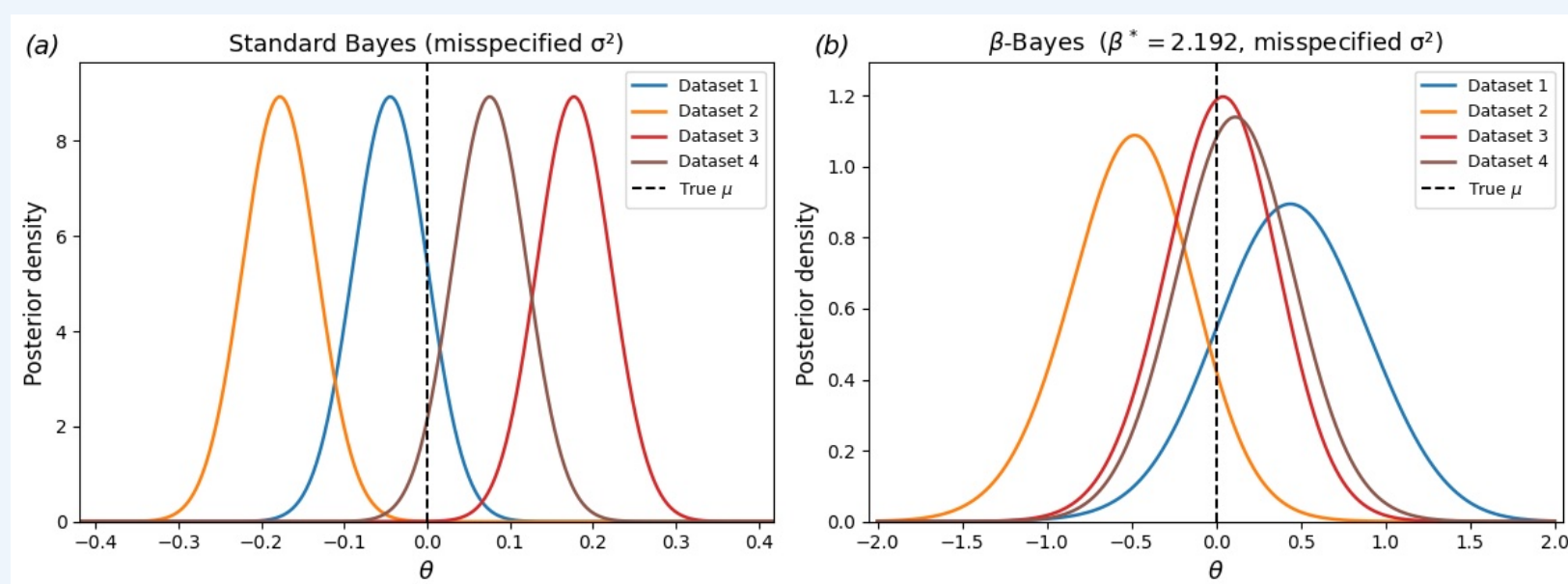


Abstract

Being a widely used statistical framework in financial settings, Bayesian inference rests on several assumptions, including that the likelihood is specified correctly. In practice, however, this assumption is often violated. This may cause posteriors computed on independent datasets of the same data generating process to **significantly disagree**, having **substantially no overlap**. This can lead to unreliable and unstable inference. We introduce a way to quantify the amount of disagreement between posteriors and establish a **target amount of overlap** — how much the posteriors computed on independent datasets of the same data generating process should overlap. We show that standard Bayesian inference **violates this threshold** under model misspecification, and introduce a method based on Generalized Bayesian Inference with β -divergence loss which can recover **well-calibrated posteriors** under the misspecified model. We theoretically show that the proposed methodology recovers the target amount of overlap under misspecification where standard Bayes fails. Empirically, we validate the framework across CAPM and Fama–French 3-factor models on three U.S. equities (JNJ, BAC, AAPL, 2012–2016), demonstrating improved stability in parameter inference and predictive inference.

