Bayesian Analysis and Arbitrage Trading

Ziren Wang, Hanxuan Lin

November 29, 2022

Outline

Outline

Arbitrage Trading

Keys of arbitrage trading

In general, arbitrageurs trade on a short-lived price difference of one asset in more than one market to make profit. In the scenario of futures, there usually exists an opportunity of spot future arbitrage.

A prediction on spot future price difference can be achieved by either non-Bayesian and Bayesian time series model. This difference works as a signal for us to determine whether to buy or sell. Here we are going to use Bayesian method to make prediction.

Arbitrage Trading

BSW Arbitrage Strategy:

1. Make a prediction on the price difference signal(PE) based on Bayesian models
2. Compare the predicted signal and current signal(S0). Under the following rule we will either long(1) or short(-1) the price difference:

$$Trade\ Signal = \begin{cases} 1, & if((S_0 - PE) < -max(Fee, r)) \\ -1, & if((S_0 - PE) > max(Fee, r)) \end{cases}$$

   where Fee is the arbitrage cost and r is market return.

3. Monitor incoming signals and update the predicted signal(ES'). If the latest signal(S') satisfies that:

$$\begin{cases} S' \geq ES', & \text{close the long position} \\ S' \leq ES', & \text{close the short position} \end{cases}$$

Outline

Simple Bayesian

Linear Regression Form[1]

$$Y_t = \beta X_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2) \tag{1}$$

Likelihood Function

$$F(Y_t|\beta, \sigma^2) = (2\pi\sigma^2)^{\frac{-T}{2}} exp(-\frac{(Y_t - \beta X_t)^T (Y_t - \beta X_t)}{2\sigma^2}) \tag{2}$$

Posterior Distribution

$$H(\beta, \sigma^2|Y_t) \propto F(\beta, \sigma^2|Y_t) \times P(\beta, \sigma^2) \tag{3}$$

by Bayes Rule P(A | $B) = \frac{P(B|A)P(A)}{P(B)}$.

[1] Here we have a OLS solution of $\hat{\beta} = (X'X)^{-1}(X'Y)$ and $\sigma^2 = \frac{\epsilon'\epsilon}{T}$

Outline

Example: AR(2)

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \epsilon \tag{4}$$

$$\beta = [\alpha, \beta_1, \beta_2], X_t = [1, Y_{t-1}, Y_{t-2}]$$

Set a normal prior for our Beta coefficients with mean = 0 and variance = 1:

1. prior for mean:

$$\begin{pmatrix} \alpha^0 \\ \beta_1^0 \\ \beta_2^0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

2. prior for variance:

$$\begin{pmatrix} \Sigma_a & 0 & 0 \\ 0 & \Sigma_{B1} & 0 \\ 0 & 0 & \Sigma_{B2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Example: AR(2)

AR(2) Model: (Continued) Set a inverse gamma prior(which is conjugate) for our variance coefficients:

1. prior for variance parameter:

$$p(\sigma^2) \sim \Gamma^{-1}(\frac{T_0}{2}, \frac{\theta_0}{2})$$

2. Here we have arbitrarily set the hyperparameters of inv-gamma to be $T_0 = 1$ and $\theta_0 = 0.1$.

Example: AR(2)

Gibbs Sampling: A normal posterior mean(M) and variance(V) could be obtained as follows [4]

1.
$$M = (\Sigma_0^{-1} + \frac{1}{\sigma^2} X_t' X_t)^{-1} (\Sigma_0^{-1} \beta_0 + \frac{1}{\sigma^2} X_t' Y_t) \tag{5}$$

2.
$$V = (\Sigma_0^{-1} + \frac{1}{\sigma^2} X_t' X_t)^{-1} \tag{6}$$

A random draw can be achieved in a simple way:

$$B^1 = M^\star + [\bar{B} \times V^{\star 1/2}]^T \tag{7}$$

by sampling a vector from standard normal distribution and transforming it as shown above.

Example: AR(2)

Gibbs Sampling:To sample from the right inverse Gamma posterior, we need to compute the T1:

$$T1 = T_0 + T\,(\textit{number of samples})$$

and scale θ_1 :

$$\theta_1 = \theta_0 + \epsilon' \epsilon^2 \tag{8}$$

We will sample T1 variables from a standard normal distribution z0 and then make the following adjustment:

$$z = \frac{\theta_1}{z0' z0}, \quad z0 \sim N(0, 1) \tag{9}$$

and z is now a draw from the correct Inverse Gamma posterior.

