**Generating Returns for Alternative Assets**

iCapital Portfolio Analytics,

Quantitative Research Group

Abstract

We develop a returns-generating process to estimate the performance of alternative investments such as private equity and hedge funds. The model uses hierarchical Bayesian techniques to address challenges posed by limited disclosure, illiquidity, and appraisal-based valuations that are common in alternative asset classes. We use a unique dataset of fund-level returns provided to us by general partners.

By jointly estimating latent economic returns and systematic factor exposures, we construct back cast and forecast returns to assess an alternative assets’ fit within a broader portfolio. Key contributions include a unique moving average process to mimic reporting lags, use of time-weighted returns within this context, and stochastic search variable selection (SSVS) to identify significant factor exposures. The model is evaluated through explanatory power and exposure significance tests. This paper aims to improve industry analytics as applied to alternative investments and support the integration of alternatives into multi-asset portfolios.

|  |
| --- |
|  |
| *Figure 1: An example of a fund’s returns as generated by our model. The fund was launched in 2019. Our model back casted 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 back cast helps investors pair the fund with a portfolio of publicly available securities.* |

# Introduction and Motivation

Preqin estimates that the weighted average asset allocation in 2023 of 1,785 private pension plans with $3.8tn AUM was 19.1% to alternatives[[1]](#footnote-2). Furthermore, in Preqin’s study of fundraising from US pensions[[2]](#footnote-3), 708 public pension funds with a total AUM of $5.9tn had an average target allocation of just under 26% to alternatives. This indicates that alternative investments as an asset class are currently large and increasing in both, size and importance.

On the investor-side, SEC rules defining what an accredited investor have recently been broadened to include “individual investors that have the knowledge and expertise to participate in [financial] markets.”. This has hastened the market expansion and liquidity of alternative investments and made it a viable option for mass affluent investors and private wealth funds[[3]](#footnote-4).

For successful inclusion of alternatives into investor portfolios, optimal asset allocation across securities is a standard part of the investment process. This requires estimates of risk, return, and covariance of the focal assets. Data for alternative asset classes are not as freely available as they are for publicly available securities. Since alternative asset classes are largely exempt from public disclosure requirements, performance metrics based on actual transactions are scarce. This limited data impedes the investment process for portfolios containing such assets.

We address these problems and other issues associated with returns on illiquid and alternative investments via our returns-generating process. We devise a returns-based model that uses hierarchical Bayesian modeling techniques to simulate the underlying economic returns for funds on iCapital’s Marketplace. The model helps investors research and educate themselves on the compatibility of a particular fund with their investment needs and risk preferences.

## Why Bayesian?

Bayesian techniques offer a more robust and flexible approach to capture the complex dynamics of alternative investment returns due to their ability to:

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

Bayesian techniques provide a principled and effective framework for handling sparse data in the context of alternative investment returns. By incorporating prior knowledge, Bayesian methods allow researchers to leverage existing information about market dynamics, factor relationships, and investor behavior to inform the estimation process. This is particularly valuable when dealing with limited track records, as the prior information can help to fill in gaps and provide a more complete picture of the underlying return generating process.

Another key advantage of Bayesian methods is their ability to update estimates as new data becomes available. Unlike traditional frequentist approaches, which often rely on fixed sample sizes and static model specifications, Bayesian techniques allow for the continuous updating of parameter estimates and return predictions as new information arrives. This dynamic updating process is especially useful in the rapidly evolving landscape of alternative investments, where market conditions, investor preferences, and regulatory environments can change quickly.

By leveraging prior knowledge, handling limited and illiquid data, and adapting to new information, Bayesian techniques offer a powerful and flexible toolkit for modeling the complex dynamics of alternative investment returns. As such, they have become increasingly popular among researchers and practitioners seeking to unlock the potential of these asset classes while managing their unique challenges.

### Short Track Records

Short track records in alternative assets returns are a problem that is well-known in the alternative assets’ literature. See, for example, Kaplan and Schoar [2005], who show that there is a dearth of performance metrics based on actual transactions. 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 alternative assets, the lack of data is a significant barrier.

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, such as short track records. The framework aligns well with the unique characteristics of alternative investments, allowing for more accurate and informed returns estimation.

A robust corpus of academic research exists to support the use of these techniques in alternative asset analysis. For instance, Korteweg and Sorensen [2017] develop a Bayesian approach to estimate the risk and return of venture capital investments, addressing the challenges of limited data and selection bias. Ang, et al [2013] use Bayesian techniques to estimate the returns of private equity investments using cash flow data from limited partners. The authors address the challenges of limited disclosure and the irregular timing of cash flows, which are common in private equity investments. Dixon and Chong [2014] show that in the scarcity of information, the complexity of selecting successful portfolio companies, together with the absence of an empirically substantiated general formulaic approach to venture capital and private equity investments, Bayesian methodologies can be used successfully. Avramov and Zhou [2010] illustrate the efficacy of using Bayesian portfolio analysis in the study of hedge fund investing, among other more public asset classes.

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.

