Portfolio efficiency tests with conditioning information - Comparing GMM and GEL estimators

Caio A. Vigo Pereira[a,1,*], Márcio P. Laurini[b,2,**]

[a] *Department of Economics - University of Kansas*
[b] *Department of Economics - FEA-RP USP*

Abstract

We evaluate the use of Generalized Empirical Likelihood (GEL) estimators in portfolio efficiency tests for asset pricing models in the presence of conditional information. Estimators from GEL family present some optimal statistical properties, such as robustness to misspecification and better properties in finite samples. Unlike GMM, the bias for GEL estimators do not increase with the number of moment conditions included, which is expected in conditional efficiency analysis. By means of Monte Carlo experiments we show that GEL estimators has better performance in the presence of data contaminations, especially under heavy tails and outliers.

Keywords: Portfolio Efficiency, Conditional Information, Efficiency Tests, GEL, GMM.
JEL: C12, C13, C58, G11, G12.

1. Introduction

In financial economics, the pivotal works of Markowitz (1952) and Sharpe (1964) on portfolio allocation are fundamental pieces that influenced a whole range of subsequent studies. The mean-variance structure had a significant impact in a number of areas, having influence not only in the finance field, in which it has been used in portfolio analysis, asset pricing and corporate finance; but also on analysis of economic policy under uncertainty, labor markets, monetary policy, as well as in hedging and even in inventory problems.

The assessment of investments funds' performance is of major importance in this context. Not only for investors, but also for fund managers, in view that his remuneration may be related with such performance. The efficiency of financial allocations plays a key role in empirical asset pricing framework, with theoretical and practical importance in financial markets. Thus, a fundamental point is to verify empirically if the allocations were efficient conditional to the full set of available information.

[*] The authors acknowledge funding from Capes, CNPq (303738/2015-4) and FAPESP (2018/04654-9).

[1] Corresponding author - 1460 Jayhawk Blvd - Snow Hall, The University of Kansas, USA. email - caiovigo@ku.edu

[2] Corresponding author - Av. dos Bandeirantes 3900, 14040-905, Ribeirão Preto, SP, Brazil. Tel.: +55-16-33290867 - email - laurini@fearp.usp.br

Approaches to construct efficiency tests under the conditional point of view has been quickly developing, being a fundamental reference the work of Ferson and Siegel (2009). The use of conditional information in efficiency tests has several advantages in relation to traditional tests. The first advantage is the incorporation of additional information in the definition of the tests. Which allows one to verify if the allocation was efficient based on the whole set of information available, and not only the information contained in the own returns and a limited set of factors. This structure allows one to verify the impact of dynamic nonlinear strategies on the efficiency of the portfolio. Which is not possible in the tests based on fixed-weight combinations of the tested asset returns, as discussed in Ferson and Siegel (2009).

Although this conditional structure of efficiency tests has several advantages in relation to traditional tests, it introduces some additional complications in terms of statistical inference. Incorporation of conditional information is accomplished through the use of an additional set of instruments in the estimation and testing procedures, which corresponds to the use of a large number of moment conditions. In general, these tests using moments conditions are performed by some version of a Generalized Method of Moments (GMM) estimator. As the GMM estimators do not impose any restrictions on the data distribution, only being based on assumptions about the moments, this method is widely used in finance. Cochrane (2005) also says that GMM structure fits naturally for the stochastic discount factors formulation of asset pricing theories, due to the easiness on the use of sample moments in the place of population moments.

This class of estimators is useful in this context, since its formulation is based only on a moment structure, and does not require the use of the complete process distribution, which is not usually known. However, the performance of these estimators and derived tests can be negatively affected under the conditions in which the conditional tests are performed. The first difficulty is the use of a large number of instruments related to the incorporation of conditional information in the efficiency tests. The use of conditional information, through additional moments, should allow the verification of the impact of the whole set of available information and non-linear strategies on the efficiency of the portfolios. However, an important result is that in instrumental variables estimation by the two-stage and iterated GMM estimators, there is a statistical bias term that is proportional to the number of moments conditions, as shown in Newey and Smith (2004). Thus, efficiency tests based on conditional information through GMM are subjected to a bias component, which grows with the number of moment conditions (conditional information) incorporated in the tests. Hence, the great advantage of conditional tests, which is the incorporation of information, is affected by the presence of this component of bias, damaging the statistical properties of these tests.

Another important limitation is the sensitivity of this method to contaminations, such as the presence of outliers and heavy tails. According to Chaussé (2010), financial data, in particular stock returns, commonly presents heavy-tailed and asymmetric distributions However, GMM estimators can be suboptimal in the presence of data contaminations. The use of higher order moment conditions makes these estimators sensitive to this effects (e.g. Anatolyev and Gospodinov (2011)), and thus these estimators are not robust to these problems.

This study analyses the use of Generalized Empirical Likelihood (GEL) to circumvent the deficiencies existing on the use of usual estimators in testing portfolio efficiency in the presence of conditional information. This class of estimators has some special characteristics that confer better statistical properties, such as robustness to outliers and heavy tails distributions, and better finite sample properties compared to the usual methods based on least squares and Generalized Method of Moments. In Generalized Empirical Likelihood and related methods the bias do not increase as the number of moment conditions grows (e.g. Anatolyev (2005)), what happens with the use of conditional information. Compared to GMM, in finite samples, GEL has a more concise asymptotic bias Newey and Smith (2004); Kitamura and Stutzer (1997). Another important feature is that some estimation methods in the GEL family of estimators have better properties in terms of robustness to contaminations such as outliers, heavy tails and other forms of incorrect specification (e.g., Anatolyev and Gospodinov (2011)).

Because of these characteristics, conditional efficiency tests based on GEL estimators should have better finite sample properties over GMM-based tests. The objective of this work is to verify these properties, and thus to analyze the performance of conditional efficiency tests based on the class of GEL estimators.

