




Multiple regression model averaging and the focused information criterion with an application to portfolio choice

Filip Klimenka & James Lewis Wolter

To cite this article: Filip Klimenka & James Lewis Wolter (2017): Multiple regression model averaging and the focused information criterion with an application to portfolio choice, Journal of Business & Economic Statistics, DOI: [10.1080/07350015.2017.1383262](https://doi.org/10.1080/07350015.2017.1383262)

To link to this article: <http://dx.doi.org/10.1080/07350015.2017.1383262>

 View supplementary material 

 Accepted author version posted online: 29 Sep 2017.

 Submit your article to this journal 

 Article views: 8

 View Crossmark data 

Multiple regression model averaging and the focused information criterion with an application to portfolio choice

Filip Klimenka^{a,b} and James Lewis Wolter^{a,b}

May 2, 2017

^aDepartment of Economics, University of Oxford

^bOxford-Man Institute of Quantitative Finance

Abstract

We consider multiple regression (MR) model averaging using the Focused Information Criterion (FIC). Our approach is motivated by the problem of implementing a mean-variance portfolio choice rule. The usual approach is to estimate parameters ignoring the intention to use them in portfolio choice. We develop an estimation method that focuses on the trading rule of interest. Asymptotic distributions of submodel estimators in the MR case are derived using a localization framework. The localization is of both regression coefficients and error covariances. Distributions of submodel estimators are used for model selection with the FIC. This allows comparison of submodels using the risk of portfolio rule estimators. FIC model averaging estimators are then characterized. This extension further improves risk properties. We show in simulations that applying these methods in the portfolio choice case results in improved estimates compared with several competitors. An application to futures data shows superior performance as well.

1 Introduction

Since the seminal work of Markowitz (1952), mathematical descriptions of optimal portfolio selection have received great academic interest. In addition, mean-variance portfolio optimization has been implemented extensively in industry practice. Many institutions directly or indirectly use these methods. These include pension funds, mutual funds, university endowments, charitable foundations and hedge funds to name a few. A substantial portion of overall investments are made

with these principles. See Garleanu and Pedersen (2013), DeMiguel, Mei and Nogales (2014) and Collin-Dufresne, Daniel, Moallemi and Saglam (2015) for recent contributions in this area.

If an investor wishes to implement a mean-variance portfolio rule, she must estimate the underlying return structure. Standard approaches separate statistical estimation from implementation. The parameters of the model are first estimated. Then estimates are plugged into policy functions when making decisions. The first stage ignores the second.

In this paper, we propose a different methodology which uses policy functions in estimation. When trying to estimate a portfolio rule, the parameters of the underlying return structure are only tangentially relevant. Our main interest lies in the optimal policy, which is a transformation of these parameters. Because of this, we develop a procedure focused on the policy, not the parameters themselves. Estimating the parameters is only a means to an end.

Our results are motivated with a mean-variance portfolio rule assuming asset returns follow a multiple regression (MR) structure. In this setup, returns equations depend on arbitrary covariates. In applications, a number of potential covariates are usually available, but we are unsure which are important. Additionally, there is often limited historical data. Which covariates should be included to optimize bias-variance trade-off in finite samples is not obvious. In our setup, this trade-off is distorted by the policy function. The importance of a particular parameter depends on how much it influences the policy. Standard model selection methods such as AIC/BIC do not account for this. Our method specifically optimizes bias-variance trade-off for the function of interest.

In the sequel, a localization structure is used to formalize the bias-variance trade-off when estimating a policy. The resulting methods are a version of the focused information criterion (FIC). The FIC was first proposed by Claeskens and Hjort (2003) in the likelihood case. Since then, it has been further developed for regression cases. Our approach is most closely related to Liu (2014). Asymptotic theory is derived for an arbitrary MR and policy rule. The portfolio choice setup considered below is a special case.

In order to implement FIC model selection, a subset of the regression coefficients are assumed local-to-zero. Specifically, coefficients have the form δ/\sqrt{T} where T is the number of observed periods and δ is an arbitrary constant. Mean-variance portfolio optimization also depends on the covariance matrix of shocks in the MR specification. A local-to-zero structure on the off-diagonal components of this covariance matrix is also assumed.

If we set a local-to-zero parameter of the MR model exactly to zero, this produces a submodel

where that parameter is fixed. By considering subsets of local-to-zero parameters we define a class of submodels: the class where given subsets of parameters are fixed at zero. Defining a class of submodels is the starting point for the FIC. Which submodels are considered is arbitrary and user determined.

The first contribution of this paper is to derive asymptotic distributions of OLS estimators of MR submodels assuming the local-to-zero parameter structure described above. In the class of submodels whose distributions are characterized, both regression parameters and error covariance matrix components can be fixed at zero. We derive the joint asymptotic distribution of submodel estimators for both sets of parameters. This extends previous FIC regression results which do not consider MR models. Previous work has not considered error covariance matrix estimates or localization of covariance terms.

Because of the \sqrt{T} localization, submodel estimators are asymptotically normal with non-zero mean and variance. The non-zero mean results from bias introduced by setting local-to-zero parameters exactly to zero. Parameter estimates are plugged into policy functions when making decisions. Parameters fixed at zero must also be plugged in. Both the estimated and fixed parameters influence the asymptotic distribution of the policy. Estimated parameters influence both bias and variance. Fixed parameters contribute only to bias. The final result is an asymptotically normal estimator of the policy with non-zero mean and variance.

The distributions described above formalize the bias-variance trade-off of submodel estimation. Different submodels and policy functions result in different mean and variance terms. Once these distributions are determined they can be used to rank submodels. In this paper, we rank submodels by the asymptotic mean square error (AMSE) of their estimated policy. Other loss functions are also possible. Estimated submodel AMSEs give the FIC ranking.

One possibility is to choose the single submodel with the best FIC ranking. However, this can be suboptimal. It may be possible to decrease AMSE by averaging submodels. In what follows, we average submodels to minimize AMSE as in Liu (2014). A closed form expression for ASME of policy function estimates is derived for the model averaging case. This is used to construct an estimator with optimal weights. There have been a number of recent papers showing the desirable properties of model averaging. Clark and McCracken (2009) average models for forecasting using a similar setup as this paper. Timmermann (2006) surveys the forecast combination literature. See Hansen (2007); Geweke and Amisano (2011); Liang, Zou, Wan and Zhang (2011) and Elliot,

Gargano and Timmermann (2013, 2015) for more examples and citations.

Simulation studies were conducted to compare the performance of the proposed FIC model averaging methodology to standard approaches. This method is compared with policies estimated with OLS, models selected by AIC/BIC, an equally-weighted model averaging scheme and single model selection using the FIC. FIC model averaging emerges as the most robust method, consistently ranked as the best or second-best across various scenarios.

An empirical application of these methods to futures markets was also conducted. Trading in several sets of commodity futures was considered using FIC model averaging. Optimal portfolios derived in Garleanu and Pedersen (2013) were estimated with rolling windows. The results show the proposed method significantly outperforms a set of alternatives in both Sharpe ratio and absolute returns.

The paper is organized as follows. Section 2 presents the model, defines submodels and describes basic estimators. Section 3 describes the localization assumptions used for model selection. Asymptotic distributions of submodel estimators of focus functions are then derived. Section 4 shows how the results from Section 3 can be used for FIC model selection. FIC model averaging estimators and their properties are then presented. Finally, a plug-in model averaging estimator is proposed. Section 5 presents simulations examining the performance of the plug-in model averaging estimator against several competitors. Section 6 presents an empirical application to commodity futures data. Section 7 concludes.

2 The model

Consider an M -dimensional predictive MR model:

$$y_{t+1} = X_t' \beta + Z_t' \gamma + \epsilon_{t+1},$$

defined with the following matrices:

$$\begin{aligned} y_{t+1}^{M \times 1} &= \begin{bmatrix} y_{1t+1} \\ \vdots \\ y_{Mt+1} \end{bmatrix}, \quad X_t^{P \times M} = \begin{bmatrix} x_{1t} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & x_{Mt} \end{bmatrix}, \quad Z_t^{K \times M} = \begin{bmatrix} z_{1t} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & z_{Mt} \end{bmatrix} \\ \epsilon_{t+1}^{M \times 1} &= \begin{bmatrix} \epsilon_{1t+1} \\ \vdots \\ \epsilon_{Mt+1} \end{bmatrix}, \quad \beta^{P \times 1} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_M \end{bmatrix}, \quad \gamma^{K \times 1} = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_M \end{bmatrix}. \end{aligned} \quad (2.1)$$

In this equation, covariates are divided into core X_t and auxiliary Z_t . For equation i , x_{it} is a $p_i \times 1$ vector of core variables. Similarly, z_{it} is a $k_i \times 1$ vector of auxiliary variables. The total number of core and auxiliary regressors in all equations are $P = \sum_{i=1}^M p_i$, $K = \sum_{i=1}^M k_i$. Core variables will always be included in estimation. The researcher is unsure which auxiliary covariates should be included. Our model selection and averaging procedures will decide this while focusing on a policy function of interest.

