BUILDING TRACKING PORTFOLIOS BASED ON A GENERALIZED INFORMATION CRITERION

Shaojun Zhang[1], Xu-Feng Niu[2] and James S. Ang[2]

[1]*Nanyang Technological University* and [2]*Florida State University*

Abstract: One important topic in financial studies is to build a tracking portfolio of stocks whose return mimics that of a chosen investment target. Statistically, this task can be accomplished by selecting an optimal constrained linear model. In this paper, we extend the Generalized Information Criterion (GIC) to constrained linear models with independently and identically distributed random errors and, more generally, with dependent errors that follow a stationary Gaussian process. The extended GIC procedure is proved to be asymptotically loss efficient and consistent under mild conditions. Simulation results show that the relative frequency of selecting the optimal constrained linear model by GIC is close to one for finite samples. We also apply GIC to build an optimal tracking portfolio for measuring the long-term impact of a corporate event on stock returns, and demonstrate empirically that it outperforms two competing methods.

Key words and phrases: Abnormal returns, average squared error loss, infinite moving average processes, stationary processes.

1. Introduction

In a number of financial studies, the essential task boils down to building a tracking portfolio of stocks whose return mimics that of a chosen investment target. For example, the sole business of an index fund is to maintain a portfolio of individual stocks so that percentage changes in value of the portfolio are approximately equal to those of the chosen index. How closely its portfolio tracks the index is an important determinant of the success of an index fund. Another example is to assess the effect of a specific corporate event on the event firm's long-term stock returns after the event has happened. Since the observed post-event return has been affected by the event, we do not know what the *status-quo* return would have been if the event had not happened. To estimate the unobservable *status-quo* return, a researcher can choose a portfolio of other stocks whose returns have moved in tandem with the event firm's returns before the event happened and use the observed post-event return on the portfolio as an estimate.

In both examples, a tracking portfolio is to be built given a desired target and a group of other stocks. Every nonempty subset of the given group of stocks

compose a portfolio that may track the target to some degree. There are as many possible tracking portfolios as the number of nonempty subsets of stocks. Among all possible portfolios, an ideal tracking portfolio should be the one whose return is equal to the target's return in every month. In reality, no portfolio tracks the target perfectly. An optimal, though not ideal, tracking portfolio would be the one whose returns are on average closest to the target's. Therefore, the task is to find such an optimal tracking portfolio.

Since return of a tracking portfolio is a weighted sum of returns on all stocks in the portfolio, building a tracking portfolio from a group of stocks is equivalent to fitting a constrained linear model with the target's return as the response variable and returns on stocks in the group as the covariates. The linear model is constrained such that all coefficients in the model sum to one. This is because the coefficient of a covariate is equal to the proportion of investment on the corresponding stock to the total investment in the tracking portfolio and the sum of all coefficients accounts for 100 percent of the total investment.

Because of the correspondence between a tracking portfolio and a constrained linear model, the task of finding the optimal tracking portfolio can be accomplished by selecting an optimal constrained linear model. In this paper, we develop a statistical procedure to do this, based on a model selection criterion known as the Generalized Information Criterion (GIC). There is considerable literature on the problem of selecting variables in the context of unconstrained linear models; see review papers by Hocking (1976) and Thompson (1978a, b) for early contributions. Miller (1990) gave a comprehensive summary of variable selection methods prior to 1990, and George (2000) reviewed the key developments in the last decade. A recent book by Burnham and Anderson (2002) gives a systematic account of the developments in model selection from the information theoretic viewpoint. The Generalized Information Criterion (GIC) we use in this paper was proposed by Rao and Wu (1989), and is a generalization of the well known Akaike's Information Criterion (AIC, Akaike (1973)) and the Bayesian Information Criterion (BIC, Schwartz (1978)).

Asymptotic properties of GIC have been investigated in various settings. For instance, Nishii (1984), Rao and Wu (1989) and Pötscher (1989) proved consistency of GIC or its asymptotic equivalents for unconstrained linear models under the assumption that there exists a finite-dimensional true model. In these studies, different assumptions were imposed on the random error terms in the linear models. Nishii (1984) and Rao and Wu (1989) assumed independently and identically distributed (i.i.d.) random errors, while Pötscher (1989) assumed that the error term is a martingale difference sequence. Shao (1997) used two criteria to evaluate asymptotic validity of a model selection procedure: consistency and asymptotic loss efficiency (see Section 3 for definitions of asymptotic loss efficiency and consistency). For unconstrained linear models with i.i.d. random

errors, Shao (1997) showed that GIC is asymptotically loss efficient regardless of the existence of a true model, and consistent if a true model exists.

In this paper, we extend GIC to constrained linear models with errors following a stationary Gaussian process. Since there is no guarantee that the target's return is completely determined by returns on a subset of stocks, the existence of a true model cannot be taken for granted. We give conditions under which the extended GIC is asymptotically loss efficient regardless of the existence of a true model, and consistent if a true model exists.

The rest of the paper is organized as follows. In Section 2, we formalize a statistical model for the problem of building an optimal tracking portfolio. In Section 3, we extend GIC to constrained linear models with dependent errors that follow a stationary Gaussian process, and we study its asymptotic properties. Section 4 reports results from a simulation study which demonstrates that the relative frequency of selecting the correct model by the GIC procedure is close to one for finite samples. In Section 5, we apply the GIC procedure to build an optimal tracking portfolio for the purpose of measuring long-term post-event abnormal returns of event firms. Performance of the GIC procedure is compared against that of two other methods empirically. It is found that the GIC procedure gives the best performance. Summary of the paper and some discussions are in Section 6. Proofs of the asymptotic properties are given in the Appendix.

2. Statistical Model

Let y_t be the return from investing in a chosen target during a unit time interval from $t - 1$ to t, i.e.,

$$y_t = \frac{\text{Target's price at } t \ - \ \text{Target's price at } t - 1}{\text{Target's price at } t - 1} \ ,$$

for $t = 1, 2, \cdots, \tau$. Let $\boldsymbol{y} = (y_1, \cdots, y_\tau)'$ be a vector of historical returns with

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{e} \ , \tag{2.1}$$

where $\boldsymbol{\mu} = E(\boldsymbol{y})$ is the mean of $\boldsymbol{y}$ and $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{\mu}$ is a vector of random variables with mean zero. Note that we allow the expected return of the target to change with time in any manner.

Suppose that m other stocks are available for building a tracking portfolio of the target. Let $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m)$ be a $\tau \times m$ matrix of rank m, where $\boldsymbol{x}_j = (x_{j1}, \cdots, x_{j\tau})'$ is a vector of historical returns on the jth stock for $j = 1, \cdots, m$. Let V be the collection of all nonempty subsets of $\{1, 2, \cdots, m\}$. Each subset $v \in V$ indexes a group of stocks. Let $\boldsymbol{X}(v)$ be the submatrix of $\boldsymbol{X}$ whose columns are returns of stocks in the subset v. To build a tracking portfolio

consisting of all stocks in the subset v, we fit the following linear model of $\boldsymbol{y}$ against the covariates:

$$\boldsymbol{y} = \boldsymbol{X}(v)\boldsymbol{\beta}(v) + \boldsymbol{e}(v) \ , \tag{2.2}$$

where the dimension of $\boldsymbol{\beta}(v)$ is equal to the size of the subset v. Note that the error term $\boldsymbol{e}(v)$ depends on v and differs from the random vector $\boldsymbol{e}$ in (2.1). More specifically, the error term $\boldsymbol{e}(v)$ is the sum of the random vector $\boldsymbol{e}$ and the model misspecification error.

In the context of building a tracking portfolio, the coefficients $\boldsymbol{\beta}(v)$ in (2.2) sum up to one. The left hand side of (2.2) is the return of investing one dollar in the target. The right hand side is the return of investing one dollar in the portfolio consisting of all stocks in the subset v, plus random noise. Each coefficient in $\boldsymbol{\beta}(v)$ is the proportion of a dollar invested in the corresponding stock, and the sum of all coefficients accounts for 100 percent of the dollar.