Firstly, we study tests robustness with the use of GMM and GEL estimators in a finite sample context. With Monte Carlo experiments we assess the effects that data contaminations, as outliers and the presence of heavy tails in the innovation structure, may cause on the results of efficiency tests. In general, we see that GEL has better performance when heavy tails are present. While, regarding the presence of outliers, both GMM and GEL may have better robustness depending on the DGP we choose to use. We also see evidence that, under the null hypothesis, the tests using either GEL or GMM estimators have a tendency to over-reject the hypothesis of efficiency in finite samples.

In the second part of the work, we evaluate how efficiency tests based on GEL and GMM estimations can lead to different decisions using real datasets. In order to better analyze this issue, we compare test results for different sample sizes and portfolios types, for two asset pricing models, as well as estimations with and without conditional information. The results indicate that, in general, efficiency tests using GEL generate lower estimates when compared to tests using the standard approach based on GMM. Moreover, for the smallest sample in use, which is the one that most resembles features of a finite sample size used in finance, we see that the efficiency tests results are conflicting among GEL and GMM methodologies.

All these results gathered may be an evidence that efficiency tests based on estimators from GEL class perform differently when compared to GMM, especially under small samples. These results extend the findings from Almeida and Garcia (2008) who obtain nonparametric estimates of bounds on the stochastic discount factors accounting for higher moments in the distribution of returns. In Almeida and Garcia (2012), the authors also developed an econometric method based on GEL that provides consistent estimators of information bounds and specification-error bounds that are based on the minimum discrepancy measures .

The structure of this paper is as follows. Next section introduces the methodology, presenting the asset pricing theory and models, as well as the econometric models for portfolios efficiency tests for both estimation methods assessed: GMM and GEL. Section 3 provides the simulations experiments to evaluate tests robustness under both methods of estimation. Section 4 presents the datasets (portfolios, factors and instruments) and empirical results. Finally, section 5 concludes. Additional results, tables and Figures are presented in Appendix.

2. Methodology

The goal is to evaluate an alternative econometric method to the predominant GMM for comparisons of portfolios efficiency in the presence of conditional information. Ferson and Siegel (2009) presented results in this direction, and this work intends to expand their results and applications testing an alternative methodology for this analysis.

The methodology to be tested is based in the Empirical Likelihood estimation. Theoretically, one can expect these estimators to have better statistical properties, such as robustness to data contaminations and better finite sample properties. One essential property of GEL estimators is the absence of some bias components found in GMM. This bias impairs the finite sample performance in the presence of a large number of moment conditions. As the conditional portfolio analysis is based on the incorporation of information in the form of additional instruments, this is the framework where we should suspect that GEL's superior finite sample properties to excel GMM's suboptimal performance, mainly due to the bias components in the latter estimator.

Notice that, when testing portfolios efficiency with the use of conditional information, one should seek to maximize the unconditional mean relative to the unconditional variance, where the portfolio composition strategies are functions of the information matrix. This is the line followed by the *unconditional mean variance efficiency* with respect to the information. It is important to compare this framework with the *conditional efficiency*, where efficiency of the mean-variance structure is evaluated under conditionals means and variance. Note that in the first case, being the approach pursued by this work, the conditional information is used in the construction of the portfolio, and then the efficiency is assessed unconditionally. These properties are explained in the next section.

2.1. Incorporating Conditional Information

Any asset pricing model may be defined following the basic pricing equation:

$$p_t = E_t(m_{t+1}x_{t+1}) \, , \tag{1}$$

where p_t is the asset price, x_{t+1} the asset payoff, and m_{t+1} is the stochastic discount factor (SDF)[3].

[3] The SDF is also known as *change of measure*, *pricing kernel*, or even as *state-price density*.

Notice that the models need to portray the prices taking into account conditional moments. This may be observed in equation (1) when it represents with the use of a conditional expectation the available set of information to the investor in period t of time. Defining Z_t as the set of available information at t, the equation (1) can be rewritten as $p_t = E(m_{t+1}x_{t+1}|Z_t)$.

According to Cochrane (2005), all asset pricing models can be reduced to distinct ways to connect the SDF to the data. Restricting only to assets in the stock class, succinctly, the payoff may be represented as $x_{t+1} = p_{t+1} + d_{t+1}$, where d_{t+1} is the dividend from the asset evaluated. For practical reasons, it is preferable to work with the gross return, i.e., $R_{t+1} \equiv \frac{x_{t+1}}{p_t}$. Thus, follows that the pricing models can be represented in accordance with the fundamental valuation equation:

$$E(m_{t+1}R_{t+1}|Z_t) = 1\ . \tag{2}$$

Assuming that exist a subset of variables $\tilde{Z}_t$ such that $\tilde{Z}_t \subset Z_t$ and multiplying both sides by the elements of $\tilde{Z}_t$ it is possible to get:

$$E_t(m_{t+1}R_{t+1} \otimes \tilde{Z}_t) = 1 \otimes \tilde{Z}_t\ , \tag{3}$$

where $\otimes$ represents a Kronecker product. If we take the unconditional expectation in (3) we get:

$$E(m_{t+1}R_{t+1} \otimes \tilde{Z}_t) = E(1 \otimes \tilde{Z}_t)\ . \tag{4}$$

For the equation (1) it is also possible to incorporate instruments and work with unconditional moments as in $E(m_{t+1}x_{t+1} \otimes \tilde{Z}_t) = E(p_t \otimes \tilde{Z}_t)$. This approach is known as *managed portfolios*, being the product $R_{t+1} \otimes \tilde{Z}_t$ denominated *scaled returns*, and the product $x_{t+1} \otimes \tilde{Z}_t$ as *scaled payoffs*. As the instrument $z_t \in \tilde{Z}_t$ is added in the pricing equation as a product, this approach may also be denominated as *multiplicative approach*. Intuitively, following Ferson and Siegel (2009), the equation (4) asks the SDF to price the dynamic strategy payoffs on average, which may also be understood in an unconditionally form.

Notice that with *managed portfolios* it is possible to incorporate conditional information and still work with unconditional moments. The main advantage of this structure is that there is no need (i) to explicit model the conditional distributions, as well as (ii) it avoids the range problem of the conditional information under assumption. If it was necessary to incorporate conditional information with the use of conditional moments, from (i) would be necessary to formulate parametric models taking the risk of incorrectly defining them; while from (ii) would be necessary to assume that all investors use the same set $\tilde{Z}_t$ of instruments that was included in the conditional model, what clearly incorporates a high degree of uncertainty.