Further, let

$$H_t = \begin{bmatrix} X_t \\ Z_t \end{bmatrix}, \quad \theta = \begin{bmatrix} \beta \\ \gamma \end{bmatrix} \quad (2.2)$$

and

$$y_{t+1} = H_t' \theta + \epsilon_{t+1}. \quad (2.3)$$

Finally, we stack observations with the following matrices

$$\begin{aligned} y &= \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}, \quad X = \begin{bmatrix} X'_0 \\ \vdots \\ X'_{T-1} \end{bmatrix}, \quad Z = \begin{bmatrix} Z'_0 \\ \vdots \\ Z'_{T-1} \end{bmatrix}, \\ \epsilon &= \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_T \end{bmatrix}, \quad H = \begin{bmatrix} X & Z \end{bmatrix}. \end{aligned} \quad (2.4)$$

All observations of (2.3) can be compactly written as:

$$\begin{aligned} y &= X\beta + Z\gamma + \epsilon, \\ &= H\theta + \epsilon. \end{aligned} \quad (2.5)$$

The portfolio choice rule considered in the sequel depends on the covariance matrix of ϵ_t ($cov(\epsilon_t) = \Omega$). Therefore, an estimator of this covariance matrix is required. All parameters in Ω are classified as either core or auxiliary. It is also possible to *a priori* restrict some covariances to zero. We rule this out for simplicity. The elements of Ω are stacked into a vector Θ for use in asymptotic results that follow. Core elements (which must contain all variances) are stacked first, followed by auxiliary covariances:

$$\Theta = \begin{bmatrix} \Phi \\ \Upsilon \end{bmatrix} \quad (2.6)$$

The notation G is used for the number of core elements and N for the number of auxiliary. The order of these parameters is not important. As with the regression coefficients, this division of parameters is used in model selection and averaging procedures described in the sequel.

2.1 Defining submodels

In order to conduct model selection and averaging, we first define submodels. Submodels are formed by taking the full model and restricting some subset of (γ, Υ) to zero. In the sequel, we consider an arbitrary set of submodels. In applications, those considered are user determined. These could

come from economic theory, computational cost restrictions or other considerations. As a starting point, we assume a set of \bar{M} distinct submodels. These are taken as given.

The subscript m is used to represent an arbitrary submodel. Let Π_m be the $K_m \times K$ selection matrix that selects the included auxiliary regressors in m . This can be constructed by removing those rows from the K -dimensional identity matrix whose order corresponds to the column order of excluded auxiliary regressors in Z . Let

$$Z_m = Z\Pi_m'$$

be the resulting matrix of retained auxiliary regressors. Submodel m thus includes all core covariates X and a subset of auxiliary covariates Z_m . This results in $P + K_m$ regressors. Define also the selection matrices

$$S_0 = \begin{bmatrix} 0_{P \times K} \\ I_K \end{bmatrix}, \quad S_m = \begin{bmatrix} I_P & 0_{P \times K_m} \\ 0_{K \times P} & \Pi_m' \end{bmatrix}. \quad (2.7)$$

Note that

$$\begin{aligned} H_m &= HS_m = \begin{bmatrix} X & Z_m \end{bmatrix}, \\ \theta_m &= S_m' \theta = \begin{bmatrix} \beta \\ \Pi_m \gamma \end{bmatrix} = \begin{bmatrix} \beta \\ \gamma_m \end{bmatrix}, \end{aligned} \quad (2.8)$$

where θ_m is the subset of regression coefficients θ which are not restricted to zero. These are the parameters that will be estimated when considering submodel m . Also note that the full model corresponds to the case $\Pi_m = I$, and the narrow model to Π_m being empty.

Let Γ_m be an $N_m \times N$ matrix that selects the included error covariances from Υ . Similar to before, this can be constructed by removing the rows from the $N \times N$ identity matrix whose order corresponds to the row order of excluded auxiliary covariances in Υ . The m th submodel's covariances can be represented as

$$\Theta_{mm} = \begin{bmatrix} \Phi_m \\ \Upsilon_{mm} \end{bmatrix} = \begin{bmatrix} \Phi_m \\ \Gamma_m \Upsilon_m \end{bmatrix}$$

where $\Phi_m = \Phi$ and $\Upsilon_m = \Upsilon$ are vectors that denote all core and all auxiliary covariances. We write

$\Theta_m = \begin{bmatrix} \Phi'_m & \Upsilon'_m \end{bmatrix}$. Covariance terms not in Θ_{mm} are set to zero in submodel m . Finally, let

$$F_0 = \begin{bmatrix} 0_{G \times N} \\ I_N \end{bmatrix}, F_m = \begin{bmatrix} I_G & 0_{G \times N_m} \\ 0_{N \times G} & \Gamma'_m \end{bmatrix} \quad (2.9)$$

be selection matrices such that

$$\Theta_{mm} = F'_m \Theta_m = \begin{bmatrix} \Phi_m \\ \Gamma_m \Upsilon_m \end{bmatrix}. \quad (2.10)$$

The full model corresponds to the case $F_m = I$.

In the context of estimation this notation is not redundant. The subscript m represents the restriction on regression parameters (through Π_m) which influences residuals. Residuals influence estimates of Θ . Because of this, in general, different submodels m_1 and m_2 will have $\hat{\Phi}_{m_1} \neq \hat{\Phi}_{m_2}$ and $\hat{\Upsilon}_{m_1} \neq \hat{\Upsilon}_{m_2}$. Υ_{mm} denotes a subset of auxiliary residual covariances Υ_m defined using Γ_m . In going from Υ_m to Υ_{mm} , certain covariance terms from Υ_m are removed. The removed terms are assumed to be zero and are not estimated. In the notation Υ_{mm} , the first subscript m relates to the restriction of parameters through Π_m which alters residuals, the second subscript m to the actual restriction of covariances through Γ_m .

To sum up, a given submodel m is characterized by selection matrices (S_m, F_m) . Parameters selected will be estimated under the submodel restrictions. The remaining parameters are set to zero in estimation.

2.2 OLS estimators

The parameters in a given submodel are estimated with OLS. The estimators for submodel m are

$$\begin{aligned}\hat{\theta}_m &= \begin{bmatrix} \hat{\beta}_m \\ \hat{\gamma}_m \end{bmatrix} = (H'_m H_m)^{-1} H'_m y \\ \hat{\Theta}_{mm} &= \begin{bmatrix} \hat{\Phi}_m \\ \hat{\Upsilon}_{mm} \end{bmatrix} = \begin{bmatrix} \hat{\Phi}_m \\ \Gamma_m \hat{\Upsilon}_m \end{bmatrix} = F'_m \begin{bmatrix} \frac{1}{T} (\hat{\epsilon}'_{im} \hat{\epsilon}_{jm})^{core} \\ \vdots \\ \frac{1}{T} (\hat{\epsilon}'_{km} \hat{\epsilon}_{lm})^{aux} \\ \vdots \end{bmatrix}, \quad i, j, k, l \in \{1, \dots, M\}\end{aligned}\quad (2.11)$$

where $\hat{\epsilon}_{im}$ denotes a vector of residuals related to a dependent variable y_i for submodel m . The coefficients of the excluded variables γ_{m^c} and covariances Υ_{mm^c} (where c denotes complement) are set to zero.

If $\Pi_m = I_K$ and $\Gamma_m = I_N$, the formulas give least squares estimators of the full model. If Π_m and Γ_m are empty, we have estimators of the narrow model where all auxiliary parameters are set to zero. Note that in submodel estimators the order of regressors and of covariances is always preserved by design.

