Improving Out-of Sample Performance of Asset Pricing Models: A Model Portfolio Approach ☆

Qing Zhou
University of Queensland, Australia

Email: q.zhou@business.uq.edu.au

Abstract

This paper draws a parallel from model combination to Markowitz's modern portfolio theory. Building upon the bias\variance trade-off framework, the paper proposes a Model Portfolio Approach and a Global Minimum Variance weighting scheme to mitigate the asset pricing model uncertainty problem. With a well-conditioned pricing error covariance estimator, our method provides improved out-of-sample pricing performance over both the single model selection method and other existing model combination weighting schemes.

Keywords: model portfolio; global minimum variance weighting; asset pricing model uncertainty; out-of-sample pricing error

> "It is a very beautiful line of reasoning. The only problem is that perhaps it is not true. (After all, nature does not have to go along with our reasoning.)"
>
> -Richard P. Feynman

1. Introduction

Developing and testing asset pricing models has been a fascinating endeavour with a long history in finance literature. Since Markowitz's modern

☆This paper is part of my PhD thesis. I gratefully acknowledge the supervision of my supervisor, Tom Smith, and the many helpful comments from my collegues at UQ business school and Australian National University.

portfolio theory (1952), many asset pricing models have been developed over the last six decades. With rational expectations and the Efficient Market Hypothesis, asset pricing models should theoretically provide us unbiased prediction of future asset return. But empirical tests in this area are quite challenging. Without knowing the ex ante beliefs in the marketplace, traditional tests of asset pricing models with ex post datasets requires a tight link between ex-ante beliefs and what factually is in the ex-post dataset (Bossaerts, 2004). But this link is too tight and the belief that realized returns are an unbiased estimate of expected returns is misplaced (Elton, 1999). The empirically poor performance of asset pricing models, especially large out-of-sample mispricing errors (Simin, 2008), poses threat to the usefulness of asset pricing models. The companion literature on asset pricing anomalies are another challenge to asset pricing models. The finding of predictive ability of conditional asset pricing models is inspiring but it may tend out to be spurious due to long haunting data snooping problem in finance (Foster et al., 1997). We are now at a stage that we are still uncertain about whether the poor performance comes from our empirical methodology or maybe models themselves are false. The question then is how can we improve asset pricing model out-of-sample performance when all models are possibly false.

In the literature, model selection methods have long been documented as a way to mitigate asset pricing model uncertainty. However, the choice of one asset pricing model to the exclusion of another is an inherently misguided strategy (O'Doherty et al., 2010). The underlying assumption of both non-Bayesian and Bayesian model selection is that the model space is complete and so that the true model is in the model space, but the truth is that we do not have such prior knowledge. Moreover, omitting useful information in other abandoned models is detrimental to accurate asset pricing. Additionally, sampling error is also a concern. The best performing model may tend out to be the worst in another sample. Despite the winner's curse problem[1], the unobservable economic structure change may also increase the risk of excluding the seemingly worse model under one regime. Just like the six monks who encountered an elephant for the first time–each monk grasping a different part of the beast and coming to a wholly different conclusion as to what an elephant is but no one giving a true picture of the elephant. Disciples of different pricing models have captured different features of the same

[1] see Hansen (2009).

financial asset price but none of them has a completely true description. A collection of all opinions provides a closer illustration to the truth.

Here in this paper, in contrast to the paradigm of single model selection methods, we propose a Model Portfolio Approach to diversify asset pricing model out-of-sample mispricing uncertainty. Our approach is in the same spirit as Modern Portfolio Theory for asset allocation (Markowitz, 1952). The core inspiration of portfolio theory is that the idiosyncratic risk can be diversified by optimally pooling a set of assets and the portfolio of assets will provide higher risk-adjusted return than any individual assets. Just like asset portfolio,which is derived under mean-variance framework by a trade-off between return and risk, we derived our optimal model portfolio by a trade-off between bias and variance. The bias\variance trade-off is the key to the success of out-of-sample prediction[2].

Our Model Portfolio Approach is directly related to the forecast combination literature. The forecast combination is firstly introduced into econometric forecasting by Bates and Granger (1969), and then extended by Granger and Ramanathan (1984), thus spawn a large literature. Some excellent reviews include Granger (1989), Clemen (1989), Diebold and Lopez (1996), Clements et al. (2002),Timmermann (2006) and Stock and Watson (2006). Recently, forecast combinations have received renewed attention in the macroeconomic forecasting lterature with respect to forecasting inflation and real output growth (e.g., Stock and Watson, 2003) Despite increasing popularity of forecast combination in economy forecasting, applications remain relatively scarce in the finance forecasting literature. Only in the recent several years, the forecasting combination has been seen in asset pricing studies Durham and Geweke (2011) provide an optimal model pooling method for S&P 500 return density forecasting by maximizing predictive log score and their result shows that the prediction probabilities of the optimal pool exceed those of the conventional models by as much as 7.75 percent. Neely et al. (2010) analyse the ability of both economic fundamentals and technical moving-average rules to forecast the monthly U.S. equity premium using out-of-sample tests and conclude that fundamental and technical approaches are complementary. Both of these two studies find improved performance of combined models in aggregate stock return forecasting. Additionally, O'Doherty et al. (2010) provide a optimal model pool for cross-section stock portfolio

[2]see Geman et al. (1992)

return prediction and find improved performance of combined asset pricing model. In their work, they only focus on five well-known asset pricing models (CAPM, Fama-French three-factor model, Carhart four-factor model, consumption CAPM, and Chen, Roll, and Ross five-factor model). Our Model Portfolio Approach contributes to this line of research by providing a new way to understand the success of model combination and a unified framework for optimal weights derivation.

In contrast to the consensus on the superiority of model combination, there is no unanimous agreement on model combination weighting. A plethora of weighting schemes have been developed in both non-Bayesian and Bayesian econometrics. But it seems that the simple arithmetic average weighting method ($1/N$ rule) outperforms the existing complicated weights most of time. Stock and Watson (2004) find that among all the competing weights, the simple $1/N$ rule gives smallest mean squared forecasting error (MSFE). The most recent work on superior simple average model combination by Issler and Lima (2009) also find that a bias-corrected simple average combining method dominate all other weights considered. This "$1/N$" puzzle has been long haunted around forecast combination practice. The common explanation of this puzzle is that the weights estimation error is too large to be offset by diversification gains due to small effective sample size[3]. Here, in our study, optimal weight is a by-product of objective function optimization along the bias\variance efficient frontier. But as out-of-sample pricing error variance and bias are unobservable, the optimization is along estimated frontier rather than true frontier. Taking the frontier estimation error problem into consideration, we propose a global minimum variance (GMV) weighting scheme, which is the weights of global minimum variance portfolio. Our weighting formula in its basic form can be unified with Granger-Bates-Ramanathan optimal weighting scheme but proved to be more general in terms of relaxing Granger-Bates-Ramanathan single model unbiasedness assumption. Moreover, by utilizing the recent development in large scale covariance matrix estimation technique, GMV weighting can be used in small effective sample size problem when model space is huge. The traditional Granger-Bates-Ramanathan OLS weighting has theoretical optimality, but due to estimation error in small samples, it usually under-perform other weighting scheme em-

[3]Number of models "N" is relatively large compared with weights estimation sample size.

pirically. Additionally, under our approach, there is no puzzle. $1/N$ rule will only be optimal when it is close to the true optimal. This result is consistent with Smith and Wallis (2009).

This paper explicitly mirrors model combination to asset portfolio theory. We do not only provide a simple way to uncover the myth of model combination, but also contribute to both model combination literature and the asset pricing model uncertainty literature with an optimal asset pricing model combination weighting scheme. Moreover, our paper can serve as a bridge between portfolio study and model combination study. The theoretical and empirical studies in these two areas will enhance the development of each other. Some peer studies have already appeared in both econometric journals and finance journals. DeMiguel et al. (2009) compares portfolio strategies which differ in the treatment of estimation risk and find that none of the strategies suggested in the literature is significantly better than simple diversification, i.e., taking the equally weighted portfolio. This puzzle has also long existed in forecast combination studies (Bunn, 1989; Clemen and Winkler, 1986; Dunis et al., 2001). To address the long standing puzzle in empirical studies, both forecast combination and asset portfolio studies propose a shrinkage weighting scheme. In forecast combination, Diebold and Pauly (1990) propose to shrink towards equal-weights. Stock and Watson (2004) also propose shrinkage towards the arithmetic average of forecasts, while most recently in portfolio construction studies, Frahm and Memmel (2010) documents the dominating feature of shrinking Markowitz weight toward equal weight. In this paper, we draw an explicit parallel of these two studies, and our unified framework will lead to more future research to further explore the similarities between these two areas.

The following sections are as follows: Section 2 defines our problem: asset pricing model uncertainty; section 3 solves our objective problem: derivation of optimal model combining weights; Section 4 provides two simulation studies to verify the advantage of Model Portfolio Approach and GMV weighting scheme; Section 5 concludes.