Problem Setup

Standard Bayes posteriors on independent data realizations of the *same* DGP can have essentially **no overlap** under model misspecification:



Background

Generalized Bayesian Inference (Bissiri, Holmes & Walker, 2016) updates beliefs even when the likelihood is wrong:

$$\pi^{(\ell)}(\theta \mid x_{1:N}) \propto \pi(\theta) \exp\left(-w \sum_{i=1}^N \ell(\theta, x_i)\right)$$

β -Bayes uses the **β -divergence loss** (Basu et al., 1998):

$$\ell^{(\beta)}(\theta, x) = -\frac{1}{\beta-1} p(x \mid \theta)^{\beta-1} + \frac{1}{\beta} \int p(z \mid \theta)^{\beta} dz$$

Larger β downweights outliers, trading robustness for data efficiency.

Methodology

Goals:

- Establish a way of measuring the **disagreement** between posteriors
- Define the **correct amount** of disagreement (the well-specified threshold)
- Show that **standard Bayes** under model misspecification violates the correct overlap
- Suggest a methodology based on **β -Bayes** which restores well-calibrated posteriors

We measure disagreement via the **average pairwise discrepancy**:

$$\bar{d}_{1/2} = \iint d_{1/2}(\Pi_x, \Pi_y) P_{\circ}(dx) P_{\circ}(dy)$$

We choose the **Rényi- $\frac{1}{2}$ discrepancy** as our measure d :

$$d_{1/2}(\Pi_X, \Pi_Y) = -2 \log \int \sqrt{\pi_X(\theta) \pi_Y(\theta)} d\theta$$

Well-Specified Threshold

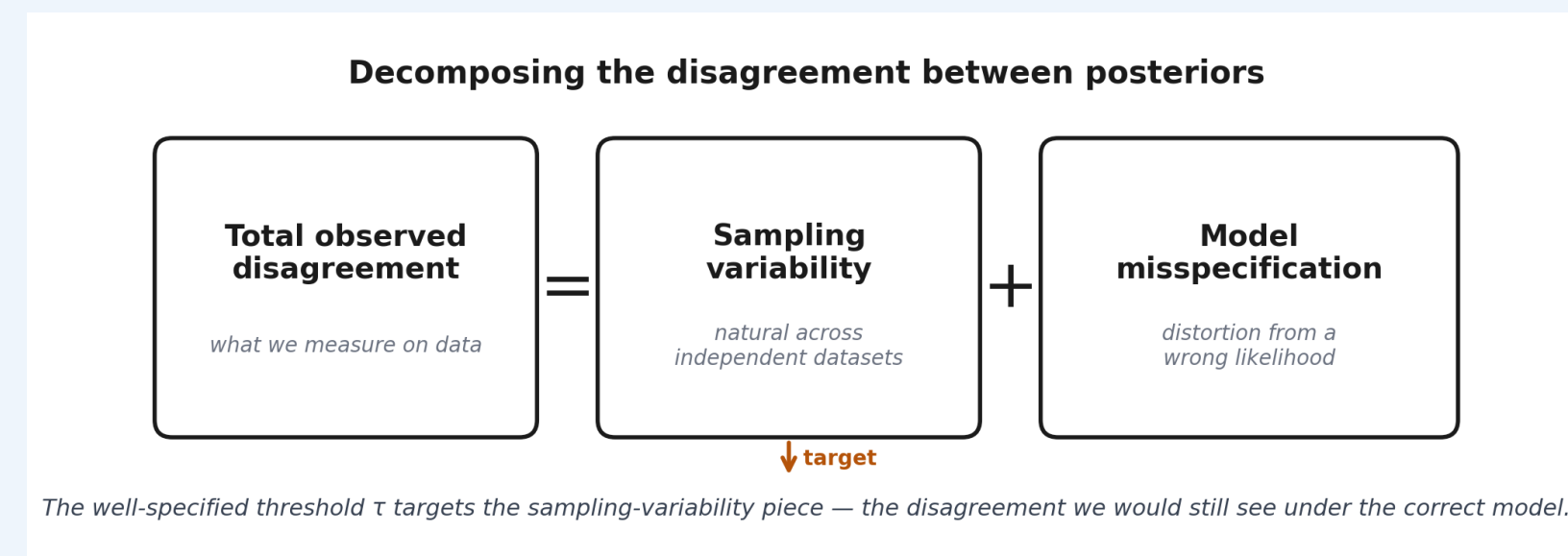


Figure 2. Decomposing observed disagreement into sampling variability and model misspecification.

