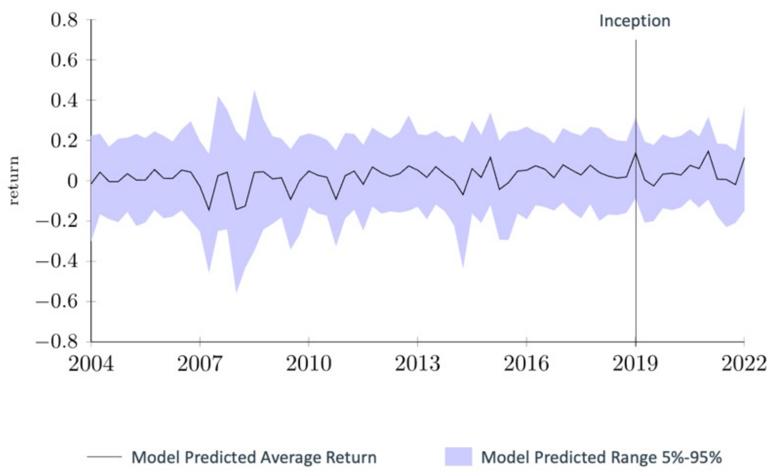
Bayesian Estimation of Alternative Asset Returns

Arnav Sheth

December 17, 2023

Contents



	Value
Model Fit (R^2)	85%
Quarters Reported (Predicted)	12 (60)

Figure 1: An example of a fund's returns as generated by GMAM 3.0. The fund was launched in 2019. GMAM 3.0 backcasted the returns to 2004 using hierarchical Bayesian techniques. These included joint estimation of de-smoothed returns and systematic factor exposures, among other model parameters. The backcast helps investors pair the fund with a portfolio of publicly available securities.

1 Introduction

Alternative investments such as private equity, hedge funds, and structured products are being traded at all-time highs. For instance, Bain & Company[1] valued private equity (PE) buyout deals for 2022 at $654 billion globally. This was the second-highest valuation since 2008, and brought the total deal volume to $2.4 trillion[2]. Additionally, total private markets AUM reached $11.7 trillion as of June 2022. Structured products also grew to reach $1.5 trillion in new issuance in 2021, according to S&P and Bloomberg states that structured products still outsize the total ETF market at $5.3tn.

All of this makes alternatives a major asset class, with institutions such as colleges, foundations, pension funds, and more investing in it. Furthermore, the SEC broadened the definition of an accredited investor to include "individual investors that have the knowledge and expertise to participate in [financial] markets." One no longer needs to pass the income requirement. iCapital has hastened the market expansion and liquidity of alternative investments and make the asset class viable for accredited investors and private wealth funds[3].

Optimal wealth allocation requires data on the risk, return, and covariance of asset classes. Unfortunately, PE is largely exempt from public disclosure requirements. The issue is particularly

[1]Source:https://www.bain.com/insights/topics/global-private-equity-report/

[2]Source:https://www.mckinsey.com/industries/private-equity-and-principal-investors/our-insights/mckinseys-private-markets-annual-review/

[3]Source:https://icapital.com/insights/practice-management/untapped-potential-alternative-investments-and-the-wealth-management-channel/

pernicious with respect to performance metrics based on actual transactions. The limited data impedes the investment process for portfolios containing such assets.

iCapital solves this problem and other issues associated with returns on illiquid and alternative investments via our Generalized Multi-Asset Model version 3.0 (henceforth, GMAM 3). GMAM 3 is a returns-based model that uses hierarchical Bayesian modeling techniques to generate the underlying economic returns for any single fund that trades on iCapital's marketplace. The model helps educate investors on fund suitability based on their investment needs and risk preferences.

This technical manual documents the important economic features of the model, and explains the economic and econometric motivation behind choices made in model construction. The intended audience for this document has a working familiarity with Bayesian hierarchical models, econometric time series, expected return beta models, and other similar topics.

1.1 Why Bayesian?

Bayesian techniques offer a robust and flexible approach for capturing the complex dynamics of returns from alternative assets due to their ability to:

- incorporate prior knowledge,
- handle limited data,
- account for illiquidity, and
- update estimates as new information becomes available.

Bayesian methodologies allow for the integration of prior beliefs or information about specific funds into the estimation of returns. The prior information effectually supplements limited data. These techniques directly incorporate uncertainty into the return estimates by treating all parameters as stochastic. The framework aligns well with the unique characteristics of alternative investments, allowing for more accurate and informed returns estimation.

It is well-known in the academic finance literature that for private equity (henceforth, PE) there is a dearth of performance metrics based on actual transactions (Kaplan and Schoar [2005]). The available time series data often relies on non-market valuations or multiyear internal rates of return (IRR), often segmented by the vintage years of the funds. For investors seeking to optimally allocate their wealth across public securities and private equity, the lack of data is a significant barrier. Even simple Markowitz-style (Markowitz [1952]) mean-variance optimization requires a sufficiently long history of returns, which is unavailable in the case of PE.

IRR is a commonly used metric in the PE world to gauge a fund's performance. It provides a convenient way by which to compare and benchmark various funds. Furthermore, it provides a computationally simple procedure to account for the timing and magnitude of cash flows since PE investments are typically illiquid and longer investment horizons. And finally, it helps investors assess the risk of a fund.

While these are good reasons to use the IRR, the measure has numerous shortcomings. Some believe that IRR assumes cash proceeds will be reinvested to achieve a particular return. This assumption does not coincide with the reality of cash flow distributions by PE. For example, if a PE fund reports a 50% IRR and has returned cash early in its life, the assumption is that the cash will be reinvested at 50%. However, it is unlikely that the fund will find such an investment opportunity every time cash is distributed.

Furthermore, IRR can distort management incentives, upwardly bias performance measures, and misrepresent volatility estimates (Phalippou [2008]). As stated in Sorensen and Jagannathan [2013], "The IRR may not exist, and it may not be unique." Even though the Modified IRR (MIRR) accounts for some well-known pitfalls in the measure, its practical implementation in a private partnership is not obvious.

GMAM 3, on the other hand, uses time-weighted returns.

$$y_t = \frac{NAV_t + Distributions_t}{NAV_{t-1} + Contributions_t}$$

Though time-weighted returns have their own caveats[4], they also mitigate many of the shortcomings of IRR. For instance, this method of calculation accounts for contributions and distributions, as well as their timing, in representing the total return on a single investment. Also, they allow for a more standardized comparison of performance across different investment vehicles and asset classes. The standardization is particularly valuable to investors seeking to combine PE with publicly available investments. Finally, time-weighted returns are a widely understood and commonly used method in the broader investment industry. Their use in measuring PE performance facilitates understanding and communication about fund performance, especially for stakeholders more familiar with traditional investment vehicles.

One of the strengths of the model's Bayesian estimation process provides measures of uncertainty around the estimated returns, and this allows the user to decide for themselves the degree to which they can rely on these estimated, time-weighed returns.

[4]They may not fully capture the complexities of private equity investments, such as the strategic timing of cash flows and the long-term, illiquid nature of these assets.

2 The Model

At its core, the model predicts fund returns from factor returns. The essence of the GMAM 3 economics returns-generating process can be modeled as:

$$x_{i,t} = f_t' \beta_{i,t} + r_t^f + \varepsilon_{i,t}$$

where f_t is a length K vector of factor returns including an intercept dummy, $\beta_{i,t}$ is a $K \times 1$ matrix of exposures, $x_{i,t}$ is the predicted return, r_t^f is the risk-free rate of return, and $\varepsilon_{i,t}$ is the error. The subscripts here are for an individual fund (i) for a given period (t). Henceforth, assume that all measures are for an individual fund, and so the i-subscript will be dropped.

We use a set of thirteen factors: Alt Commodities, Alt Hedge Fund Crowding, Alt Oil, Alt Trend, Emerging Markets, Equity Market, Equity Momentum, Equity Quality, Equity SmallCap, Equity Value, Fixed Credit, Fixed Duration, and US Dollar. Detailed information about factor construction is available upon request.

The factor modelling approach is standard, and in-of-itself does not distinguish GMAM 3.0. However, the estimation procedure is tailored to account for the unique difficulties associated with alternative assets. Specifically:

1. x_t is NOT the final return provided to reported to the investor.Alternative investment reported returns often serially correlated and lagged. The quantity x_t represents the unobservable true economic returns of the fund. The model simultaneously estimates these returns as part of the estimation procedure. To mimic the underlying characteristics of the reported returns to a greater degree[5], we model the data generating process as re-smoothing of the economic returns before they are reported to investors. Section 2.1 below discusses the economic rationale and mathematical process in greater detail. The Appendix has additional details.
2. Rather than include all of the return factors in the regression, GMAM 3.0 uses the stochastic search variable selection (SSVS) technique (George and McCulloch [1993]) to account for the likelihood that an exposure is economically meaningful. SSVS is a type of Bayesian linear regression which is useful when the number of predictors is larger than the number of observations – a problem frequently encountered in alternative investment data. See Section 2.3, and the Appendix for details.