Let $\hat{\boldsymbol{\beta}}(v)$ denote an estimate of $\boldsymbol{\beta}(v)$. Then an estimate of $\boldsymbol{\mu}$ is $\hat{\boldsymbol{\mu}}(v) = \boldsymbol{X}(v)\hat{\boldsymbol{\beta}}(v)$. The goodness of the estimate is measured by the *average squared error loss*

$$L(v) = \frac{||\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(v)||^2}{\tau} \ , \tag{2.3}$$

where $||\cdot||$ is the Euclidean norm. The objective of model selection is to find the subset v whose associated estimate $\hat{\boldsymbol{\mu}}(v)$ minimizes the average squared error loss. Once the minimizing subset v is found, the portfolio that consists of all stocks in the subset v and uses $\hat{\boldsymbol{\beta}}(v)$ as the portfolio weights is the optimal tracking portfolio.

To estimate the linearly constrained coefficients, we take the model reduction approach of Hocking (1985, Chap. 3). We briefly describe the approach for the sake of self-completeness of this paper. To simplify notation, we drop the subset index v for now.

A linear model with general linear constraints is written as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \text{subject to} \quad \boldsymbol{G}\boldsymbol{\beta} = \boldsymbol{g} \ , \tag{2.4}$$

where $\boldsymbol{y}$ is a vector of dimension τ, $\boldsymbol{X}$ is a $\tau \times m$ matrix of rank m, $\boldsymbol{\beta}$ is a vector of m coefficients, $\boldsymbol{G}$ is a $q \times m$ matrix of rank q, and $\boldsymbol{\epsilon}$ is a random vector.

An estimate of $\boldsymbol{\beta}$ can be obtained by the model reduction approach as follows. The coefficient vector $\boldsymbol{\beta}$ and the constraint matrix $\boldsymbol{G}$ are partitioned in such a way that the constraints are written as

$$\boldsymbol{G}_1\boldsymbol{\beta}_1 + \boldsymbol{G}_2\boldsymbol{\beta}_2 = \boldsymbol{g} \ ,$$

where $\boldsymbol{G}_1$ is a $q \times q$ matrix of rank q. Solving for $\boldsymbol{\beta}_1$ yields

$$\boldsymbol{\beta}_1 = \boldsymbol{G}_1^{-1}\boldsymbol{g} - \boldsymbol{G}_1^{-1}\boldsymbol{G}_2\boldsymbol{\beta}_2 \ . \tag{2.5}$$

Corresponding to the partition of $\boldsymbol{\beta}$, we partition $\boldsymbol{X}$ as $\boldsymbol{X} = (\boldsymbol{X}_1 \ \boldsymbol{X}_2)$, where $\boldsymbol{X}_1$ is a $\tau \times q$ matrix. Substituting the partition into the constrained model, we obtain the unconstrained model $\boldsymbol{y}_R = \boldsymbol{X}_R \boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$, where $\boldsymbol{y}_R = \boldsymbol{y} - \boldsymbol{X}_1 \boldsymbol{G}_1^{-1} \boldsymbol{g}$ and $\boldsymbol{X}_R = \boldsymbol{X}_2 - \boldsymbol{X}_1 \boldsymbol{G}_1^{-1} \boldsymbol{G}_2$. The least squares estimate of $\boldsymbol{\beta}_2$ is then given by $\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_R' \boldsymbol{X}_R)^{-1} \boldsymbol{X}_R' \boldsymbol{y}_R$. Substituting $\hat{\boldsymbol{\beta}}_2$ in (2.5), we get $\hat{\boldsymbol{\beta}}_1 = \boldsymbol{G}_1^{-1} \boldsymbol{g} - \boldsymbol{G}_1^{-1} \boldsymbol{G}_2 \hat{\boldsymbol{\beta}}_2$. We can write the estimate of $\boldsymbol{\beta}$ together as

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{G}_1^{-1} \boldsymbol{g} \\ 0 \end{bmatrix} + \begin{bmatrix} -\boldsymbol{G}_1^{-1} \boldsymbol{G}_2 \\ \boldsymbol{I} \end{bmatrix} \hat{\boldsymbol{\beta}}_2 \ . \tag{2.6}$$

Then the estimate of $\boldsymbol{\mu} = E(\boldsymbol{y})$ is given by

$$\hat{\boldsymbol{\mu}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{\eta} + \boldsymbol{H}\boldsymbol{y} \ , \tag{2.7}$$

where $\boldsymbol{H} = \boldsymbol{X}_R (\boldsymbol{X}_R' \boldsymbol{X}_R)^{-1} \boldsymbol{X}_R'$ and $\boldsymbol{\eta} = (\boldsymbol{I} - \boldsymbol{H}) \boldsymbol{X}_1 \boldsymbol{G}_1^{-1} \boldsymbol{g}$.

3. EGIC and its Asymptotic Properties

In the context of unconstrained linear models, numerous criteria have been proposed to select variables, see, e.g., Hocking (1976), Thompson (1978a, b), Li (1987), Miller (1990), George (2000) and the references therein. Among them, GIC or its asymptotic equivalents have been proposed and/or studied by Nishii (1984), Rao and Wu (1989), Pötscher (1989) and Shao (1997), among others. In this section we extend GIC to constrained linear models with dependent observations, and give conditions under which the extended GIC is still asymptotically loss efficient and consistent.

In the remainder of this paper, the subset index $v \in V$ and the subscript τ are used to identify quantities that depend on the subset v and those that vary with τ, respectively.

We consider constrained linear models

$$\boldsymbol{y}_\tau = \boldsymbol{X}_\tau(v)\boldsymbol{\beta}(v) + \boldsymbol{e}_\tau(v) \quad \text{subject to} \quad \boldsymbol{l}'\boldsymbol{\beta}(v) = 1 \ , \tag{3.1}$$

where v belongs to V, $\boldsymbol{l}$ is a vector of ones with the same dimension as $\boldsymbol{\beta}(v)$, and τ is the number of observed time periods. Note that the coefficients in $\boldsymbol{\beta}(v)$ vary with the subset v but not with τ. In other words, we assume that the linear relationship between $\boldsymbol{y}_\tau$ and $\boldsymbol{X}_\tau(v)$ does not change over time.

Let $\boldsymbol{\mu}_\tau = E(\boldsymbol{y}_\tau)$. A candidate model $v \in V$ is said to be *correct* if there exists $\boldsymbol{\beta}(v)$ such that $\boldsymbol{\mu}_\tau = \boldsymbol{X}_\tau(v)\boldsymbol{\beta}(v)$ for all τ. Let V^c be the collection of all correct models.

It is straightforward to decompose $\boldsymbol{y}_\tau$ as $\boldsymbol{y}_\tau = \boldsymbol{\mu}_\tau + \boldsymbol{e}_\tau$, where $\boldsymbol{e}_\tau$ is a vector of random variables with zero mean. We assume that $\{e_t\}_{t=-\infty}^\infty$ is a stationary

Gaussian process with $E(e_t) = 0$ and $E(e_t e_{t+j}) = \gamma_j$. The autocovariance matrix for $\boldsymbol{e}_\tau$, denoted as $\boldsymbol{\Psi}_\tau$, is then

$$\boldsymbol{\Psi}_\tau = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{\tau-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{\tau-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{\tau-1} & \gamma_{\tau-2} & \cdots & \gamma_0 \end{bmatrix} . \tag{3.2}$$

We further assume that the autocovariances $\{\gamma_j, j = 0, \pm 1, \ldots, \}$ are absolutely summable, that is,

$$\Upsilon \equiv \gamma_0 + 2 \sum_{j=1}^{\infty} |\gamma_j| < \infty . \tag{3.3}$$

The *Extended Generalized Information Criterion* (EGIC) selects a model that minimizes

$$\Phi_\tau(v) \equiv \frac{||\boldsymbol{y}_\tau - \hat{\boldsymbol{\mu}}_\tau(v)||^2}{\tau} + \frac{\lambda_\tau tr(\hat{\boldsymbol{\Psi}}_\tau \boldsymbol{H}_\tau(v))}{\tau} \tag{3.4}$$

over $v \in V$, where $\hat{\boldsymbol{\Psi}}_\tau$ is an estimate of $\boldsymbol{\Psi}_\tau$, $tr(\cdot)$ stands for the trace of a matrix, the hat matrix $\boldsymbol{H}_\tau(v)$ is introduced in equation (2.7), and λ_τ satisfies the conditions given in (3.6) and (3.7). Note that $\hat{\boldsymbol{\Psi}}_\tau$ does not depend on v and is obtained by fitting a linear model with all covariates included.