3 Asymptotic framework

Submodels estimated by least squares (2.11) have omitted variable bias. If γ and Υ are nonzero and fixed, the limiting distributions of the parameters restricted to zero (multiplied by \sqrt{T}) will have bias that tends to infinity. This is not useful for understanding submodel estimation in finite samples. In order to have a meaningful asymptotic bias-variance trade-off we use localization. This is a standard requirement for all FIC methods. Throughout the sequel, the data generating process is assumed to follow the full model specification and satisfy Assumption 1 below.

Assumption 1 $\gamma = \delta/\sqrt{T}$, $\Upsilon = \Delta/\sqrt{T}$, where δ and Δ are constant vectors. β and Φ are constant vectors.

As we show below, Assumption 1 ensures that submodel estimators are asymptotically normal with finite bias. Full model estimators are asymptotically unbiased. Submodel estimation also changes asymptotic covariance compared with the full model. This results in a nontrivial bias-

To proceed, let $i, j, k, l \in \{1, \dots, M\}$ and define:

$$\lambda_t = \begin{bmatrix} \begin{bmatrix} (\epsilon_{it}\epsilon_{jt} - \omega_{ij})^{core} \\ \vdots \\ G \times 1 \end{bmatrix} \\ \begin{bmatrix} (\epsilon_{kt}\epsilon_{lt} - \omega_{kl})^{aux} \\ \vdots \\ N \times 1 \end{bmatrix} \end{bmatrix}, \quad J_T = \sum_{t=1}^T \lambda_t \quad (3.1)$$

where $\omega_{ij} = E[\epsilon_{it}\epsilon_{jt}]$ are the unique elements of $cov(\epsilon_t) = \Omega$. To derive limiting distributions, we make standard high-level assumptions which are outlined in the appendix. Consequences of these assumptions useful for understanding the results that follow are given in Condition 3 below. Assumption 2 is only slightly stronger than Condition 3.

Condition 3 As $T \rightarrow \infty$:

$$\hat{Q} = T^{-1}H'H \xrightarrow{p} Q \quad (3.2)$$

and

$$\begin{bmatrix} \frac{1}{\sqrt{T}}H'\epsilon \\ \frac{1}{\sqrt{T}}J_T \end{bmatrix} \xrightarrow{d} \begin{bmatrix} R \\ J \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

In our setup, we are concerned with estimating the unconditional covariance of the errors $cov(\epsilon_t) = \Omega$. This is a common approach in systematic trading contexts where small amounts of data are used for estimation with rolling windows. The rolling window setup captures changing dynamics of regression coefficients and volatility. However, it is well known that financial data has conditional heteroskedasticity. Many other models, for example GARCH specifications, have these dynamics. Similar results to those in the sequel could be derived with this type of specification. This would require replacing the convergence of $1/\sqrt{T}J_T$ in Condition 3 with corresponding asymptotics for GARCH parameters. From this, an FIC approach could be developed following similar steps. We leave a specific description of this to future research.

Under the above assumptions, the asymptotic distribution of submodel estimators is derived in Theorem A1 of the online supplement. Similarly to Hjort and Claeskens (2003a) and Liu (2014),

these distributions can be used to conduct model selection using a focus function. A focus function is a function of the estimated parameters. For example, in the portfolio choice case considered below, the focus is the portfolio weights. In general, the focus is the object of interest. Given a particular function, the limiting distribution of a plug-in focus estimator based on submodel m can be used as the basis of model selection. This is the idea of the FIC. This program is developed for the MR case below.

3.1 Focus parameter and its limiting distribution

Consider a N_{FIC} -dimensional function of the estimated model parameters (θ, Θ) :

$$\begin{aligned}\mu &: R^{[(P+K)+(G+N)]} \rightarrow R^{N_{FIC}}, \\ \mu &= \mu(\theta, \Theta) = \mu(\beta, \gamma, \Phi, \Upsilon).\end{aligned}\tag{3.3}$$

This is the focus function, the object of interest in estimation. It is assumed to be twice continuously differentiable. Let

$$\hat{\mu}_m = \mu(\hat{\theta}_m, \hat{\Theta}_{mm})\tag{3.4}$$

denote a submodel estimator of μ . Here, estimated parameters are plugged into μ . Elements which are set to zero in submodel m are set to zero in μ . This is suppressed in the notation. Let:

$$D_{\theta, \Theta} = \begin{bmatrix} D_{\theta} \\ D_{\Theta} \end{bmatrix}\tag{3.5}$$

be a matrix of partial derivatives evaluated at the null points $(\beta', 0', \Phi', 0')'$. The null points are the limit points of the localized parameters.

The following theorem derives the asymptotic distribution of $\hat{\mu}_m$. This distribution depends on both the asymptotic distribution of estimated parameters and which parameters are fixed at zero. It is therefore not an immediate result of Theorem A1. See the proof in the online supplement for more details.

Theorem 1. Suppose that Assumptions 1-2 hold. As $T \rightarrow \infty$, we have

$$\begin{aligned} \sqrt{T} \left(\mu \left(\hat{\theta}_m, \hat{\Theta}_{mm} \right) - \mu \left(\theta, \Theta \right) \right) &\xrightarrow{d} \Lambda_m \\ &= \begin{bmatrix} D_\theta \\ D_\Theta \end{bmatrix}' \begin{bmatrix} C_m \delta + P_m R \\ L_m \Delta + U_m J \end{bmatrix} \\ &\sim \begin{bmatrix} D_\theta \\ D_\Theta \end{bmatrix}' N \left(\begin{bmatrix} C_m \delta \\ L_m \Delta \end{bmatrix}, \begin{bmatrix} P_m \Sigma_{11} P_m & P_m \Sigma_{12} U_m \\ U_m \Sigma_{21} P_m & U_m \Sigma_{22} U_m \end{bmatrix} \right) \end{aligned} \quad (3.6)$$

where C_m , P_m , L_m and U_m are matrices related to submodel m :

$$\begin{aligned} C_m &= (P_m Q - I) S_0 \\ P_m &= S_m Q_m^{-1} S_m' = S_m \left(S_m' Q S_m \right)^{-1} S_m' \\ L_m &= -F_0 \left(I - \Gamma_m' \Gamma_m \right) \\ U_m &= F_m F_m' \end{aligned}$$

Theorem 1 shows joint convergence of all submodels as any Λ_m can be expressed in terms of R and J . The size of the bias and variance in Λ_m are determined by $D_{\theta, \Theta}$ as well as δ , Δ , C_m , P_m , L_m and U_m . The parameters C_m , P_m , L_m and U_m depend on the submodel m and derivatives $D_{\theta, \Theta}$ depend on the focus μ . As a result, different submodels and focus functions result in a different mean and variance for Λ_m . The difference in distributions Λ_m formalizes bias-variance trade-off when estimating $\mu(\theta, \Theta)$ with different m . This gives the basis for a model selection criteria when estimating $\mu(\theta, \Theta)$. We now describe model selection and averaging using Λ_m .

4 FIC and Model Averaging

From Theorem 1, a loss metric for estimating $\mu(\theta, \Theta)$ using submodel m can be defined using Λ_m . We choose this loss metric to be the sum of asymptotic mean square errors (AMSE) of the N_{FIC} focus parameters:

$$SAMSE(\hat{\mu}_m) = \sum_{i=1}^{N_{FIC}} AMSE(\hat{\mu}_m^i) \quad (4.1)$$

where $\hat{\mu}_m^i$ is an element of the vector $\hat{\mu}_m$. The goal is to choose the submodel estimator $\hat{\mu}_m$ which minimizes SAMSE. This metric was chosen for simplicity. Another possible choice would be $AMSE\left(\sum_{i=1}^{N_{FIC}} \hat{\mu}_m^i\right)$. This would involve accounting for cross terms.

