Bayesian Methods in Investment Management

Prof. Svetlozar (Zari) Rachev

Chair of Econometrics, Statistics, and Mathematical Finance
School of Economics and Business Engineering
University of Karlsruhe

Bulgarian Academy of Sciences
Sofia, August 22, 2007

Presentation Outline

Outline

Reason 1: Estimation Process is Subject to Uncertainty

- Maximum likelihood estimates contain estimation errors.
- Using long data samples might alleviate estimation risk but
  - Long data samples are not always available (developing-markets assets, alternative investments, etc).
  - Distributions over long periods of time are often time-varying.
- Estimation risk is accounted for in the Bayesian setting which treats parameters as random variables.
- Bayesian estimation does not need long histories of data—it does not rely on asymptotic results (unlike maximum likelihood estimation).

Reason 1: Estimation Process is Subject to Uncertainty

- Maximum likelihood estimates contain estimation errors.
- Using long data samples might alleviate estimation risk but
  - Long data samples are not always available (developing-markets assets, alternative investments, etc).
  - Distributions over long periods of time are often time-varying.
- Estimation risk is accounted for in the Bayesian setting which treats parameters as random variables.
- Bayesian estimation does not need long histories of data—it does not rely on asymptotic results (unlike maximum likelihood estimation).

Reason 1: Estimation Process is Subject to Uncertainty

- Maximum likelihood estimates contain estimation errors.
- Using long data samples might alleviate estimation risk but
  - Long data samples are not always available (developing-markets assets, alternative investments, etc).
  - Distributions over long periods of time are often time-varying.
- Estimation risk is accounted for in the Bayesian setting which treats parameters as random variables.
- Bayesian estimation does not need long histories of data—it does not rely on asymptotic results (unlike maximum likelihood estimation).

Reason 1: Estimation Process is Subject to Uncertainty

- Maximum likelihood estimates contain estimation errors.
- Using long data samples might alleviate estimation risk but
  - Long data samples are not always available (developing-markets assets, alternative investments, etc).
  - Distributions over long periods of time are often time-varying.
- Estimation risk is accounted for in the Bayesian setting which treats parameters as random variables.
- Bayesian estimation does not need long histories of data—it does not rely on asymptotic results (unlike maximum likelihood estimation).

Reason 2: Got Views? Incorporate them

Bayesian methods provide a coherent framework for integrating views of analysts, traders, portfolio managers, and others into models. Model conclusions blend the information content of observed data with the subjective information input.

- A prominent example is the Black-Litterman model for portfolio allocation (we come back to it later).
- How do we translate views into statistical terminology? Specification of prior distributions needs care.

Reason 2: Got Views? Incorporate them

Bayesian methods provide a coherent framework for integrating views of analysts, traders, portfolio managers, and others into models. Model conclusions blend the information content of observed data with the subjective information input.

- A prominent example is the Black-Litterman model for portfolio allocation (we come back to it later).
- How do we translate views into statistical terminology? Specification of prior distributions needs care.

Reason 2: Got Views? Incorporate them

Bayesian methods provide a coherent framework for integrating views of analysts, traders, portfolio managers, and others into models. Model conclusions blend the information content of observed data with the subjective information input.

- A prominent example is the Black-Litterman model for portfolio allocation (we come back to it later).
- How do we translate views into statistical terminology? Specification of prior distributions needs care.

Reason 3: Complex Models Are Manageable in the Bayesian Setting

- The Markov Chain Monte Carlo toolbox facilitates inference about parameters and variables. Especially when distributions are not of standard form.
- Instead of a single point estimate (e.g., sample mean), one obtains the whole parameter distribution (analytically or numerically). Richer analysis and conclusions can be made.

Reason 3: Complex Models Are Manageable in the Bayesian Setting

- The Markov Chain Monte Carlo toolbox facilitates inference about parameters and variables. Especially when distributions are not of standard form.
- Instead of a single point estimate (e.g., sample mean), one obtains the whole parameter distribution (analytically or numerically). Richer analysis and conclusions can be made.

Outline

Focus on Two Areas of Application

- Portfolio Selection
  - Sensitivity of mean-variance portfolios to errors in inputs
  - Bayesian portfolio selection
  - The Black-Litterman model
- Volatility Modeling
  - Markov regime-switching GARCH modeling

Focus on Two Areas of Application

- Portfolio Selection
  - Sensitivity of mean-variance portfolios to errors in inputs
  - Bayesian portfolio selection
  - The Black-Litterman model

- Volatility Modeling
  - Markov regime-switching GARCH modeling

Outline

The Bayesian Paradigm

Bayesian researchers consider parameters as random variables, unlike classical statisticians who view them as fixed quantities.

Notation:

$\boldsymbol{R}$ = vector (matrix) of observed asset returns

$\boldsymbol{\theta}$ = parameter vector of interest

$p(\boldsymbol{R} \mid \boldsymbol{\theta})$ = likelihood function for $\boldsymbol{\theta}$

$p(\boldsymbol{\theta})$ = prior distribution of $\boldsymbol{\theta}$

The Bayesian Paradigm (Cont'd)

The Bayesian Updating Relationship

$$p(\theta \mid \boldsymbol{R}) = \frac{p(\boldsymbol{R} \mid \theta)\, p(\theta)}{\int_\theta p(\boldsymbol{R} \mid \theta)\, p(\theta)} = \frac{L(\theta)\, p(\theta)}{p(\boldsymbol{R})} \propto L(\theta)\, p(\theta)$$

$\propto$ = proportional to

Data information serves to update prior beliefs.

The Predictive Distribution

$$p\left(\widetilde{R} \mid \boldsymbol{R}\right) = \int_\theta p\left(\widetilde{R} \mid \boldsymbol{R}, \theta\right) \times p(\theta \mid \boldsymbol{R})\, d\theta$$

$\widetilde{R}$ = next-period's return

Integration over θ means that the distribution of future returns reflects parameter uncertainty.

The Bayesian Paradigm (Cont'd)

The Bayesian Updating Relationship

$$p\left(\theta \mid \boldsymbol{R}\right) = \frac{p\left(\boldsymbol{R} \mid \theta\right) p\left(\theta\right)}{\int_{\theta} p\left(\boldsymbol{R} \mid \theta\right) p\left(\theta\right)} = \frac{L\left(\theta\right) p\left(\theta\right)}{p\left(\boldsymbol{R}\right)} \propto L\left(\theta\right) p\left(\theta\right)$$

$\propto$ = proportional to

Data information serves to update prior beliefs.