In the special case when random errors $\boldsymbol{e}_\tau$ are independently and identically distributed with a normal distribution of mean zero and variance σ^2, the Extended Generalized Information Criterion can be simplified to

$$\Gamma_\tau(v) \equiv \frac{||\boldsymbol{y}_\tau - \hat{\boldsymbol{\mu}}_\tau(v)||^2}{\tau} + \frac{\lambda_\tau \hat{\sigma}_\tau^2 tr(\boldsymbol{H}_\tau(v))}{\tau} \tag{3.5}$$

over $v \in V$, where $\hat{\sigma}_\tau^2$ is an estimator of σ^2. Note that $\hat{\sigma}_\tau^2$ does not depend on v and is obtained by fitting the linear model with all m covariates included. Note that $tr(\boldsymbol{H}_\tau(v))$ is equal to the number of unconstrained coefficients in the linear model (3.1). For instance, a linear model with five covariates and one constraint has four unconstrained coefficients and $tr(\boldsymbol{H}_\tau(v)) = 4$. This simplified criterion is essentially the same as GIC discussed in Shao (1997), except that in this paper the hat matrix $\boldsymbol{H}_\tau(v)$ is a result of fitting a constrained linear model.

Shao (1997) studied the asymptotic validity of several model selection procedures with respect to two criteria: consistency and asymptotic loss efficiency. We adopt his definitions of the two criteria in this paper. Specifically, let $\check{v}_\tau$ denote the model that minimizes EGIC, $\Phi_\tau(v)$, over $v \in V$. Let v_τ^L be the model that minimizes the average squared error loss, $L_\tau(v)$, over $v \in V$. The EGIC selection procedure is said to be *consistent* if $P\{\check{v}_\tau = v_\tau^L\} \to 1$ as $\tau \to \infty$, and

to be *asymptotically loss efficient* if $L_\tau(\check{v}_\tau)/L_\tau(v_\tau^L) \overset{p}{\to} 1$ as $\tau \to \infty$, where $\overset{p}{\to}$ represents convergence in probability. Throughout this paper, all limiting processes are taken as $\tau \to \infty$.

Proofs of the following results are given in the Appendix.

Theorem 3.1. *The average squared error loss at (2.3) is $L_\tau(v) = \Delta_\tau(v) + (e_\tau' H_\tau(v) e_\tau)/\tau$, where $\Delta_\tau(v) = (||\mu_\tau - \eta_\tau(v) - H_\tau(v)\mu_\tau||^2)/\tau$. Furthermore, $\Delta_\tau(v) = 0$ when $v \in V^c$. In addition, the expected average squared error loss is $R_\tau(v) \equiv E(L_\tau(v)) = \Delta_\tau(v) + tr(\Psi_\tau H_\tau(v))/\tau$, where the matrix Ψ_τ is the autocovariance matrix at (3.2).*

Remark 3.1. The theorem shows that the average squared error loss has two components: the model misspecification error $\Delta_\tau(v)$, and the estimation error $(e_\tau' H_\tau(v) e_\tau))/\tau$ due to randomness in the observed data. Moreover, when the model v is correct, the model misspecification error $\Delta_\tau(v)$ is zero.

The following two lemmas discuss some properties of the autocovariance matrix Ψ_τ.

Lemma 3.1. *For any vectors a and b,*

$$|a' \Psi_\tau b| \leq \frac{||a||^2 + ||b||^2}{2} \Upsilon \ ,$$

where Υ is the absolute sum of autocovariances.

Lemma 3.2. *Suppose that H is an idempotent matrix of rank r. Then $tr(\Psi_\tau H) \leq r\Upsilon$ and $tr(\Psi_\tau H \Psi_\tau H) \leq (r\Upsilon)^2$.*

Theorem 3.2. *Assume that $\{e_t\}_{t=-\infty}^{\infty}$ is a stationary Gaussian process with $E(e_t) = 0$, $\gamma_j = E(e_t e_{t+j})$, and $\Upsilon \equiv \gamma_0 + 2\sum_{j=1}^{\infty} |\gamma_j| < \infty$. Assume further that $\hat{\Psi}_\tau$ used in computing the EGIC, $\Phi_\tau(v)$, is a consistent estimator of Ψ_τ and that $tr(\Psi_\tau H_\tau(v))$ converges to a finite limit as $\tau \to \infty$ for any $v \in V^c$. If*

$$\lambda_\tau \to \infty \ , \qquad \lambda_\tau/\tau \to 0 \ , \tag{3.6}$$

$$\frac{\lambda_\tau}{\tau R_\tau(v)} \to 0 \quad \text{for all} \quad v \in V - V^c \ , \tag{3.7}$$

the EGIC minimizer $\check{v}_\tau$ is asymptotically loss efficient. In addition, if V contains at least one correct model, then $\check{v}_\tau$ is consistent.

Remark 3.2. Shao (1997) proved asymptotic loss efficiency and consistency of the GIC for unconstrained linear regression models with i.i.d. random errors under condition (3.6) and

$$\liminf_{\tau \to \infty} \min_{v \in V - V^c} \Delta_\tau(v) > 0 \ . \tag{3.8}$$

It is easy to see that (3.6) and (3.8) imply (3.7) because $R_\tau(v) \geq \Delta_\tau(v)$. Shao's condition (3.8) requires that the model misspecification error $\Delta_\tau(v)$ be bounded away from zero uniformly for all incorrect models. By contrast, our condition (3.7) suggests that, as long as the model misspecification error of incorrect models tends to 0 at a rate slower than $1/\tau$, the EGIC minimizer $\check{v}_\tau$ still has these optimal asymptotic properties for certain λ_τ.

The following corollary shows that some common stochastic processes are covered by Theorem 3.2. The proof of the corollary is simple and omitted.

Corollary 3.1. *Theorem* 3.2 *is valid when* $\{e_t\}_{t=-\infty}^{\infty}$ *is an infinite moving average Gaussian process given by* $e_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$, *where* $a_t \sim$ *i.i.d.* $N(0, \sigma_a^2)$ *and* $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

Remark 3.3. The Gaussian infinite moving average process specified in Corollary 3.1 includes the stationary Gaussian processes AR(p), MA(q) and ARMA(p, q) as special cases.

Remark 3.4. It has been widely documented that many financial time series follow the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model (see, e.g., Bollerslev, Chou and Kroner (1992) and the references therein). A GARCH model characterizes the error process $\{e_t\}_{t=-\infty}^{\infty}$ by $e_t = a_t \sigma_t$, where $a_t \sim$ i.i.d. with mean 0 and variance 1, and $\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i e_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$. Under the conditions that $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$, the GARCH process $\{e_t\}$ is uncorrelated with zero mean and constant unconditional variance (see, e.g., Tsay (2001, Chap. 3)). However, it is not a Gaussian process; otherwise, being uncorrelated would imply mutual independence of $\{e_t\}$. In fact, the unconditional distribution of e_t has fatter tails than the normal distribution. Therefore, Theorem 3.2 does not apply to GARCH processes.

When the random errors are i.i.d., the GIC given in (3.5) replaces EGIC. The following corollary gives the optimal asymptotic properties of GIC for constrained linear models. Its proof is simple and omitted.

Corollary 3.2. *Assume that random errors* $\boldsymbol{e}_\tau$ *are i.i.d. with a normal distribution with mean zero and variance* σ^2 *and that the estimator* $\hat{\sigma}_\tau^2$ *is consistent. Under* (3.6) *and* (3.7), *the GIC minimizer* $\hat{v}_\tau$ *is asymptotically loss efficient. In addition, if* V *contains at least one correct model, then* $\hat{v}_\tau$ *is consistent.*

Remark 3.5. Under the condition that the estimator $\hat{\sigma}_\tau^2$ is bounded, the conclusions in Corollary 3.2 hold. A proof is available upon request.

4. Simulation Results

In this section, we carry out a simulation study for two purposes: to empirically check validity of the optimal properties of the GIC defined in (3.5), and to understand how choice of the penalty λ_n affects GIC's performance in finite samples.