2. Asset Pricing Uncertainty

2.1. The Source of Mispricing Uncertainties

The pricing kernel of any asset pricing model can be expressed as:

$$E_{t-h}[M_t R_{i,t}] = 1$$

Where M_t is stochastic discount factor(SDF), and $R_{i,t}$ is asset i's return at time t. The operator E_{t-h} is the conditional expectation conditioning on information up to time $t-h$. All asset pricing models can be unified under this framework with a specific SDF, for example, we can get the Capital asset pricing model (CAPM) by letting $M_t = a + bR_m$ (R_m is the return of a benchmark portfolio; a and$b$ are regression coefficients). But the true SDF is unobservable, thus it is uncertain which asset pricing model is the true model. There are some possibilities that none of the asset pricing models is true. Empirical studies show that both conditional and unconditional asset pricing models perform poorly, especially in the out-of-sample test (Simin, 2008). In these tests, researchers usually use realized return as a proxy for expected return, which is an improper measure (Elton, 1999). Additionally, due to econometric estimation technique, empirical approximation of an asset pricing model can also distort the pricing ability. Without a good proxy for investor's ex ante expectation, we are far to conclude whether asset pricing models are true or false. Confronted with asset pricing model uncertainty and possible incomplete existing model space, the expected performance of a selected model j may be biased toward true model with a model bias b_i^j:

$$\begin{aligned} E\{E_{t-h}^j[R_{i,t}]\} &= E_{t-h}^{True}[R_{i,t}] + b_i^j \\ E_{t-h}^j[R_{i,t}] &= E_{t-h}^{True}[R_{i,t}] + v_{i,t}^j + b_i^j \end{aligned} \tag{2.1}$$

Where $v_{i,t}^j$ is model error of a selected model j. Realized return comprises two terms, true expectation and an unexpected shock $\eta_{i,t}$ with a zero mean

$$R_{i,t} = E_{t-h}^{True}[R_{i,t}] + \eta_{i,t} \tag{2.2}$$

Empirical approximation of a selected asset pricing model j can be expressed as

$$f_{i,t}^j(X) = E_{t-h}^j[R_{i,t}] + k_i^j + \varepsilon_{i,t}^j \tag{2.3}$$

Where k_i^j is the bias term [4] and $\varepsilon_{i,t}^j$ is an error term with zero mean. Empirical pricing error is

$$e = R_{i,t} - f_{i,t}^j(X) \tag{2.4}$$

[4] A unbiased fitting might be biased for out-of-sample generalization

Plug in equation (2.1) and equation (2.3), we can have

$$e = E_{t-h}^{True}[R_{i,t}] + \eta_{i,t} - (E_{t-h}^{j}[R_{i,t}] + k_i^j + \varepsilon_{i,t}^j)$$
$$e = E_{t-h}^{True}[R_{i,t}] + \eta_{i,t} - (E_{t-h}^{True}[R_{i,t}] + v_{i,t}^j + b_i^j + k_i^j + \varepsilon_{i,t}^j)$$
$$e = \eta_{i,t} - b_i^j - v_{j,t} - k_i^j - \varepsilon_{i,t}^j \tag{2.5}$$

Expected mispricing uncertainty can be proxied by the following equation[5]:

$$EMU = E[e^2] \tag{2.6}$$

Substitute e with (2.5)

$$EMU = E[(\eta_{i,t} - v_{j,t} - k_j - b_i^j - \varepsilon_{i,t}^j)^2] \tag{2.7}$$

Expanding the equation (2.7), we can see that expected pricing risk has three components

$$EMU = Var(\eta_{i,t}) + (Var(v_{i,t}^j) + E(b_i^j)^2) + (Var(\varepsilon_{i,t}^j) + E(k_i^j)^2) \tag{2.8}$$

The three sources of mispricing uncertainty are: volatility of real asset return ($Var(\eta_{i,t})$); model risk, or the between asset pricing model uncertainty ($Var(v_{i,t}^j + E(b_i^j)^2))$), it is the sum of model variance and model bias; estimation risk ($Var(\varepsilon_{i,t}^j + E(k_i^j)^2)$) or parameter uncertainty . We show the relation between these three sources in figure(1). Obviously, the lower bound of expected pricing uncertainty of an asset pricing model is the volatility of real asset return ($Var(\eta_{i,t})$), which is irreducible in empirical modelling process but is diversifiable by forming asset portfolio. Our focus in this paper is last two reducible modelling error part.

2.2. Model Selection Risk

Existing equilibrium pricing models are not explicit about what instrumental variables form the investors' information set. Since the identity of the instruments is unknown, a plethora of papers propose various variables to explain movements in conditional expected returns. And there is little consensus on what the important conditioning variables should be. To address this kind of regression variable model uncertainty problem, researchers search

[5]The calculation is the same as mean squared forecasting error (MSFE).

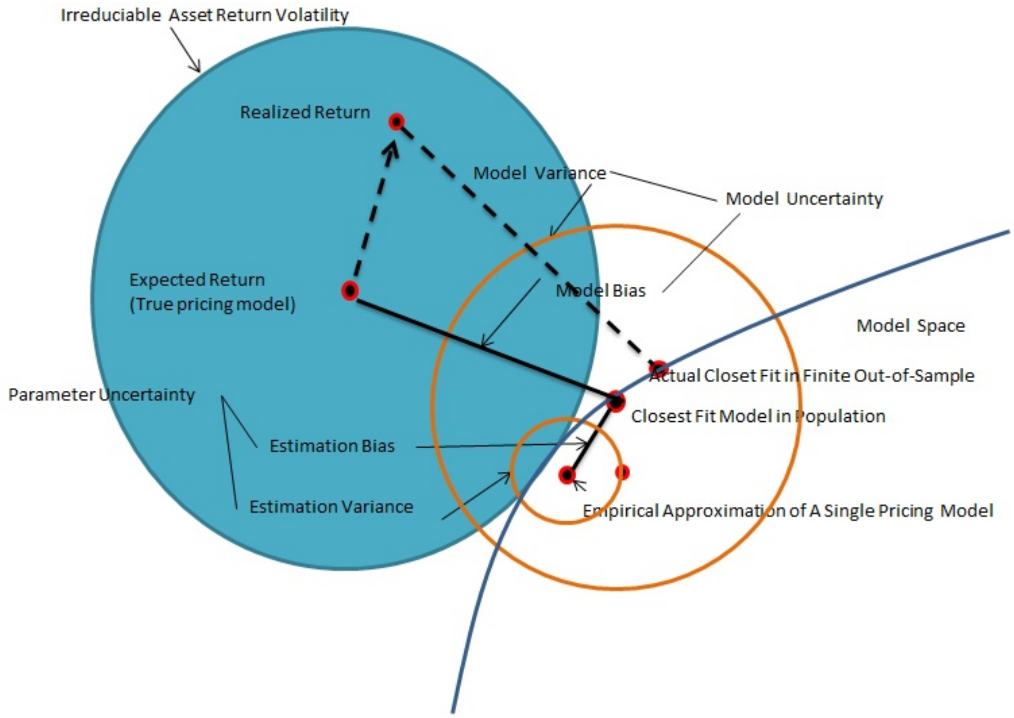

Figure 1: Sources of Asset Pricing Model Uncertainty

available information and select variables that best describe the data. They report the results as evidence of security predictability. But the results might be spurious due to data snooping (Foster et al., 1997). The data snooping fears raise researchers' awareness that common model selection methodology may not be proper in studying asset pricing model's predictive ability. Empirical results begin to show that the best selected model tend to fit a sample well but generalized poorly into another sample. The common model horse race method relying on model selection criterion (such as AIC, BIC and adjusted R^2) leads to inconsistent in-sample and out-of-sample performance, even though these criterion all have a penalty term to penalize the model complexity for in-sample over fitting. Because these criterion are good estimators of in-sample error, but not of expected out-of-sample prediction error. If the relative performance keeps consistent in two different samples, using in-sample error to select the best model does not matter much. But unfortunately, as pointed out by Hansen (2009), good in-sample fit translates into poor out-of-sample fit one-to-one when multiple models are compared in terms of their in-sample fit. Hence, model selection with these criterion is improper if we are aiming at out-of-sample model performance.

Moreover, in practice, our objective of using an asset pricing model is to identify the model which will generalise best, not in an asymptotic sense, but during a particular finite out-of-sample period. In this case, out-of-sample performance, as well as the in-sample performance, is subject to sampling error. This will approximately double the risk that the selected model will perform sub-optimally during a particular out-of-sample period. By selecting a model, we are aiming at select a model which gives best performance, but due to the sample data noise, the selected model will be positively biased with respect to the true expectation of future performance. This can be easily demonstrated by the following formula:

$$max\,(\text{performance}) + E[\text{noise}] \leqslant E[max\,(\text{performance} + \text{noise})]$$

Therefore when model space is incomplete, the single best model selection method will lose useful information contained in other seemingly inferior models, and will cause high model bias $(E(b_i^j)^2))$ and model uncertainty $(Var(v_{i,t}^j)$, leading to an overall high asset pricing uncertainty. Moreover, economic structure break will make the problem even worse. The data generating process may vary across different economic states. An asset pricing model that is proper in boom periods may be improper in bust periods. As the break point is really difficult to predict, the ideal method of selecting a single best model for each regime is thus ex ante impossible. Hence, relying on one selected model has the risk of causing large loss at the regime shifting period.

Asset pricing model uncertainty attracted renewed attention at the beginning of this millennium. Pastor and Stambaugh (2000) discusses prior mispricing uncertainty of asset pricing models and the influence on portfolio choice. Avramov (2002) investigates the role of uncertainty about the return forecasting model in choosing optimal portfolios. The approach proposed in all these studies is Bayesian Model Averaging (BMA) method. Bayesian model averaging contrasts markedly with the traditional approach of model selection. BMA averages over all the models with the posterior model probabilities. The BMA indeed proved some improvement over the past model selection methods and have several admirable properties, such as superior out-of-sample performance and limited data snooping bias. But as BMA conditions on that the model space is complete and thus the true model is in the model set Durham and Geweke (2011), given large sample size, BMA will put all weight on the true model. Therefore when our premise is that all models might be false, BMA is not a good choice. And all these Bayesian model selection studies are limited to the regress variable uncertainty. The models considered in these studies are all linear regression. But actually, the model uncertainty about asset pricing model is more broad rather than just instrument variable uncertainty. There is no theory about whether the underlying asset price process should be linear or nonlinear. We should take this functional form uncertainty into consideration as well. But existing studies rarely approach this asset pricing model functional form uncertainty.

