Norges Bank's Expert Group on Principles for Risk Adjustment of Performance Figures – Final Report

November 16, 2015

Magnus Dahlquist

Professor, Stockholm School of Economics

Christopher Polk

Professor, London School of Economics

Richard Priestley

Professor, Norwegian Business School BI

Bernt Arne Ødegaard

Professor, University of Stavanger

Contents

1. Background

Norges Bank is seeking to further develop performance reporting for the Government Pension Fund Global (GPFG). A new annual risk and return report is planned for publication in the first quarter of 2016. In order to ensure that the quality of this performance reporting follows industry and academic standards, Norges Bank has decided to seek external advice in this process. It is against this background that the Executive Board of Norges Bank has asked the members of this Expert Group[1] to provide views and advice on principles for risk adjustment of performance figures.

Objectives of the Expert Group

Norges Bank has provided us with Terms of Reference (ToR) for our work. The ToR are appended to this report. Based on this document, we have strived to provide advice that can help Norges Bank achieve what should be the main goal of the Bank's forthcoming annual risk and return report: to convey informative measures of the relative return on the GPFG to the fund's owner (the Ministry of Finance, the Norwegian Parliament and, ultimately, Norwegian citizens).

We were asked to give advice on which measures are the most appropriate in order for the owner to evaluate the fund's performance relative to its benchmark indices. In accordance with the ToR, we have concentrated our work on two classes of measures of relative returns:

> A. *Risk-adjusted performance measurement.* These measures use data from the portfolio and benchmark indices only.
>
> B. *Factor risk-adjusted performance measurement.* In addition to portfolio and benchmark data, these measures use data for so-called risk factors, including the standard market factor.

Norges Bank currently reports a set of measures of the fund's risk-adjusted excess return, corresponding to category A above and we have provided advice on what measures to report. We have also given advice on the use of factor risk adjustment measures, corresponding to category B above. This includes advice on model choice, factor construction, and the communication of such analyses in public reports.

Process and overview of report

The Expert Group met four times between September 7 and November 13 to discuss the items identified in the ToR and to finalize the report. Two members of Norges Bank's Executive Board[2] attended the meetings. To assist the Expert Group, Norges Bank Investment Management served as secretariat.

The structure of this report is aligned with the structure of the ToR. Section 2 provides a summary of

---

[1]   The Expert Group is comprised of four members: Professor Magnus Dahlquist, Professor Christopher Polk, Professor Richard Priestley, Professor Bernt Arne Ødeagaard. In addition to this assignment, the members have, and have previously had, other assignments for, or other ties to, Norges Bank and the Ministry of Finance. Professor Magnus Dahlquist is a member of Norges Bank's Research Council 2014-2017. Professor Christopher Polk has given research presentations for Norges Bank Investment Management three times: Investment Strategy conference – December 9th 2010; Academic Seminar October 15th 2012; and Norwegian Financial Research Conference – August 27th 2013. Professor Richard Priestley has received publication bonuses from the Norwegian Finance Initiative in 2012, 2014 and 2015, and the BI Norwegian Business School Department of Finance, headed by Professor Richard Priestley, has received funding for conferences from the Norwegian Finance Initiative. Professor Bernt Arne Ødegaard was employed part-time by Norges Bank from 1997 to 2013 and has co-authored academic papers with personnel currently employed by Norges Bank and the Ministry of Finance. Norges Bank was aware of these other assignments and ties prior to establishing this Expert Group.

[2]   Professor Egil Matsen and Professor Kjetil Storesletten.

the key recommendations. General recommendations on the reporting follow in section 3 *Reporting*. The recommendations on the two categories of risk-adjusted performance measures are presented in sections 4 and 5. In section 6 we discuss *other questions* in the ToR such as decomposition of return, risk and costs in the fund's reporting.

Based on the recommendations in Ang, Brandt and Denison (2014)[3], the Ministry of Finance requested advice from Norges Bank on how to extend and further develop the risk and return reporting. Norges Bank sent its reply to the Ministry on September 29th. In its meetings of September 7th and 24th, the Expert Group provided some preliminary inputs for this reply, focusing primarily on return decomposition.

---

[3]   "Review of the Active Management of the Norwegian Pension Fund Global," January 20, 2014. The report was commissioned by the Norwegian Ministry of Finance.