The Predictive Distribution

$$p\left(\widetilde{R} \mid \boldsymbol{R}\right) = \int_{\theta} p\left(\widetilde{R} \mid \boldsymbol{R}, \theta\right) \times p\left(\theta \mid \boldsymbol{R}\right) \mathrm{d}\theta$$

$\widetilde{R}$ = next-period's return

Integration over θ means that the distribution of future returns reflects parameter uncertainty.

Outline

Sensitivity of the “Classical” Mean-Variance Optimal Portfolio to Inputs

“Classical” mean-variance (MV) portfolio selection is overly sensitive to errors (small changes) in inputs (means and covariances of returns).

The reason is that the sample moments are considered to be the true moments and uncertainty about them is ignored.

Illustration

- Ten MSCI country indices are candidates for inclusion into a portfolio.
- Their daily excess returns are observed over the period Jan 2, 1998 through May 5, 2004.
- We perform “classical” mean-variance optimization to construct the optimal portfolio

Sensitivity of the "Classical" Mean-Variance Optimal Portfolio to Inputs

"Classical" mean-variance (MV) portfolio selection is overly sensitive to errors (small changes) in inputs (means and covariances of returns).

The reason is that the sample moments are considered to be the true moments and uncertainty about them is ignored.

Illustration

- Ten MSCI country indices are candidates for inclusion into a portfolio.
- Their daily excess returns are observed over the period Jan 2, 1998 through May 5, 2004.
- We perform "classical" mean-variance optimization to construct the optimal portfolio

Sensitivity of the "Classical" Mean-Variance Optimal Portfolio to Inputs

"Classical" mean-variance (MV) portfolio selection is overly sensitive to errors (small changes) in inputs (means and covariances of returns).

The reason is that the sample moments are considered to be the true moments and uncertainty about them is ignored.

Illustration

- Ten MSCI country indices are candidates for inclusion into a portfolio.
- Their daily excess returns are observed over the period Jan 2, 1998 through May 5, 2004.
- We perform "classical" mean-variance optimization to construct the optimal portfolio

Sensitivity of the "Classical" Mean-Variance Optimal Portfolio to Inputs: Illustration I

MSCI Country Index	Mean Return	St. Dev.	Correlation Matrix										Mean-Variance Optimal Weights (%)		
													1.08% Target Return	3.16% Target Return	5.24% Target Return
Denmark	2.3	20.7	1	0.51	0.51	0.47	0.48	0.54	0.55	0.51	0.51	0.46	30.2	35.0	39.8
Germany	-0.6	27.2		1	0.74	0.69	0.52	0.77	0.81	0.65	0.48	0.45	-14.7	-18.8	-23.0
Italy	2.4	23.0			1	0.68	0.57	0.76	0.81	0.62	0.50	0.47	15.5	20.8	26.0
UK	-2.3	19.6				1	0.45	0.77	0.77	0.61	0.48	0.50	37.0	31.0	25.0
Portugal	-3.1	20.0					1	0.51	0.56	0.48	0.43	0.43	19.3	12.1	5.0
Netherlands	-3.3	24.3						1	0.85	0.65	0.55	0.50	-35.1	-50.5	-65.9
France	4.0	23.6							1	0.71	0.53	0.49	22.8	42.5	62.2
Sweden	5.2	31.2								1	0.51	0.43	-7.1	-4.2	-1.3
Norway	-0.1	22.6									1	0.46	15.6	16.2	16.8
Ireland	-1.7	21.2										1	16.6	16.0	15.4

Sensitivity of the "Classical" Mean-Variance Optimal Portfolio to Inputs: Illustration II

Let's "tweak" the sample mean of Germany by 10% and observe the resulting percentage change in optimal mean-variance portfolio weights.

Percentage Changes in Optimal Portfolio Weights

"Classical" Scenario

	Denmark	Germany	Italy	UK	Portugal	Netherlands	France	Sweden	Norway	Ireland
0.04%	0.13	7.86	-1.88	25.09	-6.54	17.29	-1.88	-1.32	-3.61	-11.47
2.1%	7.9	-26.72	-22.34	24.63	17.39	29.90	8.04	65.72	-0.95	-0.6
4.2%	19.82	-16.70	-28.30	68.62	107.47	48.46	42.66	786.81	-24.15	-9.31

We can observe several extreme percentage changes in optimal weights which lack particular intuition.

Sensitivity of the “Classical” Mean-Variance Optimal Portfolio to Inputs: Illustration II

Let’s “tweak” the sample mean of Germany by 10% and observe the resulting percentage change in optimal mean-variance portfolio weights.

Percentage Changes in Optimal Portfolio Weights
“Classical” Scenario

	Denmark	Germany	Italy	UK	Portugal	Netherlands	France	Sweden	Norway	Ireland
0.04%	0.13	7.86	-1.88	25.09	-6.54	17.29	-1.88	-1.32	-3.61	-11.47
2.1%	7.9	-26.72	-22.34	24.63	17.39	29.90	8.04	65.72	-0.95	-0.6
4.2%	19.82	-16.70	-28.30	68.62	107.47	48.46	42.66	786.81	-24.15	-9.31

We can observe several extreme percentage changes in optimal weights which lack particular intuition.

Outline

Bayesian Portfolio Selection I

Portfolio selection in the Bayesian setting recognizes the uncertainty about the true means and covariances of returns.

Bayesian Mean-Variance Selection

$$\min_{\omega} \omega' \widetilde{\boldsymbol{\Sigma}} \omega$$

$$\text{s.t. } \omega' \widetilde{\boldsymbol{\mu}} = \mu^*,$$

where

$\widetilde{\boldsymbol{\Sigma}}$ = *Predictive covariance*

$\widetilde{\boldsymbol{\mu}}$ = *Predictive mean*

Bayesian Utility-Based Selection

$$\max_{\omega} E\left[U\left(\omega' \widetilde{\boldsymbol{R}}\right)\right],$$

where

$$E\left[U\left(\omega' \widetilde{\boldsymbol{R}}\right)\right] = \int U\left(\omega' \widetilde{\boldsymbol{R}}\right) \times p\left(\widetilde{R} \mid \boldsymbol{R}\right) \mathrm{d}\boldsymbol{R}.$$

Bayesian Portfolio Selection I

Portfolio selection in the Bayesian setting recognizes the uncertainty about the true means and covariances of returns.