2.3. Model Combination Approach

In contrast to the paradigm of selecting a single best model and treating it as the only true model, model combination is a way to reduce the model prediction error by averaging all models. Although model combination techniques have been well developed in all strands of statistics[6], and have been implemented in economics and many other fields such as meteorology and hydrology forecasting to model uncertainty problem, the application in finance is scarce. But actually, in finance, especially the asset pricing area, model uncertainty problem is quite substantial as the number of existing asset pricing model is large. Hence, asset pricing area will be a promising place to apply model combination techniques.

We verify the diversification gains of model combination with a simple two model case, regardless of the source of uncertainty, either because of theoretical pricing model being false or because of poor empirical approximation. The only thing of interest is the pricing error series. This diversification has the same spirit as Markowitz (1952) mean-variance approach to portfolio optimization.

Denote errors from two asset pricing models as $e_1 \sim (0, \sigma_1^2)$ and $e_2 \sim (0, \sigma_1^2)$, the correlation between two errors is ρ_{12}, the combined model error

[6]Frequentist (Pioneer work by Bates and Granger (1969), Least square method etc.), Bayesian (Min and Zellner, 1993), etc.) and recent information theoretic method (Hansen (2008), Mallow's C_p,etc.) and likelihood method by Durham and Geweke (2011). And another Bayesian fashion frequentist shrinkage method (Diebold and Pauly (1990) etc.)

is:

$$e^c = \omega e_1 + (1-\omega)e_2$$

By construction, combined error term has zero mean. The variance of the error is

$$\sigma_c^2(\omega) = \omega^2\sigma_1^2 + (1-\omega)^2\sigma_2^2 + 2\omega(1-\omega)\sigma_{12} \tag{2.9}$$

Solving first order condition of minimizing error variance

$$\omega_1^* = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}$$
$$\omega_2^* = \frac{\sigma_1^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}$$

Substituting ω^* in equation (2.9)

$$\sigma_c^2(\omega^*) = \frac{\sigma_1^2\sigma_2^2(1 - rho_{12}^2)}{\sigma_1^2 + \sigma_2^2 - 2rho_{12}\sigma_1\sigma_2} \leqslant min(\sigma_1^2, \sigma_2^2)$$

The diversification gains will only be zero when σ_1 or σ_2 is zero; or $\sigma_1 = \sigma_2$ and $\rho_{12} = 1$; or $\rho_{12} = \sigma_1/\sigma_2$.

This pooling method can be generalized to a multi model case. Similar as asset portfolio selection, we construct an optimal portfolio of models in order to reduce the expected mispricing uncertainty:

$$min\, E(e_{t+h}^2) \tag{2.10}$$

Where h denotes pricing horizon.

From equation (2.8), we can see the mispricing uncertainty comprises two components: bias and variance components. Despite the source of bias and variance, for out-of-sample pricing, we treat the model uncertainty and estimation uncertainty equally. Then we can simplify the expected mispricing uncertainty to:

$$EMU = E[e_{t+h,t}]^2 + Var(e_{t+h,t}) \tag{2.11}$$

Now our objective is as following:

$$min\, E[e_{t+h,t}]^2 + Var(e_{t+h,t}) \tag{2.12}$$

From equation (2.12), we can see that to reduce the total magnitude of mispricing uncertainty, we have to minimize both the bias and the variance. But as in reality, the sample size cannot grow to infinity, we cannot eliminate at the same time both the variance and bias. Thus a dilemma of bias/variance. We have to trade off between bias and variance. The out-of-sample performance of pricing model relies on a comprise between the two. Actually, we are always quite far away from building optimal models as there is a wide gap between the theoretical notion of consistency, an asymptotic property, and finite sample size in practice. In any finite sample, the price to pay for low bias is high variance (Geman et al., 1992).

Now we have a mapping from model combination to mean-variance portfolio construction framework. Here instead of using mean return and return risk, we trade off between bias and variance of pricing error. We introduce a certain degree of bias to exchange for a reduction of variance of model, and then a total reduction in the mispricing uncertainty. As for portfolio selection, the model weighting vector can be solved as a by-product of the minimization.

$$\omega^*_{t+h,t} = arg \min_{\omega_{t+h,t} \in W_t} E[L(e_{t+h,t}(\omega_{t+h,t})) \mid \hat{f}_{t+h,t}]$$

Here $L(e_{t+h,t}$ is simply $(e_{t+h,t})^2/$. Obviously, our optimal model combination weighting also depends on the trade-off between bias and variance.

Elliott and Timmermann (2004)show that, subject to a set of weak technical assumptions on the loss and distribution functions, the combination weights can be found as the solution to the following Taylor series expansion around $\mu_{e_{t+h,t}} = E[e_{t+h,t} \mid I_t]$, where I_t denote the available information at time t

$$\begin{aligned}\omega^*_{t+h,t} = arg \quad & \min_{\omega_{t+h,t} \in W_t} \quad L(\mu_{e_{t+h,t}}) + \frac{1}{2} L''_{\mu_e} E[(e_{t+h,t} - \mu_{e_{t+h,t}})^2 \mid I_t] \\ & + \sum_{m=3}^{\infty} L^m_{\mu_e} \sum_{i=0}^{m} \frac{1}{i!(m-i)!} E[e^{m-i}_{t+h,t} \mu^i_{e_{t+h,t}} \mid I_t]\end{aligned}$$

Where $L^k_{\mu_e} \equiv \frac{\partial^k L(e_{t+h})}{\partial \omega} \mid_{e_{t+h}} = \mu_{t+h}$. This expansion suggests that the collection of individual asset pricing model prediction $\hat{f}_{t+h}$ is useful in as far as it can predict any of the conditional moments of the prediction error distribution. Hence, if the objective is to use the asset pricing models for a better out-of-sample pricing, no matter the model is complete or not, as long as it contributes to the error moments, it is useful. We can empirically test the incremental information content of asset pricing models to see if they have marginal contribution to our model pool using model encompassing test by

Chong and Hendry (1986) and Fair and Shiller (1989, 1990). We can then construct the model set $\mathbf{F}_{m\times 1}$:

$$F_{m\times 1} = \begin{bmatrix} f_1 \\ . \\ . \\ . \\ f_m \end{bmatrix}$$

Then our combined pricing model can be formed as

$$f_c = \omega'_{m\times 1} * \mathbf{F}_{m\times 1}$$

From our derivation, empirically combined asset pricing model provides an improved out-of-sample pricing ability. Empirical success of combined asset pricing model should shed some light on the development of new theoretical asset pricing model as combined model can be seen as a way to uncover the missed pricing factor in a certain asset pricing model.

The key to out-of-sample pricing performance is the bias\variance tradeoff. Thus the direct way to construct our model portfolio is to directly find a compromise between the two terms to achieve a total mispricing uncertainty diversification.

3. Model Portfolio Approach

We draw a parallel from asset pricing model combination to asset portfolio construction, replacing the trading-off between return and risk in Markowitz's mean-variance framework, here we trade off between variance and bias and build our bias-variance framework. Therefore, we can derive our model portfolio frontier as asset portfolio frontier derivation.

3.1. Analytical Frontier Derivation

Our objective is to choose model weights (ω) which will minimize the variance ($\omega'\Sigma\omega$ and Σ is covariance matrix of model pricing errors) of the future expected pricing error with a given bias ($S = \omega' s$, and s is bias vector of single forecasting models). We derive our efficient model pool frontier as Merton(1970) does for portfolio frontier derivation:

$$\min_x \frac{1}{2}\omega'\Sigma\omega$$

s.t.

$$S = \omega' s$$

$$1 = \omega' \underline{1}$$

It is equvilant to

$$\underset{\omega}{min} \quad \frac{1}{2}\omega'\Sigma\omega + \lambda(S - \omega' s) + \gamma(1 - \omega'\underline{1})$$

Where λ and γ are Lagrange multipliers. Solving this minimization problem, we have the following necessary and sufficient first order conditions:

$$\Sigma\omega = \lambda s + \gamma\underline{1} \tag{3.1a}$$

$$S = \omega' s \tag{3.1b}$$

$$1 = \omega'\underline{1} \tag{3.1c}$$

Thus we have

$$\omega = \lambda\Sigma^{-1}b + \gamma\Sigma^{-1}\underline{1} \tag{3.2}$$

To get an exact expression for the frontier combination x, we need first solve for the Lagrange multipliers λ and γ. First, we define a number of scalars that will reduce the notational burden

$$B = s'\Sigma^{-1}s \tag{3.3a}$$

$$A = s'\Sigma^{-1}\underline{1} \tag{3.3b}$$

$$C = \underline{1}'\Sigma^{-1}\underline{1} \tag{3.3c}$$

$$D = BC - A^2 \tag{3.3d}$$

Multiplying equation (3.2) by s' to get

$$s'\omega = \lambda s'\Sigma^{-1}s + \gamma s'\Sigma^{-1}\underline{1} \tag{3.4}$$

Using equation(3.1b) and our defined scalars reduces to

$$S = \lambda B + \gamma A \tag{3.5}$$

Multiply equation(3.2) by $\underline{1}$ to get

$$\underline{1}'\omega = \lambda\underline{1}'\Sigma^{-1}s + \gamma\underline{1}'\Sigma^{-1}\underline{1} \tag{3.6}$$