2. Summary

The objective of this report is to assess risk-adjusted performance measures that can help the owner of the GPFG evaluate the results of the investment manager. Our discussion concentrates on measures that are relevant for the GPFG's fixed-income and listed equity investments. In line with the ToR for our task, we structure our discussion around views and advice that may be useful for Norges Bank's forthcoming annual risk and return report.

Academic research does not give an unambiguous and unique standard method of measuring and reporting risk-adjusted performance. As such, even though we recommend the use of certain measures, we emphasize the importance of using alternative approaches, models, and specifications.

We recommend that Norges Bank develops the performance reporting of the fund with a view to both the wider Norwegian public and a more professional audience. This can be done by dividing performance reporting into two parts aimed at the different audiences: an overall, main report for interested readers who are not necessarily finance experts, and an appendix aimed at specialist readers.

For the overall GPFG risk and return report we recommend the following: [4]

- Norges Bank should use simple, well-known, and easily communicated measures. Specifically, for absolute returns, we recommend reporting of the Sharpe ratio. Sharpe ratios for both the fund and the benchmark should be computed and compared. For relative returns, we recommend the information ratio, appraisal ratio, and Jensen's alpha. The choice of measures as well as sample periods should be constant over time.

- Norges Bank should report factor risk-adjusted performance measures based on regression analyses. These analyses should use one preferred model, marshalled by leading experts, that represents the current state of knowledge in financial research.

- When evaluating the GPFG's equity portfolio, we recommend the global version of the Fama and French (2015) 5-factor model as a proper starting point. In these analyses, the excess return should be measured relative to GPFG's equity benchmark.

- For fixed-income analyses, there is no model that is similarly well established. However, when evaluating the GPFG's bond portfolio we recommend including the default and term factors considered by Fama and French (1993), possibly extended with additional factors in line with Ang, Brandt and Denison (2014).

- Both equity and fixed-income factors should be suitably constructed so as to take the fund's investment constraints and characteristics into account.

- When analyzing the combined performance of equities and fixed-income we recommend including a limited number of fixed-income factors and using the equity risk factors of the main model.

For the appendix, we recommend the following:

- The results from several factor models should be presented in the appendix to the overall report. The objective would be to shed light on the robustness of estimated factor exposures and the alpha reported in the main report.

- As a minimum, the appendix should report estimates based on (i) the regionally based Fama-French 5-factor model, (ii) the Fama-French 3-factor model, and (iii) Carhart's (1997) 4-

---

[4]   Please see Section 4 for definitions

factor model.

- The appendix should also report results from different sample periods (since inception and 5- and 10-year rolling windows), and possibly with different versions of the factors (including both factors incorporating the fund's investment constraints and some version of the "research factors").

- The appendix should include a full description of the methodology for each of the measures presented in the overall report. To the widest extent possible, the underlying data for the analyses in both the main report and the appendix should be made publicly available.

Other recommendations:

- Norges Bank should decompose return, risk and costs into the main investment strategies, starting with the internal "operational reference portfolio", and divided into systematic factor exposures and expansion of the portfolio universe. Norges Bank should also consider further decomposing the results into three separate and more detailed levels, namely tactical allocation decisions, enhanced indexing, and security selection.

- In addition to the main report and the appendix, we recommend that Norges Bank produces a research paper with more in-depth discussion of factor risk-adjusted performance measurement, factor construction and, possibly, further alternative factor models. The aim of this research paper should be to shed light on the effect of the fund's investment constraints on the relevance of research factors in risk-adjusted performance measurement.

3. Reporting framework and structure

The GPFG was established to manage the Norwegian government's accumulated primary surpluses associated with its large and temporary petroleum revenues. The fund is managed by Norges Bank under a mandate set by the Ministry of Finance. The main aspects of the investment strategy are decided by the Norwegian Parliament and specified in the investment mandate. This mandate defines the investment universe and the fund's strategic benchmark index, which is based on widely used and publicly available indices. We note that the Ministry's mandate and the strategic benchmark index – determined outside of Norges Bank – defines the bulk of risk in the Fund.[5] Nevertheless, important strategic choices have been assigned to the Bank. Norges Bank is accountable to the owner of the fund – the Ministry of Finance – and informative reporting on the fund's risk and return is the central tool for delivering on this accountability.

Recommendations

The report on risk and return should focus on the portfolio's performance relative to the benchmark for GPFG chosen by the Ministry of Finance. Nevertheless, since most of the investment risk in the GPFG is decided by the Ministry of Finance, a report on risk and return should also communicate the consequences of the choice of GPFG benchmark.

