April 4, 2024

Sanity checks on model efficacy

This document describes a sanity check for model efficacy that uses extensive margin tests for conditional efficacy and makes meaningful probability statements on exposures.The approach identifies assets for which the model is particularly ineffective. Model analysis for such assets should not be shown on the Architect platform.

1 Motivation and Discussion

- The model currently provides the following important information to clients and advisors.
  - Return predictions (via the back-cast)
  - Factor exposures
- Determining if the model provides "valuable enough" inference about an asset entails assessing an asset's goodness of fit and the significance of exposures. 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.
- A third technique, posterior predictive checks, can provide useful insights in edge cases. These are discussed in the appendix.

2 Methodology

Architect utilizes two different sets of criteria for its primary model efficacy tests:

1. An explanatory power test measure the model's efficacy relative to a baseline mean model.
2. Exposure significant tests assess the likelihood that a model has a net positive (negative) exposure to a particular return-based risk factor.

These tests form the basis for showing G3 based analysis on the platform. The criteria is different depending on the analysis shown:

- 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 test.
- 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.

2.1 Explanatory power test

- Define the model's fit in terms of its efficacy relative to the mean model.

  - Measuring efficacy in terms of tracking error, the goal is to assess if

$$\sum_{s\in 1:S}\left(y_s-\Phi\left(F\beta+r\right)\right)^2\geq\sum_{s\in 1:S}\left(y_s-\overline{y}\right)^2$$

    where y is the data, Φ is the transformation matrix that forms quarterly predicted returns from monthly desmoothed returns, β is the exposure vector (including the intercept), $\overline{y}$ is the average reported return, F is the matrix of factor returns over the backcast, and r is the risk free rate over the back-cast.

    * Consider a two part goodness of fit test utilizing the following test statistic. While motivated by the above tracking error, the calculation for the test statistic coincides with that of $R^2_{ESS}\in(-\infty,1]$ calculated using the residual methodology:

$$R^2_{ESS}=1-\frac{\sum_{s\in 1:S}\left(y_s-\Phi\left(F\beta+r\right)\right)^2}{\sum_{s\in 1:S}\left(y_s-\overline{y}\right)^2}$$

    * 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:

$$\begin{aligned}R^2_{ext}=&p\left(R^2_{ESS}>0\right)\\=&\int_\Theta\iota\left[1-\frac{\sum_{s\in 1:S}\left(y_s-\Phi\left(F\beta+r\right)\right)^2}{\sum_{s\in 1:S}\left(y_s-\overline{y}\right)^2}>0\right]p\left(\Theta|D\right)d\Theta\end{aligned}$$

      where ι is an indicator function. High values indicate that the model is likely to impart useful information over a large portion of the parameter distribution. An asset's analysis is considered valuable enough to show on the platform if $p\left(R^2_{ESS}>0\right)>c_{ESS}$, where c_{ESS} is the minimum cutoff probability (tentatively 50%).

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

$$R^2_{ESS}=1-\frac{\sum_{s\in 1:S}\left(y_s-\mathbb{E}\left[\Phi\right]\left(F\mathbb{E}\left[\beta\right]+r\right)\right)^2}{\sum_{s\in 1:S}\left(y_s-\overline{y}\right)^2}$$

      where all expectations are taken with respect to the posterior distribution $p\left(\Theta|D\right)$. This metric does not have an obvious Bayesian interpretation, but serves as a starting point and is representative of the analytics shown on the MVP platform. For this reason, any asset for which $R^2_{ESS}<0$ automatically fails the goodness of fit test.

- Empirically this test is sensitive to the volatility of the asset relative to the prior. This is easy to see as the denominator in the above calculations are simply historical sample variance.

- A simple solution to this potential issue is to explicitly link the prior variance to the sample variance. See the appendix for details.