GMAM 3 uses Markov Chain Monte Carlo (MCMC) to compute estimate the β as well as the distributional parameters of ε_t. The final product displays only point estimates – which are calculated as expectations across draws in the MCMC process. For MCMC, each draw is a distri-

[5] This re-smoothing, while replicating the essential attributes of the reported returns more closely, does come at the cost of not reflecting the true economic changes in value over time.

bution and not a point estimate (more details on this in the Appendix). By this methodology, the expected, predicted, de-smoothed returns estimated from N simulations are given by:

$$\overline{x}_{n,t} = \frac{1}{N} \sum_{n \in 1:N} \left(f_t' \beta_{n,t} + r_t^f + \varepsilon_{n,t} \right)$$

Since this is the result of a hierarchical Bayesian regression, the true distribution of the error terms is hard to characterize precisely – it is a mixture distribution. However, we can approximately say that[6]:

$$\varepsilon_{n,t} \sim Tdist\left(0, \sigma_{\varepsilon_n}, \nu_n\right)$$

Here, $\sigma^2_{\varepsilon_n}$ is derived from model parameters sampled at each iteration and ν_n is a t-distribution degree of freedom parameter likewise sampled. In other words, the error terms approximate a t-distribution to a great degree.

2.1 Re-Smoothing

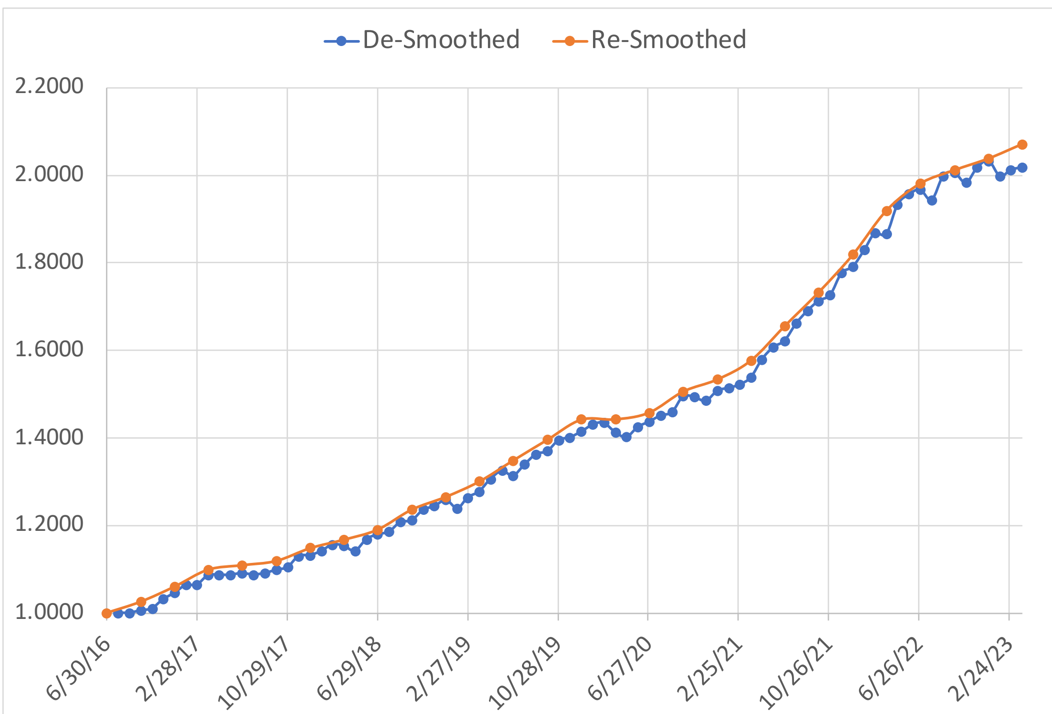


Figure 2: An example of a fund's de-smoothed (x), and re-smoothed (y) returns as generated by GMAM 3. The reported re-smoothed returns are quarterly, and the de-smoothed returns are monthly. Shown here is the cumulative return of $1 invested in each of the set of returns.

It is well-known in the finance practitioner literature that returns from alternative investment

[6]Strictly speaking, this approximates a t-distribution since the final distribution is the result of several layers of Bayesian analysis. More information about this can be found in the Appendix.

such as PE, hedge funds, or real estate funds are highly serially correlated (see, for example, Getmansky et al. [2004]). In other words, past values correlate with present values. The serial correlation occurs because of the lack of liquidity in the fund itself or some of the assets held within it. For instance, these illiquid assets may not trade frequently, leading to subjective and otherwise noisy valuations. The effect is such that when funds contain illiquid assets, their reported returns may seem steadier than their actual economic returns (returns that consider all available market information about those securities). The positive serial return correlation commonly leads to a downward bias in estimated return variance.

The effect extends to the reported returns of real estate funds(Geltner [1993]). Investors typically demand monthly or quarterly reporting. But valuations of many properties included in the funds are effectively updated only annually. Each quarter some properties have their valuations updated, and others do not. Those properties that don't get a new valuation within a quarter may carry all or a portion of their last known value into the current quarter.

Finally, Financial Accounting Standard 157, released by the FASB in 2006 during the run-up to the financial crisis, and now called Accounting Standards Code Topic 820, requires companies to mark their assets to market. The rule was a radical change from historic cost accounting and required general partners to periodically mark the assets to market. This backwards appraisal may also result in a managerial bias towards smoothing asset values.

The latent returns $x_{i,t}$, henceforth just x for simplicity of not, generated above are not smoothed. They are the true economic returns. Since GMAM 3 estimates true reported returns, we must re-smooth the latent returns to reflect the reported returns. In GMAM 3, this is done using a moving average (MA) process with smoothing parameters estimated using a Bayesian linear regression. Note that the MA is defined in an econometric sense, and not in a literal sense, although it may be interpreted as such. More specifically, the final predicted returns presented to the user are re-smoothed using the following process:

$$y = \Phi x$$

where y are the reported returns and Φ is a matrix consisting of smoothing coefficients, ϕ and $\tilde{\phi}$. More specifically, when both x and y have the same reporting frequency,

$$\Phi = \begin{bmatrix} \phi_1 & \cdots & \phi_P & \tilde{\phi}_{P+1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \phi_1 & \cdots & \phi_P & \tilde{\phi}_{P+1} & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \phi_1 & \cdots & \phi_P & \tilde{\phi}_{P+1} & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \phi_1 & \cdots & \phi_P & \tilde{\phi}_{P+1} \end{bmatrix} \tag{1}$$

ϕ is a vector of P unrestricted coefficients, and $\tilde{\phi}$ is a vector of $P+\Delta t$ restricted and unrestricted coefficients. P and Δt represent terms that map the observation period of the reported, smoothed returns y to the latent, de-smoothed returns, x. In particular,

$$\begin{aligned} y_s &= \left(\tilde{\phi}_{1:(P+\Delta t)}\right)' x_{(t[s]-P-\Delta t+1):t[s]} + \varepsilon_t^y \\ &= \left(\tilde{\phi}_{(P+1):(P+\Delta t)}\right)' x_{(t[s]-\Delta t+1):t[s]} + \phi' x_{(t[s]-P):(t[s]-\Delta t)} + \varepsilon_t^y \end{aligned}$$

The reported returns are y_s s.t. $s \in 1 : S$, and the latent returns are x_t s.t. $t \in 1 : T$. In equation (1) above, $S = T - P$, and the reported returns and de-smoothed returns have the same frequency. However, S may not always equal T, such as when the reported returns are quarterly, and the de-smoothed returns are monthly[7]. To resolve this, the following formula maps t to s:

$$t(s) = P + (s \times \Delta t)$$

For example, if the reported returns y are quarterly, and the latent returns x are monthly, $P = 3$, and $\Delta t = 3$ (since each quarter consists of three months). And so, for the first quarter, $s = 1$, and $t = 6$. When $\Delta t > 1$, the values in $\tilde{\phi}_j$ must add up to 1 for each month of the quarter (first, second, third). If not, months that fall earlier in the quarter will have a different long-run impact on NAV than months later in the quarter. For instance if we have a single quarter lag, the first month of the quarter plus the first month of the previous quarter must add to 1. This implies that the sum total of values in $\tilde{\phi}_j$ is 3 for quarterly data and 1 for monthly data.[8].

The economic assumption behind this process is that the observed fund returns y are a weighted average of the fund's economic returns x over the most recent $(P + \Delta t)$ periods, inclusive of the current period. More generally, observed fund returns follow a MA process of order $P + \Delta t$.

The restrictions used in the MA process are similar to Getmansky et al. [2004] where the observed return (R_t^0) for some period t, is a weighted average of the "true" returns (R_t^C) over the most recent $k+1$ periods: $R_t^0 = \theta_0 R_t^C + \cdots + \theta_k R_(t-k)^C$, with $\sum_{i=0}^k \theta_i = 1$ to ensure that all information is eventually incorporated into observed returns, and $\theta_i \in [0,1]$ for $i = 1, \ldots, k$. In our case the observed returns are the reported returns, y, the "true" returns are the latent returns generated using the factors, x, and the θ_i terms from Getmansky et al. [2004] are denoted as ϕ_i in our estimation process. They are generated using a multivariate normal distribution in the hierarchical Bayesian model, as follows:

$$p\left(\phi | rest\right) \sim MN\left(\phi_0, \frac{1}{\tau_y \tau_\phi} M_0^{-1}\right)$$

[7] Our factor returns are almost always monthly, as are the de-smoothed returns.