We observe that the overall requirements for the GPFG's performance and risk reporting are set out in the management mandate from the Ministry. Within this framework, we recommend that Norges Bank develops the performance reporting for the fund with a view to both the wider Norwegian public and to a more professional audience. Based on this, we recommend that performance reporting be divided into two parts aimed at the different audiences:

I. An overall GPFG performance and risk report intended for a non-specialist, but interested audience. Clarity, simplicity, and the current knowledge of academic finance should be the guiding principles in selecting which analyses to include in this overall report. In addition to documenting risk and return, the overall report should strive to communicate the strategy of the fund.

II. An appendix to the overall report intended for technically inclined readers, including experts in empirical finance. In addition to providing a detailed description of the methodology used in the overall report, the appendix should incorporate robustness analysis of the main results of the overall report, including alternative model specifications and empirical approaches.

To the widest extent possible, the underlying data for the analyses in both the main report and the appendix should be made publicly available. It is important that outside observers and researchers can verify and replicate Norges Bank's analyses. The availability of data will also have an additional, pedagogical benefit. The best way to understand these types of analyses is to undertake them.

---

[5] According to Ang et al. (2014), more than 99 percent of the variance in the fund's return between January 1998 and June 2013 can be attributed to the reference index.

4. Risk-adjusted performance measurement

Introduction

Risk-adjusted performance measures may be used by an owner to evaluate to what extent an investment manager has created value, or by a manager to evaluate potential investments prior to committing funds. We have been asked to give advice on which measures Norges Bank should use when reporting the fund's performance.

In this section, we discuss measures of risk-adjusted return that use data for portfolio and benchmark performance only. These measures can be traced back to the seminal work on the Capital Asset Pricing Model (CAPM) by Treynor (1962), Sharpe (1964), Lintner (1965a, 1965b) and Mossin (1966). A central concept of the CAPM is the "market portfolio" (where all assets enter with weight equal to the asset's value relative to the total market value of all assets). Given the assumptions of the CAPM, the market portfolio maximizes expected return per unit of risk: it is mean-variance efficient. Under the CAPM the market portfolio is the only risk factor that creates positive expected return, and investment returns should be evaluated against this portfolio. In practice, it can be difficult to define the true market portfolio. It is therefore common to use the investment manager's assigned benchmark instead of a "total market portfolio" when evaluating investment results.

Discussion

A number of different measures are available for reporting performance and various risk adjustments. In choosing which measures to use, one should consider their strengths and weaknesses from a methodological perspective, but also consider how common and well understood the various measures are. The use of simple and well-known measures will make it easier for a wider audience to understand the figures and will also facilitate comparisons with peers.

The fund should continue to report the simplest measure, i.e. return minus the benchmark return. This measure is a natural starting point because it is easy to understand and, importantly, does not rely on estimated parameters.

When turning to risk adjustments, the Sharpe ratio should be reported. This is the most common measure of a portfolio's risk-adjusted *absolute* return. It divides mean portfolio return less the risk-free rate over the sample period by the standard deviation of the return over that period. It measures return per unit of risk and is the simplest measure of the trade-off of portfolio return to total volatility.

We have observed that Norges Bank, until now, has usually reported the information ratio (IR) as a measure of the risk-adjusted *relative* performance of the GPFG. This measure divides the mean return difference between the portfolio and the benchmark over the sample period by the standard deviation of this return difference. (This standard deviation is often referred to as the "tracking error".) The IR is similar to a Sharpe ratio of the benchmark-adjusted return.

The basic measure of the portfolio's relative *risk-adjusted* return is Jensen's alpha. This gives the average return on the portfolio over and above that predicted by the CAPM, where we recall that CAPM-predicted average return depends on the portfolio's beta and the average benchmark return. A positive (negative) Jensen's alpha indicates that the investment manager has achieved a higher (lower) risk-adjusted return than the pure benchmark ("passive") exposure would have implied. When calculating this measure one key choice is to decide how to estimate the beta. We favor a simple approach of estimating a constant beta rather than allowing beta to vary over time.

Another informative measure related to Jensen's alpha is the appraisal ratio (AR). The AR is equal to the estimated alpha divided by the standard deviation of the residuals of the regression, and thus scales down the estimated alpha by the amount of diversifiable risk carried by the investment

manager. An important advantage of the AR is that it can be calculated for multi-factor alpha estimates. (We discuss models with multiple risk factors in the next section.)