Use equation(3.1c) and our defined scalars this reduces to

$$1 = \lambda A + \gamma C \tag{3.7}$$

Combining equation(3.5) and (3.7) gives us 2 equations in 2 unknowns

$$\begin{pmatrix} B & A \\ A & C \end{pmatrix} \begin{pmatrix} \lambda \\ \gamma \end{pmatrix} = \begin{pmatrix} S \\ 1 \end{pmatrix} \tag{3.8}$$

we can solve for the constants, λ andγ, as follows

$$\begin{pmatrix} \lambda \\ \gamma \end{pmatrix} = \begin{pmatrix} B & A \\ A & C \end{pmatrix}^{-1} \begin{pmatrix} S \\ 1 \end{pmatrix} \tag{3.9}$$

and thus

$$\lambda = \frac{CS - A}{D} \tag{3.10}$$

and

$$\gamma = \frac{B - AS}{D} \tag{3.11}$$

Recall equation(3.2) that the composition of a frontier portfolio is given by

$$\omega_p = \lambda \Sigma^{-1} s + \gamma \Sigma^{-1} \underline{1}. \tag{3.12}$$

Plug in the values of λ and γ to obtain

$$\omega_p = \frac{CS - A}{D} \Sigma^{-1} s + \frac{B - AS}{D} \Sigma^{-1} \underline{1} \tag{3.13}$$

Upon rearranging, we have

$$\omega_p = \frac{B\Sigma^{-1}\underline{1} - A\Sigma^{-1}s}{D} + [\frac{C\Sigma^{-1}s - A\Sigma^{-1}s}{D}]S \tag{3.14}$$

or

$$\omega_p = \underline{g} + \underline{h}S \tag{3.15}$$

Where the vectors, $\underline{g}$ and $\underline{h}$ are defined as

$$\underline{g} = \frac{B\Sigma^{-1}\underline{1} - A\Sigma^{-1}s}{D} \tag{3.16}$$

and

$$\underline{h} = \frac{C\Sigma^{-1}s - A\Sigma^{-1}\underline{1}}{D} \tag{3.17}$$

Equation(3.15) is a closed form expression for the necessary and sufficient conditions for a model pool, x_p, to be a frontier combination. The covariance of the bias between two arbitrary frontier model pools, x_p and x_q is given by

$$cov(f_p - R, f_q - R) = cov(f_p, f_q) = \omega_p' \Sigma \omega_q = \frac{C}{D}[S_p - \frac{A}{C}][S_q - \frac{A}{C}] + \frac{1}{C} \quad (3.18)$$

Thus the variance of a frontier combination is

$$\sigma_p^2 = \frac{C}{D}[S_p - \frac{A}{C}]^2 + \frac{1}{C} \quad (3.19)$$

Upon rearranging

$$\frac{\sigma_p^2}{1/C} - \frac{\frac{C}{D}[S_p - \frac{A}{C}]^2}{D/C^2} = 1 \quad (3.20)$$

Which is the equation of a hyperbola in σ, S space with a center at $(0, \frac{A}{C})$. See in figure(2).

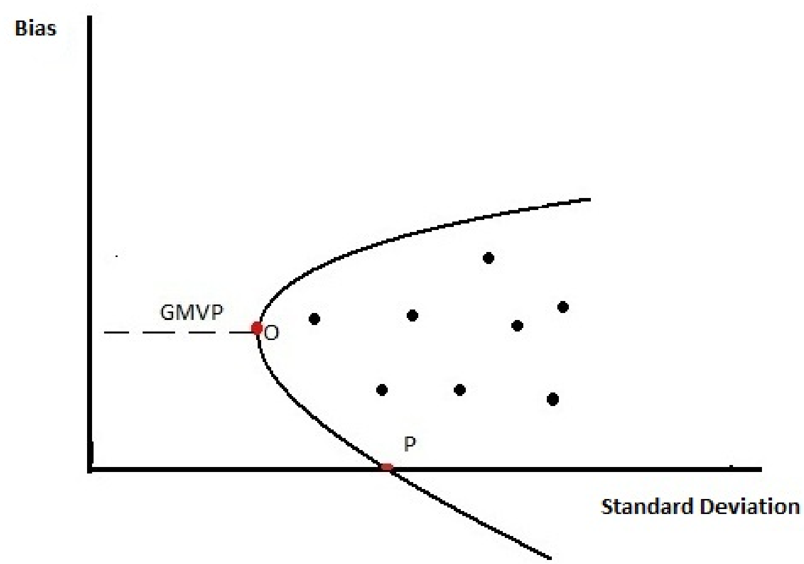

Figure 2: Theoretical Bias-Variance Frontier

Obviously, the two main inputs for our Model Portfolio Approach is the out-of-sample model pricing biases (s) and pricing error covariances (Σ). These two parameters are ex ante unobservable and thus have to estimated.

The direct way of estimating these unknowns is to randomly divide the total sample into three parts as it is shown in figure(3): a training set, a validation set, and a test set. The training set is used to fit the single models; the validation set is used to estimate single model pricing errors (both S and Σ) and model weights ; the test set is used for assessment of the generalization error of the models (Hastie et al., 2009).

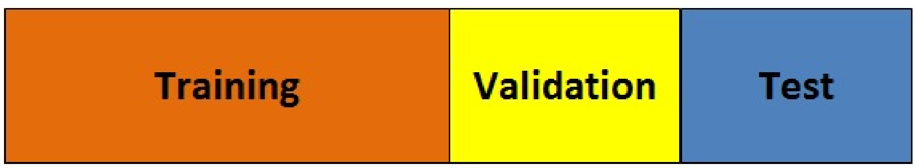

Figure 3: Three Parts of Total Sample

3.2. Global Minimum Variance Weighting Scheme

Admittedly, the three-sample division method is ideal, but if the data is insufficient, we may have to omit the validation step and then approximated analytically (AIC, BIC, MDL, SRM) or by efficient sample re-use (cross-validation and the bootstrap). But none of these substitutes are unbiased, only a separate test set will provide an unbiased estimate of test error. It seems a delimma here. But actually, we still can divide the total sample into three parts when sample size is relatively small by adjusting the parameter (s and Σ) estimation methods accordingly, as the large scale estimation techniques have already been well developed in recent years. The concern of small estimation sample size can thus be mitigated. Here in our approach, we use the three-period methods to obtain our model inputs.

The optimal model portfolio can be solved along the frontier (see figure(2)). However, in practice, our true efficient frontier will be distorted because the bias and variance have to be estimated rather than known. Thus the estimated frontier actually deviates from the true frontier as exactly what happens with the efficient portfolio frontier. As to our our bias-variance method, both bias and variance have to be estimated. The optimization is actually along the estimated frontier rather than along the true frontier. The gains from model uncertainty diversification will be reduced by the estimation error. Hence, not every ex ante optimal model pool will be ex post optimal. The actual performance is reflected on the actual frontier with test sample return and estimated weights. We demonstrate this point with a simple simulation.

Assume pricing error of different asset pricing model multivariate normally distribute, $\mathbf{e_{t+h,t}} \sim MVN(\mu, \mathbf{\Sigma})$. We simulate the true efficient frontier with true parameters μ and $\mathbf{\Sigma}$ (See the solid line in figure(4)). We then estimate the parameters $\hat{\mu}$ and $\hat{\mathbf{\Sigma}}$ and then draw the estimated efficient frontier (See the asterisk line in figure(4)). And then we draw the actual frontier, which is generated with estimated weights $\hat{\omega}$ and the true parameters(μ and $\mathbf{\Sigma}$). The actual frontier is the realization of estimated model portfolio ex post performance. It is the final real performance we can get. The actual frontier is the dotted line in figure(4). We report two typical plots from our 10,000 experiments here.

From these figures, we can see that as actual frontier is to the right of the true frontier, the estimated optimal model portfolio is ex post suboptimal. Because at the similar level of bias, we can always find a smaller variance combination. The chosen model pool is thus not ex post bias-variance efficient.

Among all the model portfolios, global minimum variance portfolio has the most convergent performance on all three frontiers. As global minimum variance combination can be estimated more accurately than other frontier combinations. The ex post optimality of global minimum variance portfolio has also been found in asset portfolio studies (Chan et al., 1999; Jagannathan and Ma, 2003; Kempf and Memmel, 2003). And the reason for the dominating performance of global minimum variance asset portfolio over tangency asset portfolio is documented in Merton (1980): Estimation of mean return is more volatile than estimation of risk(variance). For our model portfolio, the same argument also holds.

From Merton (1980), the variance of estimation on mean return and volatility are given by:

$$Var\{\hat{\mu}\} = \frac{\sigma^2}{q}$$

$$Var\{\hat{\sigma}\} = \frac{1}{2}\frac{\sigma^2}{qn}$$

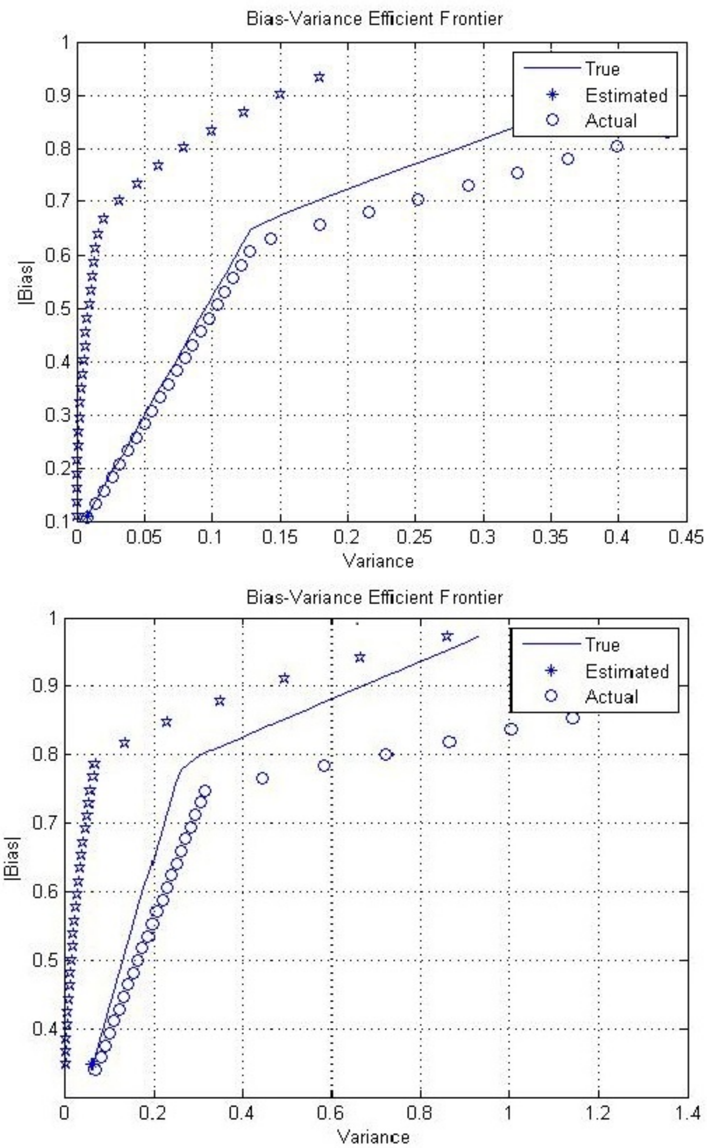

Figure 4: Simulated Frontiers

Where $\hat{\mu}$ is the estimated expected return vector and $\hat{\sigma}$ is the estimated standard deviation.$q \geq 1$ is year length of the data, $n \geq 1$ is the observation interval each year. The number of total observation is $q * n$.The larger the q, the longer the sample period, the more precise of the two estimates. For $q \to \infty$, both values go to zero, the estimation error disappear. For a finite number of years ($q < \infty$), the variance can be estimated more accurately. The relative precision is given by

$$\frac{Var\{\hat{\mu}\}}{Var\{\hat{\sigma}\}} = 2n \tag{3.21}$$

For asset return prediction, one typically use daily, weekly or monthly data,therefore,the precision ration are within a range $24 \leqq \frac{Var\{\hat{\mu}\}}{Var\{\hat{\sigma}\}} \leqq 500$. Thus for asset portfolio, volatility can be more accurately estimated than mean return.

For model portfolio, out-of-sample asset pricing bias estimation is more volatile than pricing error variance:

$$E\{e_{t+h,t}\} = E\{R_{t+h} - \hat{f}_{t+h,t}\} = E\{R_{t+h}\} - \hat{f}_{t+h,t}$$
$$Var\{e_{t+h,t}\} = Var\{R_{t+h} - \hat{f}_{t+h,t}\} = Var\{R_{t+h}\}$$

Note: $\hat{f}_{t+h,t}$ is a known quantity at the forecasting time.

And

$$Var\{E[\hat{e_{t+h,t}}]\} = Var\{E[R_{t+h}]\} = Var\{\hat{\mu}_{t+h}\} = \frac{\sigma^2}{q}$$
$$Var\{Var(\hat{e_{t+h,t}})\} = Var\{Var(\hat{f}_{t+h,t})\} = Var\{\hat{\sigma}\}$$

Substituting in equation (3.21),we can get

$$\frac{Var\{E[\hat{e_{t+h,t}}]\}}{Var\{Var(\hat{e_{t+h,t}})\}} = 2n$$

Hence, just as in the optimal asset portfolio case, we can show that the relative precision of estimation of bias and variance is in also in the range $24 \leqq \frac{VarE[\hat{e_{t+h,t}}]}{Var\{Var(\hat{e_{t+h,t}})\}} \leqq 500$. As the variance of prediction error can be estimated with smaller risk, global minimum variance combination with no dependence on the estimation of bias can be more accurately estimated and more stable. This can be simply seen from the following demonstration:

Global minimum variance model portfolio is solved by

$$\min_{\omega} \omega'\Sigma\omega, \quad s.t. \quad \omega'\mathbf{1} = 1. \tag{3.22}$$

Optimal weight only depends on the estimation of variance-covariance matrix Σ, it is independent of the information on bias. Other frontier combination weights is to solve:

$$\min_{\omega} \quad \omega'\Sigma\omega + \lambda(S - \omega's) \quad s.t. \quad \omega'\mathbf{1} = 1.$$

Optimal weight depends on the estimation of both variance-covariance matrix Σ and bias.

Here we propose a global minimum variance (GMV) weighting scheme. As with the asset portfolio case, the GMV weights is actually a by-product of the equation (3.22) minimization problem. The weight is given by

$$\omega = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$$

GMV model portfolio is on the center of the parabola (Figure 1, point O($1/C$, A/C)). A/C is the optimal bias adjustment. As can be seen from the weights calculation formula, estimation of inverse covariance matrix of prediction errors between different pricing schemes is critical to the success of the Model Portfolio Approach, or optimal weighting scheme.

The variance-covariance matrix Σ can be estimated as its sample counterpart. But for a large scale problem, as the sample become relative small, sample covariance matrix is estimated with large error. We adopt a robust covariance estimation method from Schafer et al. (2005). This method can be summarized as

$$\Sigma^* = \lambda T + (1 - \lambda)U \tag{3.23}$$

Where T and U represent the target covariance matrix and maximum likelihood covariance estimator. The shrinkage intensity λ^* is given by

$$\lambda^* = \frac{\sum_{i=1}^{p} \hat{var}(u_i) - \hat{cov}(t_i, u_i) - \hat{Bias}(u_i)(t_i - u_i)}{\sum_{i=1}^{p}(t_i - u_i)^2} \tag{3.24}$$

By producing a well-conditioned covariance estimate, we can automatically obtain an equally well-conditioned estimate of the inverse covariance (Schafer et al., 2005).The estimated covariance matrix is a weighted average of the unbiased covariance estimator and a target. The shrinkage intensity (the weights) can be solved analytically by Lediot and Wolf's (2003) theorem. This weighting method is distribution free and is not computationally intensive as MCMC, bootstrap or cross validation.

3.3. Comparsion With Other Weighting Schemes

3.3.1. Unification With Optimal Weighting

In the earlier pioneering forecast combination work, Bates and Granger (1969) discuss the optimality of combined forecast and provide a discussion of the ideal properties of a optimal combining weighting method. Bates and Granger (1969) summarize the optimal combining weights estimation into three OLS regressions:

$$\begin{aligned} &(i) \quad y = \omega' F + \varepsilon \\ &(ii) \quad y = \omega^T F + \varepsilon \quad s.t. \quad \omega' \mathbf{1} = 1 \\ &(iii) \quad y = \omega_0 + \omega' F + \varepsilon \end{aligned} \tag{3.25}$$

Where y is the forecast target. F is forecast vector which has the component prediction from single model. By regressing the realized target on the forecasts, one can obtain the coefficient estimates for ω, which are model weights for individual models. Then the combined prediction can be formed as:

$$\begin{aligned} &(i) \quad f_c = \omega' F \\ &(ii) \quad f_c = \omega^T F \quad s.t. \quad \omega' \mathbf{1} = 1 \\ &(iii) \quad f_c = \omega_0 + \omega' F \end{aligned}$$

Attention in these works are restricted to combination of two model forecasts. When extended to combination of N models by the Dickinson (1975), the model weight vector is given by:

$$\omega = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \tag{3.26}$$

Where Σ is covariance matrix of forecasting errors. This formula in this basic form is identical to our GMV weighting scheme. But under our Model Portfolio Approach, we do not require individual model unbiasedness. As can be seen from our derivation, individual models can be biased, we do not require the model to be constrained to the horizontal axis in figure(2). This restricted optimal model portfolio is located at point P in figure(2), which has larger variance than GMV combination (Point O). Thus, if a single model are biased, our method provide a better combination.

Another good property of our GMV weighting scheme is that GMV weighting explicit points out the key to the success of optimal weighting scheme. As can be seen from the weighting formula, inverse covariance matrix estimation is the critical inputs for GMV weight estimation. Thus we can concentrate our effort on improved estimation of covariance matrix. With a well estimate covariance matrix, our GMV weights can be accomodated to large model space, or small effective sample size problem, of which OLS regression weighting method is incapable.

3.3.2. A Discussion of Other Weighting Schemes

The aim of out-of-sample prediction modeling is to find a model which gives smallest test error. Model selection and model combination are two ways to achieve the same goal but with a different paradigm. With the same performance metric, model selection method selects the single best model while model combination pools all the models, but all aim at optimizing the performance metric. Therefore it is not surprising that most development in model combination is along the line of optimising different model selection criterion.

The recent Hansen's Mallows Model Averaging (2008), a weighting scheme based on minimization of Mallows criterion, is actually a direct optimization of model selection criterion C_p. The reason for using Mallows' C_p statistic is that C_p is an asymptotically unbiased estimate of both the in-sample mean-squared error (MSE) and the out-of-sample one-step-ahead mean-squared forecast error (MSFE). Geweke and Amisano's weighting method (2011) is through a maximization of predictive log score. With a binomial log likelihood, this optimization is equal to a AIC information criterion optimization which is commonly used in model selection.

Obviously, the recent development in this area is actually to find a model selection criterion which can be used as a good estimator for test error. So far the tested criterion are all from model selection. But actually, as discussed in Hastie et al. (2009), C_p, AIC and BIC are not unbiased estimators for expected out-of-sample prediction error. These are all based upon the in-sample fit of the model, penalised according to the degrees of freedom. They are not a direct estimate of out-of-sample test error. Admittedly, the estimation of test error for a particular training set is not easy in general.

