T$ is in the set of models considered, i.e., $\mathcal{M}^T \in \mathcal{M}$

Suppose the true model is

$$y = x_1\beta_1 + x_2\beta_2 + \epsilon$$

and you estimate

$$\mathcal{M}_1 : y = x_1\beta_1 + \epsilon$$

$$\mathcal{M}_2 : y = x_2\beta_2 + \epsilon$$

Comment 1: Bayesian Model Averaging

Remind: BMA is "optimal" only in an $\mathcal{M}$-closed setting, i.e., when the *true* model $\mathcal{M}^T$ is in the set of models considered, i.e., $\mathcal{M}^T \in \mathcal{M}$

Suppose the true model is

$$y = x_1\beta_1 + x_2\beta_2 + \epsilon$$

and you estimate

$$\mathcal{M}_1 : y = x_1\beta_1 + \epsilon$$

$$\mathcal{M}_2 : y = x_2\beta_2 + \epsilon$$

With BMA your prediction will be

$$\hat{y} = \omega_1 \cdot x_1\hat{\beta}_1 + \omega_2 \cdot x_2\hat{\beta}_2$$

Comment 1: Bayesian Model Averaging

Remind: BMA is "optimal" only in an $\mathcal{M}$-closed setting, i.e., when the *true* model $\mathcal{M}^T$ is in the set of models considered, i.e., $\mathcal{M}^T \in \mathcal{M}$

Suppose the true model is

$$y = x_1\beta_1 + x_2\beta_2 + \epsilon$$

and you estimate

$$\mathcal{M}_1 : y = x_1\beta_1 + \epsilon$$

$$\mathcal{M}_2 : y = x_2\beta_2 + \epsilon$$

With BMA your prediction will be

$$\hat{y} = \omega_1 \cdot x_1\hat{\beta}_1 + \omega_2 \cdot x_2\hat{\beta}_2$$

Under standard regularity conditions:

- $\omega_i \to 1$ for the model "closest" to the truth (in a Kullback-Leibler sense).
- BMA is not consistent if $\mathcal{M}^T \notin \mathcal{M}$.

Comment 1: Bayesian Model Averaging

Remind: BMA is "optimal" only in an $\mathcal{M}$-closed setting, i.e., when the *true* model $\mathcal{M}^T$ is in the set of models considered, i.e., $\mathcal{M}^T \in \mathcal{M}$

Suppose the true model is

$$y = x_1\beta_1 + x_2\beta_2 + \epsilon$$

and you estimate

$$\mathcal{M}_1 : y = x_1\beta_1 + \epsilon$$

$$\mathcal{M}_2 : y = x_2\beta_2 + \epsilon$$

With BMA your prediction will be

$$\hat{y} = \omega_1 \cdot x_1\hat{\beta}_1 + \omega_2 \cdot x_2\hat{\beta}_2$$

Under standard regularity conditions:

– $\omega_i \rightarrow 1$ for the model "closest" to the truth (in a Kullback-Leibler sense).

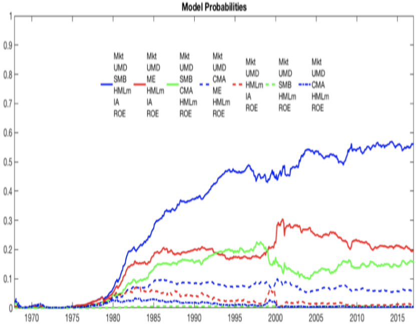
– BMA is not consistent if $\mathcal{M}^T \notin \mathcal{M}$.

Comment:

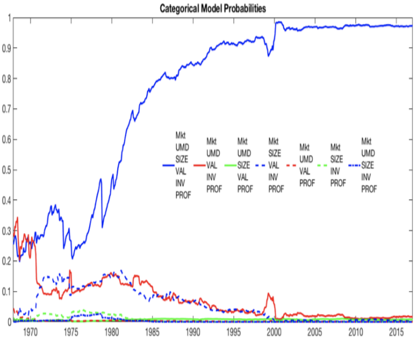
– Discuss the trade-off $\mathcal{M}$-closed vs. $\mathcal{M}$-open setting with BMA.