In section 6 below we discuss how the GPFG's return, risk and costs can be disaggregated to reflect different components of the investment strategy. In our view, Norges Bank should consider using different performance measures according to which level, or investment strategy, is being analyzed. For example, the Sharpe ratio is the most relevant risk-adjusted performance measure for the absolute performance of the GPFG, while Jensen's alpha, the IR, and the AR are appropriate measures of the relative performance of both the GPFG as a whole and various subcomponents and investment strategies of the fund.

As a guiding principle, the reported risk-adjusted performance measures should be replicable. One good way to implement this would be to make the data for the analyses publicly available, for example on the Web.

Recommendations

We recommend that the fund's public report includes the following measures, as these are simple, standard, and comprehensible for a relatively wide audience:

1. The Sharpe ratio for absolute returns.

2. The excess return, the information ratio, Jensen's alpha, and the appraisal ratio for relative returns.

In addition, we recommend that the Bank publishes a detailed explanation of the methodology applied when calculating these measures in an appendix.

5. Factor risk-adjusted performance measurement

Introduction

We have been asked to give advice on the most appropriate approach for factor risk-adjusted performance measurement. This includes advice about what factors to include in the analyses in the forthcoming report, and views on the construction and potential customization of factors. We have been asked to provide such views and advice both for the fund as a whole, and separately for the equity and fixed-income portfolios.

Broadly speaking, factor analyses can serve two main purposes: documentation of factor exposures and performance evaluation. These correspond to estimation of betas and alphas, respectively. The objective of the report has a bearing on the model choice and factor construction. In line with the ToR for our assignment, we will concentrate on performance evaluation.

Asset pricing models developed in the 1970s (e.g. Merton, 1973; Ross, 1976) allow for other systematic sources of risks than the CAPM "market factor" to influence the price of an asset. However, these models do not explicitly specify the sources of risks or priced factors. Empirical factor models typically use the CAPM as a starting point, but the asset pricing and fund performance evaluation is adjusted for the outperformance observed for the chosen factors. This is an active and developing area of research (see e.g. Harvey, 2014).

Factor risk-adjusted performance evaluations are based on time-series regressions of portfolio returns against a set of factors intended to control for exposure to systematic (i.e. priced) risk. The intercepts (alphas) from such regressions are then interpreted as performance attributable to manager value creation beyond exposure to the selected factors. When interpreting the results from such regressions one must take into consideration the characteristics of the factor series and possible investment restrictions imposed on the fund manager. Issues such as investability and whether or not trading costs are included in the factor series have implications for the interpretation of the results.

An important take-away from the discussion above is that factor risk adjustments can be useful in explaining how the excess return of a portfolio comes about, i.e. whether the excess return is due to security selection or exposure to various risk factors. However, the discussion also points to a warning regarding how performance measures based on factor risk adjustments should be interpreted: suppose a particular factor performs well and that the GPFG happens to benefit from exposure to this factor. One might argue that the GPFG should be credited for having chosen exposure to this factor. However, once one adjusts the alpha measurement for exposure to the said factor, the positive performance due to the desired factor exposure would not raise the fund's alpha.

While the interpretation of alpha from factor model estimation may be problematic, such analysis serves an additional and important purpose: the coefficient estimates in a factor model regression (the betas) are informative about how an investment manager attempts to achieve performance. Observing a positive estimate for one particular factor (e.g. "value") implies that the manager has tilted the portfolio towards value. This is important information for understanding the aggregate investment strategy of the fund.

Note also that factors such as "value" or "small cap" cannot be purchased for free. The reason is that achieving an exposure to such factors requires rebalancing of the portfolio, which in turn induces trading costs. Therefore, achieving a desired exposure to said factors at zero alpha, net of costs, can be interpreted as an accomplishment.

Evaluating performance based on factor risk adjustments is not uncontroversial, especially for a global fund. Factor regressions have weak theoretical underpinnings, and while there exists substantial empirical work on estimating factor models on U.S. stocks, much less work has been done analyzing risk factors for global portfolios. The evolving nature of the empirical literature in this area should be

kept in mind when interpreting the advice below.

In what follows we first discuss the choice of models, and then we discuss whether the factor series should be customized for GPFG. Finally we discuss interpretation issues.