[8] Note that the vector $\phi \in \tilde{\phi}$. For quarterly reported data, say $\phi = [a, b, c]'$ where $a, b, c \in \mathbb{R}$. Then, $\tilde{\phi} = [a, b, c, 1-a, 1-b, 1-c]'$ and the sum of all terms in $\tilde{\phi}$ is 3. For monthly reported data, say $\phi = [a, b, c, d, e]'$ where $a, b, c, d, e \in \mathbb{R}$. Then $\tilde{\phi} = [a, b, c, d, e, 1-a-b-c-d-e]'$ and the sum of all terms in $\tilde{\phi}$ here, is 1. This process is also described in more detail in the Appendix.

where τ_y is a global precision parameter estimated for independent measurement variance. The precision parameter has Gamma-distributed prior: $p(\tau_y) \sim Gamma\left(\alpha_{y0}, \zeta_{y0}\right)$ with given shape hyperparameter, α_{y0}, and given inverse scale hyperparameter, ζ_{y0}; τ_ϕ is a global precision multiplier parameter, distributed as $p(\tau_\phi) \sim Gamma\left(\alpha_{\phi0}, \zeta_{\phi0}\right)$ with given shape and inverse scale hyperparameters; M_0 is a $P \times P$ matrix consisting of precision hyperparameters for ϕ_0; and ϕ_0 is a hyperparameter (i.e., fed into the model).

The complete posterior distribution of y is given by:

$$p\left(y \mid rest\right) \sim MN\left(\Phi x, \frac{1}{\tau_y} I\right))$$

with Φ, x, and τ_y defined as above; and I is an identity matrix.

2.2 A Brief Digression on the Choice of Prior Distributions

The influence of well-behaved priors declines as data increases. Priors are, of course, essential in Bayesian analysis. In theory, Bayesian models exhibit a property known as 'consistency.' This means that as the sample size grows to infinity, the posterior distribution of the parameter of interest will converge to the "true" parameter value[9], assuming the model is correctly specified. In such a scenario, the choice of prior becomes less critical as the sample size increases.

That said, the use of the gamma distribution with an inverse scale parameterization above is convenient for several reasons:

1. Conjugacy: conjugate priors are advantageous for computational and analytical simplicity.
2. Positive Values: the gamma distribution is defined for positive values only.
3. Informative and Non-Informative Settings: the gamma distribution encodes both informative and highly diffuse priors. By adjusting its parameters, the priors may include strong prior beliefs or have minimal influence on the posterior, as reflecting the information available outside of the data.
4. Regularization and Stability: In Bayesian hierarchical models, the use of gamma priors can help in regularizing the estimates, particularly in complex models or in the presence of limited data. This can prevent overfitting and improve the model's stability and predictive performance.

2.3 Stochastic Search Variable Selection (SSVS)

As mentioned above, the regression performed to estimate x is a SSVS technique which defines prior regression coefficient variances consistent with inclusion or effective exclusion. In other words, it

[9] In the frequentist sense of the word.

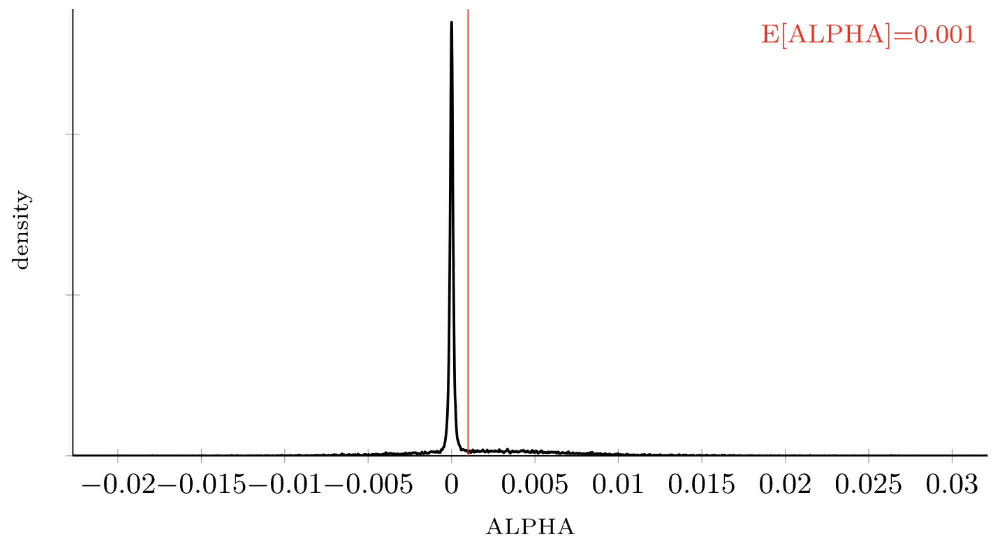


Figure 3: The probability mass function of a fund's intercept term (β_0) from the SSVS (spike and slab) regression. The spike is apparent due to limited evidence of an intercept. The probability mass is concentrated around zero with a small positive tilt. There is low evidence of an intercept term.

is a predictor selection technique. Such techniques commonly focus on which predictors to retain, though they also aim for improved predictive performance through developing an encompassing model, or model simplification without adversely affecting predictive accuracy (Piironen and Vehtari [2017]). Formal model choice is simplified for normal linear regression, as marginal likelihoods may be obtained analytically, but for many predictors, comparison of the many possible models becomes infeasible.

An imperfect analogy is stepwise regression using AIC or BIC as a selection criterion. However, when the number of regressors is large, testing of each model may be computationally prohibitive. A common workaround is a heuristic method to restrict attention to a potential subset of regressors, i.e., include or exclude variables, based on R^2 considerations (say).

SSVS uses a Bayesian approach with a normal mixture model for regression analysis. In this method, selector variables are employed to pinpoint which subsets of predictors are worth considering. It works by identifying those predictors that have a higher chance of being relevant, based on their posterior probability. Gibbs sampling[10] is standard when estimating these models. The MCMC approach samples from the distribution of all possible subsets of predictors. The subsets that show up more often in these samples are considered promising because they have a higher

[10]Gibbs sampling is a statistical technique used for generating sequences of samples from the probability distribution of multiple variables. It's a kind of Markov Chain Monte Carlo (MCMC) method. It is particularly useful in scenarios where directly sampling from the joint distribution is difficult, but sampling from the conditional distribution of each variable is feasible.

probability of being relevant.

We use the SSVS technique of George and McCulloch [1993], also known as a **spike and slab regression**. The term was coined by Mitchell and Beauchamp [1988] and referred to the prior for the regression coefficients used in their Bayesian hierarchy. This prior was chosen such that the regression parameters were mutually independent with a two-point mixture distribution made up of a uniform flat distribution (the slab) and a degenerate distribution at zero (the spike).

Ishwaran and Rao [2005] characterize a spike and slab model as being any model with a Bayesian hierarchy specified as follows:

$$
\begin{aligned}
p(y \mid X, \beta, \sigma^2) &\sim N(X\beta, \sigma^2 I_n) \\
p(\beta \mid \gamma) &\sim N(0, \Gamma) \\
\gamma := (\gamma_1, \ldots, \gamma_p)^T &\sim \pi(\cdot) \\
p(\sigma^2) &\sim \mu(\cdot)
\end{aligned}
$$

Here, $\beta = (\beta_1, \ldots, \beta_p)^T$ is the regression vector, and $\Gamma = diag(\gamma_k)_{1 \leq k \leq p}$ is its $p \times p$ hypervariance matrix. The prior π for the hypervariance γ plays a critical role in how effective the technique is for variable selection. A successful and popular choice of π are priors that make use of mixture distributions involving a spike near zero. In George and McCulloch [1993], the prior for γ_k was assumed to have a two-component distribution of the form:

$$
p\left(\gamma_k \mid \tau_k, c_k, \overline{\omega}_k\right) \overset{\text{ind}}{\sim} \left(1 - \overline{\omega}_k\right) \delta_{\tau_k}(\cdot) + \overline{\omega}_k \delta_{c_k \tau_k}(\cdot), \quad k = 1, \ldots p.
$$

The value for $\tau_k > 0$ (the spike) is chosen as some small value, where "small" is typically based on the data at hand, while $c_k > 0$, also data-specific, is chosen so that $c_k \tau_k$ (the slab) is sufficiently large. Selecting the two hyperparameters in this way allows γ_k to be small or large, and this in turn enables the posterior of β_k to shrink towards zero or be some nonzero value. The values $(\overline{\omega}_k)_{k=1}^p$ are complexity parameters that influence the likelihood of a coefficient being shrunk towards zero. In principle, each variable can have a unique complexity value, but a common practice is to set $\overline{\omega}_k = 1/2, \forall k$, in which case the prior is referred to as an indifference prior.