Definition 1 (Well-specified threshold).

Assume the likelihood is correctly specified: $P_{\circ} = P_{\theta^*}$ for some $\theta^* \in \Theta$ with $\theta^* \sim Q$. Then, the well-specified threshold τ is the expected discrepancy between posteriors computed on independent datasets drawn from P_{θ^*} :

$$\tau = \mathbb{E}_{\theta^* \sim Q} \left[\mathbb{E}_{X, Y \sim P_{\theta^*}} [d_{1/2}(\Pi_X, \Pi_Y)] \right]$$

Theoretical Results

Under Bernstein–von Mises-type regularity conditions,

$$\bar{d}_{1/2} \rightarrow \frac{1}{2} \text{Tr}(IJ^{-1}),$$

I, J = variability / sensitivity information matrices at the pseudo-true parameter.

Correctly specified $\Rightarrow I=J \Rightarrow \tau = D/2$.

Misspecified: no tuning brings standard Bayes back to $D/2$.

Existence of β^* . Under asymptotic regularity conditions, there exists β^* such that β -Bayes is asymptotically well-calibrated: $\bar{d}_{1/2} \rightarrow \tau$ as $N \rightarrow \infty$.

Financial Application: Equity Factor Models

Setup. Daily excess returns of JNJ, BAC, AAPL (2012–2016, $n=1,257$, $K=4$ blocks). Models: CAPM ($D=2$) and FF3 ($D=4$). Fat-tailed returns misspecify the Gaussian likelihood.

Finding β^* so that the well-specified threshold d^* matches the observed discrepancy $\hat{d}_{1/2}$:

Stock / Model	Exc. Kurt.	d^*	$\hat{d}_{1/2}$	β^*
JNJ / FF3	2.33	1.98	9.54	1.83
AAPL / FF3	5.94	1.98	6.81	1.39
BAC / FF3	1.83	1.98	17.90	1.95
JNJ / CAPM	2.33	0.93	7.50	2.23
AAPL / CAPM	5.94	0.93	4.04	1.61
BAC / CAPM	1.83	0.93	7.34	2.08

Average $d_{1/2}$ between posterior predictives on held-out blocks:

Stock / Model	Std. Bayes	β^*
JNJ / FF3	2.667	1.463
JNJ / CAPM	1.956	0.725
BAC / FF3	5.657	1.624
BAC / CAPM	1.883	0.718
AAPL / FF3	2.032	1.007
AAPL / CAPM	1.140	0.476

Alpha stability $\sigma(\alpha)$ across blocks:

Stock / Model	Std. Bayes	β^*	Reduction
JNJ / FF3	0.0696	0.0335	2.1×
JNJ / CAPM	0.0528	0.0115	4.6×
BAC / FF3	0.0360	0.0317	1.1×
BAC / CAPM	0.0357	0.0229	1.6×
AAPL / FF3	0.0474	0.0410	1.2×
AAPL / CAPM	0.0555	0.0523	1.1×

Portfolio signal reliability $\bar{\alpha}_{\text{ann}}/\sigma(\alpha_{\text{ann}})$:

Stock / Model	Std. Bayes	β^*
JNJ / FF3	0.418	0.458
JNJ / CAPM	0.643	2.517
BAC / FF3	0.176	1.039
BAC / CAPM	0.378	2.111
AAPL / FF3	0.294	0.433
AAPL / CAPM	0.161	0.192