To express $AMSE(\hat{\mu}_m^i)$, we define the $(1 \times N_{FIC})$ vector S^i that selects $\hat{\mu}_m^i$ from $\hat{\mu}_m$. Consequently, from Theorem 1

$$\sqrt{T}\left(\mu^i(\hat{\theta}_m, \hat{\Theta}_{mm}) - \mu^i(\theta, \Theta)\right) = S^i \sqrt{T}\left(\mu(\hat{\theta}_m, \hat{\Theta}_{mm}) - \mu(\theta, \Theta)\right), \quad (4.2)$$

and

$$\begin{aligned} AMSE(\hat{\mu}_m^i) = & S^i \left(\begin{bmatrix} D'_\theta & D'_\Theta \end{bmatrix} \begin{bmatrix} C_m \delta \delta' C'_m & C_m \delta \Delta' L'_m \\ L_m \Delta \delta' C'_m & L_m \Delta \Delta' L'_m \end{bmatrix} \begin{bmatrix} D_\theta \\ D_\Theta \end{bmatrix} \right) S^{i'} \\ & + S^i \left(\begin{bmatrix} D'_\theta & D'_\Theta \end{bmatrix} \begin{bmatrix} P_m \Sigma_{11} P_m & P_m \Sigma_{12} U_m \\ U_m \Sigma_{21} P_m & U_m \Sigma_{22} U_m \end{bmatrix} \begin{bmatrix} D_\theta \\ D_\Theta \end{bmatrix} \right) S^{i'}. \end{aligned} \quad (4.3)$$

FIC estimates SAMSE for different m s and selects the submodel with the smallest value. Consistent estimation of SAMSE is not possible as there are no consistent estimators of δ and Δ . This is due to the local asymptotic framework. However, it is possible to construct asymptotically unbiased estimators. This is done using a plug-in method with (4.3). Lemma A2 in the supplemental materials gives details.

4.1 Plug-in averaging estimator

Above we considered SAMSE of single submodels individually. However, the estimator obtained by weighting (averaging) across $\hat{\mu}_m$ for different submodels m can improve SAMSE compared with the best single submodel. This is because the set of convex combinations of $\hat{\mu}_m$ always contains each individual $\hat{\mu}_m$. A model averaging estimator can improve on individual $\hat{\mu}_m$ if the weights are chosen to minimize SAMSE. We now describe such an estimator rigorously.

Let

$$w = \begin{bmatrix} w_1 & \dots & w_{\bar{M}} \end{bmatrix}', \quad w_m \geq 0, \quad \sum_{m=1}^{\bar{M}} w_m = 1 \quad (4.4)$$

be a weight vector, where \bar{M} is the number of submodels. The weights are assumed to be positive

and sum to one. The model averaging estimator of μ is

$$\bar{\mu}(w) = \sum_{m=1}^{\bar{M}} w_m \hat{\mu}_m. \quad (4.5)$$

Theorem A3 in the supplemental material gives the asymptotic distribution of $\bar{\mu}(w)$. This distribution implies that SAMSE of the averaging estimator $\bar{\mu}(w)$ is

$$SAMSE(\bar{\mu}(w)) = \sum_{i=1}^{N_{FIC}} AMSE(\bar{\mu}^i) = w' \psi w \quad (4.6)$$

where $\bar{\mu}^i$ is i th element of $\bar{\mu}$. ψ is defined as:

$$\psi = \sum_{i=1}^{N_{FIC}} \psi^i \quad (4.7)$$

where ψ^i are $\bar{M} \times \bar{M}$ matrices corresponding to μ^i . The (m, l) th element of ψ^i is given by

$$\begin{aligned} \psi_{m,l}^i = & S^i \left(\begin{bmatrix} D'_\theta & D'_\Theta \end{bmatrix} \begin{bmatrix} C_m \delta \delta' C'_l & C_m \delta \Delta' L'_l \\ L_m \Delta \delta' C'_l & L_m \Delta \Delta' L'_l \end{bmatrix} \begin{bmatrix} D_\theta \\ D_\Theta \end{bmatrix} \right) S^{i'} \\ & + S^i \left(\begin{bmatrix} D'_\theta & D'_\Theta \end{bmatrix} \begin{bmatrix} P_m \Sigma_{11} P_l & P_m \Sigma_{12} U_l \\ U_m \Sigma_{21} P_l & U_m \Sigma_{22} U_l \end{bmatrix} \begin{bmatrix} D_\theta \\ D_\Theta \end{bmatrix} \right) S^{i'} \end{aligned} \quad (4.8)$$

and therefore $\psi_{m,l} = \sum_{i=1}^{N_{FIC}} \psi_{m,l}^i$. Note that for $m = l$, $\psi_{m,m}^i$ is simply the AMSE for μ^i and submodel m . The optimal weight vector is the value that minimizes

$$\begin{aligned} w^0 &= \arg \min w' \psi w \\ \text{s.t. } & \sum_{m=1}^{\bar{M}} w_m = 1, \quad w_m \geq 0 \end{aligned} \quad (4.9)$$

The estimator $\bar{\mu}(w^0)$ weakly dominates individual submodel estimators with respect to SAMSE.

There is no closed form solution to (4.9) when more than two submodels are present. Fortunately, the optimal weight vector can be found numerically via quadratic programming algorithms. The optimal weights are infeasible because they depend on unknown parameters δ , Δ , D_θ , D_Θ , C_m , P_m , Σ_{11} , Σ_{12} and Σ_{22} . A straightforward solution is to use sample counterparts as described in Lemma A2. This leads to data-dependent weights obtained by minimizing the sample analog

of SAMSE (i.e. minimizing $w' \hat{\psi} w$ where ψ is the sample analogue of ψ). The plug-in estimator is defined as:

$$\begin{aligned}\bar{\mu}(\hat{w}) &= \sum_{m=1}^{\bar{M}} \hat{w}_m \hat{\mu}_m \\ \hat{w} &= \arg \min w' \hat{\psi} w \\ \text{s.t. } &\sum_{m=1}^{\bar{M}} w_m = 1, w_m \geq 0\end{aligned}\tag{4.10}$$

Appendix B in the online supplement discusses the limiting distribution of $\bar{\mu}(\hat{w})$ and how it can be used for inference. The following section presents simulations and an application of this estimator to a portfolio choice problem.

5 Simulations

5.1 Optimal portfolio as a focus function

The general MR model selection and averaging theory developed in this paper was motivated by a portfolio choice problem considered in Garleanu and Pedersen (2013) (hereafter GP). GP assume asset returns follow the system:

$$\begin{aligned}r_{t+1} &= B f_t + u_{t+1} \\ f_{t+1} &= (I - C) f_t + \varepsilon_{t+1}\end{aligned}\tag{5.1}$$

with

$$\begin{aligned}E_t(u_{t+1}) &= 0, \quad \text{var}_t(u_{t+1}) = \Omega_u, \quad \text{cov}_t(u_{t+1}, \varepsilon_{t+1}) = \Omega_{u\varepsilon} \\ E_t(\varepsilon_{t+1}) &= 0, \quad \text{var}_t(\varepsilon_{t+1}) = \Omega_\varepsilon\end{aligned}\tag{5.2}$$

where r_t is a $S_1 \times 1$ vector of asset returns and f_t is a $S_2 \times 1$ vector of driving factors. This is a restricted VAR(1) specification which fits the setup of Section 2. In this case, r_{t+1} takes the place of y_{t+1} in the current notation.

In this setup, x_t is the number of shares held of the corresponding assets in r_t . In addition, r_t is defined as price changes $r_t = p_t - p_{t-1}$. This is in contrast to much of the portfolio choice literature where x_t would be portfolio weights (i.e. proportions of wealth invested in each asset) and $r_t = (p_t - p_{t-1}) / p_{t-1}$. This is needed so quadratic transaction costs can be used.

Given this specification, an investor decides which vector of asset shares x_t to hold this period

The investor chooses x_t to maximize a mean-variance portfolio problem which penalizes for risk and transaction costs (assumed to be quadratic i.e. $TC(\Delta x_t) = \frac{1}{2} \Delta x_t' \Lambda \Delta x_t$ where Λ is symmetric positive-definite). In particular, they maximize:

$$\max_{x_t} E_t \left[x_t' r_{t+1} - \frac{\gamma}{2} x_t' \Omega_u x_t - \frac{1}{2} \Delta x_t' \Lambda \Delta x_t \right] \quad (5.3)$$

where $\gamma > 0$ is a risk aversion coefficient. If transaction costs are proportional to the amount of risk ($\Lambda = \lambda \Omega_u$), GP show the solution to (5.3) is:

$$x_t = \left(1 - \frac{\gamma}{\gamma + \lambda}\right) x_{t-1} + \left(\frac{\gamma}{\gamma + \lambda}\right) Markowitz_t, \quad (5.4)$$

where

$$Markowitz_t = (\gamma \Omega_u)^{-1} B f_t. \quad (5.5)$$

The optimal policy x_t is characterized by trading partially toward the $Markowitz_t$ portfolio and holding some proportion of the current portfolio x_{t-1} . See GP Example 3 for more discussion.