[2] $\epsilon = Y - XB$, where B is a draw from conditional posterior distribution

Outline

Motivation

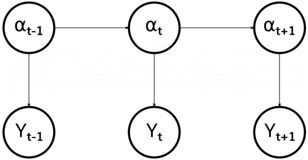
- **Time Series**
  Price is naturally a sequence, Time Dependence needs to be modeled.
- **Structural Time Series**
  - S(AR)IMA targets to covariance stationary series, requires explicitly specify the right orders of different time components (level, trend, seasonality, etc)
  - Structural time series can directly model various components (regarded as hidden states) and allow multiple seasonalities and related sequences
- **Bayesian Structural Time Series**
  - Spike-And-Slab Priors impose sparsity, and MCMC allows the estimation of probability of inclusion
  - Time Series is handled through the Kalman Filter while taking into account the common time components, which naturally have Bayesian interpretations and introduce causation beyond correlation[1]
  - Bayes model averaging combine the results and prediction calculation

BSTS Structure I

General Structural Model in State Space Form:

$$y_t = Z_t^T \alpha_t + \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, H_t) \tag{10}$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t \quad \eta_t \sim \mathcal{N}(0, Q_t) \tag{11}$$



- **(10) Observation Equation**: links the observed data y_t with the unobserved latent state α_t.
- **(11) Transition Equation**: defines how the latent state evolves over time
- Models are modular. This formulation can express very large class of models including SARIMA.

BSTS Structure II

E.g. Basic Structural Model with Regression:

$$\begin{aligned}
&observation : y_t = \mu_t + \tau_t + \beta^T \boldsymbol{x_t} + \epsilon_t \\
&trend : \mu_t = \mu_{t-1} + \delta_{t-1} + u_t \\
&slope : \delta_t = \delta_{t-1} + v_t \\
&seasonality : \tau_t = -\sum_{s=1}^{S-1} \tau_{t-s} + w_t
\end{aligned} \tag{12}$$

Formulating in the state space form:
$\alpha_t = (1, \mu_t, \delta_t, \tau_t)^T$, $Z_t = (\beta^T x_t, 1, 0, 1)^T$, $H_t = \sigma_\epsilon^2$
$\eta_t = (u_t, v_t, w_t)^T$ and $Q_t = \begin{bmatrix} \sigma_u^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_w^2 \end{bmatrix}$
The parameters are the $\sigma_\epsilon^2, \sigma_u^2, \sigma_v^2, \sigma_w^2$ and regression coefficients β

Spike And Slab Regression I

- Motivation: To impose sparsity when the length of the series is shorter than the dimension of the external predictors $\boldsymbol{x_t}$ to avoid overfitting
- Assume: $p(\beta, \gamma, \sigma_\epsilon^2) = p(\beta_\gamma | \gamma, \sigma_\epsilon^2) P(\sigma_\epsilon^2 | \gamma) p(\gamma)$
- Spike Prior: $\gamma_k = \begin{cases} 1 & \beta_k \neq 0 \\ 0 & \beta_k = 0 \end{cases}$, $\gamma \sim \prod_{k=1}^{K} \pi_k^{\gamma_k} (1 - \pi_k)^{1-\gamma_k}$

  π_k is the inclusion probability of predictor k, can be simplified to $\pi = p/K$, where p is the expected model size
- Slab Conditionals: $\beta_\gamma | \sigma_\epsilon^2, \gamma \sim \mathcal{N}(b_\gamma, \sigma_\epsilon^2 \Omega_\gamma^{-1}), \sigma_\epsilon^2 | \gamma \sim IGa(\frac{\nu}{2}, \frac{ss}{2})$
  - Can reasonably assume the prior means b is $\mathbf{0}$, or make it informative by adjusting prior sum of squares ss and prior sample size ν
  - Can impose Zellner's g-prior for prior information matrix Ω by setting $\Omega = \kappa(w\boldsymbol{X}^T\boldsymbol{X} + (1-w)diag(\boldsymbol{X}^T\boldsymbol{X}))/n$, where $\boldsymbol{X}$ is the design matrix, obtained by staking the predictors $\boldsymbol{x}_t$ in row t

Spike And Slab Regression II

Let $y_t^* = y_t - Z_t^{*T}\alpha_t$, where $Z_t^{*T}\alpha_t$ is the observation matrix with $\beta^T x_t$ set to zero. Let $\boldsymbol{y}^* = y_{1:n}^*$ so that $\boldsymbol{y}^*$ is $\boldsymbol{y}$ with the time series component substracted out. Then we have the following conditional posterior[5]:

$$\begin{aligned} \beta_\gamma | \sigma_\epsilon^2, \gamma, \boldsymbol{y}^* &\sim \mathcal{N}(\tilde{\beta_\gamma}, \sigma_\epsilon^2 V_\gamma^{-1}) \\ \sigma_\epsilon^2 | \gamma, \boldsymbol{y}^* &\sim IGa(\frac{N}{2}, \frac{SS_\gamma}{2}) \end{aligned} \tag{13}$$