### Smoothing

Another well-known issue for alternative asset analysis in the finance practitioner literature that returns from alternative investment 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 other- wise 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. For some properties, the lack of a new valuation within a quarter might result in a carry-over of their last known value into the current quarter.

Our model deals with this issue 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.

The economic assumption behind the smoothing process is simple: that the observed fund returns are a weighted average of the fund’s economic returns over the most recent periods, including the current period. This restriction is like the one in Getmansky et al [2004] where the observed return for some period, is a weighted average of the “true” returns over the most recent periods. In our case, the observed returns are the reported returns, the “true” returns are the latent returns generated using the model.

### Internal Rates of Return (IRR)

It is well-known in the academic finance literature that for PE and other alternative investments 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 PE, 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 with long 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. Many academics[[4]](#footnote-5) 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. For example, if a 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.

We, on the other hand, use time-weighted returns.

Though time-weighted returns have their own caveats[[5]](#footnote-6), 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 alts 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 fund performance facilitates understanding and communication about fund performance, especially for stakeholders more familiar with traditional investment vehicles.

# The Model

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

where is a length vector of factor returns including an intercept dummy, is a matrix of exposures, is the predicted return, is the risk-free rate of return, and is the error. The subscripts here are for an individual fund () for a given period (). Henceforth, assume that all measures are for an individual fund, and so the -subscript will be dropped.

We use a superset 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 can be found in the Appendix.

While this may seem like a conventional way to estimate returns, there are two important things to be noted about the equation above:

1. is NOT the final return provided to the investor. Alternative investment reported returns often serially correlated and lagged. The quantity 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 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 the return factors in the regression, our model 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 below, and the Appendix for details.

We use Markov Chain Monte Carlo (MCMC) to compute estimates of as well as the distributional parameters of . The final product displays only point estimates – which are calculated as averages across draws in the MCMC process. For MCMC, each draw is a distribution and not a point estimate (more details on this in the Appendix). So, the expected, predicted, de-smoothed returns estimated from simulations are given by:

Since this is the result of a hierarchical Bayesian regression, the true conditional distribution of the error terms can be characterized to arbitrary accuracy in a numerical sense:

## Re-Smoothing

|  |
| --- |
|  |
| *Figure 2: An example of a fund’s de-smoothed (), and re-smoothed () returns as generated by our model. 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.* |

As mentioned earlier, it is well-known in the finance practitioner literature that returns from alternative investment such as PE, hedge funds, or real estate funds, are highly serially correlated (see, for example, Getmansky et al. [2004]). The positive serial return correlation commonly leads to a downward bias in estimated return variance.

The latent returns , henceforth just for notational simplicity, generated above are not smoothed. They are the true economic returns. Since our model attempts to mimic true reported returns, we must re-smooth the latent returns to reflect the reported returns so that investors can appropriately compare their investments in Alts with publicly available investments. In our case, 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 returns presented to the user are re-smoothed using the following process:

where is a vector of the reported returns; is a matrix consisting of smoothing coefficients, and (defined in more detail below); and is a vector of values. When both and have the same reporting frequency:



is a vector of unrestricted coefficients, and is a vector of restricted and unrestricted coefficients. and represent terms that map the observation period of the reported, smoothed returns to the latent, de-smoothed returns .

The reported returns are s.t. , and the latent returns are s.t. . Here, may or may not equal ; for example, when the reported returns are quarterly, and the factor returns are monthly[[6]](#footnote-7). To resolve this, we have the following formula that maps to :

If reported returns ( are quarterly, and latent returns () are monthly, , and (since each quarter consists of three months). And so, for the first quarter, and . When , the values in 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 of values in is 3 for quarterly data and 1 for monthly data[[7]](#footnote-8).

In particular,

Also, the complete posterior distribution of is given by:

is a global precision parameter estimated for independent measurement variance. The precision parameter has a Gamma-distributed prior: with given shape hyperparameter, , and given inverse scale hyperparameter, . is an identity matrix.

The economic assumption behind the smoothing process is simple: that the observed fund returns () is a weighted average of the fund’s economic returns, , over the most recent () periods, including the current period. The econometric implications of this are that under the given assumption, the observed fund returns follow a MA process of order .

This restriction is similar to Getmansky et al [2004] where the observed return () for some period , is a weighted average of the “true” returns () over the most recent periods: , with to ensure that all information is eventually incorporated into observed returns, and for . In our case the observed returns are the reported returns, , the “true” returns are the latent returns generated using the factors, , and the terms (or as we notate them, terms) are generated using a multivariate normal distribution in the hierarchical Bayesian model, as follows:

Here, is as defined above, and is a global precision multiplier parameter, distributed as with given shape and inverse scale hyperparameters; is a matrix consisting of precision hyperparameters for ; and is a hyperparameter (i.e., fed into the model).

## A Brief Digression Regarding 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[[8]](#footnote-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 parametrization 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.