2.2. Estimation Methodology

The use of generalized method of moments (GMM) is fairly common in the estimation of asset pricing models. This happens primarily because with GMM there is no need to impose any distribution regarding the data, requiring only assumptions about the population moment conditions. In addition, for the multiplicative approach, its structure entails that the number of instruments must exceed the moment conditions, justifying the use of the GMM. Then, from equation (1) we can take unconditional expectations to get:

$$\begin{aligned} p_t &= E_t(m_{t+1}x_{t+1}) \\ &= E(m_{t+1}x_{t+1}|Z_t) \\ \Rightarrow \quad E(p_t) &= E(m_{t+1}x_{t+1}) \; . \end{aligned} \tag{5}$$

The asset pricing under unconditional moments must be a specific case of pricing under conditional moments. To do so, we make use the Law of Iterated Expectations.

When we use the asset pricing equation under unconditional moments, the moments conditions necessary for the estimation by GMM become evident. Isolating the terms from equation (5) we can define the errors u_t, so that $u_t = 0$, i.e., $u_t = m_{t+1}R_{t+1} - 1$. Thus, the conditions under unconditional moments can be written as:

$$E(m_{t+1}x_{t+1} - p_t) = 0 \; . \tag{6}$$

Replicating the same procedure in the equation with the gross returns:

$$E(m_{t+1}R_{t+1} - 1) = 0 \; . \tag{7}$$

For *managed portfolios*, the unconditional moments conditions are easily derived from equation (7):

$$\begin{aligned} & E[(m_{t+1}R_{t+1} - 1) \otimes \tilde{Z}_t] = 0 \\ \Rightarrow \quad & E[m_{t+1}(R_{t+1} \otimes \tilde{Z}_t) - (1 \otimes \tilde{Z}_t)] = 0 \; . \end{aligned} \tag{8}$$

Thus, the sample means of u_t are defined as:

$$g_T = \frac{1}{T}\sum_{t=1}^{T} u_t = \frac{1}{T}\sum_{t=1}^{T}[m_{t+1}R_{t+1} - 1] \; , \tag{9}$$

while for the *managed portfolios* approach the sample means of u_t are defined as:

$$g_T = \frac{1}{T}\sum_{t=1}^{T} u_t = \frac{1}{T}\sum_{t=1}^{T}[m_{t+1}(R_{t+1} \otimes \tilde{Z}_t) - (1 \otimes \tilde{Z}_t)] \; . \tag{10}$$

As the moment conditions are nothing more than the difference between observed and expected returns; then, in a graph that plots both returns, the alpha from Jensen (1968) must be the vertical distance between the points and a straight 45° line. Notice that, in order to make use of GMM, all variables that comprise the moment conditions must be jointly stationary and ergodic, besides having finite fourth moments. This case highlights the need to use the equation given in (7), instead of equation (6). In the latter one, even though the payoffs are explicit in the definition of the moments condition, as prices and dividends are expected to rise over time, this fact would cause failure in the stationarity hypothesis.

Finally, denoting by θ as the vector of parameters to be estimated, the GMM estimator can be defined as:

$$\hat{\theta}_T(\hat{W}) \equiv arg \min_{\hat{\theta}} g_T(\hat{\theta})'\hat{W}_T g_T(\hat{\theta}) \; , \tag{11}$$

where, $\hat{W}$ is the conventional positive weighting matrix $q \times q$, for q moment conditions from GMM estimation.

2.2.1. Empirical Likelihood Estimation

Smith (1997) and Owen (2001) introduced a family of estimators denominated *Generalized Empirical Likelihood* (GEL). Just like GMM, this class of estimators can be expressed in the form of moment conditions. According to Anatolyev and Gospodinov (2011), GEL is a non-parametric method having the important attractive of optimal asymptotic and finite samples properties. According to them, the class of GEL estimators leads to a better understanding regarding the properties of the estimators based on moments and allow more powerful tests, more efficient estimation of the density and distribution functions, and better bootstrap methods.

Following Anatolyev and Gospodinov (2011), consider a system of restrictions on unconditional moments, such as:

$$E[g(w, \theta_0)] = 0 \; . \tag{12}$$

where $\theta \in \Theta$ is a $k \times 1$ vector of the true parameters, w is a vector of observables, $\{w_i\}_{i=1}^n$ is a random sample, and $g(w, \theta)$ is a vector $q \times 1$ of the moments conditions. Let $\boldsymbol{p} = (p_1, p_2, \ldots, p_n)$ be a collection of probability weights assigned to each sample observation. Thus, we have the following *empirical likelihood* problem:

$$\begin{array}{ll} \max_{\boldsymbol{p},\theta} & \frac{1}{n}\sum_{i=1}^n log(p_i) \\ \text{subject to} & \sum_{i=1}^n p_i g(w_i, \theta) = 0 \\ & \sum_{i=1}^n p_i = 1 \; . \end{array}$$

From this constraint maximization we obtain the *saddlepoint problem*, given by:

$$\max_{\theta \in \Theta} \min_{\lambda} \frac{1}{n} \sum_{i=1}^{n} -log(1 + \lambda' g(w_i, \theta)) \; . \tag{13}$$

From the solution of this problem it is possible to obtain the Empirical Likelihood estimator $\hat{\theta}$ (as well as the *GEL multipliers* $\hat{\lambda}$). If the substitution is made in the *saddlepoint problem* in (13) by an arbitrary criterion that is subject to certain shape conditions, one can obtain the *GEL estimator*. To do so, let $\rho(\upsilon)$ be a strictly concave smooth function which satisfies $\rho(0) = 0$, $\partial\rho(0)/\partial\upsilon = \partial^2\rho(0)/\partial\upsilon^2 = -1$. Thus, the GEL estimator is given by $\hat{\theta}$, and the GEL multipliers by $\hat{\lambda}$, which are the solution of the *saddlepoint problem* below:

$$\min_{\theta \in \Theta} \sup_{\lambda \in \Lambda_n} \sum_{i=1}^{n} \rho(\lambda' g(w_i, \theta)) \,, \tag{14}$$

where, $\Lambda_n = \{\lambda : \lambda' m(w_i, \theta) \in \Upsilon, i = 1, \ldots, n\}$ and Υ is some open set containing zero (Newey and Smith, 2004).