4.1. Simulation setup

Simulation studies in this paper are based on historical monthly returns of five stocks. The five stocks are randomly selected and their monthly returns between January 1981 and December 1988 (inclusive) are extracted from the database distributed by the Center of Research in Securities Prices (CRSP). The five stocks are Wal Mart Stores Inc. (WMT), Dayton Hudson Corp. (DH), Mac Frugals Bargains Close Outs (MFB), Service Merchandise Inc. (SM), and Family Dollar Stores Inc (FDS). The sample autocorrelation function plots show that monthly returns of the five stocks are not autocorrelated. This is consistent with documented findings in empirical finance literature that monthly individual stock returns have insignificant autocorrelation (see, e.g., Campbell, Lo and MacKinlay (1997, Chap. 2), and the references therein). We also estimated GARCH models with these monthly returns and found that most coefficients in the conditional variance function are insignificant. Although prior studies have documented strong evidence of conditional heteroskedasticity in daily or weekly returns on individual stocks and on monthly returns on stock indices (see, e.g., Bollerslev, Chou and Kroner (1992) and the references therein), there is no such conclusive evidence for monthly individual stock returns. Therefore, it is reasonable to assume i.i.d. random errors in modeling monthly individual stock returns, and to use GIC instead of EGIC.

Monthly returns of the five stocks, denoted by $\{x_{1t}, \ldots, x_{5t}\}$, are used as independent variables in the following regression model

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \beta_5 x_{5t} + \epsilon_t , \quad t = 1, \cdots, \tau$$

$$\text{subject to} \quad \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 = 1 .$$

Given the model coefficients $(\beta_1, \beta_2, \cdots, \beta_5)$, we simulate the response variable y_t by generating i.i.d. random errors ϵ_t from a normal distribution $N(0, \sigma^2)$. We choose σ to be 0.0385 throughout the simulation, this is the sample standard deviation of the 96 monthly returns on the CRSP value weighted market index between January 1981 and December 1988.

Three sets of simulations are carried out in this study. For the first set, we fix the coefficients at (0.3,0.0,0.0,0.4,0.3) and choose the number of observations τ to be 36, 60, or 96. Two common choices of the penalty λ_τ are employed: $\lambda_\tau = \log \tau$ and $\lambda_\tau = \sqrt{\tau}$. Under each combination of τ and λ_τ, 1,000 realizations are simulated. There are $31(= 2^5 - 1)$ nonempty subsets for five covariates. Given each realization, we compute the GIC value, $\Gamma_\tau(v)$, for all 31 subsets, and identify the smallest one. We report in Table 1 the frequency of each subset being the minimizer of GIC in the 1,000 realizations.

Table 1. Frequency of each candidate model being selected by the GIC procedure, in 1000 realizations with observed returns, regression coefficients being $(0.3, 0.0, 0.0, 0.4, 0.3)$.

Candidate	$\lambda_\tau = \log(\tau)$			$\lambda_\tau = \sqrt{\tau}$		
Models	$\tau = 36$	60	96	$\tau = 36$	60	96
(1)	0	0	0	0	0	0
(2)	0	0	0	0	0	0
(3)	0	0	0	0	0	0
(4)	0	0	0	0	0	0
(5)	0	0	0	0	0	0
(1, 2)	0	0	0	0	0	0
(1, 3)	0	0	0	0	0	0
(1, 4)	0	1	0	7	4	0
(1, 5)	0	0	0	1	0	0
(2, 3)	0	0	0	0	0	0
(2, 4)	0	0	0	0	0	0
(2, 5)	0	0	0	0	0	0
(3, 4)	0	0	0	0	0	0
(3, 5)	0	0	0	0	0	0
(4, 5)	31	1	0	120	26	2
(1, 2, 3)	0	0	0	0	0	0
(1, 2, 4)	0	0	0	0	0	0
(1, 2, 5)	0	0	0	0	0	0
(1, 3, 4)	0	0	0	0	0	0
(1, 3, 5)	0	0	0	0	0	0
(1, 4, 5)	832	911	953	799	928	991
(2, 3, 4)	0	0	0	0	0	0
(2, 3, 5)	0	0	0	0	0	0
(2, 4, 5)	49	19	1	37	20	5
(3, 4, 5)	37	16	2	29	15	2
(1, 2, 3, 4)	0	0	0	0	0	0
(1, 2, 3, 5)	0	0	0	0	0	0
(1, 2, 4, 5)	23	34	24	5	4	0
(1, 3, 4, 5)	25	16	20	2	2	0
(2, 3, 4, 5)	2	0	0	0	1	0
(1, 2, 3, 4, 5)	1	2	0	0	0	0

In the second set of simulations, we experiment with larger sample sizes and let τ to be 36, 60, 96, 120, or 240. However, since we have only 96 observed monthly returns for each stock, we generate 240 values in the following way. For each stock, we compute the sample mean and sample standard deviation of the 96 observed monthly returns after omitting two extreme values at each tail. Sample means of the five stocks are 0.0349, 0.0234, 0.0200, 0.0218 and 0.0258, and sample

standard deviations are 0.0715, 0.0724, 0.0931, 0.1131 and 0.1015. We use the Kolmogorov-Smirnov Goodness-of-Fit Test to check whether returns of the five stocks are normally distributed, the p-values are 0.5, 0.5, 0.0622, 0.5 and 0.0288, respectively. Since monthly returns of the five stocks are approximately normally distributed, we generate 240 values for each stock from a normal distribution with mean and standard deviation respectively equal to the stock's sample mean and sample standard deviation. Other aspects of the simulation are the same as in the first simulation. Results from the second set of simulations are reported in Table 2.

Table 2. Frequency of each candidate model being selected by the GIC procedure, in 1000 realizations with simulated returns, regression coefficients being $(0.3, 0.0, 0.0, 0.4, 0.3)$.

Candidate	$\lambda_\tau = \log(\tau)$					$\lambda_\tau = \sqrt{\tau}$				
Models	$\tau = 36$	60	96	120	240	$\tau = 36$	60	96	120	240
(1)	0	0	0	0	0	0	0	0	0	0
(2)	0	0	0	0	0	0	0	0	0	0
(3)	0	0	0	0	0	0	0	0	0	0
(4)	0	0	0	0	0	0	0	0	0	0
(5)	0	0	0	0	0	0	0	0	0	0
(1, 2)	0	0	0	0	0	0	0	0	0	0
(1, 3)	0	0	0	0	0	0	0	0	0	0
(1, 4)	1	0	0	0	0	4	0	0	0	0
(1, 5)	0	0	0	0	0	0	0	0	0	0
(2, 3)	0	0	0	0	0	0	0	0	0	0
(2, 4)	0	0	0	0	0	0	0	0	0	0
(2, 5)	0	0	0	0	0	0	0	0	0	0
(3, 4)	0	0	0	0	0	0	0	0	0	0
(3, 5)	0	0	0	0	0	0	0	0	0	0
(4, 5)	11	0	0	0	0	33	2	0	0	0
(1, 2, 3)	0	0	0	0	0	0	0	0	0	0
(1, 2, 4)	0	0	0	0	0	4	0	0	0	0
(1, 2, 5)	0	0	0	0	0	0	0	0	0	0
(1, 3, 4)	0	0	0	0	0	0	0	0	0	0
(1, 3, 5)	0	0	0	0	0	0	0	0	0	0
(1, 4, 5)	905	946	950	967	971	951	993	999	1000	1000
(2, 3, 4)	0	0	0	0	0	0	0	0	0	0
(2, 3, 5)	0	0	0	0	0	0	0	0	0	0
(2, 4, 5)	0	0	0	0	0	1	0	0	0	0
(3, 4, 5)	2	0	0	0	0	3	0	0	0	0
(1, 2, 3, 4)	0	0	0	0	0	1	0	0	0	0
(1, 2, 3, 5)	0	0	0	0	0	0	0	0	0	0
(1, 2, 4, 5)	43	24	22	17	15	2	1	1	0	0
(1, 3, 4, 5)	32	27	26	16	14	2	4	0	0	0
(2, 3, 4, 5)	0	0	0	0	0	0	0	0	0	0
(1, 2, 3, 4, 5)	6	3	2	0	0	0	0	0	0	0

The third set of simulations is conducted to investigate the effect of coefficients' signal-to-noise ratio on the performance of the selection procedure. The only difference between the second set and the third lies in the values of the coefficients. The coefficients are equal to (0.3,0.0,0.0,0.4,0.3) in the second set and (0.2,0.0,0.0,0.7,0.1) in the third. Therefore, in the third set of simulations, the fourth coefficient has the highest signal-to-noise ratio and the fifth has the lowest among the three non-zero coefficients. Since the standard deviation of random errors is 0.0385, the fifth coefficient is within three standard deviations of zero, its signal is weak. Table 3 reports results from the third set of simulations.