## Stochastic Search Variable Selection (SSVS)

|  |
| --- |
| A graph of a person  Description automatically generated |
| *Figure 3: The probability mass function of a fund’s intercept term () 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.* |

As mentioned above, the regression performed to estimate is using the SSVS technique which defines prior regression coefficient variances consistent with inclusion or effective exclusion. In other words, it 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 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[[9]](#footnote-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 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).

|  |
| --- |
| *A graph of a graph  Description automatically generated* |
| *Figure 4: The probability mass function of a fund’s intercept term (β0) 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.* |

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

Here, is the regression vector, and is its hypervariance matrix. The prior for the hypervariance plays a critical role in how effective the technique is for variable selection. A successful and popular choice of are priors that make use of the mixture distributions involving a spike near zero. In George and McCulloch [1993], the prior for was assumed to have a two-component distribution of the form:

The value for (the spike) is chosen as some small value, where “small” is typically based on the data at hand, while , also data-specific, is chosen so that (the slab) is sufficiently large. Selecting the two hyperparameters in this way allows to be small or large, and this in turn enables the posterior of to shrink towards zero or be some nonzero value. The values 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 in which case the prior is referred to as an indifference prior.

We use a Bernoulli prior on , the selector variable:

Here, the probability of factor selection , depends on hyperparameters and . , , is a vector and selector variable used to determine which of the thirteen factors are economically significant in a probabilistic sense. Initially, each of the factors have an equally likely chance of being included[[10]](#footnote-11). Conditional on the factor being in the regression, specify a prior distribution for the regression coefficient associated with that variable. Our model uses a multivariate normal:

Here, is the prior mean of conditional on exclusion (); is the shift in the prior mean of conditional on inclusion (); is a vector – a function of which adjusts for sparsity; is a matrix, with ; and is the variance of the spike distribution as a fraction of the slab variance. And finally, , , and are all gamma-distributed with shape and inverse-scale hyperparameters as shown earlier.

|  |
| --- |
| A graph of a normal distribution  Description automatically generated |
| *Figure 5: The probability mass function of a fund index intercept term (β0) 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.* |

# Model Efficacy

Every model must be checked to determine whether the model provides valuable enough inference about an asset. In our case, we utilize two different sets of criteria for its primary model efficacy tests:

1. An explanatory power test measuring the model’s efficacy relative to a baseline mean model.
2. Significance tests for factor exposures that assess whether a model has a net positive (negative) exposure to a particular return-based risk factor.

These tests form the basis of model efficacy. The criteria are different depending on the analysis shown:

1. An asset is considered reliable for the purposes of the growth chart if it passes the goodness of fit test AND at least one coefficient OR the intercept passes the exposure credibility test.
2. An asset is considered reliable for the purposes of the factor radar if it is considered reliable for the purposes of the growth chart OR at least one coefficient, EXCLUDING the intercept, passes the exposure test.

The two avenues for reliability provide a path for showing significant risk exposures even when the model is not capturing enough of the data variation to merit showing the growth chart. If the model’s inferences on an asset fail both the goodness of fit test and the exposure test, the model is unlikely to provide enough valuable insights to be worth showing on the platform.

## Explanatory Power Test

This test defines the model’s fit in terms of its efficacy relative to the mean model. The measurement is made in terms of tracking error:

where is the data, is the matrix that transforms quarterly predicted returns from monthly de-smoothed returns, is the exposure vector (including the intercept),   is the average reported return, is the matrix of factor return data, and is the risk-free rate.

The test statistic we use is analogous to that used for calculated as a ratio of explained sum of squares to total sum of squares:

Implementation contains two components:

1. The main test incorporates the full distribution of outcomes predicted by the model. From a Bayesian perspective, this distribution of outcomes is implied by the data and the priors. The test statistic specifically computes the probability that a particular draw from the posterior is valuable, with value defined as performance better than the mean model:

where is an indicator function.

High values indicate that the model is likely to impart useful information over a large portion of the parameter distribution. We require that , indicating that the model is more likely than not predictive, as a necessary but insufficient condition for showing the growth chart.

1. The second part of the test serves as a sanity check for the first. Specifically, we confirm that the actual values used on the platform are additive over the mean model. For all assets for which Architect displays analytics, we impose the requirement:

The expectations are taken with respect to the posterior distribution, .

Any asset for which automatically fails the goodness of fit test[[11]](#footnote-12).

### Limitations

This test verifies that the model is capturing the important features of the data better than a much simpler and more parsimonious model. However, this is different than testing the model’s accuracy with respect to modeling the true data generating process.

A model that does not fit the data well relative to a mean model is unlikely to perform well in predicting returns conditional on factor exposures. However, a model which fits the data well may simply fit due to the additional degrees of freedom contained within the model.

In ordinary least squares, adding parameters always improves the fit, even if the additional relationships are spurious. In our model, the effect depends on the priors. Stronger priors imply more regularization and less of a tendency to fit to spurious signals, at the cost of biasing the results towards the prior. We mix the evidence for exposures in the data with information present in the priors, mitigating but not eliminating the tendency of regressions to fit towards spurious signals.

The possibility of a spurious fit motivates inspection of the credibility of exposures themselves, and forms the basis of the Exposure Significance Test.