Bayesian Mean-Variance Selection

$$\min_{\omega} \omega' \widetilde{\boldsymbol{\Sigma}} \omega$$

$$\text{s.t. } \omega' \widetilde{\boldsymbol{\mu}} = \mu^*,$$

where

$\widetilde{\boldsymbol{\Sigma}}$ = *Predictive covariance*

$\widetilde{\boldsymbol{\mu}}$ = *Predictive mean*

Bayesian Utility-Based Selection

$$\max_{\omega} E\left[U\left(\omega' \widetilde{\boldsymbol{R}}\right)\right],$$

where

$$E\left[U\left(\omega' \widetilde{\boldsymbol{R}}\right)\right] = \int U\left(\omega' \widetilde{\boldsymbol{R}}\right) \times p\left(\widetilde{R} \mid \boldsymbol{R}\right) \mathrm{d}\boldsymbol{R}.$$

Bayesian Portfolio Selection II

Robustness of optimal portfolio weights is improved substantially in the Bayesian setting.

Percentage Changes in Optimal Portfolio Weights

Bayesian Scenario

	Denmark	Germany	Italy	UK	Portugal	Netherlands	France	Sweden	Norway	Ireland
0.04%	-0.03	0.73	0.04	0.05	0.09	-0.22	-0.01	-0.02	-0.02	-0.01
1.9%	-0.07	1.1	0.04	0.13	0.29	-0.31	-0.05	-0.02	-0.04	-0.01
3.8%	-0.1	1.32	0.04	0.25	0.83	-0.35	-0.06	-0.02	-0.06	-0.02

Conjugate prior distributions are used to compute the optimal weights above—normal distribution for the mean vector of returns and inverted-Wishart distribution for the covariance matrix of returns.

Outline

1. Motivation
   - Main reasons to employ Bayesian methods
   - Focus on Two Areas of Application
   - The Bayesian Paradigm
2. Bayesian Portfolio Selection
   - Sensitivity to Inputs
   - Bayesian Portfolio Selection
   - The Black-Litterman Model
3. Markov Chain Monte Carlo
4. Markov Regime-Switching Models

The Black-Litterman (BL) Model: Overview

The BL model was developed in the early 1990s by the Quantitative Resources Group at Goldman Sachs.

Several features determine its appeal to practitioners:

- Investors specify views on the expected returns of as few assets as they desire
- Expected returns of assets with no views are centered on equilibrium expected returns (as determined by an equilibrium pricing model such as the CAPM)
- Corner solutions, typical to the “classical” portfolio setting (with short sales constraints), are avoided

The Black-Litterman (BL) Model: Overview

The BL model was developed in the early 1990s by the Quantitative Resources Group at Goldman Sachs.

Several features determine its appeal to practitioners:

- Investors specify views on the expected returns of as few assets as they desire
- Expected returns of assets with no views are centered on equilibrium expected returns (as determined by an equilibrium pricing model such as the CAPM)
- Corner solutions, typical to the “classical” portfolio setting (with short sales constraints), are avoided

The Black-Litterman (BL) Model: Overview

The BL model was developed in the early 1990s by the Quantitative Resources Group at Goldman Sachs.

Several features determine its appeal to practitioners:

- Investors specify views on the expected returns of as few assets as they desire
- Expected returns of assets with no views are centered on equilibrium expected returns (as determined by an equilibrium pricing model such as the CAPM)
- Corner solutions, typical to the "classical" portfolio setting (with short sales constraints), are avoided

The Black-Litterman (BL) Model: Overview

The BL model was developed in the early 1990s by the Quantitative Resources Group at Goldman Sachs.

Several features determine its appeal to practitioners:

- Investors specify views on the expected returns of as few assets as they desire
- Expected returns of assets with no views are centered on equilibrium expected returns (as determined by an equilibrium pricing model such as the CAPM)
- Corner solutions, typical to the “classical” portfolio setting (with short sales constraints), are avoided

The Black-Litterman (BL) Model: Overview (Cont'd)

The **core idea of the BL model** is the recognition that

- An investor who is market-neutral with respect to all securities should optimally hold the market
- Only when an investor is more bullish (bearish) than the market with respect to a security or when he believes there is some relative mispricing would optimal portfolio holdings differ from market holdings

The Black-Litterman (BL) Model: Overview (Cont'd)

The **core idea of the BL model** is the recognition that

- An investor who is market-neutral with respect to all securities should optimally hold the market
- Only when an investor is more bullish (bearish) than the market with respect to a security or when he believes there is some relative mispricing would optimal portfolio holdings differ from market holdings

The Black-Litterman (BL) Model: Overview (Cont'd)

The **core idea of the BL model** is the recognition that

- An investor who is market-neutral with respect to all securities should optimally hold the market
- Only when an investor is more bullish (bearish) than the market with respect to a security or when he believes there is some relative mispricing would optimal portfolio holdings differ from market holdings

The Black-Litterman Model: Equilibrium Market Information

Equilibrium risk premiums are derived from the CAPM or through reverse MV optimization:

$$\boldsymbol{\Pi} = \delta \boldsymbol{\Sigma} \omega_{\text{eq}},$$

where

δ = risk-aversion parameter
$\boldsymbol{\Sigma}$ = covariance matrix of returns
ω_{eq} = market-capitalization (unnormalized) weights

The Black-Litterman Model: Equilibrium Market Information

Equilibrium risk premiums are derived from the CAPM or through reverse MV optimization:

$$\mathbf{\Pi} = \delta \mathbf{\Sigma} \boldsymbol{\omega}_{\text{eq}},$$

where

δ = risk-aversion parameter
$\mathbf{\Sigma}$ = covariance matrix of returns
$\boldsymbol{\omega}_{\text{eq}}$ = market-capitalization (unnormalized) weights

The Black-Litterman Model: Investor's Views

Two main types of investor views are:

Absolute view. E.g., next-period's expected returns of assets A and B are 7.4% and 5.5%, respectively.