Table 3. Frequency of each candidate model being selected by the GIC procedure, in 1000 realizations with simulated returns, regression coefficients being $(0.2, 0.0, 0.0, 0.7, 0.1)$.

Candidate	$\lambda_\tau = \log(\tau)$					$\lambda_\tau = \sqrt{\tau}$				
Models	$\tau = 36$	60	96	120	240	$\tau = 36$	60	96	120	240
c(1)	0	0	0	0	0	0	0	0	0	0
c(2)	0	0	0	0	0	0	0	0	0	0
c(3)	0	0	0	0	0	0	0	0	0	0
c(4)	0	0	0	0	0	2	0	0	0	0
c(5)	0	0	0	0	0	0	0	0	0	0
c(1, 2)	0	0	0	0	0	0	0	0	0	0
c(1, 3)	0	0	0	0	0	0	0	0	0	0
c(1, 4)	434	387	222	136	14	638	658	615	508	267
c(1, 5)	0	0	0	0	0	0	0	0	0	0
c(2, 3)	0	0	0	0	0	0	0	0	0	0
c(2, 4)	9	0	0	0	0	8	5	1	0	0
c(2, 5)	0	0	0	0	0	0	0	0	0	0
c(3, 4)	0	0	0	0	0	0	0	0	0	0
c(3, 5)	0	0	0	0	0	0	0	0	0	0
c(4, 5)	131	20	0	0	0	219	68	16	1	0
c(1, 2, 3)	0	0	0	0	0	0	0	0	0	0
c(1, 2, 4)	37	36	20	15	0	21	10	8	3	0
c(1, 2, 5)	0	0	0	0	0	0	0	0	0	0
c(1, 3, 4)	26	20	14	13	3	8	8	6	3	0
c(1, 3, 5)	0	0	0	0	0	0	0	0	0	0
c(1, 4, 5)	297	473	690	790	942	93	246	351	484	733
c(2, 3, 4)	0	0	0	0	0	0	0	0	0	0
c(2, 3, 5)	0	0	0	0	0	0	0	0	0	0
c(2, 4, 5)	8	10	3	0	0	3	1	1	0	0
c(3, 4, 5)	8	1	1	0	0	0	0	1	1	0
c(1, 2, 3, 4)	6	1	2	0	0	1	0	0	0	0
c(1, 2, 3, 5)	0	0	0	0	0	0	0	0	0	0
c(1, 2, 4, 5)	17	26	27	19	15	4	3	1	0	0
c(1, 3, 4, 5)	25	25	19	25	26	3	1	0	0	0
c(2, 3, 4, 5)	0	0	0	0	0	0	0	0	0	0
c(1, 2, 3, 4, 5)	2	1	2	2	0	0	0	0	0	0

4.2. Results

Both Table 1 and Table 2 show that, as the number of observations τ increases, the relative frequency that the GIC selection procedure picks the correct model (1, 4, 5) gets closer to 1. It validates consistency of the GIC procedure. Both tables show that the logarithm penalty often overestimates the model, whereas the square-root penalty both overestimates and underestimates frequently.

Both tables also show that the probability of selecting the correct model under the square-root penalty goes to 1 faster than under the logarithm penalty. This phenomenon can be explained by theoretical bounds on the convergence rates of GIC's error probability given in Shao (1998) and Zhang (1993). They show that the rate at which the probability of choosing wrong models by GIC goes to zero is an inverse function of the penalty λ_τ. Since $\sqrt{\tau}$ increases faster than $\log(\tau)$, the error probability goes to zero faster under the penalty $\sqrt{\tau}$ than that under the penalty $\log(\tau)$. Practically, this suggests that the square-root penalty is preferable when the sample size ranges from moderate to large. For small samples, the square-root penalty does not seem to have an advantage over the logarithm penalty.

Table 3 confirms the above observations in the first two tables. In addition, it shows that signal-to-noise ratio has an important effect on which model the GIC procedure chooses. Since the signal of the fifth covariate is weak, the GIC procedure has difficulty in picking up the fifth covariate and tends to choose the model with only the first and fourth covariates. From a practical point of view, since the magnitude of the fifth coefficient is small relative to not only other non-zero coefficients but also the possible values of random noise, the difference between the true model and the model with only the first and the fourth covariate is not likely to be significant. Table 3 also shows the difference in performance between the logarithm penalty and the square-root penalty when the sample size ranges from moderate to large. For example, for the sample size of 120 (or 240), the relative frequency of selecting the correct model is 79% (or 94%) for the logarithm penalty and 48% (or 73%) for the square-root penalty. In summary, the square-root penalty has more difficulty than the logarithm penalty in selecting the correct model when some coefficients have weak signal-to-noise ratio. The simulation result suggests that the logarithm penalty is more appropriate in applications where a signal-to-noise ratio is likely to be weak.

5. Application

In this section, we apply GIC to build a tracking portfolio for the purpose of measuring long-term post-event abnormal stock return. A great interest in

learning the long-term impact of corporate events has recently arisen among finance researchers and has generated a still growing literature. The evidence for the existence of long-term post-event abnormal stock returns challenges the belief that the U.S. stock market is efficient, and motivates research in behavioral finance. See Fama (1998) for a summary of the literature with references. In these studies, the most important job is to precisely estimate the event firm's unobservable *status-quo* post-event returns, which is the return the event firm would have got if the event had not happened. In the following, we compare the performance of three estimation methods.

5.1. Measures of abnormal return

The three-year buy-and-hold abnormal return of firm i is

$$AR_i = R_i - BR_i , \tag{5.1}$$

where R_i is the buy-and-hold return of firm i over a time period of three years, and BR_i is a specific benchmark return over the same period. The period of three years is a common choice in finance literature. The benchmark return serves as an estimate of the unobservable *status-quo* return that an event firm would have had over the three years following the event month if the event had not happened. The three-year buy-and-hold return of firm i is computed by compounding monthly returns, i.e., $R_i = \prod_{t=1}^{36}(1 + r_{it}) - 1$, where r_{it} is firm i's return in month t.

We consider three benchmarks. The *first benchmark* is a size and book-to-market ratio matched portfolio widely used in finance literature, see, e.g., Dharan and Ikenberry (1995), Desai and Jain (1997), Barber and Lyon (1997), Lyon, Barber and Tsai (1999) and Mitchell and Stafford (2000). To identify the size and book-to-market ratio matched portfolio of an event firm, we construct 70 reference portfolios on the basis of firm size and book-to-market ratio and choose the one that includes the event firm as the matched portfolio. The 70 reference portfolios are formed as follows.

Step 1. At the end of June of year t, we calculate firm size as price per share multiplied by shares outstanding, sort all NYSE firms by firm size into ten portfolios of equal size, and then place each AMEX/Nasdaq firm in the portfolio whose range of firm sizes covers the firm's size.

Step 2. We partition the smallest size decile portfolio into five subportfolios of equal size on the basis of firm size rankings of all firms in the portfolio without regard to listing exchange, so that we have 14 firm size portfolios.

Step 3. We divide each of the 14 portfolios into five subportfolios of equal size by ranking all firms in the portfolio by their book-to-market ratios at the end of year

$t-1$, so that we end up with 70 reference portfolios. A firm's book-to-market ratio at the end of year $t-1$ is equal to the ratio of the book common equity (COMPUSTAT data item 60) at the end of the firm's fiscal year ending in year $t-1$ over the firm's market common equity at the end of December of year $t-1$. Throughout the procedure, we include only stocks with ordinary common equity shares (CRSP share code 11) and exclude firms of negative book common equity whenever book equity is needed.

The benchmark return based on a size and book-to-market ratio matched portfolio is computed as follows:

$$BR_i^{SZBM} = \prod_{t=1}^{36} \left[1 + \frac{\sum_{j=1}^{n_t} r_{jt}}{n_t} \right] - 1 , \qquad (5.2)$$

where r_{jt} is the monthly return of firm j in month t and n_t is the number of firms in month t. We label the first benchmark as B1:SZBM and denote its return by the superscript "SZBM".