## Exposures Significance Test

As exposures to factors drive much of the identification of other model parameters, highly credible exposures generally coincide with model efficacy. Specifically, Architect tests if the data and priros imply a high probability that the exposure is greater (less than) zero. The test statistic for each coefficient is given by:

is a cutoff value (currently 90%). The test can be easily calculated from the posterior distribution for all coefficients and the intercept.

For the purposes of the growth chart, funds must pass the previously described goodness of fit test AND pass the exposure test for at least one coefficient or the intercept. Inclusion of the intercept provides a path for funds with highly hedged exposures to credible excess returns and appear in the growth chart.

For the purposes of the factor radar, funds must either meet the criteria for the growth chart OR pass the exposure test for at least one coefficient excluding the intercept.

A model that fails goodness of fit tests may still provide useful information on exposures. For instance, the model might fail to capture the moving average components, or certain parameters may have very high plausible ranges given the data. This motivates showing the factor radar for certain funds where predictive test fails and display of the growth chart is not appropriate.

# Conclusion

Our model represents a significant advancement in the quantitative analysis of alternative investments, such as private equity and hedge funds. By employing hierarchical Bayesian modeling techniques, we effectively address the challenges posed by limited and noisy data, illiquidity, and the need for regular updates as new information becomes available.

The model's ability to estimate unobservable true economic returns and re-smooth them to mimic reported returns provides investors with valuable insights into fund performance, enabling more informed decision-making. The use of time-weighted returns and the incorporation of stochastic search variable selection (SSVS) further enhance the model's robustness and adaptability to various market conditions and investment scenarios.

Our model efficacy tests, including explanatory power tests and significance tests for factor exposures, ensure that the model's outputs are reliable and actionable. The detailed mathematical formulas and econometric techniques documented in this manual demonstrate the model's scientific foundation and promote transparency.

As alternative investments continue to gain prominence in the portfolios of institutional and individual investors, the need for sophisticated quantitative tools like ours will only grow. By providing a systematic and data-driven approach to evaluating alternative investment funds, we contribute to the democratization of these asset classes and supports the ongoing expansion and liquidity of the alternative investment market.

# References

1. Andrew Ang, Bingxu Chen, William N. Goetzmann, and Ludovic Phalippou. 2018. Estimating private equity returns from limited partner cash flows. Journal of Finance, 73 (4): 1751-1783.
2. Matthew Dixon and Jike Chong. 2014. A Bayesian Approach to Ranking Private Companies Based on Predictive Indicators. AI Communications, 7 (2): 173–188
3. Doron Avramov and Guofu Zhou. 2010. Bayesian Portfolio Analysis. Review of Financial Economics, 2:25-47.
4. David Geltner. Estimating market values from appraised values without assuming an efficient market. The Journal of Real Estate Research, 8(3):324–345, 1993.
5. Edward I. George and Robert E. McCulloch. Variable selection via gibbs sampling. Journal of the American Statistical Association, 88(423):881–889, 1993.
6. 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.
7. Hemant Ishwaran and J. Sunil Rao. Spike and slab variable selection: frequentist and bayesian strategies. The Annals of Statistics, 33(2):730–733, 2005.
8. Steven N Kaplan and Antoinette Schoar. Private equity performance: Returns, persistence, and capital flows. The Journal of Finance, 60(4):1791–1823, 2005.
9. Harry M. Markowitz. Portfolio selection. Journal of Finance, 7(1):77–91, 1952.
10. T.J. Mitchell and J.J. Beauchamp. Bayesian variable selection in linear regression. Journal of the American Statistical Association, 83:1023–1036, 1988.
11. Ludovic Phalippou. The hazards of using irr to measure performance: The case of private

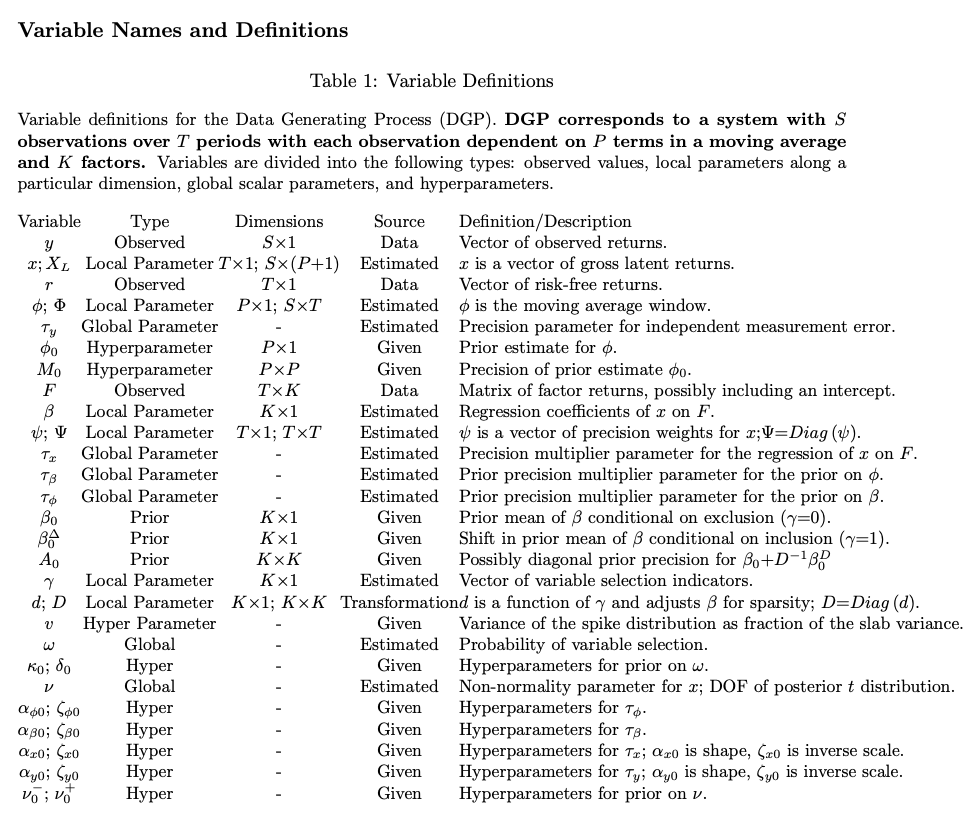
equity. 2008. https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=1111796.