Relative view. E.g., asset C is expected to outperform assets A and B by 2% next period

Views are expressed as the returns on *view portfolios* composed of the securities involved in the respective views.

The absolute views above correspond to two view portfolios—one long in asset A and the another long in asset B.

The relative view above is usually expressed by means of a *zero-investment* view portfolio which is long in the security expected to outperform (C) and short in the securities expected to underperform (A and B). Different portfolio weighting schemes are possible.

The Black-Litterman Model: Investor's Views

Two main types of investor views are:

Absolute view. E.g., next-period's expected returns of assets A and B are 7.4% and 5.5%, respectively.

Relative view. E.g., asset C is expected to outperform assets A and B by 2% next period

Views are expressed as the returns on *view portfolios* composed of the securities involved in the respective views.

The absolute views above correspond to two view portfolios—one long in asset A and the another long in asset B.

The relative view above is usually expressed by means of a *zero-investment* view portfolio which is long in the security expected to outperform (C) and short in the securities expected to underperform (A and B). Different portfolio weighting schemes are possible.

The Black-Litterman Model: Investor's Views

Two main types of investor views are:

Absolute view. E.g., next-period's expected returns of assets A and B are 7.4% and 5.5%, respectively.

Relative view. E.g., asset C is expected to outperform assets A and B by 2% next period

Views are expressed as the returns on *view portfolios* composed of the securities involved in the respective views.

The absolute views above correspond to two view portfolios—one long in asset A and the another long in asset B.

The relative view above is usually expressed by means of a *zero-investment* view portfolio which is long in the security expected to outperform (C) and short in the securities expected to underperform (A and B). Different portfolio weighting schemes are possible.

The Black-Litterman Model: Investor's Views

Two main types of investor views are:

Absolute view. E.g., next-period's expected returns of assets A and B are 7.4% and 5.5%, respectively.

Relative view. E.g., asset C is expected to outperform assets A and B by 2% next period

Views are expressed as the returns on *view portfolios* composed of the securities involved in the respective views.

The absolute views above correspond to two view portfolios—one long in asset A and the another long in asset B.

The relative view above is usually expressed by means of a *zero-investment* view portfolio which is long in the security expected to outperform (C) and short in the securities expected to underperform (A and B). Different portfolio weighting schemes are possible.

The Black-Litterman Model: Investor's Views

Two main types of investor views are:

Absolute view. E.g., next-period's expected returns of assets A and B are 7.4% and 5.5%, respectively.

Relative view. E.g., asset C is expected to outperform assets A and B by 2% next period

Views are expressed as the returns on *view portfolios* composed of the securities involved in the respective views.

The absolute views above correspond to two view portfolios—one long in asset A and the another long in asset B.

The relative view above is usually expressed by means of a *zero-investment* view portfolio which is long in the security expected to outperform (C) and short in the securities expected to underperform (A and B). Different portfolio weighting schemes are possible.

The Black-Litterman Model: Distributional Assumptions

The covariance matrix of returns, $\mathbf{\Sigma}$, is estimated outside of the model and considered given.

Assumptions about expected returns vector, μ:

Market Info Source: $\mu \sim \mathrm{N}(\mathbf{\Pi}, \tau\mathbf{\Sigma})$,

where τ = scaling factor

Subjective Info Source: $\boldsymbol{P}\mu \sim \mathrm{N}(\boldsymbol{Q}, \Omega)$,

where

$\boldsymbol{P}$ = matrix containing view portfolios
$\boldsymbol{Q}$ = vector of views on expected returns
Ω = prior covariance matrix

The Black-Litterman Model: Distributional Assumptions

The covariance matrix of returns, $\boldsymbol{\Sigma}$, is estimated outside of the model and considered given.

Assumptions about expected returns vector, $\boldsymbol{\mu}$:

Market Info Source: $\boldsymbol{\mu} \sim \mathrm{N}\left(\boldsymbol{\Pi}, \tau\boldsymbol{\Sigma}\right)$,

where τ = scaling factor

Subjective Info Source: $\boldsymbol{P\mu} \sim \mathrm{N}(\boldsymbol{Q}, \Omega)$,

where

$\boldsymbol{P}$ = matrix containing view portfolios
$\boldsymbol{Q}$ = vector of views on expected returns
Ω = prior covariance matrix

The Black-Litterman Model: Distributional Assumptions

The covariance matrix of returns, $\boldsymbol{\Sigma}$, is estimated outside of the model and considered given.

Assumptions about expected returns vector, $\boldsymbol{\mu}$:

Market Info Source: $\boldsymbol{\mu} \sim \mathrm{N}\left(\boldsymbol{\Pi}, \tau\boldsymbol{\Sigma}\right)$,

where τ = scaling factor

Subjective Info Source: $\boldsymbol{P}\boldsymbol{\mu} \sim \mathrm{N}(\boldsymbol{Q}, \Omega)$,

where

$\boldsymbol{P}$ = matrix containing view portfolios
$\boldsymbol{Q}$ = vector of views on expected returns
Ω = prior covariance matrix

The Black-Litterman Model: Distributional Assumptions

The covariance matrix of returns, $\boldsymbol{\Sigma}$, is estimated outside of the model and considered given.