GMAM 3.0 uses a Bernoulli prior on γ_k, the selector variable:

$$
p\left(\gamma_k\right) \sim Bern\left(\omega\right)
$$

Here, the probability of factor selection $p\left(\omega\right) \sim Beta\left(\kappa_0, \delta_0\right)$ depends on hyperparameters κ_0 and δ_0. γ_k for $k \in 1 : K$ is a selector variable used to determine which of the thirteen factors are economically significant in a probabilistic sense. Initially, each factor has an equally likely chance of

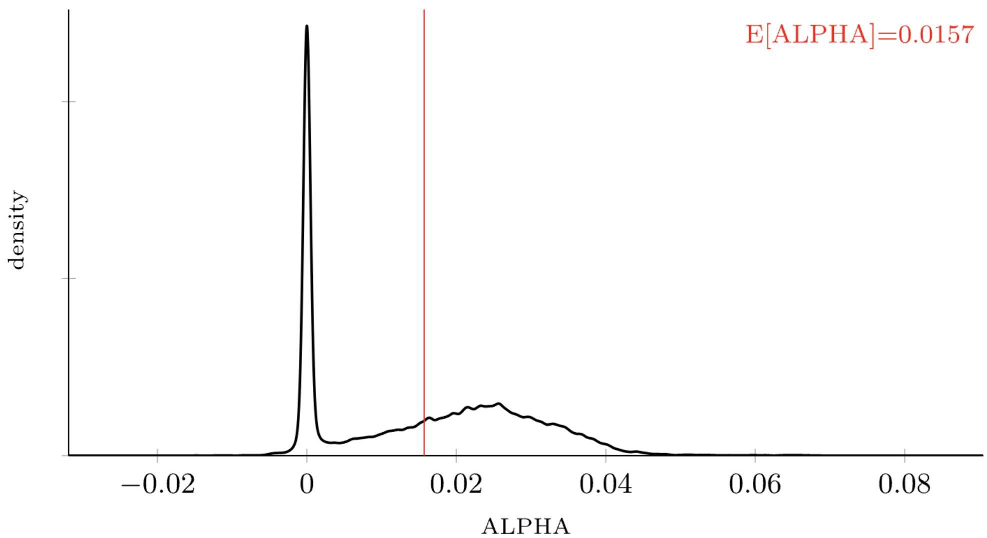


Figure 4: The probability mass function of a fund's intercept term (β_0) from the SSVS (spike and slab) regression. Though we see a spike here, the bulk is in the long right tail representing substantial upside and a significantly greater weight of evidence for a positive alpha. This fund has over two years of history.

being included.[11] Conditional on the factor selection, specify a prior distribution for the regression coefficient associated with that variable. GMAM 3.0 uses a multivariate normal:

$$p\left(\beta|rest\right) \sim MN\left(\beta_0 + D^{-1}\beta_0^{\Delta}, \frac{1}{\tau_x\tau_y\tau_\beta}\left[DA_0D\right]^{-1}\right)$$
$$d_k = \sqrt{\gamma_k + (1-\gamma_k)\frac{1}{v^2})}$$

Here, β_0 is the prior mean of β conditional on exclusion (i.e., when $\gamma_k = 0$); β_0^{Δ} is the shift in the prior mean of β conditional on inclusion (i.e., when $\gamma_k = 1$); d_k for $k \in 1:K$ is a $K \times 1$ vector – a function of γ which adjusts β for sparsity; D is a $K \times K$ matrix with $D = diag(d_k)$; and ν is the variance of the spike distribution as a fraction of the slab variance. And finally, τ_x, τ_y, and τ_β are all gamma-distributed with shape and inverse-scale hyperparameters as shown earlier.

3 Conclusion

We use a hierarchical Bayesian model to generate fund returns when data is sparse and noisy. The data generation process includes de-smoothed returns, generated using a spike and slab regression, and re-smoothed returns used to mimic reported returns. Our technique provides a systematic and rigorous way by which fund performance can be quantified.

[11] Subject to a zero exposure prior.

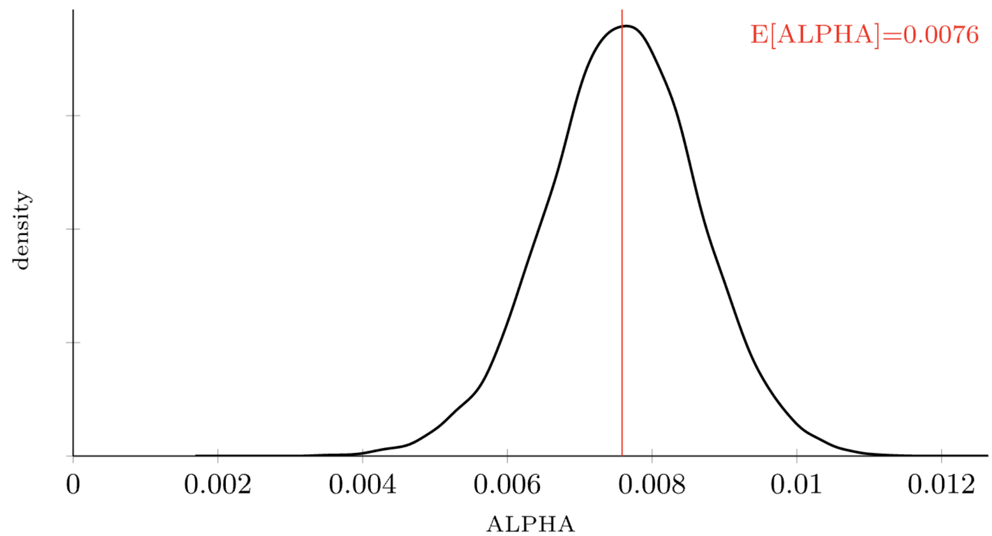


Figure 5: The probability mass function of a fund index intercept term (β_0) from the SSVS (spike and slab) regression. This index has almost twenty years of history and clearly shows positive alpha, though its magnitude is still uncertain.

References

[1] David Geltner. Estimating market values from appraised values without assuming an efficient market. *The Journal of Real Estate Research*, 8(3):324–345, 1993.

[2] Edward I. George and Robert E. McCulloch. Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889, 1993.

[3] Mila Getmansky, Andrew W. Lo, and Igor Makarov. An econometric model of serial correlation and illiquidity in hedge fund returns. *Journal of Financial Economics*, 74(3):529–609, 2004.

[4] Hemant Ishwaran and J. Sunil Rao. Spike and slab variable selection: frequentist and bayesian strategies. *The Annals of Statistics*, 33(2):730–733, 2005.

[5] Steven N Kaplan and Antoinette Schoar. Private equity performance: Returns, persistence, and capital flows. *The Journal of Finance*, 60(4):1791–1823, 2005.

[6] Harry M. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[7] T.J. Mitchell and J.J. Beauchamp. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83:1023–1036, 1988.

[8] Ludovic Phalippou. The hazards of using irr to measure performance: The case of private equity. 2008. URL `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1111796`.

[9] Juho Piironen and Aki Vehtari. Comparison of bayesian predictive methods for model selection. *Statistics and Computing*, 27:711–735, 2017.

[10] Morten Sorensen and Ravi Jagannathan. The public market equivalent and private equity performance, 2013. URL `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2347972`.

Appendix

The complete generative process is given by:

$$
\begin{aligned}
p\,(y|rest) &\sim MN\left(X_L R\phi + x_S, \frac{1}{\tau_y} I\right) \\
p\,(x|rest) &\sim MN\left(F\beta + r, \frac{1}{\tau_x \tau_y} \Psi^{-1}\right) \\
p\,(\phi|rest) &\sim MN\left(\phi_0, \frac{1}{\tau_y \tau_\phi} M_0^{-1}\right) \\
p\,(\beta|rest) &\sim MN\left(\beta_0 + D^{-1}\beta_0^{\Delta}, \frac{1}{\tau_x \tau_y \tau_\beta} \left[D A_0 D\right]^{-1}\right) \\
d_k &= (\gamma_k + (1-\gamma_k)\frac{1}{v^2})^{0.5} \\
p\,(\gamma_k) &\sim Bern\,(\omega) \\
p\,(\omega) &\sim Beta\,(\kappa_0, \delta_0) \\
p\,(\tau_y) &\sim Gamma\,(\alpha_{y0}, \zeta_{y0}) \\
p\,(\tau_x) &\sim Gamma\,(\alpha_{x0}, \zeta_{x0}) \\
p\,(\psi_t) &\sim Gamma\,(\nu/2, \nu/2) \\
p\,(\nu) &\sim Gamma\,(\alpha_{\nu 0}, \zeta_{\nu 0}) \\
p\,(\tau_\phi) &\sim Gamma\,(\alpha_{\phi 0}, \zeta_{\phi 0}) \\
p\,(\tau_\beta) &\sim Gamma\,(\alpha_{\beta 0}, \zeta_{\beta 0})
\end{aligned}
$$

Note that here y can be written in two forms using matrix notation. When x and y have the same frequency:

$$y = \Phi x = X_L R\phi + x_S$$

Where

$$
X_L R\phi + x_S = \begin{bmatrix} x_1 & x_2 & \cdots & x_{P-1} & x_P & x_{P+1} \\ x_2 & x_3 & \cdots & x_P & x_{P+1} & x_{P+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{T-P-1} & x_{T-P} & \cdots & x_{T-3} & x_{T-2} & x_{T-1} \\ x_{T-P} & x_{T-P+1} & \cdots & x_{T-2} & x_{T-1} & x_T \end{bmatrix} [R] \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{P-1} \\ \phi_P \end{bmatrix} + \begin{bmatrix} x_{P+1} \\ x_{P+2} \\ \vdots \\ x_{T-1} \\ x_T \end{bmatrix}
$$

And

$$
R = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -1 & -1 & \cdots & -1 & -1 \end{bmatrix}
$$

To generalize to other frequencies, recall that $t\left[s\right] \equiv P + s * \Delta t$ and define:

$$\begin{aligned}
\Phi_{sj} \equiv & \begin{cases} \phi_{P-(t[s]-\Delta t-j)} & 1 \leq P - (t\left[s\right] - \Delta t - j) \leq P \\ 1 - \left(\iota^{\phi}_{\Delta t-(t[s]-j)}\right)' \phi & t\left[s\right] - \Delta t < j \leq t\left[s\right] \\ 0 & \text{otherwise} \end{cases} \\
X_{Lsj} \equiv & x_{t[s]-(P+\Delta t-j)} \\
\iota^{\phi}_{pl} \equiv & \begin{cases} 1 & p - l \mod \Delta t = 0 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

The above matrix formulation (3) can be generalized by only including rows of X_L where $t \in \{t\left[1...S\right]\}$ and adjusting the restriction matrix. The general version is then:

$$y = \Phi x = X_L R\phi + x_S, \quad \text{s.t. } x_S = X_L \iota_{\Delta t}$$

where $\iota_{\Delta t}$ is a vector of P zeros followed by Δt ones, making x_S the sum of the last Δt columns of X_L.