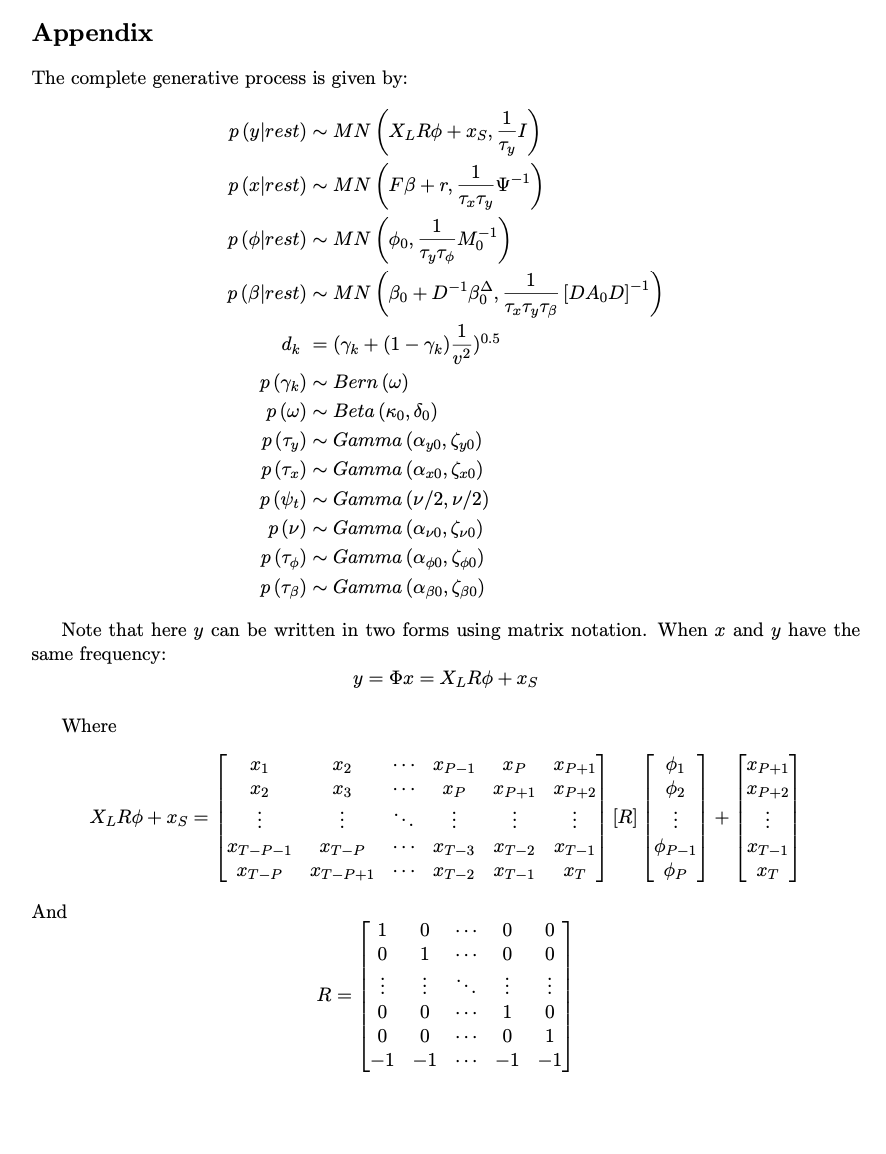
1. Juho Piironen and Aki Vehtari. Comparison of bayesian predictive methods for model selection. Statistics and Computing, 27:711–735, 2017.
2. Morten Sorensen and Ravi Jagannathan. The public market equivalent and private equity performance, 2013. https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=2347972.