GEL moments conditions can be modified to incorporate serially correlated data. This approach is know as *smoothed generalized empirical likelihood (SGEL)* (Kitamura, 1997; Kitamura and Stutzer, 1997; Smith, 1997, 2000). Let $\{w_t\}_{t=1}^{n}$ be a strictly stationary and ergodic time series. The smoothed moment conditions can be written as:

$$g_t^w(\theta) = \sum_{j=t-n}^{t-1} \omega(j)(g(w, \theta_0)) \,, \tag{15}$$

and the system of weights is given by $\sum_{j=-\infty}^{\infty} \omega(j) = 1$ and $\omega(j) = \frac{1}{b} K\,(j/b)$, where $K(u) : (-\infty, \infty) \to \mathbb{R}$ is a symmetric, continuously differentiable kernel function, with $K(0) \neq 0$ and $\int K(u) du = 1$, and b is a bandwidth. Replacing the moments $g(w, \theta_0)$ in the *saddle point problem* in (14) by our smoothed moment $g_t^w(\theta)$ given in equation (15) we can obtain the $\hat{\theta}_{SGEL}$ estimator.

Anatolyev and Gospodinov (2011) demonstrate how both estimators class, GEL and GMM, have equivalent asymptotic properties. The authors point out the asymptotic normality of GEL estimators. To do so, they invoke the Central Limit Theorem and show that for $\sqrt{n} \sum_{i=1}^{m} g(w, \theta_0)$ implies that the GEL estimator θ:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Omega_\theta) \,. \tag{16}$$

Thus, under overidentification conditions, the elements of the multiplier estimator $\hat{\lambda}$ are linearly dependent. The authors also show that the GEL estimators of θ_0 and λ_0 are asymptotically independent.

However, even if GEL and GMM estimators have identical asymptotic properties, in finite samples they exhibit different behaviors. According to Newey and Smith (2004), a good way to evaluate the bias of these estimators is through the analysis of how many terms compose the analytical expressions of the second order asymptotic bias, and also by the precise manner in which their magnitudes are related to the number of moments constraints from the model.

Therefore, the examination should focus on the analysis of the higher-orders asymptotic bias expressions. Newey and Smith (2004) derive this higher-order asymptotic bias for the i.i.d. case using a random sample for both GEL and GMM estimators. They conclude that GEL estimation is preferable to GMM because GEL has one less term in its second order asymptotic bias expression. Moreover, the authors also demonstrate a practical implication when there is a considerable quantity of instruments. Under this situation would not be recommended to select many instruments on a GMM estimation to avoid inflating the bias.

Anatolyev (2005) get similar conclusions when comparing the second order asymptotic bias for GEL and GMM estimators in time series models. In summary, estimations based via GEL implies that, in opposition to GMM, the bias should not increase as the number of moment conditions grows.

Hence, using the criterion of bias assessment from Newey and Smith (2004), GEL estimator shows to be the best compared to GMM estimator. According to Anatolyev and Gospodinov (2011), this higher order asymptotic superiority from GEL estimator is usually attributed to its one-step nature, as opposed to the multi-step of GMM.

Succinctly, one can say that the estimation by GEL method seeks to minimize the distance between the vector of probabilities p and the empirical density $1/n$ in equation (13). Each of the estimators which are within the GEL class use different metrics to measure the distance. Owen (2001) defines the *Empirical Likelihood* (EL), where $\rho(\upsilon) = ln(1-\upsilon)$. Kitamura and Stutzer (1997) developed the estimator *exponential tilting* (ET), where $\rho(\upsilon) = -exp(\upsilon)$. Finally, we have the *continuous updated estimator* (CUE), where $\rho(\upsilon)$ is a quadratic function. The CUE was developed by Hansen et al. (1996), but it was Newey and Smith (2004) who showed that this estimator can also be classified in the GEL family.

2.2.2. HAC Estimation

The equation (11) that defines the GMM estimator may have an alternative representation. Let θ_0 be the true value, and assume θ_0 is an interior point of the parameter space Θ, so that $\theta \in \Theta \subset \mathbb{R}^p$. Let w_t be a vector of observed data, the estimator $\hat{\theta}(\hat{W})$ can be defined as the solution to the p first-order conditions associated with:

$$G_T(\hat{\theta}_T)' W_T g_T(\hat{\theta}_T) = 0 \, , \tag{17}$$

where $G_T(\hat{\theta}_T) = T^{-1} \sum_{j=1}^{t} \partial g(w_j, \theta)/\partial \theta'$, $g(\cdot)$ is a $q \times 1$ vector with q moment conditions, and $q \geq p$.

As the asset returns may not be an i.i.d. process, this fact creates the need to work under serial correlation or even with dependence on returns. Thus, we need to work with estimators that have robust properties for these deviations. One possibility to overcome these situations is to use estimators based on the long-run covariance matrix. To this end, under the assumption of weak stationarity and ergodicity, the long-run covariance matrix Ω can be defined as the optimal matrix W from the equation (11), as follows:

$$W^* = \{lim_{n \to \infty} Var(\sqrt{T} \bar{g}_T(\theta_0)) \equiv \Omega(\theta_0)\}^{-1} \tag{18}$$

with,

$$\Omega(\theta_0) = \sum_{j=-\infty}^{\infty} \gamma(j) \tag{19}$$

where $\gamma(j)$ are the autocovariances defined by $\gamma(j) = E\left[\left(w_t - E(w_t)\right)\left(w_{t-j} - E(w_{t-j})\right)\right]$ for the j-th order. For the multivariate version, the long-run covariance matrix Ω has the widely known expression,

$$\boldsymbol{\Omega}(\boldsymbol{\theta}_0) = \sum_{j=-\infty}^{\infty} \boldsymbol{\Gamma}_j = \boldsymbol{\Gamma}_0 + \sum_{j=1}^{\infty} \left(\boldsymbol{\Gamma}_j + \boldsymbol{\Gamma}_j'\right) = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' \,, \tag{20}$$

where $\boldsymbol{\Lambda}$ is a lower triangular matrix given by the Cholesky decomposition of $\boldsymbol{\Omega}(\boldsymbol{\theta}_0)$, and $\boldsymbol{\Gamma_j}$ is the j-th order autocovariance matrix defined by:

$$\boldsymbol{\Gamma}_j = E(\boldsymbol{g}_t\boldsymbol{g}_{t-j}'), \quad j = 0, \pm 1, \pm 2, \ldots \tag{21}$$