The problem of estimating x_t can be naturally cast in the FIC framework developed in Section 2. In this case, the focus function μ is the vector of portfolio shares x_t and parameters θ and Θ correspond to B and Ω_u . Uncertainty about B corresponds to the situation where an investor wants to implement x_t , but is unsure which potential factors f_t to include. Uncertainty about Ω_u corresponds to the problem of estimating covariance between assets, possibly setting some terms to zero to avoid estimation error. The outlined regressions are usually estimated with relatively small sample sizes to deal with model instability. As a result, we have a standard bias-variance trade-off interpretation: adding more factors allows for more complicated dynamics, but small sample sizes mean this significantly increases estimator variance.

A number of studies have concluded that simple trading strategies can dominate more complicated ones. This is because of estimation error from trying to recover complicated asset dynamics from small samples (see DeMiguel, Nogales and Uppal (2014) and DeMiguel, Garlappi and Uppal (2009)). Ideally, an estimation method would explicitly account for increased variance in more complicated models and increased bias in simpler ones. The best model would be chosen by optimizing the trade-off between these competing errors. The trade-off should account for the final object

of interest, the portfolio trading rule x_t . The focused information criterion and model averaging methods developed above were designed with this in mind. We now apply these methods to the portfolio choice situation in simulations and an empirical application.

5.2 Simulation specification and results

We illustrate the performance of our methods in simulations designed to mimic our application. A portfolio of five assets is considered. The portfolio size is restricted by computational cost. Solving the optimization problem for submodel weights in FIC model averaging is the main constraint. Our setup outlined below corresponds to 2048 submodels. Even this moderately sized case represents a considerable computational burden when thousands of simulations are needed. Simulations took approximately a week on a 60 core computing cluster. Analyzing systems with a larger number of assets and factors is an important problem. Presently, we focus on a simple and realistic case to provide insight into the method's performance. Our simulations are described as follows:

1. Portfolio shares are estimated using different amounts of data (window size). The values $T = \{100, 150, 200, 250\}$ are considered. Small samples sizes such as these are standard in practice.
2. First, we simulate factors of sample size T . Each asset r_{it+1} is assumed to have two factors: f_{it}^s , $s = 1, 2$. These represent short-term (f_{it}^1) and long-term (f_{it}^2) information, short-term being less persistent. The factors are unique across assets and follow AR(1) processes:

$$f_{it+1}^s = (1 - c_i^s) f_{it}^s + \varepsilon_{it+1}^s.$$

The errors follow $\varepsilon_{it+1}^s \sim N(0, \text{var}(\varepsilon_{it+1}^s))$. Parameters c_i^1 , c_i^2 and $\text{var}(\varepsilon_{it+1}^s)$ are assumed equal for all $i \in \{1, \dots, 5\}$. The values of these parameters are reported in Table 1.

3. We use the simulated factors to obtain a random sample of returns that follow a specific version of (5.1):

$$r_{it+1} = b_{1i} f_{it}^1 + b_{2i} f_{it}^2 + u_{it+1}$$

where $u_{it+1} \sim N(0, \Omega_{u,ii})$ and the coefficients b_{1i} , b_{2i} are equal for all i . In addition, returns are assumed normalized by their unconditional standard deviation. This is why 0.9 is chosen

4. We consider 6 different scenarios for pairs b_{1i}, b_{2i} that are reported in Table 1. These correspond to different degrees of predictability of returns. All pairs are chosen so the predictive mean is significantly smaller than the error variance. The signal-to-noise ratio has an approximate maxima of $1/3$ over all specifications. This is the empirically relevant situation.
5. We apply FIC and FIC model averaging (denoted FICav) to estimate x_t . This requires estimating b_{1i}, b_{2i} and Ω_u . These parameters correspond to θ and Θ described in section 2.
6. To implement FIC, we must specify which variables are core and auxiliary. It is assumed all b_{1i}, b_{2i} and all covariance terms in Ω_u are auxiliary. For b_{1i}, b_{2i} , submodels are formed from all subsets of these values in each of the 5 return equations. For Ω_u , submodels are formed from two specifications: (1) Ω_u is unconstrained and (2) Ω_u is diagonal. More complicated submodels are not considered for computational reasons.
7. Estimation with FIC also requires derivatives of μ with respect to parameters. Using the closed-form expression for x_t (5.4) the required derivatives are:

$$\begin{aligned}\frac{\partial x_t}{\partial \Omega_{u,ij}} &= -\left(\frac{1}{\gamma+\lambda}\right) \left[\Omega_u^{-1} \frac{\partial \Omega_u}{\partial \Omega_{u,ij}} \Omega_u^{-1} \right] B f_t \\ \frac{\partial x_t}{\partial b_{ji}} &= \left(\frac{1}{\gamma+\lambda}\right) \Omega_u^{-1} \frac{\partial B}{\partial b_{ji}} f_t\end{aligned}\tag{5.6}$$

The factors f_t used in the focus are the final period values in the simulated data. Therefore, f_t is random across simulations. This better represents rolling window trading strategies implemented in applications.

8. Computing the focus function - portfolio shares (5.4) - requires hyper-parameters. Following GP, the absolute risk aversion is set to $\gamma = 10^{-9}$, which is considered reasonable for a large asset manager. The transaction cost parameter is $\lambda = 10^{-6}$, a number viewed as conservative.
9. A number of competitors are also computed for comparison: (1) full-model OLS (OLS); (2) AIC and BIC selected models (AIC and BIC); (3) to see if there is value added in optimizing submodel weights, final portfolios are estimated by equally weighting the portfolios from all submodels (EW); (4) portfolios are estimated with FIC but no model averaging (FIC).

In all scenarios	$c_i^1 = 0.5$	$c_i^2 = 0.1$
	$\Omega_{u,ii} = 0.9$	$\Omega_{u,ij} = 0.4, i \neq j$
	$var(\varepsilon_{it+1}^1) = 20$	$var(\varepsilon_{it+1}^2) = 5$
	$\Omega_{\varepsilon,ij} = 4, i \neq j$	$\Omega_{u\varepsilon,ij} = 0$
	b_{1i}	b_{2i}
Scenario 1	0.01	-0.01
Scenario 2	0.01	-0.005
Scenario 3	0.01	-0.001
Scenario 4	0	-0.01
Scenario 5	0	-0.005
Scenario 6	0	-0.001

Table 1: Simulated Parameters

Each simulation is repeated 500 times. The average SMSE for the estimated portfolios relative to the optimal portfolios is our performance metric. The results are presented in Table 2.

The consistently best performers over all simulations are FICav and EW. Both methods have good results in all specifications. For smaller window sizes $T = 100, 150$, FICav clearly outperforms EW. For larger sizes $T = 200, 250$ the results are more mixed. EW tends to perform better when return coefficients are larger and predictability is higher. When predictability is lower, FICav is best. When predictability is lowest, FICav beats all methods by a significant margin.

These results suggest FICav does well when there is uncertainty about the relevance of factors given the amount of data available. With more data, there is less uncertainty with higher coefficients. When the coefficients are lower, the question is less resolved. In more complicated models with a larger number of covariates, more data will be needed to resolve this uncertainty.

EW is similar to equally weighted forecast combinations. This approach often does well in forecasting. See Timmermann (2006) for a survey and Claeskens, Magnus, Vasnev and Wang (2016) for a theoretical explanation. It is interesting we find strong performance for EW in our portfolio choice case.

In daily trading, a larger number of observations such as $T = 200, 250$ will usually be used. In these cases, we expect predictability to be low. With weekly or monthly data, predictability is higher, but our window size will often be lower. This would be in keeping with $T = 100, 150$. These two situations correspond to the cases where FICav performs best compared to competitors. Therefore, our simulations suggest that FICav is likely to perform well in many systematic trading contexts. This is what we find in our empirical application.