2.1.1 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 add much value.
  - However, a model which fits the data well may simply fit due to the additional degrees of freedom contained within the model. In OLS, adding parameters always improves the fit, even if the additional relationships are spurious. In G3, 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. With respect to factor exposures, G3 will weight the data against the priors, mitigating but not completely eliminating the tendency of regressions to fit towards spurious signals. For this reason, a G3 model that fits the data better than the mean model is not necessarily superior to the mean model.
- Such limitations motivate the need for additional evidence that the model is capturing useful information.

3 Exposure significance test

- A test for significant exposures determines if the exposure information, in-of-itself, is valuable enough to warrant showing the factor radar and exposure analysis to clients. The test statistic checks if the data implies a high probability that the exposure is greater (less than) zero. The test statistic for each coefficient is given by:

$$\begin{aligned} c_{exp} \leq & p_{\beta k} \\ & s.t. \\ p_{\beta k} \equiv & \max\left[p\left(\beta_k > 0\right), 1 - p\left(\beta_k > 0\right)\right] \\ p\left(\beta_k > 0\right) \equiv & \int \iota\left[\beta_k > 0\right] p\left(\Theta | D\right) d\Theta \end{aligned}$$

  and c_{exp} 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.
- 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 the second path to showing the factor radar.

4 Sample Application

The methodology was applied within the research POC to the following assets. Note that no asset specific priors were deployed, as these sorts of tests are used primarily to motivate the case for asset specific priors. The p_{coef} column gives the maximum posterior credibility for a non-intercept coefficient being greater (less) than zero. The p_{inter} gives the same value but only for the intercept. R^2_{ext} provides the result of the extensive margin test. R^2_{ESS} provides the R^2 calculated using the error sum of squares methodology using the point value expectations. μ_{ppc} gives the posterior predictive quantile of the cumulative fund return.

Research POC Label	frequency	p_{coef}	p_{inter}	R^2_{ext}	show chart?	show radar?	ppc μ
amgpantheonfundllc	monthly	0.99	0.96	0.95	yes	yes	0.934
apollodebtsolutionsbdcclassi	monthly	0.78	0.86	0.98	no	no	0.530
aresindustrialreitclasst	monthly	0.83	0.69	0.01	no	no	0.995
aresprivatemarketsclassi	monthly	0.62	0.57	0.60	no	no	0.916
aresrealestateincometrust	monthly	0.96	1.00	0.75	yes	yes	0.755
atlasenhancedfundlp	monthly	1.00	1.00	1.00	yes	yes	0.548
baringsprivatecreditmonthlyonly	monthly	0.80	1.00	0.70	yes	yes	0.574
blackstoneprivatecreditfund	monthly	1.00	1.00	1.00	yes	yes	0.174
blackstonerealestateincometrustinc	monthly	0.95	1.00	0.97	yes	yes	0.489
brookfieldreiticapitaloffshoreaccessfundsp1	monthly	0.90	0.56	0.81	yes	yes	0.901
campbellabsolutereturnprogram	monthly	1.00	0.57	1.00	yes	yes	0.609
canyonbalancedhedgefocusfundlp	monthly	1.00	0.88	1.00	yes	yes	0.701
carlyletacticalprivatecreditfund	monthly	0.96	0.93	1.00	yes	yes	0.239
fscreditrealestateincometrustreit	monthly	1.00	1.00	0.33	no	yes	0.196
goldentreemasterfund	monthly	1.00	0.78	1.00	yes	yes	0.826
hamiltonlaneprivateassetsfund	monthly	0.98	1.00	1.00	yes	yes	0.423
icapitalapolloalignedalternativeslpclassc	monthly	0.70	0.60	0.95	no	no	0.766
icapitalcoopersquareaccessfundlp	monthly	1.00	0.55	1.00	yes	yes	0.236
icapitaldoublelineopportunisticfundltd	monthly	1.00	0.71	1.00	yes	yes	0.528
icapitalhgvoraaccessfundlp	monthly	1.00	0.84	1.00	yes	yes	0.763
icapitalincomeopportunitiesfundlp	monthly	1.00	1.00	1.00	yes	yes	0.182
icapitalkingstreetcapitalaccessfundlp	monthly	1.00	0.97	1.00	yes	yes	0.892
icapitalkkrprivatemarketsfund	monthly	1.00	0.60	1.00	yes	yes	0.983
icapitalmillenniumfundlp	monthly	1.00	1.00	1.00	yes	yes	0.406
icapitalmultistrategyfundlp	monthly	1.00	1.00	0.74	yes	yes	0.750
icapitalnewalphaaccessfunduslp	monthly	1.00	0.63	1.00	yes	yes	0.629
icapitaloffshorestrategieshoneycomboffshorefundsp	monthly	1.00	0.51	1.00	yes	yes	0.510
icapitalrenaissanceidgefundltd	monthly	0.97	0.61	0.94	yes	yes	0.158