The *second benchmark* is a portfolio of the ten firms that have the largest sample correlation coefficients with the event firm among all firms in the size and book-and-market ratio matched portfolio. To identify the ten firms, we first choose the size and book-and-market ratio matched portfolio for the event firm as described above, and identify all firms in the portfolio that have returns in the five years before the event month and in the three years after the event month. We then calculate the sample correlation coefficient between each identified firm and the event firm using the 60 monthly returns in the pre-event five years. At last, we choose the ten firms that have the largest correlation coefficients. We label the second benchmark as B2:MC10 for the most correlated ten, and compute the three-year post-event benchmark return as follows:

$$BR_i^{MC10} = \sum_{j=1}^{10} \frac{\left[\prod_{t=1}^{36} (1 + r_{jt}) \right] - 1}{10} , \qquad (5.3)$$

where r_{jt} is the monthly return of firm j in month t. The benchmark return is the return of investing equally in the ten most correlated firms over the three years starting with the event month.

The *third benchmark* is obtained using the GIC procedure. After obtaining the ten most correlated firms as described above, we apply the GIC selection procedure with the event firm as the response variable and the ten firms as the covariates. We use the 60 pre-event monthly returns and the logarithm penalty in the selection procedure. There are $1023 (= 2^{10} - 1)$ possible subsets for ten covariates. We select the subset that minimizes the GIC criterion and form a

tracking portfolio accordingly. We label the third benchmark as B3:GIC and compute the three-year benchmark return as

$$BR_i^{GIC} = \sum_{j=1}^{n_i} w_j \left[\prod_{t=1}^{36}(1 + r_{jt}) - 1 \right] , \qquad (5.4)$$

where r_{jt} is the monthly return of firm j in month t, n_i is the number of firms in the GIC optimal tracking portfolio, and w_j is the optimal weight of firm j in the GIC optimal tracking portfolio.

5.2. Empirical assessment of performance

To assess the performance of the three benchmarks, we employ a procedure that uses actual security return data to examine the characteristics of abnormal returns produced by the three benchmarks. This type of procedure has been used widely in finance literature to compare performance of various methodologies for measuring abnormal returns, see, e.g., Brown and Warner (1980), Kothari and Warner (1997), Barber and Lyon (1997) and Lyon, Barber and Tsai (1999).

In this procedure, we randomly choose, with replacement, a sample of 200 event months between July 1984 and December 1994, inclusively. For each chosen event month, we then randomly choose, without replacement, an event firm that has returns in the five years before the event month and in the three years after the event month (there are generally several thousand firms in each month that satisfy the requirement). We apply the above three benchmarks to compute three-year post-event abnormal returns for each event firm. Table 4 reports sample mean, median, standard deviation (St. D.), inter-quartile range (IQR), skewness coefficient and kurtosis coefficient of 200 abnormal returns under each benchmark. As evident in Table 4, abnormal returns under different benchmarks have different distributions. In particular, benchmarks B1:SZBM and B2:MC10 produce large negative medians and highly positive skewness coefficients.

Since the 200 event firms are randomly selected and not many of the 200 event months were supposed to experience any event, they are expected to have zero abnormal returns. In other words, the 200 abnormal returns are expected to concentrate around zero. We apply three statistical tests to test whether the central tendency of these abnormal returns is around zero: Student's t test, Fisher's distribution$-$free sign test, and Wilcoxon's signed rank test (see, e.g., Hollander and Wolf (1999, Chap. 3)). P-values from all three tests are reported in the last three columns of Table 4. The t test shows that none of the three benchmarks produce a significant mean abnormal return. The sign test tells that benchmarks B1:SZBM and B2:MC10 produce significantly negative medians but the benchmark B3:GIC has an insignificant median. The rank test reveals that

benchmark B1:SZBM has a significantly negative median whereas benchmarks B2:MC10 and B3:GIC have insignificant medians.

Table 4.Comparison between the three benchmarks based on 200 randomly selected firms.

	Descriptive Statistics						p-values		
	mean	median	St.D.	IQR	skewness	kurtosis	t	sign	rank
B1:SZBM	-0.022	-0.256	1.180	0.957	1.437	6.982	0.794	0.009	0.023
B2:MC10	0.054	-0.147	1.145	0.995	1.389	6.775	0.503	0.028	0.169
B3:GIC	0.084	-0.030	1.381	0.941	0.115	9.856	0.390	0.621	0.900

As the three tests tell different stories, what conclusion might we draw from the empirical evidence? Summary statistics in Table 4 suggest that both normality and symmetry are violated by abnormal returns for the benchmarks B1:SZBM and B2:MC10. The sign test does not rely on the assumption of normality or symmetry of the underlying distribution and, according to it, only benchmark B3:GIC produces abnormal returns that concentrate around zero, consistent with our expectation. We suggest that benchmark B3:GIC gives more precise estimates of the unobservable *status-quo* post-event returns than do the other two benchmarks.

6. Discussion

In this paper, EGIC for dependent observations has not been employed in simulation and empirical studies because only monthly stock returns are involved. It is well documented in empirical finance literature that monthly stock returns have insignificant autocorrelation, while daily returns appear to be negatively autocorrelated (see, e.g., Campbell, Lo and MacKinlay (1997, Chap. 2), and the references therein). EGIC may be more appropriate in a study on how to build optimal portfolios to track daily movements of a chosen financial index. A study on tracking a financial index also provides an opportunity to investigate choice of penalties empirically. For instance, we may construct two index funds with the logarithm penalty and the square-root penalty respectively, and then compare performance of the two index funds in terms of how closely each fund tracks the target index.

Another direction of future research is to apply the GIC procedure to measure the long-term post-event abnormal returns of firms that have experienced real corporate events. Such applications will reveal the impact of different types of events on stock returns and improve our understanding of financial markets.

Acknowledgements

Appendix. Proofs of the Main Results

Proof of Theorem 3.1. The loss function $L_\tau(v)$ can be decomposed as

$$
\begin{aligned}
\tau L_\tau(v) &= ||\boldsymbol{\mu}_\tau - \hat{\boldsymbol{\mu}}_\tau(v)||^2 = ||\boldsymbol{\mu}_\tau - (\boldsymbol{\eta}_\tau(v) + \boldsymbol{H}_\tau(v)\boldsymbol{y}_\tau)||^2 \\
&= ||\boldsymbol{\mu}_\tau - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau - \boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau||^2 \\
&= ||\boldsymbol{\mu}_\tau - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau||^2 + \boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau \\
&\quad - 2\boldsymbol{e}_\tau'\boldsymbol{H}_\tau'(v)[\boldsymbol{\mu}_\tau - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau - \boldsymbol{\eta}_\tau(v)] \\
&= ||\boldsymbol{\mu}_\tau - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau||^2 + \boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau \ .
\end{aligned}
$$

The second and third equalities hold because of (2.7) and (2.1), respectively. The fourth equality holds because $\boldsymbol{H}_\tau'(v)\boldsymbol{H}_\tau(v) = \boldsymbol{H}_\tau(v)$. The last equality holds because

$$
\boldsymbol{H}_\tau'(v)[\boldsymbol{\mu}_\tau - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau - \boldsymbol{\eta}_\tau(v)] = \boldsymbol{H}_\tau'(v)(I - \boldsymbol{H}_\tau(v))(\boldsymbol{\mu}_\tau - \boldsymbol{X}_1(v)\boldsymbol{G}_1^{-1}\boldsymbol{g}(v)) = 0 \ .
$$

For $v \in V^c$, since $\boldsymbol{\mu}_\tau = \boldsymbol{X}_\tau(v)\boldsymbol{\beta}_\tau(v)$ and $E(\boldsymbol{e}_\tau) = 0$, the least squares estimate $\hat{\beta}$ at (2.7) is unbiased. By taking expectation on both sides of equation (2.7), we obtain $\boldsymbol{\mu}_\tau = \boldsymbol{X}_\tau(v)\boldsymbol{\beta}_\tau(v) = \boldsymbol{\eta}_\tau(v) + \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau$. Therefore, $\Delta_\tau(v) = ||\boldsymbol{\mu}_\tau - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau||^2/\tau = 0$.