Variable Names and Definitions

Table 1: Variable Definitions

Variable definitions for the Data Generating Process (DGP). **DGP corresponds to a system with S observations over T periods with each observation dependent on P terms in a moving average and K factors.** Variables are divided into the following types: observed values, local parameters along a particular dimension, global scalar parameters, and hyperparameters.

Variable	Type	Dimensions	Source	Definition/Description
y	Observed	$S{\times}1$	Data	Vector of observed returns.
$x; X_L$	Local Parameter	$T{\times}1$; $S{\times}(P{+}1)$	Estimated	x is a vector of gross latent returns.
r	Observed	$T{\times}1$	Data	Vector of risk-free returns.
ϕ; Φ	Local Parameter	$P{\times}1$; $S{\times}T$	Estimated	ϕ is the moving average window.
τ_y	Global Parameter	-	Estimated	Precision parameter for independent measurement error.
ϕ_0	Hyperparameter	$P{\times}1$	Given	Prior estimate for ϕ.
M_0	Hyperparameter	$P{\times}P$	Given	Precision of prior estimate ϕ_0.
F	Observed	$T{\times}K$	Data	Matrix of factor returns, possibly including an intercept.
β	Local Parameter	$K{\times}1$	Estimated	Regression coefficients of x on F.
ψ; Ψ	Local Parameter	$T{\times}1$; $T{\times}T$	Estimated	ψ is a vector of precision weights for x;$\Psi{=}Diag\,(\psi)$.
τ_x	Global Parameter	-	Estimated	Precision multiplier parameter for the regression of x on F.
τ_β	Global Parameter	-	Estimated	Prior precision multiplier parameter for the prior on ϕ.
τ_ϕ	Global Parameter	-	Estimated	Prior precision multiplier parameter for the prior on β.
β_0	Prior	$K{\times}1$	Given	Prior mean of β conditional on exclusion ($\gamma{=}0$).
β_0^Δ	Prior	$K{\times}1$	Given	Shift in prior mean of β conditional on inclusion ($\gamma{=}1$).
A_0	Prior	$K{\times}K$	Given	Possibly diagonal prior precision for $\beta_0{+}D^{-1}\beta_0^D$
γ	Local Parameter	$K{\times}1$	Estimated	Vector of variable selection indicators.
d; D	Local Parameter	$K{\times}1$; $K{\times}K$	Transformation	d is a function of γ and adjusts β for sparsity; $D{=}Diag\,(d)$.
v	Hyper Parameter	-	Given	Variance of the spike distribution as fraction of the slab variance.
ω	Global	-	Estimated	Probability of variable selection.
κ_0; δ_0	Hyper	-	Given	Hyperparameters for prior on ω.
ν	Global	-	Estimated	Non-normality parameter for x; DOF of posterior t distribution.
$\alpha_{\phi 0}$; $\zeta_{\phi 0}$	Hyper	-	Given	Hyperparameters for τ_ϕ.
$\alpha_{\beta 0}$; $\zeta_{\beta 0}$	Hyper	-	Given	Hyperparameters for τ_β.
α_{x0}; ζ_{x0}	Hyper	-	Given	Hyperparameters for τ_x; α_{x0} is shape, ζ_{x0} is inverse scale.
α_{y0}; ζ_{y0}	Hyper	-	Given	Hyperparameters for τ_y; α_{y0} is shape, ζ_{y0} is inverse scale.
ν_0^-; ν_0^+	Hyper	-	Given	Hyperparameters for prior on ν.

Complete Posterior Distribution

The complete posterior distribution is given by:

$$
\begin{aligned}
p\left(\Theta|y,F,r\right) &\propto p\left(y|x,\gamma,\omega,\beta,\phi,\tau_x,\tau_y,\tau_\phi,\tau_\beta,\psi,\nu,F\right)\times p\left(x|\beta,\phi,\tau_x,\tau_y,\psi,F\right)\\
&\times p\left(\phi|\tau_y,\tau_\phi\right)\times p\left(\beta|\gamma,\tau_x,\tau_y,\tau_\beta\right)\times p\left(\gamma|\omega\right)\times p\left(\omega\right)\\
&\times p\left(\psi|\nu\right)\times p\left(\nu\right)\times p\left(\tau_x\right)\times p\left(\tau_y\right)\times p\left(\tau_\phi\right)\times p\left(\tau_\beta\right)\\
&= MN\left(y;\Phi x,\frac{1}{\tau_y}I\right)\times MN\left(x;F\beta+r,\frac{1}{\tau_x\tau_y}\Psi^{-1}\right)\\
&\times MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\times MN\left(\beta;\beta_0+D^{-1}\beta_0^{\Delta},\frac{1}{\tau_x\tau_y\tau_\beta}\left[DA_0D\right]^{-1}\right)\\
&\times\prod_{k=1}^{K}Bern\left(\gamma_k;\omega\right)\times Beta\left(\omega;\kappa_0,\delta_0\right)\\
&\times\prod_{t=1}^{T}Gamma\left(\psi;\nu/2,\nu/2\right)\times Gamma\left(\nu;\alpha_{\nu 0},\zeta_{\nu 0}\right)\\
&\times Gamma\left(\tau_x;\alpha_{x0},\zeta_{x0}\right)\times Gamma\left(\tau_y;\alpha_{y0},\zeta_{y0}\right)\\
&\times Gamma\left(\tau_\phi;\alpha_{\phi 0};\zeta_{\phi 0}\right)\times Gamma\left(\tau_\beta;\alpha_{\beta 0},\zeta_{\beta 0}\right)
\end{aligned}
$$

Posterior of ϕ

First, define:

$$
\begin{aligned}
\tilde{y} &\equiv y - x_S\\
\tilde{X}_L &\equiv X_L R
\end{aligned}
$$

$$
\log p\left(\phi|rest\right) = \log MN\left(\phi;\mu_\phi,\Lambda_\phi^{-1}\right)+c_4^{\phi}
$$

where

$$
\begin{aligned}
\Lambda_\phi &= \tau_y\left(\tilde{X}_L'\tilde{X}_L + \tau_\phi M_0\right)\\
\mu_\phi &= \tau_y \Lambda_\phi^{-1}\left(\tilde{X}_L'\tilde{y} + \tau_\phi M_0'\phi_0\right)\\
c_1^\phi &= \frac{S+P}{2}\log\left(\frac{\tau_y}{2\pi}\right) + \frac{P}{2}\log\tau_\phi + \frac{1}{2}\log det\left(M_0\right)\\
&\quad + \log\Bigg[MN\left(x; F\beta + r, \frac{T}{\tau_x\tau_y}\Psi^{-1}\right)\times MN\left(\beta;\beta_0 + D^{-1}\beta_0^\Delta, \frac{1}{\tau_x\tau_y\tau_\beta}\left[DA_0D\right]^{-1}\right)\\
&\quad \times\prod_{k=1}^{K} Bern\left(\gamma_k;\omega\right)\times Beta\left(\omega;\kappa_0,\delta_0\right)\times\prod_{t=1}^{T} Gamma\left(\psi;\nu/2,\nu/2\right)\\
&\quad \times Gamma\left(\nu;\alpha_{\nu 0},\zeta_{\nu 0}\right)\times Gamma\left(\tau_x;\alpha_{x0},\zeta_{x0}\right)\times Gamma\left(\tau_y;\alpha_{y0},\zeta_{y0}\right)\\
&\quad \times Gamma\left(\tau_\beta;\alpha_{\beta 0},\zeta_{\beta 0}\right)\times Gamma\left(\tau_\phi;\alpha_{\phi 0},\zeta_{\phi 0}\right)\Bigg] + c^{ev}\\
c_2^\phi &= c_1^\phi - \frac{\tau_y}{2}\left[\tau_\phi\phi_0'M_0\phi_0 + \tilde{y}'\tilde{y}\right]\\
c_3^\phi &= c_2^\phi + \frac{1}{2}\mu_\phi'\Lambda_\phi\mu_\phi\\
c_4^\phi &= c_3^\phi + \frac{P}{2}\log 2\pi - \frac{1}{2}\log det\Lambda_\phi\\
c_{ev} &= -\log p\left(y_t\right)
\end{aligned}
$$

The last constant makes the full posterior into a valid probability distribution.