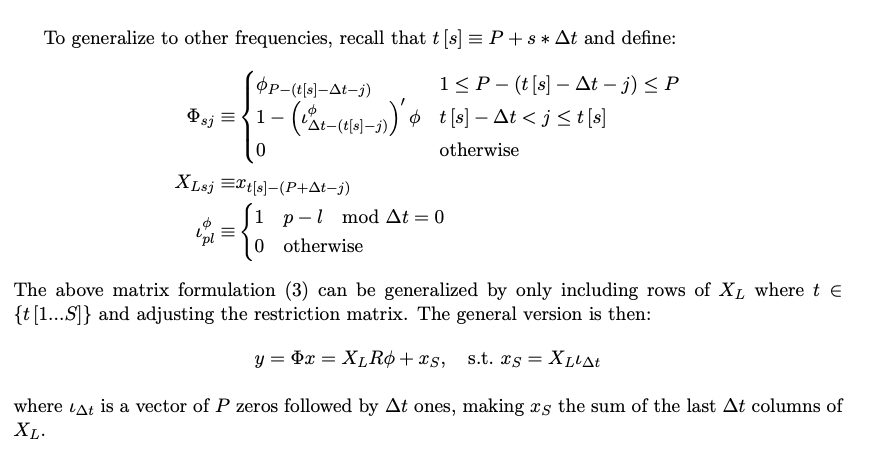
# Factor Definitions

|  |  |
| --- | --- |
| Factor Name | Index |
| Alt Commodities | Bloomberg Commodity Index |
| Alt HF Crowding | *Difference Between:*  Barclays Form 13 Filing;  Russell 3000 Total Return Index |
| Alt Oil | *Average of:*  Middle East Crude Oil;  WTI Crude Oil;  Brent Crude Oil |
| Alt Trend | Credit Suisse Managed Futures Index |
| Emerging Markets | *Difference Between:*  MSCI Emerging Markets Index (PR);  MSCI ACWI (PR) |
| Equity Market | MSCI ACWI (TR) |
| Equity Momentum | *Difference Between:*  MSCI ACWI Momentum NR USD;  MSCI ACWI IMI |
| Equity Quality | *Difference Between:*  MSCI ACWI Quality NR USD;  MSCI ACWI IMI |
| Equity SmallCap | *Difference Between:*  S&P BMI Global Small-cap PR;  S&P Global Broad Market PR |
| Equity Value | *Difference Between:*  S&P BMI Global Value PR;  S&P Global Broad Market PR |
| Fixed Credit | *Difference Between:*  Bloomberg US Corporate High Yield Index;  Bloomberg Barclays US Aggregate Bond Index |
| Fixed Duration | Bloomberg U.S. Treasury: 7-10 Year Total Return Index Value Unhedged |
| US Dollar | US Dollar Index |