Model choice

Given the state of the literature in finance, we believe it might be useful for Norges Bank to consider multiple models in its factor analyses. However, to keep the main risk and return report transparent, and using simplicity as a guiding principle, the Bank should focus on one main model in the main report. In addition, we recommend that the Bank produces an appendix to the report, incorporating several alternative factor models and alternative specifications. An important motivation for publishing such an appendix is to shed light on the robustness of the estimated factor exposures and alphas. Considering several alternative models in the appendix might also mitigate the concerns that an unambiguous consensus model has yet to emerge in the empirical academic literature, and the fact that the literature has focused more on the cross-section of returns on U.S. stocks than on global stocks, so the global cross-section of returns is arguably less well understood.

As a general principle, Norges Bank should base its factor analyses on a model that represents the current state of knowledge in financial research and which to the largest possible extent can be viewed as a reasonable consensus model in empirical finance. However, it is conceivable that at some point in the future the consensus will shift. This implies that, on the one hand, the Bank must be prepared to develop its main model as the research frontier in empirical asset pricing moves forward. On the other hand, it is valuable to preserve consistency in how results are reported over time. This speaks for focusing on a consensus model and sticking to this model over time, unless the academic literature converges on a model which is unambiguously better than the initial model. In due time, the Executive Board of Norges Bank should evaluate whether the chosen benchmark model has held up relative to the research frontier in empirical finance. Moreover, the analysis in the technical appendix should evolve over time, reflecting the research frontier.

For analyses of the GPFG's <u>equity</u> portfolio, we believe that the international version of the Fama and French (2015) 5-factor model represents the most suited version among the frontier models. Eugene Fama and Kenneth French are pioneers in the literature of factor risk adjustment and we view their recent 5-factor model as a reasonable candidate for a consensus model.

As we write this report, it is unclear whether all factor series included in the 5-factor model will be available to Norges Bank prior to its publication of the first risk and return report. If access to satisfactory data precludes implementation of the 5-factor model, the Bank must choose an older main model for the first report. If it comes to this, we recommend that Fama and French's (1993) 3-factor model serves as the main model in the report for 2015. Norges Bank should then state explicitly that this model will be replaced by the 5-factor model once its (research-) factor series become publicly available and/or external providers can produce satisfactory data.

The results in Fama and French (2015) suggest that the 5-factor model performs better when estimated on a regional basis (North America, Europe, etc.) than when estimated for the world as a whole, using global factors. Hence, we recommend that the appendix includes regional versions of the FF5 model. This would involve running separate FF5 factor regressions for each region and averaging the estimates across regions and using the regional market caps in the GPFG's benchmark as weights. Note that it could also be interesting to consider the regional results separately, especially for the U.S. The reason is that the lion's share of academic research on factor analysis has focused on U.S. data. Focusing on results for the GPFG in the U.S. market can therefore make it easier to relate to factor regressions in the academic literature.

For <u>fixed-income</u> investments, there is less empirical academic work and less consensus on the choice of factor model. In particular, there is no factor model for fixed income which enjoys a standing

comparable to that of the Fama-French models discussed above for equities. However, default and term factors have been used by academic researchers for a long time. Additional factors to be included when analyzing the GPFG's bond portfolio should be based on the results in Ang, Brandt and Denison (2014). We believe it is not necessary to include a market factor when analyzing the bond portfolio in isolation.

When adjusting for differences in the default premium, recent work by Asvanunt and Richardson (2015) suggests adjusting for differences in the duration of safe versus risky bonds. In order to keep the main report as simple as possible (and the analysis replicable), this kind of adjustment could be placed in the appendix.

In the analysis of the performance of GPFG as a whole, including both equities and bonds, we would caution against including a large number of factors, as a parsimonious model is preferable. As such, the number of fixed-income factors should be limited. We would recommend using the equity market as the market factor for the combined portfolio, along with the other equity risk factors.

Factor exposures could vary over time. Taking such time variation into account would require richer estimation techniques. Such techniques might capture true strategic changes, such as changes in factor exposures. To limit the complexity, we advise against using such techniques in the main report. While interesting and potentially informative, such analysis would be better suited to a research paper.