Posterior of x

$$\log p\left(x|rest\right) = \log MN\left(x;\mu_x,\Lambda_x^{-1}\right) + c_4^x$$

where

$$
\begin{aligned}
\Lambda_x &= \tau_y \left(\Phi'\Phi + \tau_x \Psi\right) \\
\mu_x &= \tau_y \Lambda_x^{-1} \left(\Phi' y + \tau_x \Psi \left(r + F\beta\right)\right) \\
c_1^x &= \frac{S+T}{2} \log\left(\frac{\tau_y}{2\pi}\right) + \frac{T}{2} \log \tau_x + \frac{1}{2} \log Det\left(\Psi\right) \\
&\quad + \log \Bigg[MN\left(\phi; \phi_0, \frac{1}{\tau_y \tau_\phi} M_0^{-1}\right) \times MN\left(\beta; \beta_0 + D^{-1}\beta_0^{\Delta}, \frac{1}{\tau_x \tau_y \tau_\beta} \left[D A_0 D\right]^{-1}\right) \\
&\quad \times \prod_{k=1}^{K} Bern\left(\gamma_k; \omega\right) \times Beta\left(\omega; \kappa_0, \delta_0\right) \times \prod_{t=1}^{T} Gamma\left(\psi; \nu/2, \nu/2\right) \times Gamma\left(\nu; \alpha_{\nu 0}, \zeta_{\nu 0}\right) \\
&\quad \times Gamma\left(\tau_x; \alpha_{x0}, \zeta_{x0}\right) \times Gamma\left(\tau_y; \alpha_{y0}, \zeta_{y0}\right) \\
&\quad \times Gamma\left(\tau_x; \alpha_{x0}, \zeta_{x0}\right) \times Gamma\left(\tau_y; \alpha_{y0}, \zeta_{y0}\right) \\
&\quad \times Gamma\left(\tau_\beta; \alpha_{\beta 0}, \zeta_{\beta 0}\right) \times Gamma\left(\tau_\phi; \alpha_{\phi 0}, \zeta_{\phi 0}\right) \Bigg] + c^{ev} \\
c_2^x &= c_1^x - \frac{\tau_y \tau_x}{2} \left(r + F\beta\right)' \Psi \left(r + F\beta\right) - \frac{\tau_y}{2} y' y \\
c_3^x &= c_2^x + \frac{\mu_x' \Lambda_x \mu_x}{2} \\
c_4^x &= c_3^x + \frac{T}{2} \log 2\pi - \log det \Lambda_x
\end{aligned}
$$

Posterior of τ_y

- Let $\tilde{\beta} \equiv \beta - \beta_0$. Then the conditional posterior is:

$$
\log p\left(\tau_y | rest\right) = \log Gamma\left(\tau_y; \alpha_y, \zeta_y\right) + c_2^{\tau_y}
$$

  where

$$
\begin{aligned}
\alpha_y &= \frac{S+T+P+K}{2} + \alpha_{y0} \\
\zeta_y &= \zeta_{y0} + \frac{1}{2}\left(\tilde{y} - \tilde{X}_L\phi\right)'\left(\tilde{y} - \tilde{X}_L\phi\right) + \frac{\tau_\phi}{2}\left(\phi-\phi_0\right)' M_0 \left(\phi-\phi_0\right) \\
&+ \frac{\tau_x}{2}\left((x-r)-F\beta\right)' \Psi \left((x-r)-F\beta\right) \\
&+ \frac{\tau_x\tau_\beta}{2}\left(\tilde{\beta} - D^{-1}\beta_0^{\Delta},\right)' DA_0D\left(\tilde{\beta} - D^{-1}\beta_0^{\Delta}\right) \\
c_1^{\tau_y} &= \alpha_{y0}\log\zeta_{y0} - \log\Gamma\left(\alpha_{y0}\right) - \frac{S+T+K+P}{2}\log 2\pi + \frac{T+K}{2}\log\tau_x \\
&+ \frac{1}{2}\log detM_0 + \frac{1}{2}\log det\Psi + \frac{1}{2}\log det\left(DA_0D\right) + \frac{P}{2}\log\tau_\phi + \frac{K}{2}\log\tau_\beta \\
&+ \log\Bigg(\prod_{k=1}^{K} Bern\left(\gamma_k;\omega\right) \times Beta\left(\omega;\kappa_0,\delta_0\right) \times \prod_{t=1}^{T} Gamma\left(\psi;\nu/2,\nu/2\right) \\
&\times Gamma\left(\nu;\alpha_{\nu0},\zeta_{\nu0}\right) \times Gamma\left(\tau_x;\alpha_{x0},\zeta_{x0}\right) \\
&\times Gamma\left(\tau_\beta;\alpha_{\beta0},\zeta_{\beta0}\right) \times Gamma\left(\tau_\phi;\alpha_{\phi0},\zeta_{\phi0}\right)\Bigg) + c^{ev} \\
c_2^{\tau_y} &= c_1^{\tau_y} + \log\Gamma\left(\alpha_y\right) - \alpha_y\log\zeta_y
\end{aligned}
$$

Posterior for τ_x

$$
\log p\left(\tau_x|rest\right) = \log Gamma\left(\tau_x;\alpha_x,\zeta_x\right) + c_2^{\tau_x}
$$

where

$$
\begin{aligned}
\alpha_x &= \frac{T+K}{2} + \alpha_{0x} \\
\zeta_x &= \frac{\tau_y}{2}\left((x-r)-F\beta\right)' \Psi \left((x-r)-F\beta\right) + \frac{\tau_y\tau_\beta}{2}\left(\tilde{\beta} - D^{-1}\beta_0^{\Delta}\right)' DA_0D\left(\tilde{\beta} - D^{-1}\beta_0^{\Delta}\right) + \zeta_{0x} \\
c_1 &= \frac{T+K}{2}\log\frac{\tau_y}{2\pi} + \frac{1}{2}\log detDA_0D + \frac{1}{2}\log det\Psi + \frac{K}{2}\log\tau_\beta \\
&+ \alpha_{x0}\log\zeta_{x0} - \log\Gamma\left(\alpha_{x0}\right) \\
&+ \log\Bigg(MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right) \times MN\left(y;\Phi x,\frac{1}{\tau_y}I\right) \\
&\times \prod_{k=1}^{K} Bern\left(\gamma_k;\omega\right) \times Beta\left(\omega;\kappa_0,\delta_0\right) \\
&\times \prod_{t=1}^{T} Gamma\left(\psi;\nu/2,\nu/2\right) \times Gamma\left(\nu;\alpha_{\nu0},\zeta_{\nu0}\right) \times Gamma\left(\tau_y;\alpha_{y0},\zeta_{y0}\right) \\
&\times Gamma\left(\tau_\beta;\alpha_{\beta0},\zeta_{\beta0}\right) \times Gamma\left(\tau_\phi;\alpha_{\phi0},\zeta_{\phi0}\right)\Bigg) + c^{ev} \\
c_2 &= c_1 + \log\Gamma\left(\alpha_x\right) - \alpha_x\log\zeta_x
\end{aligned}
$$

Posterior for τ_ϕ

$$\log p\left(\tau_\phi|rest\right) = \log Gamma\left(\tau_\phi;\alpha_\phi,\zeta_\phi\right) + c_2^{\tau\phi}$$

where

$$\begin{aligned}
\alpha_\phi &= \alpha_{\phi_0} + \frac{P}{2}\\
\zeta_\phi &= \zeta_{\phi_0} + \frac{\tau_y}{2}\left(\phi-\phi_0\right)' M_0\left(\phi-\phi_0\right)\\
c_1^{\tau\phi} &= \alpha_{\phi_0}\log\zeta_{\phi_0} - \log\Gamma\left(\alpha_{\phi_0}\right) + \frac{1}{2}\log Det\left(M_0\right) + \frac{P}{2}\log\left(\frac{\tau_y}{2\pi}\right)\\
&\quad + \log\left[MN\left(y;\Phi x,\frac{1}{\tau_y}I\right)\times MN\left(x;F\beta+r,\frac{1}{\tau_x\tau_y}\Psi^{-1}\right)\times MN\left(\beta;\beta_0+D^{-1}\beta_0^{\Delta},\frac{1}{\tau_x\tau_y\tau_\beta}\left[DA_0D\right]^{-1}\right)\right.\\
&\quad \times\prod_{k=1}^{K} Bern\left(\gamma_k;\omega\right)\times Beta\left(\omega;\kappa_0,\delta_0\right)\times\prod_{t=1}^{T} Gamma\left(\psi;\nu/2,\nu/2\right)\times Gamma\left(\tau_\beta;\tau_{\beta 0},\zeta_{\beta 0}\right)\\
&\quad \left.\times Gamma\left(\nu;\alpha_{\nu 0},\zeta_{\nu 0}\right)\times Gamma\left(\tau_x;\alpha_{x0},\zeta_{x0}\right)\times Gamma\left(\tau_y;\alpha_{y0},\zeta_{y0}\right)\right] + c^{ev}\\
c_2^{\tau\phi} &= c_1^{\tau\phi} - \alpha_\phi\log\zeta_\phi + \log\Gamma\left(\alpha_\phi\right)
\end{aligned}$$