		FIC _{av}	EW	OLS	AIC	BIC	FIC
$T = 100$							
b_{1i}	b_{2i}						
0.01	-0.01	4.5302	6.3279	34.2575	17.4270	15.6629	24.5871
0.01	-0.005	3.6701	4.3433	36.9764	16.0701	14.7769	23.9436
0.01	-0.001	3.4690	4.7763	33.0678	17.6854	15.3047	21.0974
0	-0.01	1.0732	3.2116	34.2426	11.4254	10.4932	18.3754
0	-0.005	0.3010	2.8001	33.4551	12.8679	10.7594	18.1785
0	-0.001	0.0780	2.8566	36.3061	12.2589	10.7077	17.1469
$T = 150$							
b_{1i}	b_{2i}						
0.01	-0.01	4.3333	5.2271	21.4836	10.7591	9.9172	15.2787
0.01	-0.005	3.5131	3.4910	22.4661	11.8565	10.3562	16.3633
0.01	-0.001	3.5139	3.2722	21.0513	9.7811	8.7249	15.2467
0	-0.01	1.0393	2.2289	21.1134	8.8578	8.1197	13.1943
0	-0.005	0.2728	1.6981	21.0552	7.5322	7.0410	11.8398
0	-0.001	0.0530	1.7006	20.8802	8.4218	7.2766	12.5089
$T = 200$							
b_{1i}	b_{2i}						
0.01	-0.01	4.0787	4.0548	13.9658	8.8504	8.0542	12.5504
0.01	-0.005	3.6966	3.4619	12.9879	7.7561	6.9376	11.6288
0.01	-0.001	3.3242	2.6768	15.0313	7.9720	7.7051	11.1033
0	-0.01	1.0003	1.8802	12.6276	6.1816	5.4085	8.8025
0	-0.005	0.2666	1.3682	16.1418	7.1397	6.1273	9.4356
0	-0.001	0.0399	1.2818	15.3056	6.3143	5.0851	8.8245
$T = 250$							
b_{1i}	b_{2i}						
0.01	-0.01	4.2335	4.0230	11.3642	7.2625	6.8187	11.4659
0.01	-0.005	3.4736	2.9863	10.7567	6.9468	6.6358	8.7341
0.01	-0.001	3.3656	2.3416	13.1793	7.2718	6.5057	11.0019
0	-0.01	1.0147	1.6919	11.9479	6.6895	5.7949	9.1400
0	-0.005	0.2737	0.9149	11.9601	4.6237	3.9844	6.2438
0	-0.001	0.0289	0.8120	11.4313	4.2774	3.7218	7.1031

Table 2: Simulated Average Portfolio Sum of Mean Squared Errors, 10^{10}

6 Commodity Futures Trading Application

In this section, FIC model averaging is applied to commodity futures trading. Fifteen highly liquid commodity futures contracts are considered: Soybean Oil (BO), Crude Oil (CL), Corn (C_), Gold

(GC), Goldman Sachs Commodity Index (GI), Copper (HG), Orange Juice (JO), Coffee (KC), Lumber (LB), Live Cattle (LC), Live Hogs (LH), Natural Gas (NG), Rough Rice (NR), Platinum (PL) and Silver (SI). Daily data on prices is taken from the Pinnacle continuous futures database. The dates considered are Jan 3, 2012 - July 1, 2016. All prices are in US dollars

All prices are first differenced then standardized by their sample standard deviation. This is standard in practice. Next, two predictive factors were constructed for each asset. These are the previous five day and twenty day averages of the standardized differenced prices. This setup is able to capture short term momentum in prices with a longer term reversal. These choices mimic our simulations and are a common approach in practice. Many other choices for factors are also possible, but are not explored here.

Twelve different portfolios were considered, each consisting of five assets drawn from the pool of fifteen described above. These twelve portfolios are presented in Table 3. Different portfolios are considered to explore the stability of the estimation approach and trading results. For each portfolio, rolling window estimation and trading was conducted over the sample period. The window size used was $T = 200$ observations. The estimated model was predictive regressions of the standardized price changes with the relevant factors. All hyper-parameter and submodel choices are the same as in simulations.

Transaction costs were assumed to have the quadratic form described in Section 5. As in GP, the covariance matrix used to compute transaction costs was the estimated residual covariance matrix for the entire observed sample. This was computed using non-normalized prices. In real trading, transaction costs would be observed, not computed as in this stylized setup.

Trading is as in GP. An agent starts with no shares in any of the assets. Each period, she trades to hold the estimated optimal portfolio. Profits are computed from price changes minus transaction costs. Sharpe ratios and total profits are then derived. These are our performance measures. Note that our trading is out-of-sample whereas GP considers in-sample results.

Total profits in millions and annualized Sharpe ratios from all twelve portfolios are presented in Table 3. Averages over all portfolios for each estimator are also reported. Estimation took approximately five days on a 60 core computing cluster. However, estimating a single portfolio position takes approximately 30 minutes. Implementing daily trading is clearly feasible.

In keeping with our simulations, the FICav and EW estimators are by far the best. All other methods lose significant amounts of money and have negative Sharpe ratios on average. The FICav

estimator has both a positive average profit (112.2) and average Sharpe ratio (0.2729). EW has a small average Sharpe ratio (0.0059) and significantly negative average profit (-575.3).

FICav takes significantly smaller positions than the other estimators. This can be seen from profits. Other estimators take large positions which yield high profits if they pay off. However, they frequently result in large losses. FICav takes smaller positions and its profits and losses are relatively small. To illustrate this, for Portfolio 1, Figures 7.1 and 7.2 show cumulative profits and Figures 7.3 and 7.4 portfolio positions for FICav and EW over the trading period. The positions of EW are much larger than FICav. Their respective largest values are approximately five million and one million shares. For this portfolio, EW results in larger profits but a much smaller Sharpe ratio. This is partially a consequence of larger volatility from larger positions and position changes.

FICav only loses money in three of the twelve portfolios and losses are small. In particular, they are small compared to profits from FICav in other portfolios. The three losses are -73.6, -16.8 and -1.1 whereas its three largest profits are 206.1, 286.8 and 457.5. EW has losses in half the portfolios and they can be large. The three largest losses from EW are -5,751.8, -5,472.5 and -1,197.8 whereas its three largest profits are 2,547.6, 1,540.4 and 991.8. On balance, FICav is the most robust procedure with the best results. It has the highest average Sharpe ratio by a significant margin. It is the only procedure that profits on average over all portfolios.

OLS, AIC, BIC and FIC all have very poor results on average. To explore the robustness of this finding, we also estimate trading strategies for OLS, AIC, BIC and EW with expanding windows. FIC and FICav procedures are designed for small samples sizes and therefore these estimates are not computed. The results for OLS, AIC and BIC are greatly improved. However, all still have significantly negative average profits and Sharpe ratios. EW improves as well. Its average Sharpe ratio (0.0782) is larger and average profits less negative (-124.1). These values are still significantly worse than FICav with rolling windows.

The results from expanding window estimation are not surprising. The additional data means more stable estimators, but they are not accurate enough to produce large profits or Sharpe ratios on average. The underlying asset dynamics in our sample are likely changing through time. If the regression coefficients and error covariance structure are changing, expanding window estimates will converge to an average of these parameters. Rolling windows can potentially capture the changing structure and profit from it. Our results suggest FICav is able to do this.

We make no claim that our setup is optimal. However, the results show that FICav improves

7 Conclusion

In this paper, model averaging estimation of an MR using the FIC was considered. Regression coefficients and error covariance terms were assumed to be localized. This allowed for model selection and averaging focused on particular functions of the underlying parameters. The FIC approach chooses the submodel which optimizes bias-variance trade-off when estimating a focus function. Model averaging takes the FIC idea and averages over submodels in order to improve AMSE. This is a rigorous way of accounting for uncertainty about which aspects of the model should be included to minimize estimation error. The model averaging estimator was shown in simulations to outperform several competing methods. An application to futures data showed similar improvements.