These model combination studies so far narrow their focus only to one kind of model selection methods. But actually, there is another "family" of model selection approaches which relies in some sense on out-of-sample testing. The simplest form of this approach is simple "validation" in which the model is selected which performs best on a particular out-of-sample set. This method is commonly used in the "early stopping" approach to divide the whole sample into sub-samples, and then use the validation sample prediction error as a direct estimation for test sample error. A more computationally intensive form of this approach is "k-fold cross validation", in which the data set is divided into k sub-samples and the performance on each sub-sample is estimated with respect to the model optimised on all other sub-samples. The extreme of this approach is full "leave one out" cross-validation (Wahba and Wold, 1975). A related but distinct approach is the use of "bootstrap" re-sampling (Efron and Tibshirani, 1993) of the data to obtain unbiased estimates of prediction error. Motivated by these direct prediction error estimation method, our GMV weighting can be seen as a direct optimization of out-of-sample prediction error and thus should have smallest realized out-of-sample prediction error.

Among the many sophisticated weighting schemes, simple arithmetic average weighting scheme, picking a set of models and then giving them all equal weight, has the most robust out-of-sample performance. The most consensus on the explanation for this puzzle is that diversification gain from weights estimation is not large enough to offset the estimation loss, but with a enlarged sample size, weights estimation will be worthwhile. With this notion, the Bayesian style frequentist shrinkage weighting scheme have been developed. This method is like a combination of combining weights, a weight average weight between estimated weight and a target reference weight. The reference weight is usually independent of sample data. "1/N" are most often used as a shrinkage target. Actually, shrinkage weight has the same effect as impose a shrinkage structure on covariance matrix estimator (Frahm and Memmel, 2010).

By grafting Schafer and Strimmer' s shrinkage covariance estimator to our GMV weighting scheme, GMV can be applied to both large sample size problem and small sample size problem. And as GMV weights is solved from our Model Portfolio Approach, it is directly minimize out-of-sample test error. Therefore it should have improved performance over other weighting schemes.

3.3.3. The Choice of Weighting Schemes

GMV weighting has advantageous merits in most aspects but it is not flawless. To better implement it, the underlying assumption is important.

The first fact is that GMV is optimal on the premise that it is highly possible that all model considered are false. Because if the true model is in the model space, it should have a weight equal to 1 while other models should get a zero weighting. But due to estimation error, the model weight on true model can just be close to 1. In this case, Bayesian model averaging might be a better choice as it assume a complete model space which comprises true model.

Another fact is that, just as with the asset portfolio case, model portfolio diversification efficiency gain depends on the correlation between model errors. The gain is an increasing function of between model correlation ρ. Thus "1/N" weighting will dominate if models are weakly correlated.

For asset pricing model uncertainty, our choice of the weighting scheme depends on the asset pricing models at hand. As most of asset pricing models are derived under same utility maximization, they may be well correlated and with large financial data set, GMV weighting can be a better weighting scheme for the asset pricing model portfolio. Empirical test of this argument will be conducted in future work.

4. Monte Carlo Experiment

In this section, we design two Monte Carlo simulation experiments to demonstrate our analytical results in previous sections: The improvement of model portfolio over single models and the optimality of GMV weighting scheme.

We make no assumption about the underlying asset pricing model , the only input of our method is out-of-sample pricing errors. Thus for simulation studies here, we only need to generate error series. These error series can be seen as from any asset pricing models, such as CAPM, APT, Fama-French three factor models and consumption based asset pricing models, or more broadly, forecasts from fundamental analysis and technical analysis. Also it can be forecasts from survey of investors or analysts, as long as we can estimate the forecasting error and the related moments.

4.1. Comparsion between Single Model and Model Portfolio

This first simulation is a simple demonstration of the advantage of combined models. Firstly, we draw true return and single model forecasts from a joint normal distribution:

$$\begin{pmatrix} R_{t+h} \\ \hat{\mathbf{R}}_{t+h,t} \end{pmatrix} \sim \mathcal{MVN}\left(\begin{pmatrix} \mu_{R_{t+h}} \\ \mu_{\hat{\mathbf{R}}_{t+h,t}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{R_{t+h}} & \sigma'_{\mathbf{R}\hat{\mathbf{R}}_{\mathbf{t+h,t}}} \\ \sigma'_{\mathbf{R}\hat{\mathbf{R}}_{\mathbf{t+h,t}}} & \mathbf{\Sigma}_{\hat{\mathbf{R}}\hat{\mathbf{R}}_{t+h,t}} \end{pmatrix} \right) \tag{4.1}$$

Where $\hat{\mathbf{R}}_{t+h,t}$ is a $n \times 1$ vector denotes N forecasts from individual asset pricing models. True return generated from normal distribution $N(\mu_{R_{t+h}}.\sigma^2_{R_{t+h}})$. $\mathbf{\Sigma}_{\hat{\mathbf{R}}\hat{\mathbf{R}}_{t+h,t}}$ is covariance matrix of pricing error series from individual pricing models. $\sigma'_{\mathbf{R}\hat{\mathbf{R}}_{\mathbf{t+h,t}}}$ is covariance matrix of single forecasts and the true return. We fix a sample size $T = 100$, and split the sample into two parts: the first 60 draws are used for in sample training, and the last 40 observations are used for out-of-sample test. We fix the number of model $N = 4$. Thus this simulation is done with large effective sample size. Here, for simulation study, we can assume in-sample model fitting has already been done, because our approach has no requirement of the in-sample estimation methods. Thus we only need two periods for simulation, one for test error and model weights estimation, and the other one for out-of-sample test. However, in empirical application, we need three sub-samples, so one more estimation sub-sample to be specified.

With different specification of variance-covariance structure, we have four experiments: For the first one, single pricing models have same variance and same correlation coefficients with true return, $\rho = 0.25$; in the second one, we assume 4 single pricing models have same variance but different correlation coefficients with true return, ρ are assigned to $0.75, 0.55, 0.35, 0.15$ respectively. The last two simulation assume single pricing models have different variance and then give same specification for the correlation between true return and single pricing model as the first two simulations.

We generate combined forecast using our GMV weighting scheme. The error covariance are estimated with the first 60 observations and forecast weights are estimated by minimizing mean square forecasting error (MSFE). We do all the simulations 10,000 times, and then compare the performance of combined pricing model with the single pricing models both in-sample and out-of-sample. We use relative mean squared forecasting error(RMSFE) to evaluate the performance of combined model. It is calculated as:

$$RMEFE = \frac{MSFE_i}{MSFE_c} \quad i = 1, 2, 3, 4.$$

Where $MSFE_i$ is MSFE of single pricing model i and $MSFE_c$ is combined pricing model MSFE. We compute the average RMSFE from our 10,000 simulations. Table(.1) shows result of in-sample test, and table(.2) reports result of out-of-sample test.

[Insert table .1 here]

[Insert table .2 here]

From the simulation results, we can see that across all scenarios, model portfolio out-perform all single models. Model Portfolio Approach is a better way to mitigate asset pricing model uncertainty.

4.2. Comparsion between Different Weighting Schemes

4.2.1. Experiment Design

Following our model set-up, real asset return can be decomposed into expected return and unexpected return. In our study here, we assume that true conditional expectation of return is $E_{t-1}(y_t) = \alpha_0 + \alpha_1 y_{t-1}$, then we generate our data with a Gaussian autoregressive AR(1) process:

$$R_t = \alpha_0 + \alpha_1 R_{t-1} + \xi_t \tag{4.2}$$

$$\xi_t \sim N(0,1), \quad \alpha_0 = 0, \quad \text{and} \quad \alpha_1 = 0.5 \tag{4.3}$$

Where ξ_t is an unpredictable component, thus the minimum pricing error of a model is unity in the specification here. The total deviation of an asset pricing model from the true expectation comprise a bias and an idiosyncratic pricing error, while the deviation from realized actual return include an extra unavoidable error due to the variance of return itself. Empirical approximation of an asset pricing model is:

$$f_{i,t} = \hat{\alpha}_0 + \hat{\alpha}_1 R_{t-1} + \eta_{i,t}$$
$$t = 1, ..., T \quad T \text{ is sample length}$$
$$i = 1, .., N \quad N \text{ is the number of asset pricing models or forecasts}$$

Where $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are estimates of α_0 and α_1. $\eta_{i,t}$ is drawn from a multivariate normal distribution $\eta \sim (B, \Sigma)$. B is mean bias vector of individual pricing model, with elements $B_i = \beta B_{i-1} + u_i$, and $u_i \sim$ i.i.d. $U(a, b)$, $0 < \beta < 1$.

Follow Issler and Lima (2009), we specify a spatial dependence in bias by letting $\beta = 0.5$. And the aggregate average bias is $\frac{a+b}{2(1-\beta)}$. We consider a zero average bias by letting $a = -0.5, b = 0.5$. Σ is the variance covariance matrix of pricing errors. We use this model to generate our single pricing model forecasts.

We generate an overall sample of size T, We then divide T to 3 consecutive periods, where time is indexed $t = 1, 2, ...T_1, ..., T_2, ...T$. The first sub-period E is the initial estimation sample, or training sample as it is usually called, where the models are fitted and α and β are estimated. The number of observations in it is $E = T_1$, comprising $(t = 1, 2, ...T_1)$.The second one is validation sample, where the prediction error of each model and the relevant bias-adjust term and combining weights are estimated. It has $V = T_2 - T_1$ observations. The last period is test sample with $Te = T - T_2$ observations, where the performance of different models are evaluated. In our experiment here, we fix estimation sample size $T = 200$, validation sample size $V = 50$ and test sample size $Te = 50$. We then adjust the total number of models. We use N as the number of models. We give a value of 3 and 40 to N. The value of N/V determines the effective sample size for model weights estimation. $N = 3$ with $V = 50$ can be used as a large sample, and $N = 40$ with $V = 50$ is a relatively small sample size[7].

Inspired by asset portfolio studies, we suspect that estimation error is not the only reason for "1/N" puzzle, underlying pricing errors correlation should be another factor to consider. To test our intuition here, we simulate both highly and weakly correlated error cases under both small effective and large effective sample size. For the high correlation, we restrict the correlation coefficients to be in a range of $[0.9, 1]$, while for the weakly correlated error, we restrict the correlation coefficients to be within $[0, 0.25]$.