Posterior for τ_β

$$\log p\left(\tau_\beta|rest\right) = \log Gamma\left(\tau_\beta;\alpha_\beta,\zeta_\beta\right) + c_2^{\tau\beta}$$

where

$$\begin{aligned}
\alpha_\beta \equiv&\alpha_{\beta 0} + \frac{K}{2}\\
\zeta_\beta \equiv&\zeta_{\beta 0} + \frac{\tau_x\tau_y}{2}\left(\tilde\beta - D^{-1}\beta_0^{\Delta}\right)' DA_0D\left(\tilde\beta - D^{-1}\beta_0^{\Delta}\right)\\
c_1^{\tau\beta} \equiv&\alpha_{\beta 0}\log\zeta_{\beta 0} - \log\Gamma\left(\alpha_{\beta 0}\right) + \frac{K}{2}\log\frac{\tau_x\tau_y}{2\pi} + \frac{1}{2}\log detDA_0D\\
&\quad + \log\left(MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\times MN\left(y;\Phi x,\frac{1}{\tau_y}I\right)\times MN\left(x;F\beta+r,\frac{1}{\tau_x\tau_y}\Psi^{-1}\right)\right.\\
&\quad \times\prod_{k=1}^{K} Bern\left(\gamma_k;\omega\right)\times Beta\left(\omega;\kappa_0,\delta_0\right)\\
&\quad \times\prod_{t=1}^{T} Gamma\left(\psi;\nu/2,\nu/2\right)\times Gamma\left(\nu;\alpha_{\nu 0},\zeta_{\nu 0}\right)\times Gamma\left(\tau_\phi;\tau_{\phi 0},\zeta_{\phi 0}\right)\\
&\quad \left.\times Gamma\left(\tau_y;\alpha_{y0},\zeta_{y0}\right)\times Gamma\left(\tau_x;\alpha_{x0},\zeta_{x0}\right)\right) + c^{ev}\\
c_2^{\tau\beta} \equiv&c_1^{\tau\beta} - \alpha_\beta\log\zeta_\beta + \log\Gamma\left(\alpha_\beta\right)
\end{aligned}$$

Posterior for β

First, define $\tilde{\beta}_0^{\Delta} \equiv \beta_0 - D^{-1}\beta_0$. The the conditional posterior $p(\beta|rest)$ is:

$$\log p(\beta|rest) = \log MN(\beta; \mu_\beta, \Lambda_\beta) + c_4^\beta$$

where

$$\begin{aligned}
\Lambda_\beta \equiv& \tau_x\tau_y\left[F'\Psi F + \tau_\beta D A_0 D\right] \\
\mu_\beta \equiv& \tau_x\tau_y\Lambda_\beta^{-1}\left[F'\Psi(x-r) + \tau_\beta D A_0\left(D\beta_0 + \beta_0^\Delta\right)\right] \\
c_1^\beta \equiv& \frac{T+K}{2}\log\frac{\tau_x\tau_y}{2\pi} + \frac{1}{2}\log det D A_0 D + \frac{1}{2}\log det\Psi + \frac{K}{2}\log\tau_\beta \\
&+ \log\Bigg(MN\left(\phi; \phi_0, \frac{1}{\tau_y\tau_\phi}M_0^{-1}\right) \times MN\left(y; \Phi x, \frac{1}{\tau_y}I\right) \\
&\times \prod_{k=1}^{K} Bern(\gamma_k; \omega) \times Beta(\omega; \kappa_0, \delta_0) \\
&\times \prod_{t=1}^{T} Gamma(\psi; \nu/2, \nu/2) \times Gamma(\nu; \alpha_{\nu 0}, \zeta_{\nu 0}) \\
&\times Gamma(\tau_y; \alpha_{y0}, \zeta_{y0}) \times Gamma(\tau_x; \alpha_{x0}, \zeta_{x0}) \\
&\times Gamma(\tau_\beta; \alpha_{\beta 0}, \zeta_{\beta 0}) \times Gamma(\tau_\phi; \alpha_{\phi 0}, \zeta_{\phi 0})\Bigg) + c^{ev} \\
c_2^\beta =& c_1^\beta - \frac{\tau_x\tau_y}{2}\left((x-r)'\Psi(x-r) + \tau_\beta\left(D\beta_0 + \beta_0^\Delta\right)' A_0\left(D\beta_0 + \beta_0^\Delta\right)\right) \\
c_3^\beta =& c_2^\beta + \frac{\mu_\beta'\Lambda_\beta\mu_\beta}{2} \\
c_4^\beta =& c_3^\beta - \frac{1}{2}\log det\Lambda_\beta + \frac{K}{2}\log 2\pi
\end{aligned}$$

Note that while $\beta_{0k} + \beta_{0k}^\Delta$ is the prior on the mean of variable k conditional on selection, β_{0k} is not the prior on the mean if variable k is not selected. While the difference is rarely material, it can be corrected. To derive the correction, let $\dot{\beta}_{0k}$ be the desired mean conditional on no selection and $\dot{\beta}_{0k} + \dot{\beta}_{0k}^\Delta$ is the prior mean conditional on selection. Then solve for the value conditional on $\gamma = 0$ and $\gamma = 1$:

$$\begin{aligned}
\dot{\beta}_{0k} &= \beta_{0k} + v\beta_0^\Delta \\
\dot{\beta}_{0k} + \dot{\beta}_{0k}^\Delta &= \beta_{0k} + \beta_{0k}^\Delta
\end{aligned}$$

Therefore:

$$\begin{aligned}
\beta_{0k} &= \dot{\beta}_{0k} - \frac{v}{1-v}\dot{\beta}_{0k}^\Delta \\
\beta_{0k}^\Delta &= \frac{1}{1-v}\dot{\beta}_{0k}^\Delta
\end{aligned}$$

Posterior for γ

What follows is the conditional posterior for γ_k in the scenario where each γ_k is conditionally independent of the other values of γ (denoted as γ_{-k}). In other words, $p\left(\gamma_k|\gamma_{-k}, rest\right) = p\left(\gamma_k|rest\right)$. Note that this does not imply unconditionally that $\gamma_k \perp \gamma_{-k}$ as other variables (e.g. β) influence both γ_k and γ_{-k}.

Practically this implies A_0 is diagonal, such that $a_0 \equiv diag\left(A_0\right)$. Also recall $d_k^2 \equiv \gamma_k + \frac{1-\gamma_k}{v^2}$. As the only discrete distribution, the derivation for $p\left(\gamma\right)$ proceeds somewhat differently than others. The distribution for p_k with a conditionally independent prior for β is given by $\frac{\tilde{p}(\gamma_k=1)}{\tilde{p}_k(\gamma_k=0)+\tilde{p}_k(\gamma_k=1)}$.

$$\log p\left(\gamma_k\right) = \gamma_k \log p_{\gamma k} + \left(1-\gamma_k\right)\log\left(1-p_{\gamma k}\right) + c_2^{\gamma_k}$$

where

$$\begin{aligned}
p_{\gamma k} =& \frac{\tilde{p}_{\gamma k}|_1}{\tilde{p}_{\gamma k}|_0 + \tilde{p}_{\gamma k}|_1} \\
\tilde{p}_{\gamma k}|_1 \equiv& \exp\left(-\frac{\tau_x\tau_y\tau_\beta a_{0k}}{2}\left(\tilde{\beta}_k^2 - 2\beta_{0k}^{\Delta}\tilde{\beta}_k\right)\right)\omega \\
\tilde{p}_{\gamma k}|_0 \equiv& \exp\left(-\frac{\tau_x\tau_y\tau_\beta a_{0k}}{2}\left(\frac{\tilde{\beta}_k^2}{v^2} - \frac{2\beta_{0k}^{\Delta}\tilde{\beta}_k}{v}\right)\right)\frac{1-\omega}{v} \\
c_1^{\gamma k} \equiv& \frac{1}{2}\log\frac{a_{0k}\tau_x\tau_y\tau_\beta}{2\pi} \\
&+\log\left(MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\times MN\left(y;\Phi x,\frac{1}{\tau_y}I\right)\right. \\
&\times MN\left(x;F\beta+r,\frac{1}{\tau_x\tau_y}\Psi^{-1}\right) \\
&\times\prod_{j=1,j\neq k}^{K}\left(N\left(\beta_j;\frac{\beta_0^{\Delta}}{d_j}+\beta_0,\frac{1}{\tau_x\tau_y\tau_\beta a_{0j}d_j^2}\right)\times Bern\left(\gamma_j;\omega\right)\right)\times Beta\left(\omega;\kappa_0,\delta_0\right) \\
&\times\prod_{t=1}^{T}Gamma\left(\psi;\nu/2,\nu/2\right)\times Gamma\left(\nu;\alpha_{\nu 0},\zeta_{\nu 0}\right) \\
&\times Gamma\left(\tau_y;\alpha_{y0},\zeta_{y0}\right)\times Gamma\left(\tau_x;\alpha_{x0},\zeta_{x0}\right) \\
&\left.\times Gamma\left(\tau_\beta;\alpha_{\beta 0},\zeta_{\beta 0}\right)\times Gamma\left(\tau_\phi;\alpha_{\phi 0},\zeta_{\phi 0}\right)\right) + c^{ev} \\
c_2^{\gamma k} \equiv& c_1^{\gamma k} - \frac{\tau_x\tau_y\tau_\beta a_{0k}\left(\beta_{0k}^{\Delta}\right)^2}{2} \\
c_3^{\gamma k} \equiv& c_2^{\gamma k} + \log\left(\tilde{p}_{\gamma k}|_1 + \tilde{p}_{\gamma k}|_0\right)
\end{aligned}$$

Note that the normalization is accounted for in $c_3^{\gamma k}$. The normalization is fully revealed as the true probabilities must add to one. Similarly, the approximate distribution for any γ_k is given by $\frac{\tilde{q}_k(\gamma_k=1)}{\tilde{q}_k(\gamma_k=0)+\tilde{q}_k(\gamma_k=1)}$.