Assumptions about expected returns vector, $\boldsymbol{\mu}$:

Market Info Source: $\boldsymbol{\mu} \sim \mathrm{N}(\boldsymbol{\Pi}, \tau\boldsymbol{\Sigma})$,

where τ = scaling factor

Subjective Info Source: $\boldsymbol{P\mu} \sim \mathrm{N}(\boldsymbol{Q}, \boldsymbol{\Omega})$,

where

$\boldsymbol{P}$ = matrix containing view portfolios
$\boldsymbol{Q}$ = vector of views on expected returns
$\boldsymbol{\Omega}$ = prior covariance matrix

The Black-Litterman Model: Predictive Moments

It can be shown that the posterior mean and covariance of returns in the Black-Litterman model are, respectively:

$$\boldsymbol{M} = \left((\tau\boldsymbol{\Sigma})^{-1} + \boldsymbol{P}'\Omega^{-1}\boldsymbol{P}\right)^{-1}\left((\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\Pi} + \boldsymbol{P}'\Omega^{-1}\boldsymbol{Q}\right)$$

and

$$\boldsymbol{V} = \left((\tau\boldsymbol{\Sigma})^{-1} + \boldsymbol{P}'\Omega^{-1}\boldsymbol{P}\right)^{-1}.$$

The predictive moments of returns are then

$\boldsymbol{M}$ = Predictive mean of returns

$\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \boldsymbol{V}$ = Predictive covariance of returns

The Black-Litterman Model: Predictive Moments

It can be shown that the posterior mean and covariance of returns in the Black-Litterman model are, respectively:

$$\boldsymbol{M} = \left((\tau\boldsymbol{\Sigma})^{-1} + \boldsymbol{P}'\Omega^{-1}\boldsymbol{P}\right)^{-1}\left((\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\Pi} + \boldsymbol{P}'\Omega^{-1}\boldsymbol{Q}\right)$$

and

$$\boldsymbol{V} = \left((\tau\boldsymbol{\Sigma})^{-1} + \boldsymbol{P}'\Omega^{-1}\boldsymbol{P}\right)^{-1}.$$

The predictive moments of returns are then

$\boldsymbol{M}$ = Predictive mean of returns

$\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \boldsymbol{V}$ = Predictive covariance of returns

The Black-Litterman Model: Predictive Moments

It can be shown that the posterior mean and covariance of returns in the Black-Litterman model are, respectively:

$$\boldsymbol{M} = \left((\tau\boldsymbol{\Sigma})^{-1} + \boldsymbol{P}'\Omega^{-1}\boldsymbol{P}\right)^{-1}\left((\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\Pi} + \boldsymbol{P}'\Omega^{-1}\boldsymbol{Q}\right)$$

and

$$\boldsymbol{V} = \left((\tau\boldsymbol{\Sigma})^{-1} + \boldsymbol{P}'\Omega^{-1}\boldsymbol{P}\right)^{-1}.$$

The predictive moments of returns are then

$\boldsymbol{M}$ = Predictive mean of returns

$\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \boldsymbol{V}$ = Predictive covariance of returns

The Black-Litterman Model: Predictive Moments

It can be shown that the posterior mean and covariance of returns in the Black-Litterman model are, respectively:

$$\boldsymbol{M} = \left((\tau\boldsymbol{\Sigma})^{-1} + \boldsymbol{P}'\Omega^{-1}\boldsymbol{P}\right)^{-1}\left((\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\Pi} + \boldsymbol{P}'\Omega^{-1}\boldsymbol{Q}\right)$$

and

$$\boldsymbol{V} = \left((\tau\boldsymbol{\Sigma})^{-1} + \boldsymbol{P}'\Omega^{-1}\boldsymbol{P}\right)^{-1}.$$

The predictive moments of returns are then

$\boldsymbol{M}$ = Predictive mean of returns

$\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \boldsymbol{V}$ = Predictive covariance of returns

The Black-Litterman Model: Illustration I

- Consider the eight constituents of MSCI World Index with largest market capitalization as of Jan 2, 1990
- Daily returns are observed for the period Jan 2, 1990 through Dec 31, 2003
- *Absolute view*: Japan will return 10% on an annual basis
- *Relative view*: Switzerland will outperform U.S. by 5% on an annual basis

$$P = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad Q = \begin{pmatrix} 0.05 \\ 0.1 \end{pmatrix}$$

$\Omega = \{\omega_{ij}\}_{i,j=1}^{2}$, where

$$\omega_{11} = 0.0001, \quad \omega_{22} = 0.0004, \quad \text{and} \quad \omega_{12} = 0$$

The Black-Litterman Model: Illustration I

- Consider the eight constituents of MSCI World Index with largest market capitalization as of Jan 2, 1990
- Daily returns are observed for the period Jan 2, 1990 through Dec 31, 2003
- *Absolute view*: Japan will return 10% on an annual basis
- *Relative view*: Switzerland will outperform U.S. by 5% on an annual basis

$$\boldsymbol{P} = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad , \quad \boldsymbol{Q} = \begin{pmatrix} 0.05 \\ 0.1 \end{pmatrix}$$

$\boldsymbol{\Omega} = \{\omega_{ij}\}_{i,j=1}^{2}$, where

$$\omega_{11} = 0.0001, \quad \omega_{22} = 0.0004, \quad \text{and} \quad \omega_{12} = 0$$

The Black-Litterman Model: Illustration II

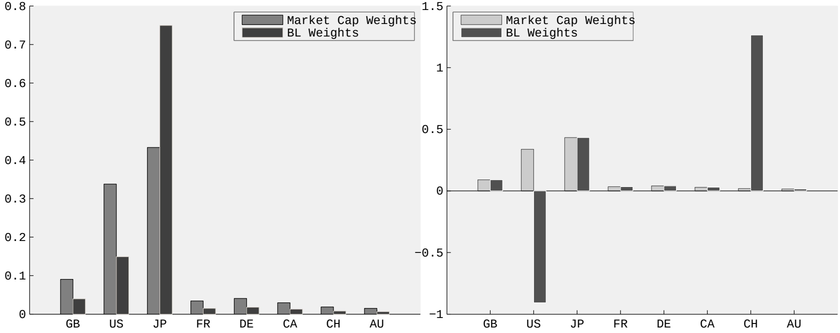
The predictive expected returns are computed:

	GB	US	JP	FR	DE	CA	CH	AU
Absolute View Only	0.0271	0.022	0.0986	0.0324	0.0358	0.0216	0.0278	0.0256
Relative View Only	0.0291	-0.0026	0.0492	0.0368	0.0397	0.0068	0.0458	0.0174
Both Views	0.0292	0.0175	0.0987	0.0353	0.0388	0.0196	0.0334	0.0263

Expected returns are expressed on an annual basis. Since securities are correlated, views on a few assets are propagated through the expected returns on all assets.

The Black-Litterman Model: Illustration III

The Bayesian mean-variance optimal portfolio weights are shown in the plots below.



The plot on the left-hand side corresponds to the absolute view, while the plot on the right-hand side corresponds to the relative view.