		FICav	EW	OLS	AIC	BIC	FIC
Portfolio 1	SR	1.0342	0.5557	0.4402	0.0441	0.1115	0.7564
JO,GI,KC,LH,NG	P	206.1	621.3	3,754.8	186.7	501.1	2,135.8
Portfolio 2	SR	0.2936	0.9189	-0.2766	-0.2721	-0.2800	0.7296
GI,KC,LH,NG,NR	P	102.5	2,547.6	-2,630.9	-1,274.1	-1,268.8	3,232.8
Portfolio 3	SR	-0.1870	0.4816	-1.2718	-0.7025	-0.6611	-0.8151
KC,LH,NG,NR,PL	P	-73.6	991.8	-8,631.7	-2,662.2	-2,201.4	-3,065.7
Portfolio 4	SR	0.6677	0.2322	-2.7551	-3.7402	-3.9668	-3.6116
LH,NG,NR,PL,SI	P	204.6	1,540.4	-31,730.2	-22,222.0	-19,295.1	-20,856.1
Portfolio 5	SR	0.0704	-0.8377	-2.4394	-2.8137	-3.4324	-3.3537
NG,NR,PL,SI,GC	P	27.1	-5,472.5	-50,279.3	-34,459.9	-35,213.2	-36,227.2
Portfolio 6	SR	0.5541	-0.0584	-2.6477	-2.3153	-3.1962	-4.2987
NR,PL,SI,GC,C_	P	457.5	-203.9	-55,454.1	-39,242.2	-36,904.2	-34,437.7
Portfolio 7	SR	0.5217	-0.1010	-2.3199	-2.6520	-2.5933	-2.7356
PL,SI,GC,C_,BO	P	286.8	-429.2	-53,527.3	-38,604.2	-35,863.3	-33,880.3
Portfolio 8	SR	0.2552	-0.9061	-2.5898	-1.6496	-2.6866	-3.1087
SI,GC,C_,BO,HG	P	125.7	-5,751.8	-61,228.4	-20,815.2	-23,463.3	-38,867.9
Portfolio 9	SR	0.1273	-0.8251	-0.1509	-0.8352	-1.1462	-0.7846
GC,C_,BO,HG,CL	P	12.4	-1,197.9	-716.3	-1,683.0	-1,998.0	-2,529.9
Portfolio 10	SR	-0.3113	-0.0809	-0.1863	-0.6080	-1.0001	-0.5645
C_,BO,HG,CL,LC	P	-16.8	-99.2	-712.2	-984.9	-1,172.1	-1,923.0
Portfolio 11	SR	-0.0382	0.3402	-0.1308	-0.1076	-0.2884	0.7154
BO,HG,CL,LC,LB	P	-1.1	167.8	-204.1	-81.2	-223.9	710.6
Portfolio 12	SR	0.2873	0.3514	0.0030	-0.4915	-0.4410	0.4564
HG,CL,LC,LB,JO	P	14.8	382.2	5.1	-267.3	-311.2	516.4
Averages	SR	0.2729	0.0059	-1.1938	-1.3453	-1.6317	-1.3846
	P	112.2	-575.3	-21779.6	-13509.1	-13117.8	-13766.0

Table 3: Commodity Futures Trading Results, Rolling Window; annualized Sharpe ratio (SR), cumulative profits in millions (P)

		EW	OLS	AIC	BIC
Portfolio 1	SR	0.3496	0.1283	0.1017	-0.6619
JO,GI,KC,LH,NG	P	515.0	400.2	188.5	-454.6
Portfolio 2	SR	0.1673	-0.4520	-0.7410	-0.7530
GI,KC,LH,NG,NR	P	121.1	-3,624.6	-2,740.1	-2,773.8
Portfolio 3	SR	0.3316	-0.5918	-0.3031	-0.3328
KC,LH,NG,NR,PL	P	336.4	-4,371.0	-333.1	-290.3
Portfolio 4	SR	0.2254	-0.7753	-0.6906	-0.8527
LH,NG,NR,PL,SI	P	584.8	-6,152.9	-5,351.5	-1,939.6
Portfolio 5	SR	-0.4323	-0.5088	0.0795	0.1605
NG,NR,PL,SI,GC	P	-1,340.7	-5,511.8	1,278.1	1,763.8
Portfolio 6	SR	-0.3085	-0.5881	0.2212	0.1535
NR,PL,SI,GC,C_	P	-1,235.6	-6,230.6	3,796.8	1,699.2
Portfolio 7	SR	0.3246	-0.1508	0.1340	0.1047
PL,SI,GC,C_,BO	P	279.1	-1,492.1	1,550.9	1,166.9
Portfolio 8	SR	-0.2931	-0.4872	-0.0431	-0.0682
SI,GC,C_,BO,HG	P	-530.4	-4,825.7	-480.6	-780.9
Portfolio 9	SR	-0.3962	0.0060	-0.0174	-0.0174
GC,C_,BO,HG,CL	P	-770.0	24.2	-64.0	-64.0
Portfolio 10	SR	0.2322	0.2950	-0.2853	-0.2005
C_,BO,HG,CL,LC	P	47.8	432.0	-102.1	-63.0
Portfolio 11	SR	0.1834	-0.3662	-0.0874	-0.2000
BO,HG,CL,LC,LB	P	52.0	-334.6	-29.1	-73.7
Portfolio 12	SR	0.5545	-0.5773	-0.6725	-0.7679
HG,CL,LC,LB,JO	P	451.9	-571.1	-218.4	-239.7
Averages	SR	0.0782	-0.3390	-0.1920	-0.2863
	P	-124.1	-2688.2	-208.7	-170.8

Table 4: Commodity Futures Trading Results, Expanding Window; annualized Sharpe ratio (SR), cumulative profits in millions (P)