The expression for the expected average squared error is

$$
\begin{aligned}
R_\tau(v) &= E(L_\tau(v)) = \Delta_\tau(v) + E((\boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau))/\tau \\
&= \Delta_\tau(v) + tr(E(\boldsymbol{e}_\tau\boldsymbol{e}_\tau')\boldsymbol{H}_\tau(v))/\tau = \Delta_\tau(v) + tr(\boldsymbol{\Psi}_\tau\boldsymbol{H}_\tau(v))/\tau \ .
\end{aligned}
$$

Proof of Lemma 3.1. Because of the special structure of $\boldsymbol{\Psi}_\tau$, we have

$$
\begin{aligned}
|\boldsymbol{a}'\boldsymbol{\Psi}_\tau\boldsymbol{b}| &= \left|\sum_{k=1}^{\tau}\sum_{l=1}^{\tau} a_k b_l \gamma_{k-l}\right| = \left|\gamma_0\sum_{l=1}^{\tau} a_l b_l + \sum_{i=1}^{\tau-1}\left(\gamma_i\sum_{l=1}^{\tau-i}(a_{l+i}b_l + a_l b_{l+i})\right)\right| \\
&\leq \gamma_0\sum_{l=1}^{\tau}|a_l b_l| + \sum_{i=1}^{\tau-1}\left(|\gamma_i|\sum_{l=1}^{\tau-i}(|a_{l+i}b_l| + |a_l b_{l+i}|)\right) \ .
\end{aligned}
$$

Notice that for any $i \in \{1, 2, \cdots, \tau - 1\}$,

$$
\sum_{l=1}^{\tau-i} a_{l+i}b_l \leq \sum_{l=1}^{\tau-i} \frac{a_{l+i}^2 + b_l^2}{2} \leq \frac{1}{2}\left(\sum_{l=1}^{\tau} a_l^2 + \sum_{l=1}^{\tau} b_l^2\right) = \frac{||\boldsymbol{a}||^2 + ||\boldsymbol{b}||^2}{2} \ ,
$$

and similarly $\sum_{l=1}^{\tau-i} a_l b_{l+i} \le \frac{1}{2}(||\boldsymbol{a}||^2 + ||\boldsymbol{b}||^2)$. We thus obtain

$$|\boldsymbol{a}'\boldsymbol{\Psi}_\tau \boldsymbol{b}| \le \frac{||\boldsymbol{a}||^2 + ||\boldsymbol{b}||^2}{2}(\gamma_0 + 2\sum_{j=0}^{\infty}|\gamma_j|) = \frac{||\boldsymbol{a}||^2 + ||\boldsymbol{b}||^2}{2}\Upsilon .$$

Proof of Lemma 3.2. Let $\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{I}_r & 0 \\ 0 & 0 \end{bmatrix}$ be a $\tau \times \tau$ matrix where $\boldsymbol{I}_r$ represents the identity matrix of dimension r. Since $\boldsymbol{H}$ is an idempotent matrix of rank r, there exists a $\tau \times \tau$ orthogonal matrix $\boldsymbol{C}$ such that $\boldsymbol{C}'\boldsymbol{H}\boldsymbol{C} = \boldsymbol{\Lambda}$. Then we have $tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}) = tr(\boldsymbol{\Psi}_\tau \boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}') = tr(\boldsymbol{C}'\boldsymbol{\Psi}_\tau \boldsymbol{C}\boldsymbol{\Lambda}) = \sum_{k=1}^{r} \boldsymbol{c}_k'\boldsymbol{\Psi}_\tau \boldsymbol{c}_k$, where $\boldsymbol{c}_k$ is the kth column vector of the matrix C. Since C is orthogonal, $\boldsymbol{c}_k'\boldsymbol{c}_k = 1$ for $k = 1, 2, \cdots, \tau$, and thus we know $tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}) \le r\Upsilon$ by Lemma 3.1.

Notice that $tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}) = tr(\boldsymbol{\Lambda}\boldsymbol{C}'\boldsymbol{\Psi}_\tau \boldsymbol{C}\boldsymbol{\Lambda})$ and $\boldsymbol{\Lambda}\boldsymbol{C}'\boldsymbol{\Psi}_\tau \boldsymbol{C}\boldsymbol{\Lambda}$ is symmetric and non-negative definite. Therefore we have $tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}\boldsymbol{\Psi}_\tau \boldsymbol{H}) \le [tr(\boldsymbol{\Psi}_\tau \boldsymbol{H})]^2 \le (r\Upsilon)^2$.

Proof of Theorem 3.2. We use arguments similar to those in Li (1987) and Shao (1997).

First, note that the EGIC procedure is to minimize

$$\Phi_\tau(v) = \frac{S_\tau(v)}{\tau} + \frac{\lambda_\tau tr(\hat{\boldsymbol{\Psi}}_\tau \boldsymbol{H}_\tau(v))}{\tau}$$
$$= \frac{||\boldsymbol{e}_\tau||^2 + ||\boldsymbol{\mu}_\tau(v) - \hat{\boldsymbol{\mu}}_\tau(v)||^2 + 2\boldsymbol{e}_\tau'(\boldsymbol{\mu}_\tau(v) - \hat{\boldsymbol{\mu}}_\tau(v))}{\tau} + \frac{\lambda_\tau tr(\hat{\boldsymbol{\Psi}}_\tau \boldsymbol{H}_\tau(v))}{\tau} .$$

For $v \in V^c$, since $\boldsymbol{\mu}_\tau(v) - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v) = 0$, we have $L_\tau(v) = (\boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau)/\tau$ and thus

$$\Phi_\tau(v) = \frac{||\boldsymbol{e}_\tau||^2}{\tau} + \frac{\lambda_\tau tr(\hat{\boldsymbol{\Psi}}_\tau \boldsymbol{H}_\tau(v))}{\tau} - \frac{\boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau}{\tau} . \tag{A.1}$$

For $v \in V - V^c$, we have

$$\Phi_\tau(v) = \frac{||\boldsymbol{e}_\tau||^2}{\tau} + \frac{||\boldsymbol{\mu}_\tau(v) - \hat{\boldsymbol{\mu}}_\tau(v)||^2}{\tau} + \frac{\lambda_\tau tr(\hat{\boldsymbol{\Psi}}_\tau \boldsymbol{H}_\tau(v)) - 2tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}_\tau(v))}{\tau}$$
$$+ \frac{2[tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}_\tau(v)) - \boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau]}{\tau} + \frac{2\boldsymbol{e}_\tau'[\boldsymbol{\mu}_\tau(v) - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v)]}{\tau}$$
$$= \frac{||\boldsymbol{e}_\tau||^2}{\tau} + L_\tau(v) + o_p(L_\tau(v)), \tag{A.2}$$

where the last equality holds uniformly in $v \in V - V^c$. To establish the last equality, it suffices to show that

$$\max_{v \in V - V^c} \frac{\boldsymbol{e}_\tau'[\boldsymbol{\mu}_\tau(v) - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v)]}{\tau R_\tau(v)} \xrightarrow{p} 0 , \tag{A.3}$$

$$\max_{v \in V - V^c} \frac{tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}_\tau(v)) - \boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau}{\tau R_\tau(v)} \xrightarrow{p} 0 , \tag{A.4}$$

$$\max_{v \in V - V^c} \frac{\lambda_\tau tr(\hat{\boldsymbol{\Psi}}_\tau \boldsymbol{H}_\tau(v)) - 2tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}_\tau(v))}{\tau R_\tau(v)} \xrightarrow{p} 0 \ , \tag{A.5}$$

$$\max_{v \in V - V^c} \left| \frac{L_\tau(v)}{R_\tau(v)} - 1 \right| \xrightarrow{p} 0 \ . \tag{A.6}$$

We prove (A.3) first. Given any $\varepsilon > 0$, by Chebyshev's inequality and Theorem 3.1, we have

$$P \left\{ \max_{v \in V - V^c} \left| \frac{\boldsymbol{e}_\tau'[\boldsymbol{\mu}_\tau(v) - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v)]}{\tau R_\tau(v)} \right| > \varepsilon \right\}$$

$$\leq \sum_{v \in V - V^c} \frac{E[\boldsymbol{e}_\tau'(\boldsymbol{\mu}_\tau(v) - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v))]^2}{[\tau R_\tau(v)\varepsilon]^2}$$

$$\leq \sum_{v \in V - V^c} \frac{\Upsilon \|\boldsymbol{\mu}_\tau(v) - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v)\|^2}{[\tau R_\tau(v)\varepsilon]^2} \ \leq \ \frac{\Upsilon}{\varepsilon^2} \sum_{v \in V - V^c} \frac{1}{\tau R_\tau(v)} \ .$$

Since the last term tends to 0 under (3.7), we obtain (A.3).