# Variable Names and Definitions







A math equations and formulas

Description automatically generated

A math equation with black text

Description automatically generated

A math equations and formulas

Description automatically generated

A math equations on a white background

Description automatically generated

A math equations and formulas

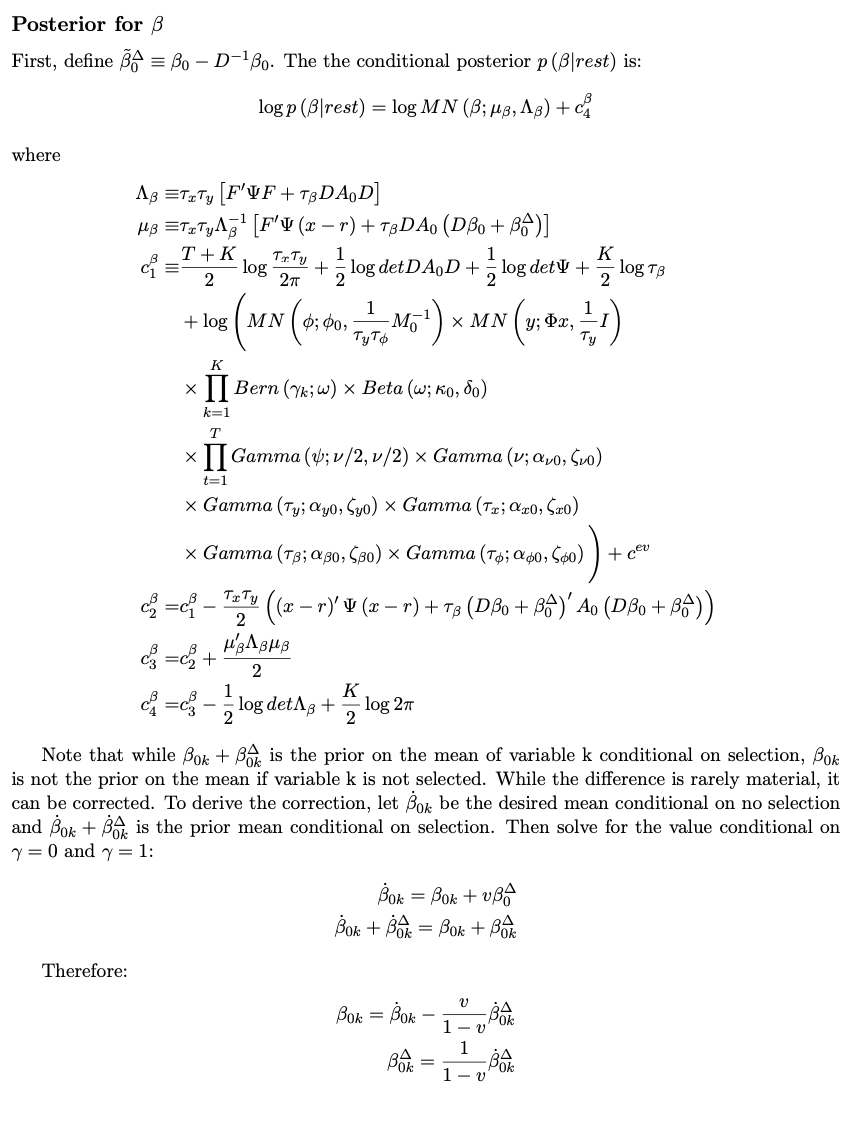
Description automatically generated

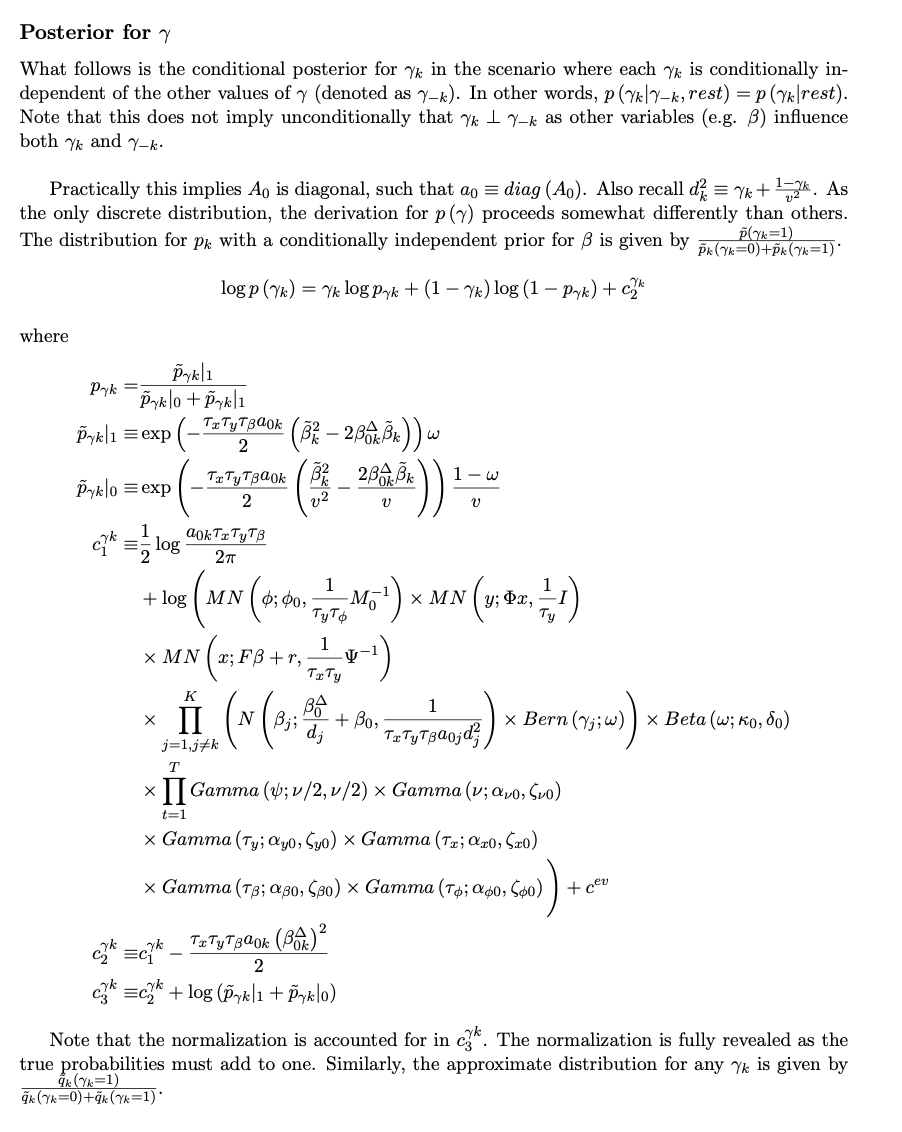
A math equations and formulas

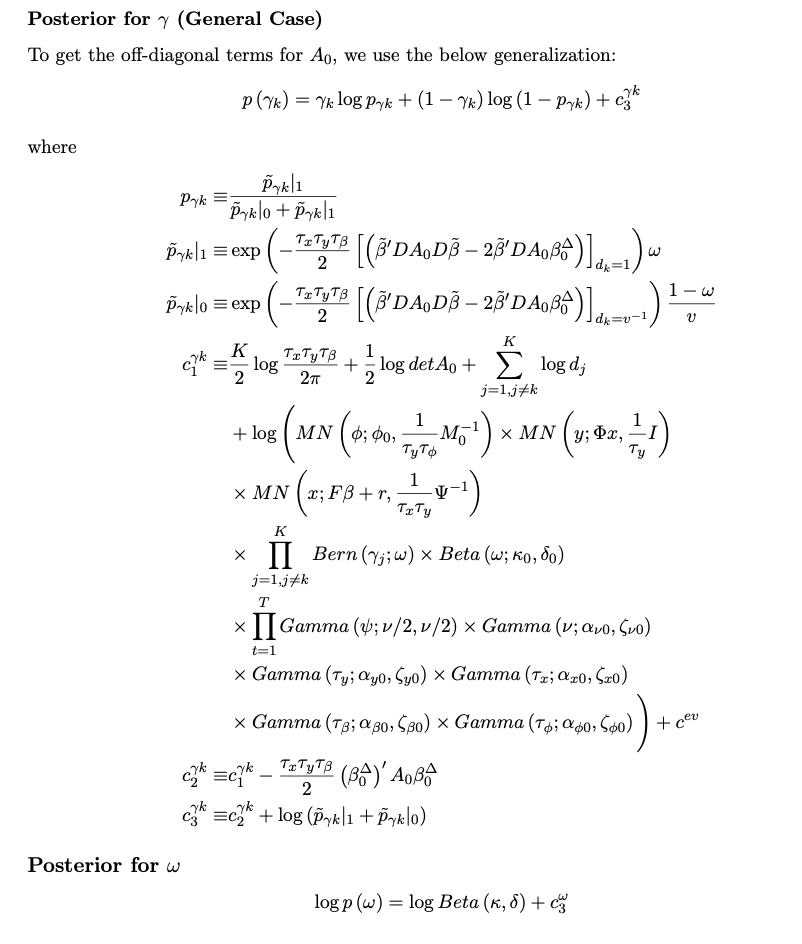
Description automatically generated with medium confidence

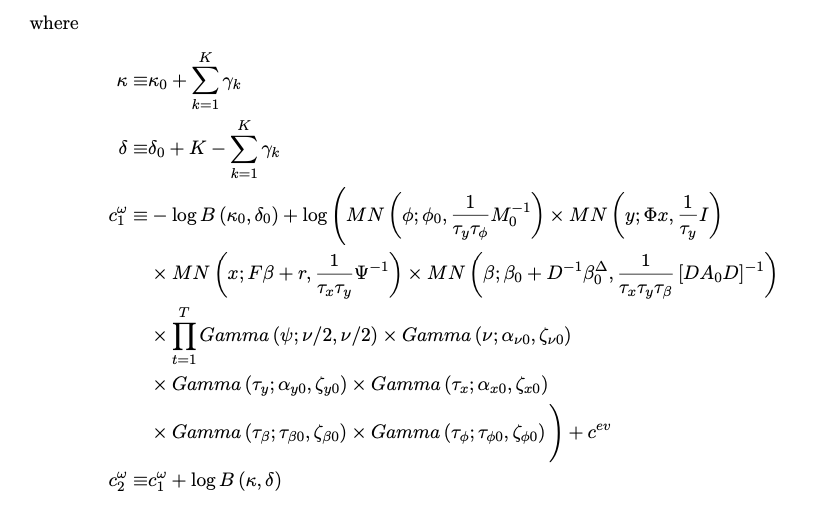
A table of math equations

Description automatically generated









A black text on a white background

Description automatically generated

A math equations and formulas

Description automatically generated

A math equation with numbers and symbols

Description automatically generated

A math equations and formulas

Description automatically generated

1. https://www.preqin.com/news/kpers-raises-alternatives-allocation-cap-as-funding-shortfalls-persist [↑](#footnote-ref-2)
2. https://www.preqin.com/insights/research/reports/fundraising-from-us-pensions-a-guide-to-raising-capital [↑](#footnote-ref-3)