We do all these four simulations 50,000 times, and then we generate combined model under "1/N","GMV", "OLS" and "Shrinkage" weighting schemes. For all these weighting methods, we consider both bias-corrected and no-bias-adjustment combined pricing models.

4.2.2. Simulation Results

We evaluate performance of different weighting schemes by average asset pricing risk or expected mispricing uncertainty, which is calculated as MSFE,

[7]There is no unanimous specification for small and large sample size

and also variance of model weights which is proxied by average variance of model weights. Results are reported in table .3.

[Insert table .3 here]

From table .3, we can see that, as what we predict, the small sample effective size is not the only reason for "$1/N$" puzzle, correlation between errors also account for the relative under performance of theoretical optimal weights. In our small effective sample size scenario, when errors are highly correlated, $1/N$ is not the best combination, instead, GMV weighting scheme has the best performance. As OLS is inferior for large scale estimation problem, here for large model number, all three OLS perform the worst. Obviously, $1/N$ is optimal when it is close to the true optimal. Only in this case, diversification gains cannot offset the loss of optimal weights estimation error, and thus $1/N$ offer a better weighting choice. For large sample size, as optimal weights can be relatively more accurately estimated, optimal weights from OLS should perform best. We can see in table .3, all three OLS weighting give smaller asset pricing risk. GMV weighting has similar performance as OLS, but has a much lower weights variance. GMV weighting is more stable than OLS. The reason for the improved performance of GMV weighting, is because by using more robust covariance estimator, GMV has similar large effective sample size properties as OLS and better features for small effective sample size problem.

Across all the weighting schemes, the bias-corrected weighting is inferior to the relevant weighting without bias adjustment. The reason is as what we proved in the previous section. The bias is more difficult to predict than the variance of error and thus more volatile. And usually, the in-sample unbiasedness does not lead to unbiasedness in out-of-sample period.

5. Conclusion

While appreciating the beauty of asset pricing model theory, we still lack the confidence of concluding whether asset pricing models are true or false empirically. But fortunately, for practical implementation of asset pricing models, as our aim is to find a model which will generate well into a finite independent sample, we do not require zero single model error. Actually we can improve our modelling performance even when all models are false but contain useful incremental information.

The explicit mapping of model portfolio approach to asset portfolio theory paves a finance oriented approach to asset pricing model uncertainty problem. By pooling asset pricing models, the model portfolio diversifies the single pricing model uncertainty and thus reduces the overall mispricing uncertainty. A by product of model portfolio optimization is a model weighting scheme. As our effort is on ex post model pool performance, we advocate a GMV weighting scheme which is a compromise between estimation error and diversification benefit. The simulation results show that in large sample, GMV weighting has the identically good performance as optimal Bates-Granger-Ramanathan OLS weights but more stable, and it also has improved performance in small samples. A natural factor from our approach to blame for "$1/N$" puzzle is the correlation between errors and corresponding covariance matrix estimation. If covariance matrix can be accurately estimated, there is no puzzle. $1/N$ rule will only dominate when it is actually optimal.

What finance gives us is not only models but also ways of thinking and viewing many other problems. Our work is the first to draw an explicit parallel from model combination to portfolio theory. This work paves the way for future research to fully appreciate wisedom of these two areas. A natural extension of our study is to apply the finance hedging idea into model portfolio approach. Additionally, model portfolio is not just a statistical trick, it implies that model uncertainty might be a mispricing factor in existing asset pricing models and thus allow statistic arbitrage opportunities.

References

Avramov, D., 2002. Stock return predictability and model uncertainty. Journal of Financial Economics 64 (3), 423–458.

Bates, J. M., Granger, C. W. J., 1969. The combination of forecasts. OR 20 (4), pp. 451–468.
URL `http://www.jstor.org/stable/3008764`

Bossaerts, P., 2004. The Paradox of Asset Pricing. Frontiers of Economic Research. Princeton University Press.
URL `http://books.google.com.au/books?id=Z7ChQgAACAAJ`

Bunn, D., 1989. Forecasting with more than one model. Journal of Forecasting 8 (3), 161–166.

Chan, L., Karceski, J., Lakonishok, J., 1999. On portfolio optimization: Forecasting covariances and choosing the risk model. Tech. rep., National Bureau of Economic Research.

Chong, Y., Hendry, D., 1986. Econometric evaluation of linear macroeconomic models. The Review of Economic Studies 53 (4), 671–690.

Clemen, R., Winkler, R., 1986. Combining economic forecasts. Journal of Business & Economic Statistics, 39–46.

Clemen, R. T., 1989. Combining forecasts: A review and annotated bibliography. International Journal of Forecasting 5 (4), 559 – 583.
URL http://www.sciencedirect.com/science/article/pii/0169207089900125

Clements, M., Hendry, D., College, N., 2002. Pooling of Forecasts. Discussion papers in economics. Nuffield College.
URL http://books.google.com.au/books?id=gaHySAAACAAJ

DeMiguel, V., Garlappi, L., Uppal, R., 2009. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? Review of Financial Studies 22 (5), 1915–1953.

Dickinson, J., 1975. Some comments on the combination of forecasts. Operational Research Quarterly, 205–210.

Diebold, F., Pauly, P., 1990. The use of prior information in forecast combination. International Journal of Forecasting 6 (4), 503–508.

Diebold, F. X., Lopez, J. A., March 1996. Forecast evaluation and combination. Working Paper 192, National Bureau of Economic Research.
URL http://www.nber.org/papers/t0192

Dunis, C., Laws, J., Chauvin, S., 2001. The use of market data and model combination to improve forecast accuracy. Development in Forecasts Combination and Portfolio Choice (Wiley, Oxford).

Durham, G., Geweke, J., 2011. Improving asset price prediction when all models are false. Manuscript, University of Technology, Sydney. URL http://www. censoc. uts. edu. au/pdfs/geweke_papers/gp_working_5b. pdf.

Efron, B., Tibshirani, R., 1993. An introduction to the bootstrap. Vol. 57. Chapman & Hall/CRC.

Elton, E. J., 1999. Expected return, realized return, and asset pricing tests. The Journal of Finance 54 (4), pp. 1199–1220.
URL http://www.jstor.org/stable/798003

Fair, R., Shiller, R., 1989. The informational content of ex ante forecasts.

Fair, R., Shiller, R., 1990. Comparing information in forecasts from econometric models. The American Economic Review, 375–389.

Foster, F. D., Smith, T., Whaley, R. E., 1997. Assessing goodness-of-fit of asset pricing models: The distribution of the maximal r2. The Journal of Finance 52 (2), pp. 591–607.
URL http://www.jstor.org/stable/2329491

Frahm, G., Memmel, C., 2010. Dominating estimators for minimum-variance portfolios. Journal of Econometrics 159 (2), 289–302.

Geman, S., Bienenstock, E., Doursat, R., 1992. Neural networks and the bias/variance dilemma. Neural computation 4 (1), 1–58.

Granger, C. W. J., 1989. Invited review combining forecastsï£ · twenty years later. Journal of Forecasting 8 (3), 167–173.
URL http://dx.doi.org/10.1002/for.3980080303

Granger, C. W. J., Ramanathan, R., 1984. Improved methods of combining forecasts. Journal of Forecasting 3 (2), 197–204.
URL http://dx.doi.org/10.1002/for.3980030207

Hansen, B., 2008. Least-squares forecast averaging. Journal of Econometrics 146 (2), 342–350.

Hansen, P., 2009. In-sample fit and out-of-sample fit: Their joint distribution and its implications for model selection. Tech. rep., Working Paper.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer Series in Statistics. Springer.
URL http://books.google.com.au/books?id=tVIjmNS3Ob8C

Issler, J., Lima, L., 2009. A panel data approach to economic forecasting: The bias-corrected average forecast. Journal of Econometrics 152 (2), 153–164.

Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. The Journal of Finance 58 (4), 1651–1684.

Kempf, A., Memmel, C., 2003. On the estimation of the global minimum variance portfolio.

Ledoit, O., Wolf, M., d'Economia i Empresa, U. P. F. D., 2003. Honey, I shrunk the sample covariance matrix. Universitat Pompeu Fabra.

Markowitz, H., 1952. Portfolio selection. The Journal of Finance 7 (1), pp. 77–91.
URL http://www.jstor.org/stable/2975974

Merton, R., 1980. On estimating the expected return on the market* 1:: An exploratory investigation. Journal of Financial Economics 8 (4), 323–361.

Min, C., Zellner, A., 1993. Bayesian and non-bayesian methods for combining models and forecasts with applications to forecasting international growth rates. Journal of Econometrics 56 (1-2), 89–118.

Neely, C., Rapach, D., Tu, J., Zhou, G., 2010. Out-of-sample equity premium prediction: economic fundamentals vs. moving-average rules. Journal of Finance.

O'Doherty, M., Savin, N., Tiwari, A., 2010. Modeling the cross section of stock returns: A model pooling approach.

Pastor, L., Stambaugh, R., 2000. Comparing asset pricing models: an investment perspective. Journal of Financial Economics 56 (3), 335–381.

Schafer, J., Strimmer, K., et al., 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Statistical applications in genetics and molecular biology 4 (1), 32.