Non-Normality of Returns and Non-Normality of Parameters

- The assumption of data normality is not realistic, especially for data with frequency higher than monthly
- So might be the usual assumptions for prior distribution (of the mean), often guided by analytical convenience (so-called "conjugate priors")
- Making realistic assumptions comes at the expense of non-standard posterior and predictive distributions. Employ numerical computational methods (MCMC).
- Next, we provide a brief overview of MCMC and illustrate it with an application to a regime-switching GARCH model.

Non-Normality of Returns and Non-Normality of Parameters

- The assumption of data normality is not realistic, especially for data with frequency higher than monthly
- So might be the usual assumptions for prior distribution (of the mean), often guided by analytical convenience (so-called "conjugate priors")
- Making realistic assumptions comes at the expense of non-standard posterior and predictive distributions. Employ numerical computational methods (MCMC).
- Next, we provide a brief overview of MCMC and illustrate it with an application to a regime-switching GARCH model.

Non-Normality of Returns and Non-Normality of Parameters

- The assumption of data normality is not realistic, especially for data with frequency higher than monthly
- So might be the usual assumptions for prior distribution (of the mean), often guided by analytical convenience (so-called "conjugate priors")
- Making realistic assumptions comes at the expense of non-standard posterior and predictive distributions. Employ numerical computational methods (MCMC).
- Next, we provide a brief overview of MCMC and illustrate it with an application to a regime-switching GARCH model.

Non-Normality of Returns and Non-Normality of Parameters

- The assumption of data normality is not realistic, especially for data with frequency higher than monthly
- So might be the usual assumptions for prior distribution (of the mean), often guided by analytical convenience (so-called “conjugate priors”)
- Making realistic assumptions comes at the expense of non-standard posterior and predictive distributions. Employ numerical computational methods (MCMC).
- Next, we provide a brief overview of MCMC and illustrate it with an application to a regime-switching GARCH model.

Markov Chain Monte Carlo (MCMC)

Basic Idea of MCMC

- Construct and simulate a Markov chain, whose stationary distribution is the posterior (predictive) distribution, $p\,(\theta \mid \boldsymbol{R})$
- The simulation output from the chain is a sample of (nearly) identical (but not independent) draws from the posterior (predictive) distribution
- The *Metropolis-Hastings algorithm* is the core of MCMC. Other algorithms, such as the Gibbs sampler, are variants of it.

Metropolis-Hastings (MH) Algorithm

- Denote by $p(\theta \mid \boldsymbol{R})$ the unnormalized posterior density from which direct sampling is not possible.
- Suppose that there is a density which closely approximates the p and denote it by $q(\theta)$ ("the proposal density"). It may or may not depend on the previous draw in the iterative MH algorithm ($\theta^{(t-1)}$).
- The MH algorithm consists of the steps below. At iteration t,
  - Draw θ^* from $q(\theta)$
  - Accept the draw with probability

$$a(\theta^*, \theta^{(t-1)}) = \min\left\{1, \frac{p(\theta^*) \,/\, q(\theta^* \mid \theta^{(t-1)})}{p(\theta^{(t-1)}) \,/\, q(\theta^{(t-1)} \mid \theta^*)}\right\} \quad (1)$$

Metropolis-Hastings (MH) Algorithm

- Denote by $p(\theta \mid \boldsymbol{R})$ the unnormalized posterior density from which direct sampling is not possible.
- Suppose that there is a density which closely approximates the p and denote it by $q(\theta)$ ("the proposal density"). It may or may not depend on the previous draw in the iterative MH algorithm ($\theta^{(t-1)}$).
- The MH algorithm consists of the steps below. At iteration t,
  - Draw θ^* from $q(\theta)$
  - Accept the draw with probability

$$a(\theta^*, \theta^{(t-1)}) = \min\left\{1, \frac{p(\theta^*) \,/\, q(\theta^* \mid \theta^{(t-1)})}{p(\theta^{(t-1)}) \,/\, q(\theta^{(t-1)} \mid \theta^*)}\right\} \quad (1)$$

Metropolis-Hastings (MH) Algorithm

- Denote by $p(\theta \mid \boldsymbol{R})$ the unnormalized posterior density from which direct sampling is not possible.
- Suppose that there is a density which closely approximates the p and denote it by $q(\theta)$ ("the proposal density"). It may or may not depend on the previous draw in the iterative MH algorithm ($\theta^{(t-1)}$).
- The MH algorithm consists of the steps below. At iteration t,
  - Draw θ^* from $q(\theta)$
  - Accept the draw with probability

$$a(\theta^*, \theta^{(t-1)}) = \min\left\{1, \frac{p(\theta^*) / q(\theta^* \mid \theta^{(t-1)})}{p(\theta^{(t-1)}) / q(\theta^{(t-1)} \mid \theta^*)}\right\} \quad (1)$$

Metropolis-Hastings (MH) Algorithm

- Denote by $p(\theta \mid \boldsymbol{R})$ the unnormalized posterior density from which direct sampling is not possible.
- Suppose that there is a density which closely approximates the p and denote it by $q(\theta)$ ("the proposal density"). It may or may not depend on the previous draw in the iterative MH algorithm ($\theta^{(t-1)}$).
- The MH algorithm consists of the steps below. At iteration t,
  - Draw θ^* from $q(\theta)$
  - Accept the draw with probability

$$a(\theta^*, \theta^{(t-1)}) = \min\left\{1, \frac{p(\theta^*) \,/\, q(\theta^* \mid \theta^{(t-1)})}{p(\theta^{(t-1)}) \,/\, q(\theta^{(t-1)} \mid \theta^*)}\right\} \quad (1)$$

Metropolis-Hastings (MH) Algorithm

- Denote by $p(\theta \mid \boldsymbol{R})$ the unnormalized posterior density from which direct sampling is not possible.
- Suppose that there is a density which closely approximates the p and denote it by $q(\theta)$ ("the proposal density"). It may or may not depend on the previous draw in the iterative MH algorithm ($\theta^{(t-1)}$).
- The MH algorithm consists of the steps below. At iteration t,
  - Draw θ^* from $q(\theta)$
  - Accept the draw with probability

$$a(\theta^*, \theta^{(t-1)}) = \min\left\{1, \frac{p(\theta^*) \,/\, q(\theta^* \mid \theta^{(t-1)})}{p(\theta^{(t-1)}) \,/\, q(\theta^{(t-1)} \mid \theta^*)}\right\} \quad (1)$$

Markov Regime-Switching (MS) Volatility Models

- The volatility process may not be constant through time. There are two broad categories of changes in the volatility parameters:
  - Structural breaks—permanent changes in the volatility parameters; e.g. stock market crashes, changes in data collection practices, etc.
  - Regime changes—reversible transitions of the parameters among a finite number of states of the world; e.g. business cycles
- Regime-switching model estimation is generally quite complex. Maximum likelihood estimation could be problematic and burdensome.
- Bayesian estimation uses the MCMC toolbox and deals elegantly with complexities. Computationally-intensive.