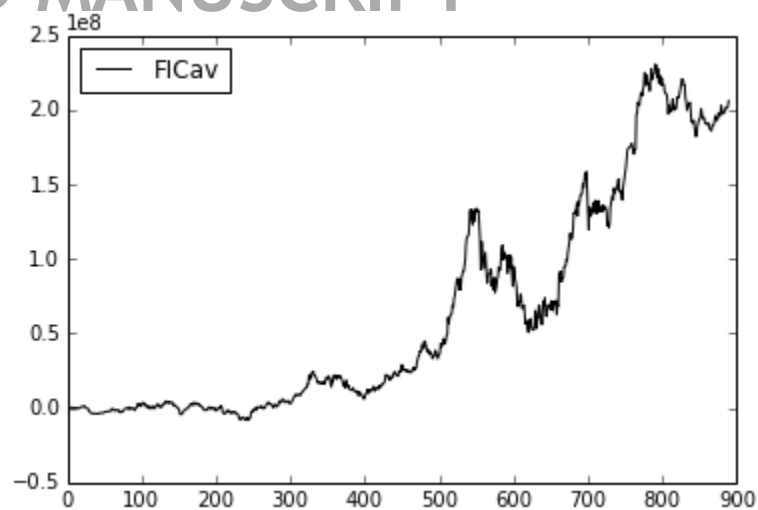


Figure 7.1: FICav Account Curve, Portfolio 1

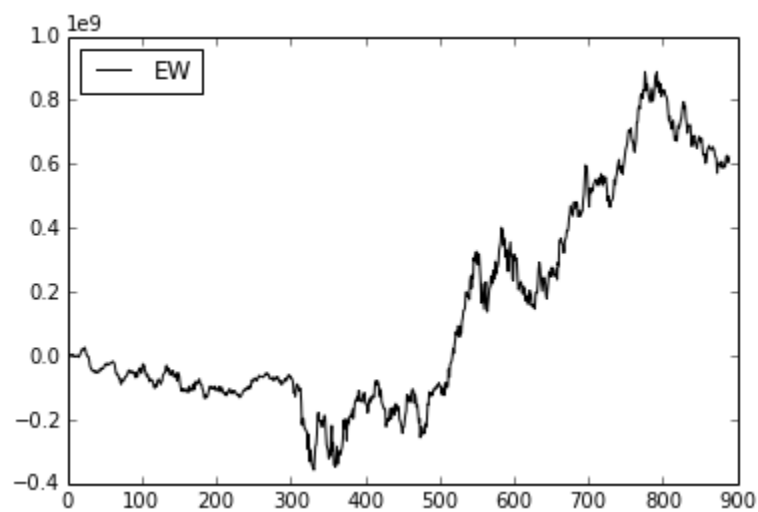


Figure 7.2: EW Account Curve, Portfolio 1

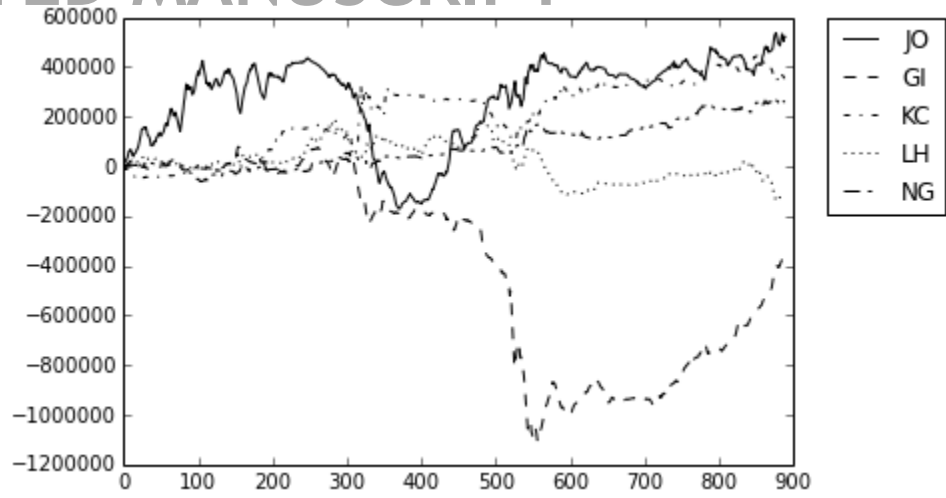


Figure 7.3: FICav Portfolio Positions, Portfolio 1

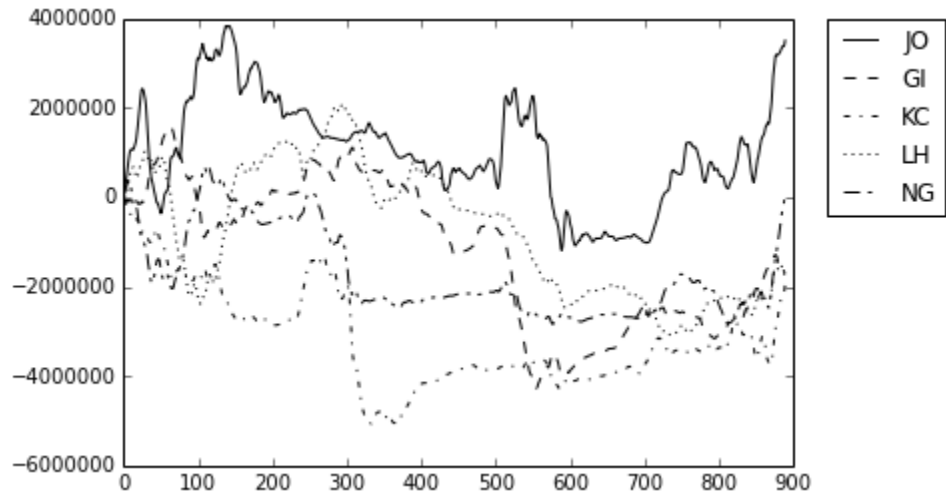


Figure 7.4: EW Portfolio Positions, Portfolio 1

8 Appendix

Assumption 2. As $T \rightarrow \infty$, for $i, j \in 1, \dots, M$:

$$T^{-1}H_i'H_j \xrightarrow{p} Q_{ij} = \begin{bmatrix} Q_{X_i'X_j} & Q_{X_i'Z_j} \\ Q_{Z_i'X_j} & Q_{Z_i'Z_j} \end{bmatrix}$$

$$\begin{aligned} \begin{bmatrix} \frac{1}{\sqrt{T}} H_i' \epsilon_j \\ \frac{1}{\sqrt{T}} J_T \end{bmatrix} &= \begin{bmatrix} \frac{1}{\sqrt{T}} X_i' \epsilon_j \\ \frac{1}{\sqrt{T}} Z_i' \epsilon_j \\ \frac{1}{\sqrt{T}} J_T \end{bmatrix} \xrightarrow{d} \begin{bmatrix} R_{ij} \\ J \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{H_i' \epsilon_j \epsilon_j' H_i} & \Sigma_{H_i' \epsilon_j J} \\ \Sigma_{J \epsilon_j' H_i} & \Sigma_{22} \end{bmatrix} \right) \\ &= N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{X_i' \epsilon_j \epsilon_j' X_i} & \Sigma_{X_i' \epsilon_j \epsilon_j' Z_i} & \Sigma_{X_i' \epsilon_j J} \\ \Sigma_{Z_i' \epsilon_j \epsilon_j' X_i} & \Sigma_{Z_i' \epsilon_j \epsilon_j' Z_i} & \Sigma_{Z_i' \epsilon_j J} \\ \Sigma_{J' \epsilon_j' X_i} & \Sigma_{J' \epsilon_j' Z_i} & \Sigma_{22} \end{bmatrix} \right) \end{aligned}$$

where $\Sigma_{H_i' \epsilon_j \epsilon_j' H_i} = E \left(H_i' \epsilon_j \epsilon_j' H_i \right)$ and similarly for other objects in the covariance matrix.

Assumption 2 implies Condition 3 in the main text. Note that Assumption 2 is more general than that needed for the expressions of $T^{-1} H' H$ and

$$\begin{bmatrix} \frac{1}{\sqrt{T}} H' \epsilon \\ \frac{1}{\sqrt{T}} J_T \end{bmatrix}$$

which only contain elements of Q_{ii} and $\frac{1}{\sqrt{T}} H_i' \epsilon_i$. This generality is needed for deriving the limiting expressions for the covariances in Theorem A1.

Assumption 2 is weak. Convergence of $1/\sqrt{T} H_i' \epsilon_j$ could be shown with standard martingale methods in many situations (Hall and Heyde (1980)). Convergence of $1/\sqrt{T} J_T$ will follow from moment conditions on the error terms and a standard central limit theorem. Other primitive assumptions leading to these results are also possible.

References

- [1] Claeskens, G., Hjort, N.L. (2003) 'The focused information criterion', *Journal of American Statistical Association*, 98, 900–916.
- [2] Claeskens, G., Magnus, J.R., Vasnev, A.L., Wang, W. (2016) 'The forecast combination puzzle: A simple theoretical explanation', *International Journal of Forecasting*, 32, 754–762.

- [3] Clark, T.E., McCracken, M.W. (2009) 'Combining Forecasts from Nested Models', *Oxford Bulletin of Economics and Statistics*, 73 (3), 303–329.
- [4] Collin-Dufresne, P., Daniel, K., Moallemi, C., Saglam, M. (2015) 'Dynamic Asset Allocation with Predictable Returns and Transaction Costs', available at [SSRN: <http://ssrn.com/abstract=2618910> or <http://dx.doi.org/10.2139/ssrn.2618910>].
- [5] DeMiguel, V., Garlappi, L., Uppal, R. (2009) 'Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?', *The Review of Financial Studies*, 22, 1915-1953.
- [6] DeMiguel, V., Nogales, F., Uppal, F. (2014) 'Stock Return Serial Dependence and Out-of-Sample Portfolio Performance', *Review of Financial Studies*, 27, 1031-1073.
- [7] DeMiguel, V., Mei, X., Nogales, F. (2014) 'Multiperiod Portfolio Optimization with Many Risky Assets and General Transaction Costs', LBS working paper.
- [8] Elliot, G., Gargano, A., Timmermann, A. (2013) 'Complete Subset Regressions', *Journal of Econometrics*, 177, 357-373.
- [9] Elliot, G., Gargano, A., Timmermann, A. (2015) 'Complete Subset Regressions with Large - Dimensional Sets of Predictors', *Journal of Economic Dynamics and Control*, 54, 86 - 110.
- [10] Garleanu, N., Pedersen, L. (2013) 'Dynamic Trading with Predictable Returns and Transaction Costs', *Journal of Finance*, 68, 2309-2340.
- [11] Geweke, J., Amisano, G. (2011) 'Optimal prediction pools', *Journal of Econometrics*, 164, 130-141.
- [12] Hall, P., Heyde, C.C. (1980) *Martingale Limit Theory and its Applications*, New York: Academic Press
- [13] Hamilton, J. (1994) *Time Series Analysis*, Princeton: Princeton University Press.
- [14] Hansen, B. (2007) 'Least Squares Model Averaging', *Econometrica*, 75, 1175-1189.
- [15] Hjort, N.L., Claeskens, G. (2003a) 'Frequentist model average estimators', *Journal of American Statistical Association*, 98, 879–899.

- [16] Hjort, N.L., Claeskens, G. (2003b) 'Rejoinder to the focused information criterion and frequentist model average estimators', *Journal of American Statistical Association*, 98, 938–945
- [17] Liang, H., Zou, G., Wan, A., Zhang, X. (2011) 'Optimal weight choice for frequentist model average estimators', *Journal of American Statistical Association*, 106, 1053–1066.
- [18] Liu, C. (2014) 'Distribution theory of the least squares averaging estimator', *Journal of Econometrics*, 186, 142–159.
- [19] Markowitz, H. (1952) 'Portfolio Selection', *Journal of Finance*, 7, 77–91.
- [20] Timmerman, A. (2006) 'Forecast Combinations'. In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), *Handbook of Economic Forecasting*. Elsevier, Amsterdam, pp. 135–195.