Simin, T., 2008. The poor predictive performance of asset pricing models. The Journal of Financial and Quantitative Analysis 43 (2), pp. 355–380.
URL http://www.jstor.org/stable/27647353

Smith, J., Wallis, K. F., 2009. A simple explanation of the forecast combination puzzle*. Oxford Bulletin of Economics and Statistics 71 (3), 331–355.
URL http://dx.doi.org/10.1111/j.1468-0084.2008.00541.x

Stock, J. H., Watson, M. W., 2003. Forecasting output and inflation: The role of asset prices. Journal of Economic Literature 41 (3), pp. 788–829.
URL http://www.jstor.org/stable/3217532

Stock, J. H., Watson, M. W., 2004. Combination forecasts of output growth in a seven-country data set. Journal of Forecasting 23 (6), 405–430.
URL http://dx.doi.org/10.1002/for.928

Stock, J. H., Watson, M. W., 2006. Chapter 10 forecasting with many predictors. Vol. 1 of Handbook of Economic Forecasting. Elsevier, pp. 515 – 554.
URL http://www.sciencedirect.com/science/article/pii/S1574070605010104

Timmermann, A., 2006. Chapter 4 forecast combinations. Vol. 1 of Handbook of Economic Forecasting. Elsevier, pp. 135 – 196.
URL http://www.sciencedirect.com/science/article/pii/S1574070605010049

Wahba, G., Wold, S., 1975. A completely automatic french curve: Fitting spline functions by cross validation. Communications in Statistics-Theory and Methods 4 (1), 1–17.

Appendix

Table .1: In-Sample Performance Evaluation

Asset Pricing Model	Variance of Error	Bias	RSMSE
Different variance and different correlation			
Model 1	0.3065	-0.0282	13986.1323
Model 2	0.3994	-0.0151	25410.6878
Model 3	0.4806	-0.01413	27436.7238
Model 4	0.5788	-0.0250	42897.0610
GMV Weights	0.0518	8.74E-18	1.0000
Different variance and same correlation			
Model 1	0.5481	0.0091	16.8721
Model 2	0.5320	0.0031	14.2641
Model 3	0.5471	0.0307	14.2846
Model 4	0.5522	0.0120	14.9493
GMV Weights	0.3004	5.1105e-18	1.0000
Same variance and different correlation			
Model 1	0.3182	-0.0036	26578.6374
Model 2	0.4078	0.0228	36036.7888
Model 3	0.4985	0.0036	45745.5714
Model 4	0.5914	0.0350	53988.2495
GMV Weights	0.1153	0.0000	1.0000
Same variance and same correlation			
Model 1	0.6353	-0.0075	133.0049
Model 2	0.6483	0.0010	155.2658
Model 3	0.3306	-0.0141	21.2260
Model 4	0.3686	-0.0120	37.7622
GMV Weights	0.2877	0.0000	1.0000

Table .2: Out-of-Sample Performance Evaluation

Asset Pricing Model	Variance of Error	Bias	RSMSE
Different variance and different correlation			
Model 1	0.3021	-0.0564	11.8351
Model 2	0.3975	-0.0364	9.5486
Model 3	0.4765	-0.0275	12.6878
Model 4	0.5661	-0.0523	11.7616
GMV Weights	0.0569	0.0550	1.0000
Different variance and same correlation			
Model 1	0.5505	0.0212	3.5465
Model 2	0.5357	0.0090	2.7700
Model 3	0.5434	0.0637	2.7733
Model 4	0.5582	0.0295	2.6118
GMV Weights	0.3396	0.0843	1.0000
Same variance and different correlation			
Model 1	0.3175	-0.0107	1.5967
Model 2	0.4032	0.0467	2.0771
Model 3	0.4933	0.0089	2.7988
Model 4	0.5801	0.0721	2.9204
GMV Weights	0.1279	-0.2802	1.0000
Same variance and same correlation			
Model 1	0.6345	-0.0241	3.5945
Model 2	0.6517	-0.0084	3.9561
Model 3	0.3287	-0.0349	1.3858
Model 4	0.3668	-0.0334	2.5248
GMV Weights	0.3203	0.0447	1.0000

Table .3: Weighting Schemes Comparsion

Small Effective Sample Size with Weakly Correlated Model Error										
		Bias-Corrected1/N	1/N	Bias-Corrected GMV	GMV	OlS1	OLS2	OLS3	Bias-Corrected Shrinkage	Shrinkage
MSFE	Mean	1.56	1.48	2.44	2.17	7.94	8.41	6.33	1.56	1.48
	Variance	0.07	0.07	0.33	0.16	34.79	21.47	16.28	0.07	0.07
	Skewness	-0.49	-0.25	0.36	-0.05	1.32	1.04	1.34	-0.49	-0.25
	Kurtosis	1.94	1.72	2.64	1.41	3.33	2.64	3.37	1.94	1.72
	0.01 Quantile	1.11	1.10	1.56	1.60	3.07	3.59	3.10	1.11	1.10
	0.25 Quantile	1.26	1.18	2.06	1.80	4.81	5.68	3.65	1.26	1.18
	0.50 Quantile	1.63	1.52	2.44	2.16	5.24	6.47	4.58	1.63	1.52
	0.75 Quantile	1.72	1.68	2.78	2.55	8.88	9.15	7.85	1.72	1.68
	0.99 Quantile	1.92	1.86	3.56	2.65	20.73	16.64	15.20	1.92	1.86
Average Variance of Weights		0.00	0.00	0.05	0.05	4.70	6.55	3.87	0.00	0.00
Small Effective Sample Size with Highly Correlated Model Error										
		Bias-Corrected1/N	1/N	Bias-Corrected GMV	GMV	OlS1	OLS2	OLS3	Bias-Corrected Shrinkage	Shrinkage
MSFE	Mean	1.73	1.71	1.81	1.60	367.40	1131.24	33.34	1.73	1.71
	Variance	0.11	0.11	0.35	0.13	303865.85	7762946.57	2444.81	0.11	0.11
	Skewness	0.14	0.29	0.56	0.22	1.38	2.44	1.77	0.14	0.29
	Kurtosis	1.80	1.86	2.29	2.05	3.82	7.32	4.89	1.80	1.86
	0.01 Quantile	1.27	1.27	1.11	1.09	4.19	3.27	2.07	1.27	1.27
	0.25 Quantile	1.50	1.49	1.22	1.22	8.43	3.94	5.81	1.50	1.49
	0.50 Quantile	1.70	1.66	1.74	1.62	16.43	12.63	8.57	1.70	1.66
	0.75 Quantile	2.05	2.05	2.36	1.75	595.62	370.65	43.83	2.05	2.05
	0.99 Quantile	2.24	2.25	2.92	2.20	1648.25	8855.83	156.39	2.24	2.25
Average Variance of Weights		0.00	0.00	0.03	0.03	811699001.31	628897.22	3216856.73	0.00	0.00
Large Effective Sample Size with Weakly Correlated Model Error										
		Bias-Corrected1/N	1/N	Bias-Corrected GMV	GMV	OlS1	OLS2	OLS3	Bias-Corrected Shrinkage	Shrinkage
MSFE	Mean	1.67	1.49	1.75	1.51	1.13	1.47	1.11	1.70	1.49
	Variance	0.19	0.10	0.28	0.11	0.06	0.11	0.06	0.24	0.11
	Skewness	0.95	0.46	1.35	0.51	0.70	0.53	0.60	1.24	0.48
	Kurtosis	4.71	3.34	6.52	3.45	4.29	3.52	3.85	6.05	3.41
	0.01 Quantile	0.90	0.86	0.90	0.85	0.65	0.83	0.64	0.90	0.85
	0.25 Quantile	1.36	1.27	1.39	1.27	0.95	1.24	0.94	1.37	1.26
	0.50 Quantile	1.61	1.47	1.66	1.48	1.11	1.44	1.09	1.62	1.46
	0.75 Quantile	1.90	1.70	2.01	1.71	1.28	1.67	1.26	1.95	1.69
	0.99 Quantile	3.00	2.35	3.49	2.40	1.84	2.36	1.78	3.28	2.36
Average Variance of Weights		0.00	0.00	0.09	0.09	0.69	1.05	0.20	0.05	0.05
Large Effective Sample Size with Highly Correlated Model Error										
		Bias-Corrected1/N	1/N	Bias-Corrected GMV	GMV	OlS1	OLS2	OLS3	Bias-Corrected Shrinkage	Shrinkage
MSFE	Mean	2.10	1.86	2.09	1.78	1.21	1.55	1.19	2.09	1.78
	Variance	0.26	0.16	0.23	0.16	0.05	0.09	0.04	0.23	0.16
	Skewness	0.19	0.10	0.03	0.01	0.13	0.08	-0.05	0.03	0.01
	Kurtosis	1.78	1.58	2.29	2.13	2.44	1.82	2.86	2.29	2.13
	0.01 Quantile	1.34	1.34	1.29	1.12	0.81	1.11	0.81	1.29	1.12
	0.25 Quantile	1.77	1.55	1.88	1.59	1.09	1.31	1.09	1.88	1.59
	0.50 Quantile	1.96	1.75	2.08	1.72	1.16	1.57	1.17	2.08	1.72
	0.75 Quantile	2.60	2.21	2.31	2.15	1.31	1.80	1.28	2.31	2.15
	0.99 Quantile	2.84	2.44	2.86	2.39	1.57	2.03	1.53	2.86	2.39
Average Variance of Weights		0.00	0.00	0.26	0.26	94.18	263.79	4.60	0.26	0.26

Note: $1/N$ weighting: Assign equal weights to individual models and N is the number of models and bias correction term is estimated as $1/N\{1/V \sum_{t=T_1}^{T_2} R_t - f_t\}$. GMV weighting: Weights are estimated with formula (3.2) and bias adjustment is estimated as $\frac{A}{C} = \frac{s'\Sigma^{-1}\underline{1}}{\underline{1}'\Sigma^{-1}\underline{1}}$ (Σ is estimated using Schafer and Strimmer (2005) technique); Three OLS weightings are as the three regression in equation (3.25); Shrinkage weighting use the dominating shrinkage estimator from Frahm and Memmel (2010).