3. Source: https://icapital.com/insights/practice-management/untapped-potential-alternative-investments-and-the-wealth-management-channel/ [↑](#footnote-ref-4)
4. See, for example: https://pages.stern.nyu.edu/~adamodar/pdfiles/ovhds/ch5.pdf#page=43 [↑](#footnote-ref-5)
5. They may not fully capture the complexities of PE investments, such as the strategic timing of cash flows and the long-term, illiquid nature of these assets. [↑](#footnote-ref-6)
6. Our factor returns are almost always monthly, as are the de-smoothed returns. [↑](#footnote-ref-7)
7. Note that the vector . For quarterly reported data, say where . Then   and the sum of all terms in   is 3. For monthly reported data, say where . Then   and the sum of all terms in here, is 1. [↑](#footnote-ref-8)
8. In the frequentist sense of the word. [↑](#footnote-ref-9)
9. 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. [↑](#footnote-ref-10)
10. Subject to a zero-exposure prior. [↑](#footnote-ref-11)
11. Empirically, this test is sensitive to the volatility of the asset relative to the prior. Note that the denominator in the calculations above are simply historical sample variance. A simple solution to this potential issue is to explicitly link the prior variance to the sample variance. This is the default approach in Architect as of April 10, 2024. [↑](#footnote-ref-12)