Details about the estimation

Two details about the estimation of factor models deserve to be mentioned specifically. First, the common approach for implementing factor models in mutual fund evaluations is to let the left-hand variable in the regressions be the return of the fund in excess of a risk-free rate. This could be done for the GPFG. However, the Ministry of Finance has assigned a benchmark for the fund and the goal is to measure the value added over and above the benchmark. Hence, it becomes natural to let the left-hand variable in the regressions be the excess return of the fund's portfolio relative to the benchmark. An advantage is that the right-hand side of the factor regressions comprises the various factors as originally specified. In particular, the market factor should be the return of the market portfolio associated with the chosen factor model. Thus, when estimating the FF5 model, one should use the FF5 market factor, possibly customized for investability (see the next section), net of the risk-free interest rate. If one were to use a different market return on the right-hand side, then the estimated model would no longer be a FF5 model but something else.

Note though that this approach may have a disadvantage: since the return on the fund's benchmark index will differ from the FF5 market portfolio, the benchmark index itself might be associated with an "alpha".[6] The alpha estimated for GPFG using the approach discussed in the previous paragraph will incorporate the alpha associated with the benchmark index. A reasonable interpretation would then be that the "true" alpha is the difference between the fund's estimated alpha and the alpha associated with the benchmark. We recommend that the issue of the alpha inherent in the benchmark index be thoroughly investigated in the research paper.

Second, in mutual fund evaluation it is standard to consider fund returns net of costs (trading costs, expenses, etc.) when computing performance measures. For example, in a factor model regression, the excess return (the left-hand side of the regression) is measured after costs. We recommend that Norges Bank use the same approach as its starting point in the risk and return report. However, care should be taken to adjust the costs for elements accruing as a consequence of the owner's desired strategy. These elements include the costs of investing in the factors and the costs associated with

---

[6] To see this, note that if one were to estimate a factor model (e.g. FF5) with the return on the benchmark in excess of the risk-free rate on the left-hand side, the regression would likely generate an alpha different from zero.

possible changes in the benchmark. The fund's true alpha is thus the alpha estimated net of all costs, plus an assessment of how much extra it would cost for a representative fund to implement the benchmark index, relative to a pure stationary index strategy.

We recommend Norges Bank to follow standard practice for incorporating costs. This applies both to the factor regressions and to the risk-adjusted performance measures discussed in section 4. The fund should also report costs separately and try, as far as possible, to compare the actual costs to a reasonable assessment of the costs of implementing the desired factor exposure and index changes.

Investability and customization of factor series

A key insight from the academic literature is that in order to interpret an alpha estimate as a performance measure, it is necessary that the factors used in the regressions are investable for the portfolio manager. If these are to be investable, the construction of factor series must take into account the restrictions on the fund, such as no short-selling, risk budget constraints, maximum holdings in individual companies, and sufficient trading volume. For example, we would not advise including momentum as a factor when evaluating the performance of the *total* GPFG equity portfolio. This is because investing in the momentum factor would require frequent and extensive rebalancing of the portfolio. For a large fund such as the GPFG, it would therefore be practically impossible to obtain a profitable systematic exposure to this factor, at the fund level. However, it is interesting to examine if components of the fund, e.g. security selection, load on momentum.

Factor series customized to the investability constraints of a given fund are not readily available in the public domain. One approach for deriving such series could be that NBIM calculates these series itself. Another approach could be to ask a third-party service provider to produce factors deemed to be investable for the fund, and then make these factor series publicly available.

We suggest that Norges Bank publishes a research paper that explains the rationale for the customization of the factors. Such a research report should also shed light on how the factor customization and, hence, the trading constraints influence the measured performance (alpha) of the GPFG.

We suggest the research paper also includes a discussion on international variation in accounting standards. Such variation may influence the analysis. For example, book value depends on the depreciation allowances assumed in each country. Moreover, when constructing any factor portfolio (e.g. a value tilt) for the global portfolio, this might involve large tilts in terms of countries and sectors.[7]

The research paper could also consider adding additional alternative factor models, for example the one featured in recent work by Hou, Xue and Zhang (2015a, 2015b). Their competing model has performed well on U.S. data. However, we are not aware of any published work estimating this model on non-U.S. data, so for now this model seems better suited to the research paper.