To prove (A.4), given any $\varepsilon > 0$, by Chebyshev's inequality and Lemma 3.2 we have

$$P \left\{ \max_{v \in V - V^c} \left| \frac{tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}_\tau(v)) - \boldsymbol{e}_\tau' \boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau}{\tau R_\tau(v)} \right| > \varepsilon \right\}$$

$$\leq \sum_{v \in V - V^c} \frac{E[tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}_\tau(v)) - \boldsymbol{e}_\tau' \boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau]^2}{[\tau R_\tau(v)\varepsilon]^2}$$

$$= \sum_{v \in V - V^c} \frac{\mathrm{Var}(\boldsymbol{e}_\tau' \boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau)}{[\tau R_\tau(v)\varepsilon]^2} \ \leq \ \left(\frac{m\Upsilon}{\varepsilon} \right)^2 \sum_{v \in V - V^c} \frac{1}{(\tau R_\tau(v))^2} \ .$$

Since $R_\tau(v) > 0$, the last term goes to zero under (3.7), (A.4) holds.

Then, since $tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}_\tau(v))$ is bounded according to Lemma 3.2 and $\hat{\boldsymbol{\Psi}}_\tau$ is a consistent estimator of $\boldsymbol{\Psi}_\tau$, (A.5) holds under (3.7). Finally, (A.6) is equivalent to (A.4) since

$$\left| \frac{L_\tau(v)}{R_\tau(v)} - 1 \right| = \frac{|L_\tau(v) - R_\tau(v)|}{R_\tau(v)} = \frac{|\boldsymbol{e}_\tau' \boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau - tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}_\tau(v))|}{\tau R_\tau(v)} \ .$$

We thus conclude the proof of (A.2).

Next we show the asymptotic loss efficiency and consistency of the EGIC minimizer $\check{v}_\tau$, using (A.1) and (A.2). When V^c is empty, we know from (A.2) that the minimizer of $\Phi_\tau(v)$, $\check{v}_\tau$, is asymptotically equal to the minimizer of $L_\tau(v)$, that is, $\check{v}_\tau$ is asymptotic loss efficient.

When V^c is not empty, we can show that, for any $v^c \in V^c$,

$$\Phi_\tau(v^c) - \frac{\|\boldsymbol{e}_\tau\|^2}{\tau} = o_p(L_\tau(v)) \tag{A.7}$$

uniformly in $v \in V - V^c$, using arguments similar to those in the proofs of (A.5) and (A.4). Equation (A.7), together with (A.2), imply that $\check{v}_\tau$ will always belong to V^c asymptotically if V^c is not empty. Using arguments as in the proof of (A.4), we can further prove

$$\max_{v \in V^c} \frac{e'_\tau H_\tau(v) e_\tau}{\lambda_\tau tr(\hat{\Psi}_\tau H_\tau(v))} \to_p 0 .$$

Then we know from (A.1), for $v^c \in V^c$, $\Phi_\tau(v) - \|e_\tau\|^2/\tau$ is asymptotically dominated by the term $\lambda_\tau tr(\hat{\Psi}_\tau H_\tau(v))/\tau$. Under the assumption that $\hat{\Psi}_\tau$ is a consistent estimator of Ψ_τ and that $tr(\Psi_\tau H_\tau(v))$ converges to a finite limit as $\tau \to \infty$ for any $v \in V^c$, the dominating term $\lambda_\tau tr(\hat{\Psi}_\tau H_\tau(v))/\tau$ has the same minimizer as $L_\tau(v) = e'_\tau H_\tau(v) e_\tau/\tau$ asymptotically. Therefore we obtain

$$P\{\check{v}_\tau \in V^c \text{ but } \check{v}_\tau \neq v_\tau^L\} \to 0 , \tag{A.8}$$

which means that $\check{v}_\tau$ is asymptotically loss efficient when V^c is not empty. Equation (A.8) also implies that $P\{\check{v}_\tau = v_\tau^L\} \to 1$ when V^c is not empty, that is, $\check{v}_\tau$ is consistent.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International symposium on Information Theory* (Edited by B. N. Petrov and F. Csáki), 267-281. Akadémiai Kiado, Budapest.

Barber, B. M. and Lyon, J. D. (1997). Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. *J. Finan. Econom.* **43**, 341-372.

Bollerslev, T. Chou R. Y. and Kroner K. F. (1992). ARCH modeling in finance. *J. Econometrics* **52**, 5-59.

Brown, S. J. and Warner, J. B. (1980). Measuring security price performance. *J. Finan. Econom.* **8**, 205-258.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Inference: A Practical Information-Theoretic Approach.* 2nd edition. Springer Verlag, New York.

Campbell, J. Y. Lo, A. W. and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets.* Princeton University Press, Princeton, New Jersey.

Desai, H. and Jain, P. C. (1997). Long-run common stock returns following splits and reverse splits. *J. Business* **70**, 409-433.

Dharan, B. G. and Ikenberry, D. L. (1995). The long-run negative drift of post-listing stock returns. *J. Finance* **50**, 1547-1574.

Fama, E. (1998). Market efficiency, long-term returns, and behavioral finance. *J. Finan. Econom.* **49**, 283-306.

George, E. I. (2000). The variable selection problem. *J. Amer. Statist. Assoc.* **95**, 1304-1308.

Hocking, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics* **32**, 1-49.

Hocking, R. R. (1985). *The Analysis of Linear Models.* Brooks/Cole Publishing Company.

Hollander, M. and Wolfe D. A. (1999). *Nonparametric Statistical Methods.* 2nd edition. John Wiley, New York.

Kothari, S. P. and Warner, J. B. (1997). Measuring long-horizon security price performance. *J. Financ. Econom.* **43**, 301-339.

Li, K.-C. (1987). Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: Discrete index set. *Ann. Statist.* **15**, 958-975.

Lyon, J. D. Barber B. M. and Tsai C. L. (1999). Improved methods for tests of long-run abnormal stock returns. *J. Finance* **54**, 165-201.

Miller, A. (1990). *Subset Selection in Regression.* Chapman and Hall, London.

Mitchell, M. L. and Stafford E. (2000). Managerial decisions and long-term stock price performance. *J. Business* **73**, 287-329.

Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *Ann. Statist.* **12**, 758-765.

Pötscher, B. M. (1989). Model selection under nonstationary: autoregressive models and stochastic linear regression models. *Ann. Statist.* **17**, 1257-1274.

Rao, C. R. and Wu, Y. (1989). A strongly consistent procedure for model selection in a regression problem. *Biometrika* **76**, 369-374.

Schwartz, G. (1978). Estimating the dimensions of a model. *Ann. Statist.* **6**, 461-464.

Shao, J. (1997). An asymptotic theory for linear model selection. *Statist. Sinica* **7**, 221-264.

Shao, J. (1998). Convergence rates of the generalized information criterion. *J. Nonparametr. Statist.* **9**, 217-225.

Thompson, M. L. (1978a). Selection of variables in multiple regression: Part I, A review and evaluation. *Internat. Statist. Rev.* **46**, 1-19.

Thompson, M. L. (1978b). Selection of variables in multiple regression: Part II, Chosen procedures, computations and examples. *Internat. Statist. Rev.* **46**, 126-146.

Tsay, R. S. (2001). *Analysis of Financial Time Series.* John Wiley, New York.

Zhang, P. (1993). On the convergence rate of model selection criteria. *Comm. Statist.* **22**, 2765-2775.

Division of Banking and Finance, Nanyang Business School, Nanyang Technological University, S3-B1C-91 Nanyang Avenue, Singapore 659798.

E-mail: asjzhang@ntu.edu.sg

Department of Statistics, Florida State University, Tallahassee, FL 32306-4330, U.S.A.

E-mail: niu@stat.fsu.edu

Department of Finance, College of Business, Florida State University, Tallahassee, FL 32306-1110, U.S.A.

E-mail: jang@cob.fsu.edu