icapitalsegpartnersfundlp	monthly	1.00	0.86	1.00	yes	yes	0.709
icapitalshaccessfundlp	monthly	1.00	0.84	1.00	yes	yes	0.752
icapitalsorobanopportunitiesfundlp	monthly	1.00	0.65	1.00	yes	yes	0.804
icapitalthirdpointfundlp	monthly	1.00	0.81	1.00	yes	yes	0.580
icapitalworldquantmillenniumsealslp	monthly	0.86	0.55	0.99	no	no	0.377
mackayshieldshighincomeopportunitieslp	monthly	1.00	0.58	1.00	yes	yes	0.818
nuveenglobalcitiesreitinc	monthly	0.94	0.93	0.87	yes	yes	0.789
oakstreetnetleasetrustadvisoryclassi	monthly	0.70	0.61	0.87	no	no	0.699
owlrockcoreincomecorp	monthly	1.00	1.00	1.00	yes	yes	0.327
owlrocktechnologyincomecorp	monthly	0.87	0.68	0.99	no	no	0.729
ozdpiihedgefocusfundlp	monthly	0.98	0.56	1.00	yes	yes	0.648
pcapclassi	monthly	0.84	1.00	1.00	yes	yes	0.513
renaissanceinstitutionalequitiesfundllc	monthly	1.00	0.88	1.00	yes	yes	0.648
spfhedgefocusfundltd	monthly	0.97	1.00	1.00	yes	yes	0.265
starwoodrealestateincometrustincsreit	monthly	0.91	0.79	0.50	yes	yes	0.863
steelecreekcapitalcorporation	monthly	0.86	0.57	0.95	no	no	0.559
stepstoneprivatemarketssprim	monthly	0.67	0.74	0.08	no	no	0.949
tcapclassi	monthly	0.74	0.97	1.00	yes	yes	0.199
veritionmultistrategyfundcomposite	monthly	1.00	1.00	0.72	yes	yes	0.765
wmasystematicequityalphalsonshorefundlp	monthly	0.99	0.96	1.00	yes	yes	0.435
worldquantmillenniumwmqsgeaeonshorefundp	monthly	1.00	0.67	1.00	yes	yes	0.855
agdirectlendingcomposite	quarterly	0.90	1.00	0.87	yes	yes	0.368
blackrockprivateinvestmentsfund	quarterly	0.98	0.55	1.00	yes	yes	0.645
carlylecreditsolutionscars	quarterly	0.91	1.00	1.00	yes	yes	0.511
carlyleproperty	quarterly	0.91	0.99	0.93	yes	yes	0.569
clearlakecapitalpartnersviicapitallp	quarterly	0.93	0.55	0.89	yes	yes	0.622
genericpe5867100	quarterly	1.00	1.00	1.00	yes	yes	0.379
hlsfvicapitalaccessfundlp	quarterly	0.98	0.82	0.87	yes	yes	0.698
icapitalapolloaccordivaccessfundlp	quarterly	0.79	0.60	0.94	no	no	0.408
icapitalbcpviiiaccessfundlp	quarterly	0.96	0.63	0.99	yes	yes	0.584
icapitalblackstonegrowthaccessfundlp	quarterly	1.00	0.50	1.00	yes	yes	0.499
icapitalclearlakecapitalpartnersvaccessfundlp	quarterly	1.00	0.96	0.93	yes	yes	0.623
icapitalpofivaccessfundlp	quarterly	0.97	0.56	0.91	yes	yes	0.747
icapitalspreviiaccessfundlp	quarterly	0.89	0.75	0.62	no	no	0.736
icapitalvintageivaccessfunduslp	quarterly	0.98	0.54	1.00	yes	yes	0.583
icapitalvintagevaccessfunduslp	quarterly	0.99	0.52	1.00	yes	yes	0.543
jpminfrastructureinvestmentsfundiif	quarterly	0.91	1.00	0.96	yes	yes	0.362
nuveenchurchillprivateecoinvestmentsfundpcap	quarterly	1.00	0.60	1.00	yes	yes	0.698
msimpecoi	quarterly	0.80	0.97	0.81	yes	yes	0.636
msimpecoii	quarterly	0.78	0.84	0.99	no	no	0.554