An important estimator class for $\boldsymbol{\Omega}(\boldsymbol{\theta}_0)$ matrix are the non-parametric estimators. As the autocovariances are unknown; we can replace them by their sample autocovariances:

$$\begin{aligned}
&\hat{\boldsymbol{\Gamma}}_j = T^{-1}\textstyle\sum_{t=j+1}^{T} \hat{\boldsymbol{g}}_t\hat{\boldsymbol{g}}_{t-j}', && j = 0, 1, \ldots, T-1 \\
&\hat{\boldsymbol{\Gamma}}_j = T^{-1}\textstyle\sum_{t=j+1}^{T} \boldsymbol{g}(w_t, \hat{\boldsymbol{\theta}}_T)\boldsymbol{g}(w_{t-j}, \hat{\boldsymbol{\theta}}_T)', && \text{for } j \geq 0, \\
&\hat{\boldsymbol{\Gamma}}_j = \hat{\boldsymbol{\Gamma}}'_{-j}, && \text{for } j < 0 \,.
\end{aligned} \tag{22}$$

While it is possible to estimate the corresponding samples, the estimator (22) is not consistent because the number of parameters grows in proportion to the sample size. To overcome this difficulty, Newey and West (1987) and Andrews (1991) formulated the widely used class of non-parametric estimators for the optimal long-run covariance matrix consistent to heteroskedasticity and autocorrelation (HAC) defined by:

$$\hat{\boldsymbol{\Omega}}_{HAC}(\boldsymbol{\theta}_0) = \sum_{j=-(T-1)}^{T-1} k(j/b)\hat{\boldsymbol{\Gamma}}_j \,, \tag{23}$$

where $k(\cdot)$ is a kernel function $k : \mathbb{R} \to \mathbb{R}$ satisfying the properties $k(x) = k(-x)$, $k(0) = 1$, $|k(x)| \leq 1$, being $k(x)$ continuous at $x = 0$ and $\int_{-\infty}^{\infty} k^2(x)dx < \infty$. The necessary conditions for $\boldsymbol{\Omega}_{HAC}(\boldsymbol{\theta}_0)$ to be consistent requires that the bandwidth b grows in a lower rate when compared to the sample size, so that $b \to \infty$ and $b/T \to 0$ when $T \to \infty$.

Some options have been proposed for the kernel function, as well as for the bandwidth in (23). There are a variety of choices for the HAC matrix. Andrews (1991) and Newey and West (1987) propose some possibilities for the kernel function and procedures for the bandwidth selection. Notice that the asymptotic properties of the GMM are not affected by the choice of a kernel or a specific bandwidth.

The estimation of the optimal long-run covariance matrix consistent to heteroskedasticity and autocorrelation $\boldsymbol{\Omega}_{HAC}(\boldsymbol{\theta}_0)$ should be used in the estimation of the parameters via GMM methods. Therefore, the GMM estimator $\hat{\theta}_T$ can be defined as:

$$\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\Omega}}_{HAC}(\boldsymbol{\theta}_0)) \equiv arg \min_{\hat{\boldsymbol{\theta}}} \boldsymbol{g}_T(\hat{\boldsymbol{\theta}})'\hat{\boldsymbol{\Omega}}_{HAC}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{g}_T(\hat{\boldsymbol{\theta}}) \,. \tag{24}$$

2.3. Tests of Efficiency

To test portfolios efficiency within the mean-variance framework, we need to predefine the asset pricing model in use. We chose the two most common linear pricing factors models used in this type of analysis: the CAPM from Sharpe (1964) and Lintner (1965), in addition to the three-factor of Fama and French (1993). It is important to mention that this work has no interest to assess whether any specific asset pricing model correctly price assets; but, as different estimation methodologies may impact on inference regarding efficiency for each model.

For the CAPM model, the SDF can be defined as:

$$m_{t+1} = a + bR^W_{t+1} \ , \tag{25}$$

where a and b are constants. The R^W is defined as the *wealth portfolio*, where generally is used a proxy that reflects the market behavior for empirical works. For the Fama-French three-factor model the SDF is defined as:

$$m_{t+1} = a + b_1 Mkt_{t+1} + +b_2 SMB_{t+1} + +b_3 HML_{t+1} \ , \tag{26}$$

where a and b's are constants, Mkt defines the return of a market proxy, SMB *small minus big* factor, and HML *high minus low* factor. Both models, CAPM and Fama-French, may also be derived in beta approach. To see this fact, before it is necessary to demonstrate that there is a connection between the stochastic discount factor representation in (1) and the beta representation. More precisely, one can say that both formats are equivalent and carry the same information, so that it is possible to move from one representation to another with no loss. At this point, we want to show that $p_t = E_t(m_{t+1}x_{t+1}) \ \Rightarrow \ E(R_{i,t+1}) = \alpha + \beta_{i,m}\lambda_m$. For simplicity, take equation (2) without incorporating conditional information and apply the decomposition of covariance,

$$\begin{aligned} & 1 = E(m_{t+1}R_{i,t+1}) \\ \Rightarrow \ & 1 = E(m_{t+1})E(R_{i,t+1}) + Cov(m_{t+1}, R_{i,t+1}) \ . \end{aligned} \tag{27}$$