Posterior for γ (General Case)

To get the off-diagonal terms for A_0, we use the below generalization:

$$p(\gamma_k) = \gamma_k \log p_{\gamma k} + (1-\gamma_k)\log(1-p_{\gamma k}) + c_3^{\gamma k}$$

where

$$\begin{aligned}
p_{\gamma k} \equiv& \frac{\tilde{p}_{\gamma k}|_1}{\tilde{p}_{\gamma k}|_0 + \tilde{p}_{\gamma k}|_1} \\
\tilde{p}_{\gamma k}|_1 \equiv& \exp\left(-\frac{\tau_x\tau_y\tau_\beta}{2}\left[\left(\tilde{\beta}' D A_0 D \tilde{\beta} - 2\tilde{\beta}' D A_0 \beta_0^\Delta\right)\right]_{d_k=1}\right)\omega \\
\tilde{p}_{\gamma k}|_0 \equiv& \exp\left(-\frac{\tau_x\tau_y\tau_\beta}{2}\left[\left(\tilde{\beta}' D A_0 D \tilde{\beta} - 2\tilde{\beta}' D A_0 \beta_0^\Delta\right)\right]_{d_k=v^{-1}}\right)\frac{1-\omega}{v} \\
c_1^{\gamma k} \equiv& \frac{K}{2}\log\frac{\tau_x\tau_y\tau_\beta}{2\pi} + \frac{1}{2}\log detA_0 + \sum_{j=1, j\neq k}^{K}\log d_j \\
&+ \log\left(MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\times MN\left(y;\Phi x,\frac{1}{\tau_y}I\right)\right. \\
&\times MN\left(x;F\beta + r,\frac{1}{\tau_x\tau_y}\Psi^{-1}\right) \\
&\times \prod_{j=1, j\neq k}^{K} Bern(\gamma_j;\omega)\times Beta(\omega;\kappa_0,\delta_0) \\
&\times \prod_{t=1}^{T} Gamma(\psi;\nu/2,\nu/2)\times Gamma(\nu;\alpha_{\nu 0},\zeta_{\nu 0}) \\
&\times Gamma(\tau_y;\alpha_{y0},\zeta_{y0})\times Gamma(\tau_x;\alpha_{x0},\zeta_{x0}) \\
&\left.\times Gamma(\tau_\beta;\alpha_{\beta 0},\zeta_{\beta 0})\times Gamma(\tau_\phi;\alpha_{\phi 0},\zeta_{\phi 0})\right) + c^{ev} \\
c_2^{\gamma k} \equiv& c_1^{\gamma k} - \frac{\tau_x\tau_y\tau_\beta}{2}\left(\beta_0^\Delta\right)' A_0 \beta_0^\Delta \\
c_3^{\gamma k} \equiv& c_2^{\gamma k} + \log\left(\tilde{p}_{\gamma k}|_1 + \tilde{p}_{\gamma k}|_0\right)
\end{aligned}$$

Posterior for ω

$$\log p(\omega) = \log Beta(\kappa,\delta) + c_3^\omega$$

where

$$
\begin{aligned}
\kappa \equiv & \kappa_0 + \sum_{k=1}^{K} \gamma_k \\
\delta \equiv & \delta_0 + K - \sum_{k=1}^{K} \gamma_k \\
c_1^{\omega} \equiv & -\log B\left(\kappa_0, \delta_0\right) + \log \left(MN\left(\phi; \phi_0, \frac{1}{\tau_y \tau_\phi} M_0^{-1}\right) \times MN\left(y; \Phi x, \frac{1}{\tau_y} I\right) \right. \\
& \times MN\left(x; F\beta + r, \frac{1}{\tau_x \tau_y} \Psi^{-1}\right) \times MN\left(\beta; \beta_0 + D^{-1}\beta_0^{\Delta}, \frac{1}{\tau_x \tau_y \tau_\beta} \left[D A_0 D\right]^{-1}\right) \\
& \times \prod_{t=1}^{T} Gamma\left(\psi; \nu/2, \nu/2\right) \times Gamma\left(\nu; \alpha_{\nu 0}, \zeta_{\nu 0}\right) \\
& \times Gamma\left(\tau_y; \alpha_{y0}, \zeta_{y0}\right) \times Gamma\left(\tau_x; \alpha_{x0}, \zeta_{x0}\right) \\
& \left. \times Gamma\left(\tau_\beta; \tau_{\beta 0}, \zeta_{\beta 0}\right) \times Gamma\left(\tau_\phi; \tau_{\phi 0}, \zeta_{\phi 0}\right) \right) + c^{ev} \\
c_2^{\omega} \equiv & c_1^{\omega} + \log B\left(\kappa, \delta\right)
\end{aligned}
$$

Posterior for ψ_t

Conditional posterior:

$$
\log p\left(\psi_t\right) = \log Gamma\left(\psi_t, \alpha_{\psi t}, \zeta_{\psi t}\right) + c_2^{\psi t}
$$

where

$$
\begin{aligned}
\alpha_{\psi t} \equiv& \frac{\nu+1}{2} \\
\zeta_{\psi t} \equiv& \frac{\nu}{2} + \frac{\tau_x \tau_y}{2} \left((x_t - r_t) - f_t' \beta \right)^2 \\
c_1^{\psi t} \equiv& \frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma \left(\frac{\nu}{2} \right) + \frac{1}{2} \log \left(\frac{\tau_x \tau_y}{2\pi} \right) \\
&+ \log \left(\prod_{j=1, j\neq t}^{T} \left[N \left(x_t; f_t' \beta + r, \frac{1}{\tau_x \tau_y \psi_t} \right) \times Gamma \left(\psi_j; \frac{\nu}{2}, \frac{\nu}{2} \right) \right] \right. \\
&\times MN \left(\phi; \phi_0, \frac{1}{\tau_y \tau_\phi} M_0^{-1} \right) \times MN \left(y; \Phi x, \frac{1}{\tau_y} I \right) \\
&\times MN \left(\beta; \beta_0 + D^{-1} \beta_0^{\Delta}, \frac{1}{\tau_x \tau_y \tau_\beta} \left[D A_0 D \right]^{-1} \right) \\
&\times \prod_{k=1}^{K} Bern \left(\gamma_k; \omega \right) \times Beta \left(\omega; \kappa_0, \delta_0 \right) \times Gamma \left(\nu; \alpha_{\nu 0}, \zeta_{\nu 0} \right) \\
&\times Gamma \left(\tau_y; \alpha_{y0}, \zeta_{y0} \right) \times Gamma \left(\tau_x; \alpha_{x0}, \zeta_{x0} \right) \\
&\left. \times Gamma \left(\tau_\beta; \alpha_{\beta 0}, \zeta_{\beta 0} \right) \times Gamma \left(\tau_\phi; \alpha_{\phi 0}, \zeta_{\phi 0} \right) \right) + c^{ev} \\
c_2^{\psi t} \equiv& c_1^{\psi t} - \alpha_{\psi t} \log \zeta_{\psi t} + \log \Gamma \left(\alpha_{\psi t} \right)
\end{aligned}
$$

Posterior for ν

Conditional:

$$
p(\nu) = \left(\frac{\nu}{2} \right)^{\frac{T\nu}{2} + \alpha_{\nu 0} - 1} \Gamma^{-T} \left(\frac{\nu}{2} \right) \exp \left(\frac{\nu \eta_1}{2} \right) \eta_2 \exp c_3^{\nu}
$$

where

$$
\begin{aligned}
\eta_1 \equiv& \sum_{t \in 1:T} (\log \psi_t - \psi_t) - 2\zeta_{\nu 0} \\
\eta_2 \equiv& \left[\int_{\nu^-}^{\infty} \left(\frac{\nu}{2} \right)^{\frac{T\nu}{2} + \alpha_{\nu 0} - 1} \Gamma^{-T} \left(\frac{\nu}{2} \right) \exp \left(\frac{\nu \eta_1}{2} \right) d\nu \right]^{-1} \\
c_1^{\nu} \equiv& \alpha_{\nu 0} \log \zeta_{\nu 0} - \log \Gamma (\alpha_{\nu 0}) - \log \int_{\nu -}^{\infty} [1 - Gamma (z; \alpha_{\nu 0}, \zeta_{\nu 0})] \, dz \\
& + \log \left(MN \left(y; \Phi x, \frac{1}{\tau_y} I \right) \times MN \left(\phi; \phi_0, \frac{1}{\tau_y \tau_\phi} M_0^{-1} \right) \right. \\
& \times MN \left(\beta; \beta_0 + D^{-1} \beta_0^{\Delta}, \frac{1}{\tau_x \tau_y \tau_\beta} [D A_0 D]^{-1} \right) \times MN \left(x; F\beta + r, \frac{1}{\tau_x \tau_y} \Psi^{-1} \right) \\
& \times \prod_{k=1}^{K} Bern (\gamma_k; \omega) \times Beta (\omega; \kappa_0, \delta_0) \\
& \times Gamma (\tau_x; \alpha_{y0}, \zeta_{y0}) \times Gamma (\tau_y; \alpha_{y0}, \zeta_{y0}) \\
& \left. \left. \times Gamma (\tau_\beta; \alpha_{\beta 0}, \zeta_{\beta 0}) \times Gamma (\tau_\phi; \alpha_{\phi 0}, \zeta_{\phi 0}) \right) \right] + c^{ev} \\
c_2^{\nu} \equiv& c_1^{\nu} + (\alpha_{\nu 0} - 1) \log 2 - \sum_{t \in 1:T} \log \psi_t \\
c_3^{\nu} \equiv& c_2^{\nu} - \log \eta_2
\end{aligned}
$$