Markov Regime-Switching (MS) Volatility Models

- The volatility process may not be constant through time. There are two broad categories of changes in the volatility parameters:
  - Structural breaks—permanent changes in the volatility parameters; e.g. stock market crashes, changes in data collection practices, etc.
  - Regime changes—reversible transitions of the parameters among a finite number of states of the world; e.g. business cycles
- Regime-switching model estimation is generally quite complex. Maximum likelihood estimation could be problematic and burdensome.
- Bayesian estimation uses the MCMC toolbox and deals elegantly with complexities. Computationally-intensive.

Markov Regime-Switching (MS) Volatility Models

- The volatility process may not be constant through time. There are two broad categories of changes in the volatility parameters:
  - Structural breaks—permanent changes in the volatility parameters; e.g. stock market crashes, changes in data collection practices, etc.
  - Regime changes—reversible transitions of the parameters among a finite number of states of the world; e.g. business cycles
- Regime-switching model estimation is generally quite complex. Maximum likelihood estimation could be problematic and burdensome.
- Bayesian estimation uses the MCMC toolbox and deals elegantly with complexities. Computationally-intensive.

Markov Regime-Switching (MS) Volatility Models

- The volatility process may not be constant through time. There are two broad categories of changes in the volatility parameters:
  - Structural breaks—permanent changes in the volatility parameters; e.g. stock market crashes, changes in data collection practices, etc.
  - Regime changes—reversible transitions of the parameters among a finite number of states of the world; e.g. business cycles
- Regime-switching model estimation is generally quite complex. Maximum likelihood estimation could be problematic and burdensome.
- Bayesian estimation uses the MCMC toolbox and deals elegantly with complexities. Computationally-intensive.

MS GARCH(1,1) Process

Assume the following simple dynamics for the return and the volatility:

$$
\begin{aligned}
r_t &= \mathbf{X}_t\gamma + \sigma_{t|t-1}\epsilon_t \\
\sigma^2_{t|t-1} &= \omega + \alpha u^2_{t-1} + \beta\sigma^2_{t-1|t-2}
\end{aligned}
$$

- If regimes are present in the volatility dynamics but are unaccounted for, volatility forecasts would overestimate volatility in low-volatility states and overestimate volatility in high-volatility states.
- Flexible specification of volatility dynamics: *all three GARCH parameters change across regimes*. Intuition: the way variance responds to past return shocks and volatility varies across regimes.

MS GARCH(1,1) Process

Assume the following simple dynamics for the return and the volatility:

$$\begin{aligned} r_t &= \boldsymbol{X}_t\gamma + \sigma_{t|t-1}\epsilon_t \\ \sigma^2_{t|t-1} &= \omega + \alpha u^2_{t-1} + \beta\sigma^2_{t-1|t-2} \end{aligned}$$

- If regimes are present in the volatility dynamics but are unaccounted for, volatility forecasts would overestimate volatility in low-volatility states and overestimate volatility in high-volatility states.
- Flexible specification of volatility dynamics: *all three GARCH parameters change across regimes*. Intuition: the way variance responds to past return shocks and volatility varies across regimes.

MS GARCH(1,1) Process

Assume the following simple dynamics for the return and the volatility:

$$
\begin{aligned}
r_t &= \boldsymbol{X}_t \gamma + \sigma_{t|t-1} \epsilon_t \\
\sigma^2_{t|t-1} &= \omega + \alpha u^2_{t-1} + \beta \sigma^2_{t-1|t-2}
\end{aligned}
$$

- If regimes are present in the volatility dynamics but are unaccounted for, volatility forecasts would overestimate volatility in low-volatility states and overestimate volatility in high-volatility states.
- Flexible specification of volatility dynamics: *all three GARCH parameters change across regimes*. Intuition: the way variance responds to past return shocks and volatility varies across regimes.

MS GARCH(1,1) Process: Volatility Specification

- Regimes are usually modeled as being driven endogenously by an unobserved state variable. Denote it by S_t. In a three-regime scenario, $S_t = \{1, 2, 3\}$, for $t = 1, \ldots, T$.
- Conditional volatility is then expressed as

$$\sigma^2_{t|t-1} = \omega_{S_t} + \alpha_{S_t} u^2_{t-1} + \beta_{S_t} \sigma^2_{t-1|t-2}.$$

- *Conditional on the regime path*, $\boldsymbol{S} = \{S_t\}_{t=1}^T$, the volatility dynamics is that of a simple GARCH(1,1) process.

MS GARCH(1,1) Process: Volatility Specification

- Regimes are usually modeled as being driven endogenously by an unobserved state variable. Denote it by S_t. In a three-regime scenario, $S_t = \{1, 2, 3\}$, for $t = 1, \ldots, T$.
- Conditional volatility is then expressed as

$$\sigma^2_{t|t-1} = \omega_{S_t} + \alpha_{S_t} u^2_{t-1} + \beta_{S_t} \sigma^2_{t-1|t-2}.$$

- *Conditional on the regime path*, $\mathbf{S} = \{S_t\}_{t=1}^T$, the volatility dynamics is that of a simple GARCH(1,1) process.

MS GARCH(1,1) Process: Volatility Specification

- Regimes are usually modeled as being driven endogenously by an unobserved state variable. Denote it by S_t. In a three-regime scenario, $S_t = \{1, 2, 3\}$, for $t = 1, \ldots, T$.
- Conditional volatility is then expressed as

$$\sigma^2_{t|t-1} = \omega_{S_t} + \alpha_{S_t} u^2_{t-1} + \beta_{S_t} \sigma^2_{t-1|t-2}.$$

- *Conditional on the regime path*, $\boldsymbol{S} = \{S_t\}_{t=1}^{T}$, the volatility dynamics is that of a simple GARCH(1,1) process.