Thus,

$$\begin{aligned} & E(R_{i,t+1}) = \frac{1}{E(m_{t+1})} - \frac{Cov(m_{t+1}, R_{i,t+1})}{E(m_{t+1})} \\ \Rightarrow \ & E(R_{i,t+1}) = \frac{1}{E(m_{t+1})} + \left(\frac{Cov(m_{t+1}, R_{i,t+1})}{Var(m_{t+1})}\right)\left(-\frac{Var(m_{t+1})}{E(m_{t+1})}\right) \\ \Rightarrow \ & E(R_{i,t+1}) = R^f + \beta_{i,m}\lambda_m \ , \end{aligned} \tag{28}$$

so that we get the beta representation, where $R^f = \dfrac{1}{E(m_{t+1})}$ denotes a risk-free rate (or zero-beta rate) if present, $\beta_{i,m} = \dfrac{Cov(m_{t+1}, R_{i,t+1})}{Var(m_{t+1})}$, i.e., the regression coefficient of the return $R_{i,t+1}$ on m (SDF), and $\lambda_m = -\dfrac{Var(m_{t+1})}{E(m_{t+1})}$. The pricing model in beta format seeks to explain the variation in average returns across assets to express that expected return should be proportional to the regression coefficient $\beta_{i,m}$. Note that, in this format betas $\beta_{i,m}$ are explanatory variables varying for each asset i, while R^f and λ_m represent, respectively, the intercept and the common slope for all assets i in a cross-section regression. In this approach, assets with higher betas should get higher average returns. Thus, $\beta_{i,\pi}$ is interpreted as the amount of risk that the asset i is exposed to the risk factor π, and the term λ_π is interpreted as the price of such risk exposure.

For simplicity, define $\mathbf{f}_t$ as a vector with dimension $K \times 1$ for the factors that compose each one of the models, and assume that from now on we are working with excess returns ($R^e_{i,t+1} = R_{i,t+1} - R^f_{t+1}$), where it will not be used anymore the superscript notation e. For the Fama-French model, the vector $\mathbf{f}_t$ has dimension 3×1, while for the CAPM it is a scalar. Then, one can write:

$$R_{i,t} = \alpha + \beta_i \mathbf{f}_t + \varepsilon_t, \qquad t = 1, \dots, T \quad ; \text{and} \quad i = 1, \dots, N \,, \tag{29}$$

where, for practical reasons $R_{i,t}$ and $\mathbf{f}_t$ already represent, respectively, the excess returns for the N securities and for the K factors in both models in a sample size T. For a system with N assets we have the following statistical structure for these models:

$$\begin{aligned} &\boldsymbol{R}_t = \boldsymbol{\alpha} + \boldsymbol{\beta} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t \\ \\ &E[\boldsymbol{\varepsilon}_t] = 0 \\ &E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'_t] = \boldsymbol{\Sigma} \\ &Cov[\boldsymbol{f}_t, \boldsymbol{\varepsilon}'_t] = \mathbf{0} \,, \end{aligned} \tag{30}$$

where $\mathbf{R}_t$, $\boldsymbol{\alpha}$ e $\boldsymbol{\varepsilon}_t$ have $N \times 1$ dimension; while $\boldsymbol{f}_t$ has $K \times 1$ dimension, and $\boldsymbol{\beta}$ is a $N \times K$ matrix. Defining $\boldsymbol{\Sigma}$ as the covariance matrix of the disturbances ε_t. The theoretical framework for these asset pricing models implies that the vector $\boldsymbol{\alpha} = \mathbf{0}$. In this case, under the correct pricing assumption, these models can be written as:

$$E(\boldsymbol{R}) = \boldsymbol{\beta} E(\boldsymbol{f}) \,. \tag{31}$$

Therefore, the portfolio defined by K factors derived from a linear pricing model is said to be efficient only when the N estimated intercepts are not jointly statistically significant. One can even say that the N intercepts of regressions must be equal to u_t, i.e., the pricing errors. The test of efficiency to assess whether all pricing errors u_t are jointly equal to zero can be done through a Wald test. The null and alternative hypotheses are given by:

$$\begin{aligned} &H_0 : \boldsymbol{\alpha} = \mathbf{0} \\ &H_A : \boldsymbol{\alpha} \neq \mathbf{0} \,. \end{aligned} \tag{32}$$

Whereas the test statistic is given by:

$$J_{Wald} = \hat{\boldsymbol{\alpha}}' [\boldsymbol{Cov}(\hat{\boldsymbol{\alpha}})]^{-\mathbf{1}} \hat{\boldsymbol{\alpha}} \,, \tag{33}$$

so that under the null hypothesis J_{Wald} must have distribution χ^2 with N degrees of freedom. However, one should remember of the limitation from the Wald test that underlies in the large sample distribution theory. According to Cochrane (2005), the test remains valid asymptotically even if the factor is stochastic and the covariance matrix of the disturbances $\boldsymbol{\Sigma}$ is estimated. If, on the one hand, there is no need to assume that the errors are normally distributed; on the other, this test ignores the sources of variation in finite samples. Supported on the Central Limit Theorem, the test is based primarily on the fact that $\hat{\boldsymbol{\alpha}}$ has Normal distribution.