– In an $\mathcal{M}$-open setting model weights are the outcome of the decision problem (see Ch.9. Bernardo and Smith 1994), i.e., models can be selected based on the portfolio utility.

Comment 2: Real-Time Portfolio Allocation



(a) Model Probabilities

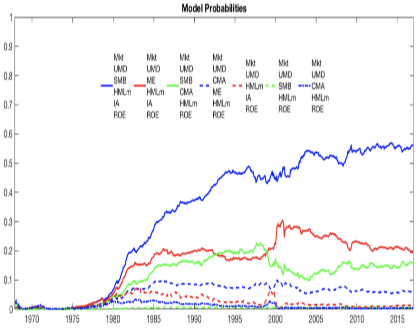


(b) Categorial Model Probabilities

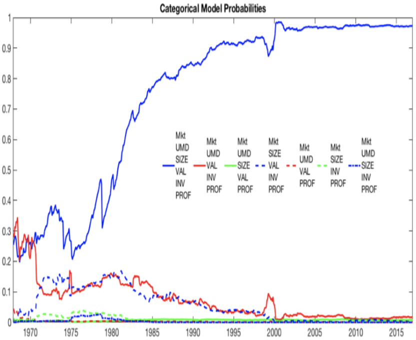
In the paper:

- Expanding window; initial 20 years of data, from 01:1967 to 12:1986, then recursively add observations until the end of the sample.
- Allocations convergence towards the "best" model (in relative terms).

Comment 2: Real-Time Portfolio Allocation



(c) Model Probabilities



(d) Categorial Model Probabilities

Comment:

- Rolling window; initial 20 years of data, from 01:1967 to 12:1986, then keep the same window until the end of the sample.
- Allocations won't convergence any more towards the "best" model.

Comment 3: Forming Optimal Portfolios

An investor maximizes wealth

$$\max_{\boldsymbol{w}} E_t \left[U \left(W_{t+1} \right) \right]$$

by choosing among N zero-cost risky portolios and a riskless asset, s.t.,

$$W_{t+1} = W_t \cdot \left[R_t^f + \boldsymbol{w}_t' f_{t+1} \right]$$

If $U \left(W_{t+1} \right)$ is quadratic only the first two moments matter

Comment 3: Forming Optimal Portfolios

An investor maximizes wealth

$$\max_{\boldsymbol{w}} E_t \left[U \left(W_{t+1} \right) \right]$$

by choosing among N zero-cost risky portolios and a riskless asset, s.t.,

$$W_{t+1} = W_t \cdot \left[R_t^f + \boldsymbol{w}'_t f_{t+1} \right]$$

If $U\left(W_{t+1}\right)$ is e.g., power utility, also the higher moments matter (see, e.g., Guidolin and Timmermann 2008).

Comment 3: Forming Optimal Portfolios

An investor maximizes wealth

$$\max_{\boldsymbol{w}} E_t \left[U \left(W_{t+1} \right) \right]$$

by choosing among N zero-cost risky portolios and a riskless asset, s.t.,

$$W_{t+1} = W_t \cdot \left[R_t^f + \boldsymbol{w}_t' f_{t+1} \right]$$

Recall that

$$p \left(f_{t+1} \right) = \sum_{j=1}^{M} \omega_j \cdot p \left(f_{t+1} | \mathcal{M}_j \right)$$

i.e, BMA is a mixture of predictive densities $=$ approximate any shape.

Comment 3: Forming Optimal Portfolios

An investor maximizes wealth

$$\max_{\boldsymbol{w}} E_t \left[U \left(W_{t+1} \right) \right]$$

by choosing among N zero-cost risky portolios and a riskless asset, s.t.,

$$W_{t+1} = W_t \cdot \left[R_t^f + \boldsymbol{w}'_t f_{t+1} \right]$$

Recall that

$$p \left(f_{t+1} \right) = \sum_{j=1}^{M} \omega_j \cdot p \left(f_{t+1} | \mathcal{M}_j \right)$$

i.e, BMA is a mixture of predictive densities = approximate any shape.

Comment:

– Perhaps look at how optimal portfolio investments change from higher moments within a power utility context.

Conclusion

Very interesting paper, recommended reading!!