Where the sufficient statistics can be written as:

$$\begin{aligned} &V_\gamma = (\boldsymbol{X}^T\boldsymbol{X})_\gamma + \Omega_\gamma, \quad \tilde{\beta_\gamma} = V_\gamma^{-1}(\boldsymbol{X}_\gamma^T\boldsymbol{y}^* + \Omega_\gamma b_\gamma) \\ &N = \nu + n, \quad SS_\gamma = ss + \boldsymbol{y}^{*\boldsymbol{T}}\boldsymbol{y}^* + b_\gamma^T\Omega_\gamma b_\gamma - \tilde{\beta_\gamma^T} V_\gamma \tilde{\beta_\gamma} \end{aligned} \tag{14}$$

Spike And Slab Regression III

Since the conditional are conjugate, we can marginalize over β_γ and $\frac{1}{\sigma_\epsilon^2}$ to obtain [3]:

$$\gamma | \boldsymbol{y}^* \sim C(\boldsymbol{y}^*) \frac{|\Omega_\gamma|^{\frac{1}{2}}}{|V_\gamma|^{\frac{1}{2}}} \frac{p(\gamma)}{SS_\gamma^{\frac{N}{2}-1}} \tag{15}$$

Where $C(\boldsymbol{y}^*)$ is a normalizing constant that depends on $\boldsymbol{y}^*$ but not γ

Notice: The above posterior distribution of γ places positive probability on coefficients being zero (as opposed to probability density like lasso). Therefore, the sparsity in this model is a feature of the full posterior distribution, and not simply the value at the mode.

Parameter Learning and Forecasting

Let θ denote the set of model parameters other than β and σ_ϵ^2. We can deploy MCMC with the repetitive steps:

1. Simulate $\alpha \sim p(\alpha|\boldsymbol{y}, \theta, \beta, \sigma_\epsilon^2)$ by Kalman smoother [2]
2. Simulate $\theta \sim p(\theta|\boldsymbol{y}, \alpha, \beta, \sigma_\epsilon^2)$
3. Simulate β, σ_ϵ^2 from a MC with stationary distribution $p(\beta, \sigma_\epsilon^2|\boldsymbol{y}, \alpha, \theta)$

Then by definition the posterior predictive distribution of $\tilde{y}$ is:

$$p(\tilde{y}|\boldsymbol{y}) = \int p(\tilde{y}|\phi)p(\phi|\boldsymbol{y})d\phi \tag{16}$$

Where $\phi = (\theta, \beta, \sigma_\epsilon^2, \alpha)$ is all components that we need to simulate. Since multiple simulations yield multiple ϕ, we then can get multiple sample from $p(\tilde{y}|\boldsymbol{y})$ too. Finally we can summarize it as a point estimation (e.g. $E(\tilde{y}|\boldsymbol{y})$) over all samples.

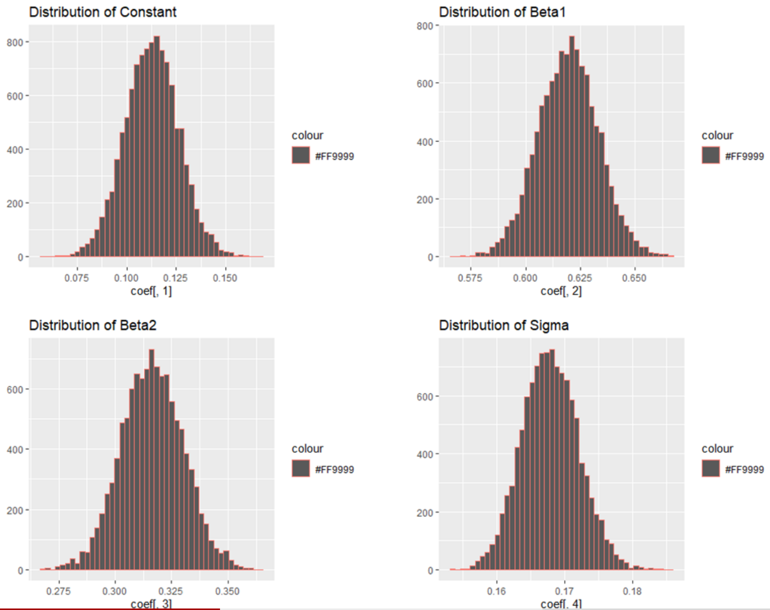
Outline

Experiment: Dataset

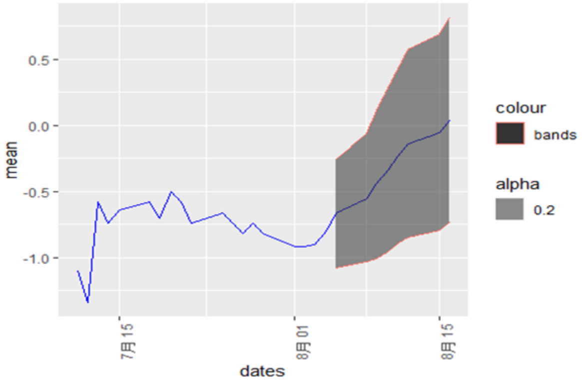
- **Data**: Daily Gold Future Price and 5 associated potential predictors
- **Time Horizon**: 10 years, August 2012 - August 2022
- **Purpose**: Trade when the price difference between the near-month contract and the far-month contract is forecasted "large enough"