5 Appendix: Posterior predictive checks

- Poster predictive checks analyze how rare the scenario presented by the data is relative to what would be predicted by the model.
- For example, consider the mean lifetime return of a fund. If the historical return of the fund is 17%, and this is the 80th percentile outcome as predicted by the model, then the model's predicted return is lower than the historical return, although the historical return is still plausible within the output of the model. However if it were the 99.9th percentile, then the modeled return is contradicting the data.
  - Specifically, a posterior predictive check on the mean lifetime return would consider:

$$p\left(\overline{y} \leq \overline{\hat{y}}\right) = \int_{\Theta} \iota\left(\overline{y} \leq \overline{\hat{y}}\right) p\left(\Theta|D\right) d\Theta$$

  - Technical note: In this case $\hat{y}$ is NOT the marginal distribution of the expectation of y given the model input. It is the marginal distribution of y itself given the parameters and data generating process. The difference is the expectation of idiosyncratic variance is zero, and therefore, the probability distribution of y is wider than the probability distribution of the expectation of y.
- These checks consider model fit, which is already covered on a relative basis in terms of mean squared error for the extensive margin tests. Yet the two measures consider fit from two different perspectives: on an absolute basis in the posterior predictive checks, and on a relative basis to the mean model for the extensive margin tests. There is no guarantee that the tests will agree:
  - A fund which passes the mean return posterior check and fails the extensive margin check is capturing the average return but not any better than the mean model.
  - A fund that passes the extensive margin check and fails the posterior check is likely missing important components of the data's variation, but still captures enough to be additive over the mean model.
- Setting the acceptable posterior predictive mean return percentile between 1% and 99%, among the 69 alternative assets analyzed, only KKR has a return outside the 98% credible interval for returns AND passes the growth chart test.
- Special case study: KKR/iDirect
  - For example, the since inception return of the KKR private markets fund is outside the distribution, with a realized cumulative return of 72% and an expected modeled return of 27%. Yet, the variance of the fund is 2.2% (monthly) as compared with a tracking error of 1.8% relative to the model. Therefore, the model fits the data significantly better than the mean model. The model attributes about half of the fund's performance to systematic features, and attributes the remainder to idiosyncratic variation. Inspection of the the track record shows that in four non-consecutive months, the fund outperformed the model by a combined 29%. Such "lumpy" distributions are not easily explained by systematic factors.
  - The model captures other features of the data well, enough to improve upon the mean model in practically all draws from the posterior distribution. Interpreting this result, about half of the fund's performance is explained by idiosyncratic factors outside of the model. The model

therefore captures particular features of the data well, enough to improve upon the mean model in practically all draws from the posterior distribution. In other words, the model adds significant value in explaining the results of KKR, despite being an incomplete representation.

– See section 6.3 of Gelman et al 2014 for a further discussion.