Gibbons et al. (1989) derive the finite sample distribution of the null hypothesis in which the alphas are jointly equal to zero. In contrast to the J_{Wald} test, this test recognizes sample variation in the estimated covariance matrix of the disturbances $\hat{\boldsymbol{\Sigma}}$. However, the test requires that the errors are normally distributed, homoskedastic and uncorrelated. This test is defined by:

$$J_{GRS} = \frac{T-N-K}{N}\left(1 + E_T(\boldsymbol{f})'\hat{\boldsymbol{\Omega}}^{-1}E_T(\boldsymbol{f})\right)^{-1}\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\alpha}}\,, \tag{34}$$

where it is used the same notation from Hansen and Singleton (1982) in which $E_T(\cdot)$ represents the sample mean, and

$$\hat{\Omega} = \frac{1}{T}\textstyle\sum_{t=1}^{T}[\boldsymbol{f}_t - E_T(\boldsymbol{f})][\boldsymbol{f}_t - E_T(\boldsymbol{f})]'$$

$$\hat{\Sigma} = \frac{1}{T}\textstyle\sum_{t=1}^{T}\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t'\,.$$

Therefore, under i.i.d. and normally distributed errors, the statistic test J_{GRS} has a non-conditional distribution as an F with N degrees of freedom in the numerator, and $T-N-K$ degrees of freedom in the denominator. Note that, assuming $\varepsilon_t \sim N.I.D.$ one can show that $\hat{\boldsymbol{\alpha}}$ has Normal distribution, and $\hat{\boldsymbol{\Sigma}}$ has a Wishart distribution. Precisely:

$$\begin{aligned}&\hat{\boldsymbol{\alpha}} \sim N\left(\boldsymbol{\alpha}, \frac{1}{T}\left[1 + E_T(\boldsymbol{f})'\hat{\boldsymbol{\Omega}}^{-1}E_T(\boldsymbol{f})\right]\boldsymbol{\Sigma}\right)\\ &T\hat{\boldsymbol{\Sigma}} \sim W_N\left(T-2, \boldsymbol{\Sigma}\right)\,,\end{aligned} \tag{35}$$

which, being the Wishart distribution a multivariate χ^2, implies that $\hat{\boldsymbol{\alpha}}'[\boldsymbol{Cov}(\hat{\boldsymbol{\alpha}})]^{-1}\hat{\boldsymbol{\alpha}}$ should result in an F distribution.

2.4. Empirical Strategy

Our goal is to evaluate how GMM and GEL behave when used for efficiency tests in asset pricing models. We analyze both estimators under different specifications for sample sizes, asset pricing models and portfolios formations. Specifically, we are interested in assess their robustness under (i) finite samples, (ii) data contaminations[4], and (iii) with increasing number moments conditions.

To capture higher moments conditions, we make use of the unconditional mean variance efficiency with respect to conditional information approach (Ferson and Siegel, 2009). We

[4]Such as presence of outliers and heavy tails in the data.

use the *managed portfolios* structure presented in Section 2.1, in which the moment conditions are derived in equation (8). The conditional information is added through lagged instruments of the state of the economy. In this situation, the instruments are only used in a multiplicative form in the sample moments. We compare this approach when no information regarding the state of the economy (lagged variables) is used. In this case, with no conditional information, we estimate the pricing equation given in (7).

Firstly, we start the assessment using a simulation approach. In order to evaluate robustness, we analyze the statistical properties of the efficiency tests using GMM and GEL estimations in a finite sample context. To this end, four Monte Carlo experiments are performed to examine the tests robustness to disturbances in the data series when both estimators are used. In particular, we analyze the impact of heavy tails and outliers.

In the second part, we empirically assess both estimators using real data. The methods are evaluated under different time intervals (T) and various portfolios sizes (N). In addition, the analysis is done for both asset pricing models already mentioned: the CAPM and the Fama-French three-factor, using the efficiency tests defined according to section 2.3.

The efficiency tests considered are the Wald and the GRS tests, presented in equations (33) and (34) respectively. A HAC estimator for the covariance matrix is used for the GMM, as presented in equation (24). The GEL estimator is used following Section 2.2.1. In order to deal with serially correlated data, we use smoothed moment conditions as in (15).

3. Evaluating Robustness with Monte Carlo Simulations

In this section we are interested in evaluate the robustness properties of the test statistics using GMM and GEL estimators in a finite sample context. The main goal here is to analyze the size of the Wald and GRS tests under different specifications. The robustness properties are of special interest, since contaminations such as heavy tails and outliers may be present in this type of data.

In our Monte Carlo experiments we restrict the DGP of the artificial returns to be efficient. This is done defining our generating process to be a function of a specific number of factors with no intercept (i.e., setting $\alpha = 0$). Defining different processes for the disturbance term in this DGP, we can generate data with certain features that we are interested to assess. We construct four different scenarios to try to incorporate some patterns seen in real financial data. Then, we analyze the robustness of the estimators through the size properties of tests presented in the previous section.

In order to build a dataset of artificial returns, we use the actual returns from the 6 portfolios formed on size and book-to-market and the factors from Fama-French three-factor model. Seeking to analyze the behavior of our estimators in a finite sample context, we set the sample size to $T = 120$ [5]. We work with *managed portfolios* to assess the impact of a higher number of moment conditions during the estimation process. We use a set

[5] We use monthly data spanning the last 120 months (10 years) prior to Dec/2014.

of 5 standard instruments commonly employed in this type of analysis to form our set of conditional information. In Section 4 we explain in details the portfolios, the factors and the lagged instruments used in this work.

For each portfolio $i = 1, 2, \ldots, 6$ we run OLS regressions of the excess returns $R_{i,t}$ on the three factors from Fama-French model as follows:

$$R_{i,t} = \beta_{i,1} Mkt_t + \beta_{i,2} SMB_t + \beta_{i,3} HML_t + \varepsilon_{i,t}, \quad t = 1, \ldots, T \quad ; \quad i = 1, \ldots, N \, . \tag{36}$$

Thus, for each one the six portfolios, we obtain 3 estimated coefficients of the parameters $\beta_{i,1}, \beta_{i,2}, \beta_{i,3}$. Using these estimates, we build 6 artificial series of returns with 120 observations each defining a process for the disturbance $\hat{\varepsilon}_{i,t}^{\text{Sim}*}$. In summary, our simulations share the following common structure:

$$R_{i,t}^{\text{Sim}*} = \hat{\beta}_{i,1}^{OLS} Mkt_t + \hat{\beta}_{i,2}^{OLS} SMB_t + \hat{\beta}_{i,3}^{OLS} HML_t + \hat{\varepsilon}_{i,t}^{\text{Sim}*}, \quad t = 1, \ldots, 120 \quad ; \quad i = 1, \ldots, 6 \, . \tag{37}$$

All four scenarios have this generating process, where just the disturbance term $\hat{\varepsilon}_{i,t}^{\text{Sim}*}$ is what differentiates each one of them. We carry out 500 artificial returns datasets simulations for each one of the four scenarios. We chose to run 500 simulations, mainly because of the computational burden related to the estimation of the parameters for the efficiency tests (especially GEL method has a high computational cost). Below we describe the four different scenarios in consideration when we define different processes for the error.