Interpretation issues

The long horizon and global nature of the GPFG's investment universe imply that it is not trivial to determine the correct risk-free rate for the fund. An estimate of the risk-free rate series is required both in factor analysis and when calculating ratios as discussed in section 4. When measuring the risk-free rate, one must take a stand on issues involving the country basket, nominal versus real, short-term versus long-term, etc. We emphasize that the choice of the risk-free rate proxy should be transparent

---

[7]   This is discussed in earlier work by Fama and French on international size and value portfolios. For example, in Fama and French (2012): "The global portfolios use global size breaks, but to mitigate any effects of differences in accounting rules across the four regions, we use each region's B/M break points to allocate its stocks to the global portfolios." They also discuss related issues with size and momentum.

to external observers. Rolling over short-term bonds in the main currencies could provide a simple first step. However, any bespoke construction of the risk-free rate would take the analysis a step away from the simple and widely used model. As such, this would ideally be dealt with in the research paper.

We would also encourage Norges Bank to develop research on factors corresponding to hedge fund strategies such as various arbitrage trades, carry and volatility. This could be included in the research paper discussed above.

Recommendations

1. Factor risk-adjusted performance measures for equity should be based on the international version of the Fama and French (2015) 5-factor model.

2. For fixed-income, there is no model that is similarly well established, but we recommend the default and term factors considered by Fama and French (1993), possibly expanded with additional factors in line with Ang, Brandt and Denison (2014). In our view it is not necessary to include a market factor in the (isolated) analysis of the fixed-income portfolio.

3. In the analysis of total performance of the fund's equity and fixed-income investments, the number of fixed-income factors should be limited. Norges Bank should use the equity market as the market factor for the combined equity and fixed-income portfolio of the fund.

4. Results from several factor models should be presented in the appendix to the public report. These should include a regional Fama-French 5-factor model, a Fama-French 3-factor model, a Carhart 4-factor model, and a version with the "research factors", i.e. a model where the original Fama-French factors are used (instead of the benchmark factor series where trading restrictions are taken into account). Alpha estimates from the main text and the appendix should be compared and discussed.

5. In addition to the new annual risk and return report, we recommend that Norges Bank writes a research paper with more in-depth discussion of factor risk-adjusted performance measurement and factor construction.

6. Other Questions

Decomposition of management costs, risk and return

Transaction costs, administration costs, and revenues from securities lending need to be considered in the set-up of the analyses and their interpretation. We would advise Norges Bank to pursue these topics in the research paper.

Norges Bank should decompose return, risk and costs on the main investment strategies, starting with the internal "operational reference portfolio", and divided into systematic factor exposures and expansion of the portfolio universe. The Bank should further decompose the results and costs for three separate and more detailed levels: tactical allocation, enhanced indexing, and security selection. Examples that can be expanded and developed are provided in the GPFG historical performance and risk review from March 2014, figures 11 and 12. This could be a useful tool for understanding the origin of the overall results.

We note that there are potential downsides to extremely detailed reporting, for example if it discloses trading strategies. Moreover, it may adversely affect the incentives of portfolio managers, which may be reluctant to reveal profitable investment strategies and also less willing to implement such strategies. Highly detailed reporting could also affect negatively Norges Bank's ability to recruit and retain personnel. These drawbacks are, however, likely to be minor at the aggregation level of reporting suggested above, and must be weighed against the owner's benefits of more detailed reporting.

Return, risk and cost decomposition should be reported with necessary explanations. When reporting results for each strategy, one should recognize that there might be important synergies between them. For example, there are complementarities between enhanced indexing and security lending.

Also, public reports should not give the impression that all strategies are expected to have positive excess returns in all periods. The nature of being a risk factor means that, even though the factor has a positive risk premium in the long run, there will be periods when the factor has negative realized returns. These periods can last for years.

Sample periods

For the Sharpe ratio, information ratio, appraisal ratio and alpha discussed in section 3, the current public reporting sample periods would be reasonable. For regressions, longer sample periods would be preferable. A natural starting point is to consistently report different time periods, for example (1) since inception, (2) rolling 10-year windows, (3) rolling 5-year windows. To the extent that non-standard sample periods are used, it is important to clearly motivate them. Changes in investment strategy and data availability might be reasons to report results for non-standard periods. Since the internal operational reference portfolio has existed for less than five years, Norges Bank should consider using a three year sample period for some of the analyses in the *first* risk and return report.

Recommendations

1. We recommend that Norges Bank decomposes return, risk and costs on the main investment strategies in the public reports.

2. We recommend that the Bank reports factor regression results for sample periods corresponding to rolling 5-year and 10-year windows, in addition to results since the fund's inception.

References

Asvanunt, A. and S. Richardson. 2015. The Credit Risk Premium. AQR Capital Management, working paper.

Carhart, M. 1997. On Persistence in Mutual Fund Performance. Journal of Finance 52: 57-82.