Variable	Meaning
ETF_H	Gold ETF Position (oz)
ATM_O_VOL	1-month at-the-money option volatility
SPEC_LONG	Gold speculative net long volume
SPOT_INTL	International gold price
ETF_VOL	Gold Miners ETF Volatility Index
AU_S.SHF	Gold near-month contract price
AU_SHF	Gold far-month contract price

Experiment: AR(2) Estimation

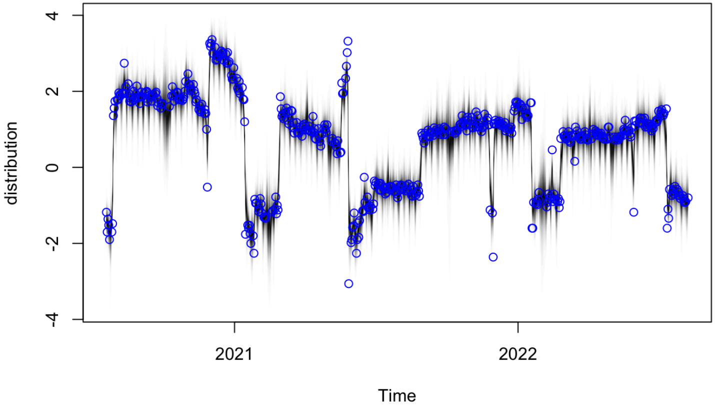


Experiment: AR(2) Forecasting



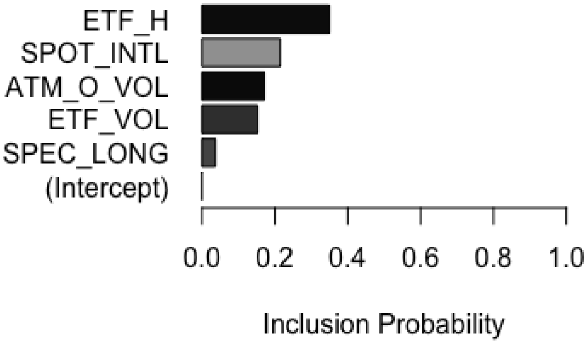
Here we made a prediction of 10 business days forward, and show the 90% confidence interval.

Experiment: BSTS I



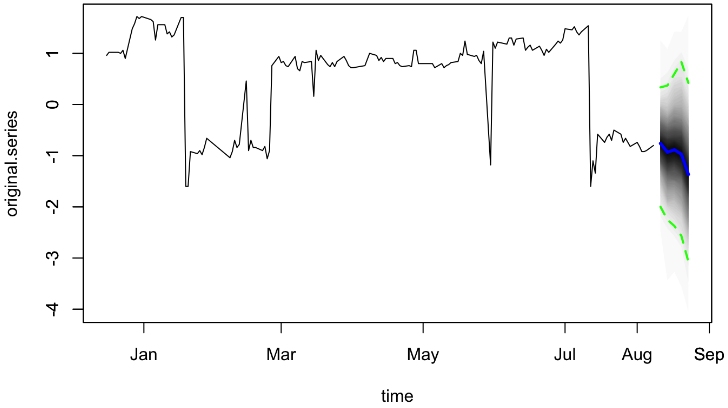
Posterior distribution $p(Z_t^T \alpha_t | \boldsymbol{y})$. The blue point is the actual observation.

Experiment: BSTS II



Posterior inclusion probability of each of predictor.

Experiment: BSTS III



Posterior 5-step forecasting. The blue curve is the actual observation.

Kay H Brodersen, Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L Scott.
Inferring causal impact using bayesian structural time-series models.
The Annals of Applied Statistics, pages 247–274, 2015.

James Durbin and Siem Jan Koopman.
A simple and efficient simulation smoother for state space time series analysis.
Biometrika, 89(3):603–616, 2002.

Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin.
Bayesian data analysis.
Chapman and Hall/CRC, 1995.

James Douglas Hamilton.
Time series analysis.
Princeton university press, 2020.

Steven L Scott and Hal R Varian.
Predicting the present with bayesian structural time series.

International Journal of Mathematical Modelling and Numerical Optimisation, 5(1-2):4–23, 2014.