MS GARCH(1,1) Process: Specification of State Variable Dynamics

- In regime-switching models, the state variable, S_t, follows a first-order (discrete) Markov chain.
- Its dynamics, in the three-regime case, is governed by the matrix of transition probabilities,

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix},$$

  where $\pi_{i,j} = P(S_t = i \mid S_t = j)$.
- The main estimation difficulty comes from the path dependence of the volatility dynamics

MS GARCH(1,1) Process: Specification of State Variable Dynamics

- In regime-switching models, the state variable, S_t, follows a first-order (discrete) Markov chain.
- Its dynamics, in the three-regime case, is governed by the matrix of transition probabilities,

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix},$$

where $\pi_{i,j} = P(S_t = i \mid S_t = j)$.

- The main estimation difficulty comes from the path dependence of the volatility dynamics

MS GARCH(1,1) Process: Specification of State Variable Dynamics

- In regime-switching models, the state variable, S_t, follows a first-order (discrete) Markov chain.
- Its dynamics, in the three-regime case, is governed by the matrix of transition probabilities,

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix},$$

where $\pi_{i,j} = P(S_t = i \mid S_t = j)$.

- The main estimation difficulty comes from the path dependence of the volatility dynamics

MS GARCH(1,1) Process: Estimation

- MS GARCH estimation in the classical (frequentist) setting involves integrating out the unobserved state variable, S_t. Difficult task.
- The Bayesian framework deals with unobserved (hidden) variables by simulating them together with all other model parameters.
- Employ MCMC for simulation from the posterior distributions.

MS GARCH(1,1) Process: Estimation

- MS GARCH estimation in the classical (frequentist) setting involves integrating out the unobserved state variable, S_t. Difficult task.
- The Bayesian framework deals with unobserved (hidden) variables by simulating them together with all other model parameters.
- Employ MCMC for simulation from the posterior distributions.

MS GARCH(1,1) Process: Illustration

- Stock returns (MSCI Canada Country Index) are assumed to follow Student's *t* distribution with ν degrees of freedom. The factor returns, $\boldsymbol{X}_t$, are extracted with principal components analysis.
- The parameters to estimate in the model and their prior distributions are:
  - $\gamma \sim \mathrm{N}\left(\gamma_0, \boldsymbol{V}_0\right)$
  - $\boldsymbol{\theta}_i \equiv (\omega_i, \alpha_i, \gamma_i) \sim \mathrm{N}\left(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right), i = 1, 2, 3$
  - $\boldsymbol{\pi}_i \equiv (\pi_{i1}, \pi_{i2}, \pi_{i3}) \sim \mathrm{Dirichlet}\left(a_{i1}, a_{i2}, a_{i3}\right), i = 1, 2, 3$
  - $\nu \sim \exp(\lambda)$

MS GARCH(1,1) Process: Illustration

- Stock returns (MSCI Canada Country Index) are assumed to follow Student's *t* distribution with ν degrees of freedom. The factor returns, $\boldsymbol{X}_t$, are extracted with principal components analysis.
- The parameters to estimate in the model and their prior distributions are:
    - $\gamma \sim \text{N}\,(\gamma_0, \boldsymbol{V}_0)$
    - $\boldsymbol{\theta}_i \equiv (\omega_i, \alpha_i, \gamma_i) \sim \text{N}\,(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1, 2, 3$
    - $\boldsymbol{\pi}_i \equiv (\pi_{i1}, \pi_{i2}, \pi_{i3}) \sim \text{Dirichlet}\,(a_{i1}, a_{i2}, a_{i3})$, $i = 1, 2, 3$
    - $\nu \sim \exp(\lambda)$

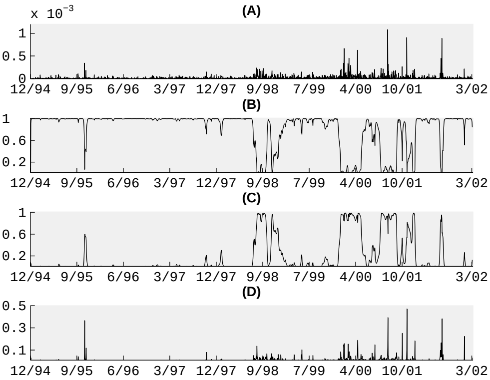
MS GARCH(1,1) Process: Parameter Posterior Means

The joint posterior distribution of the regime-switching model parameters is computed using a block-structure version of the MH algorithm. The posterior means are given below:

	Conditional Volatility Parameters						Transition Probabilities		
Regime 1	ω_1	1.4e-5	α_1	0.03	β_1	0.03	0.98	0.01	0.01
Regime 2	ω_2	6.8e-5	α_2	0.07	β_2	0.34	0.04	0.93	0.03
Regime 3	ω_3	0.7	α_3	0.73	β_3	0.54	0.46	0.35	0.19

Regression Parameters						**Degrees of Freedom**
γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	ν
-9.5e-5	0.3	-0.08	-0.64	0.55	0.07	29.28

MS GARCH(1,1) Process: Posterior Regime Probabilities



(A) shows the squared return innovations; (B), (C), and (D) show the posterior probabilities of regimes 1, 2, and 3.

Summary

- The Bayesian framework accounts for uncertainty, provides modeling flexibility, and allows for views to be incorporated into the investment management decision-making process
- Efficient computational algorithms for dealing with complicated models are constantly being developed
- One of the most challenging areas remains prior distributions specification

Summary

- The Bayesian framework accounts for uncertainty, provides modeling flexibility, and allows for views to be incorporated into the investment management decision-making process
- Efficient computational algorithms for dealing with complicated models are constantly being developed
- One of the most challenging areas remains prior distributions specification

Summary

- The Bayesian framework accounts for uncertainty, provides modeling flexibility, and allows for views to be incorporated into the investment management decision-making process
- Efficient computational algorithms for dealing with complicated models are constantly being developed
- One of the most challenging areas remains prior distributions specification

Thank you for your attention.