Fama, E. and K. French. 1993. Common Risk Factors in the Returns on Stocks and Bonds. Journal of Financial Economics 33: 3-56.

Fama, E. and K. French. 2015. International Tests of a Five-Factor Asset Pricing Model. Tuck School of Business Working Paper No. 2622782.

Harvey, C., L. Liu and H. Zhu. 2014. … and the Cross-Section of Expected Return. NBER Working Paper No. 20592.

Hou, K., C. Xue and L. Zhang. 2015a. Digesting Anomalies: An Investment Approach. Review of Financial Studies 28: 650-705.

Hou, K., C. Xue and L. Zhang. 2015b. A Comparison of New Factor Models. NBER Working Paper No. 20682.

Lintner, J. 1965a. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. Review of Economics and Statistics 47: 13-37.

Lintner, J. 1965b. Security Prices, Risk and Maximal Gains from Diversification. Journal of Finance 20: 587-615.

Merton, R. 1973. An Intertemporal Capital Asset Pricing Model. Econometrica 41: 867-87.

Mossin, J. 1966. Equilibrium in a Capital Asset Market. Econometrica 35: 768-83.

Ross, S. 1976. The Arbitrage Theory of Capital Asset Pricing. Journal of Economic Theory 13: 341-60.

Sharpe, W. 1966. Mutual Fund Performance. Journal of Business 39: 119-38.

Treynor, J. 1962. Toward a Theory of Market Value of Risky Assets. Unpublished manuscript.

Attachment: Terms of Reference

Terms of Reference - Advice regarding principles for risk adjustment of performance figures

1. Purpose

The Executive Board has decided to invite an Expert Group to assess risk-adjusted return measures and provide advice that may be used to further develop the Government Pension Fund Global ('the fund') reporting.

2. Definitions

We distinguish between A) Risk-adjusted excess return (using returns data from the portfolio and the benchmark indices), B) Factor risk-adjusted excess return (using data from A, as well as return series of risk factors).

3. Overview of questions

The Executive Board wishes to obtain input on the following questions:

A - Risk-adjusted performance measurement:

NBIM currently reports a set of measures on the fund's risk-adjusted excess return. The Expert Group is asked to give advice on which measures are the most appropriate.

B - Factor risk-adjusted performance measurement

- Evaluation of standard OLS-regression of returns on a set of factors as a method to evaluate and draw conclusions on the stock selection skills.
- Advice on econometric approach to analysis of factor risk exposure
- Should time variation in the factor exposures be explicitly addressed in an analysis of stock selection skill/alpha?
- Equity:
    - Choice of factors:
        - What set of factors would be a general consensus choice in order to measure the pure contribution to portfolio performance from stock selection in international equity portfolios?
        - What set of factors is relevant for NBIM?
    - Factor construction:
        - Should the factors included in a factor regression represent an investable, passive alternative to active management?
        - Should the equity universe and regional weighting used in the construction of factors be adapted to NBIM's investment universe and regional allocation of the GPFG benchmark portfolio?

- Fixed income: Which model may be used as the «standard model» for factor risk-adjusted performance of the fixed income portfolio? What factor return series / data should be used?
- Should the analyses be adapted to the fund's reporting currency (currency basket). If so, how?

C- Other questions
- Treatment of management costs (transaction costs and administration costs/fees)
- Inputs to October 1 letter to Ministry of Finance regarding principles of return decomposition
- Sample for estimating the various excess return measures:
  a) What should be the time span of the sample?
  b) With what frequency should returns data be measured
- How can reporting affect incentives?

4. Data sources for the various performance measurements

Monthly return series for the fund, the equity investments and the fixed income investments, as well as the respective benchmark indices. Factor return series - either publicly available or customized according to the Expert Group's specification.

5. Work schedule

NBIM (CRO) will be serve as secretariat and produce data, calculations and prepare drafts, to the extent the Expert Group requests such contributions.

The Expert Group is asked to provide preliminary advice to the Executive Board by October 1, supplemented by a short final report by 1 November if feasible. The final report will be made available to the public.

Currently, we assume the following schedule:

- 1 Start-up meeting
- 2 Work meetings
- 1 Meeting to provide assessments and advice to the Executive Board

Including hours worked in between these meetings, we estimate a total time use of approximately 30 hours. Members of the Expert Group will be compensated for the hours spent on this assignment.