Scenario 1 - Gaussian Shocks: The first scenario is our baseline. We seek to assess the efficiency tests for both estimators (GMM and GEL) in the presence of Gaussian innovations. The generating process for $\hat{\varepsilon}_{i,t}^{\text{Sim}*}$ is defined by:

$$\begin{gathered} \hat{\varepsilon}_{i,t}^{\text{Sim}*} = \hat{\xi}_{i,t}^{\text{Sim1}}, \quad t = 1, \ldots, 120 \quad ; \quad i = 1, \ldots, 6 \\ \hat{\xi}_{i,t}^{\text{Sim1}} \sim N(0, \hat{\sigma}_i^{2\ \text{OLS}}) \, . \end{gathered} \tag{38}$$

Scenario 2 - Shocks from a t distribution: In the second scenario we want to evaluate the efficiency tests under the presence of heavy tails. As heavy tails are characterized by more extreme values in the disturbance term, then an appropriate way to model this phenomenon is to use innovations drawn from a t-Student distribution. We set the parameter of this distribution to have 4 degrees of freedom in order to have fatter tails. The DGP for $\hat{\varepsilon}_{i,t}^{\text{Sim}*}$ is given by:

$$\begin{gathered} \hat{\varepsilon}_{i,t}^{\text{Sim}*} = \hat{\nu}_{i,t}^{\text{Sim2}}, \quad t = 1, \ldots, 120 \quad ; \quad i = 1, \ldots, 6 \\ \hat{\nu}_{i,t}^{\text{Sim2}} \sim t(4) \, . \end{gathered} \tag{39}$$

Scenario 3 - Outlier on a fixed date: The third and fourth simulation scenarios seek to evaluate the Wald and GRS tests when outliers are present in data. In the third case, we model the generating process to plug a large magnitude shock on a fixed date in our sample. Arbitrarily, we chose to add an error in the middle of the sample, i.e., when $t = 60$. Following the structure of the previous scenarios, the beta coefficients of each asset in the portfolio are estimated by OLS, and when $t = T/2 = 60$ there is a negative shock of 5 standard deviations randomly drawn from a Normal distribution with variance calculated using the original data. In this case, $\hat{\varepsilon}_{i,t}^{\text{Sim}*}$ is defined as:

$$\begin{gathered} \hat{\varepsilon}_{i,t}^{\text{Sim}*} = \mathbb{1}_{t=T/2}(\hat{\kappa}_{i,t}^{\text{Sim3}}), \quad t = 1, \ldots, 120 \quad ; \quad i = 1, \ldots, 6 \\ \mathbb{1}_{t=T/2}(\kappa_{i,t}^{\text{Sim3}}) = \begin{cases} -\hat{\kappa}_{i,t}^{\text{Sim3}} & \text{, if } t = T/2 \\ 0 & \text{, if } t \neq T/2 \end{cases} \\ \hat{\kappa}_{i,t}^{\text{Sim3}} \sim N(0, 5\hat{\sigma}_i^{2\ \text{OLS}}) \,. \end{gathered} \tag{40}$$

Scenario 4 - Outlier with 5% probability: The fourth scenario takes another direction to simulate outliers. We use a probability process of extreme events, arbitrarily assuming that the probability of an outlier to occur in each period is 5%. In case of success, we add an outlier with 5 standard deviations randomly drawn from a Normal distribution with variance estimated from the original data. In this case, the DGP of $\hat{\varepsilon}_{i,t}^{\text{Sim}*}$ is given by:

$$\begin{gathered} \hat{\varepsilon}_{i,t}^{\text{Sim}*} = \hat{\xi}_{i,t}^{\text{Sim4}} - \mathbb{1}_{\hat{p}_{i,t}<0.05}(\hat{\kappa}_{i,t}^{\text{Sim4}}), \quad t = 1, \ldots, 120 \quad ; \quad i = 1, \ldots, 6 \\ \mathbb{1}_{\hat{p}_{i,t}<0.05}(\kappa_{i,t}^{\text{Sim4}}) = \begin{cases} \hat{\kappa}_{i,t}^{\text{Sim4}} & \text{, if } \hat{p}_{i,t}^{\text{Sim4}} < 0.05 \\ 0 & \text{, if } \hat{p}_{i,t}^{\text{Sim4}} \geq 0.05 \end{cases} \\ \hat{p}_{i,t}^{\text{Sim4}} \sim \text{unif}(0,1) \\ \hat{\xi}_{i,t}^{\text{Sim4}} \sim N(0, \hat{\sigma}_i^{2\ \text{OLS}}) \\ \hat{\kappa}_{i,t}^{\text{Sim4}} \sim N(0, 5\hat{\sigma}_i^{2\ \text{OLS}}) \,. \end{gathered} \tag{41}$$

3.0.1. Sampling Distributions of the Test Statistics

To analyze the results of the Monte Carlo experiments, we use the graphical method proposed by Davidson and MacKinnon (1998). This method is based on the empirical distribution function (EDF) of p-values from Wald and GRS tests, i.e., J_{Wald} and J_{GRS}. Denote the EDF by:

$$\begin{gathered} \hat{F}(x_i) \equiv \tfrac{1}{N} \textstyle\sum_{j=1}^{N} \mathbb{1}_{p_j \leq x_i} \\ \mathbb{1}_{p_j \leq x_i} = \begin{cases} 1 & \text{, if } p_j^* \leq x_i \\ 0 & \text{, if } p_j^* > x_i \,, \end{cases} \end{gathered} \tag{42}$$

where p_j^*, is the p-value of the J tests, i.e., either $p_j^{Wald} \equiv p(J_{Wald})$, or $p_j^{GRS} \equiv p(J_{GRS})$.

Davidson and MacKinnon (1998) propose the *P-value plot*. This graph plots $\hat{F}(x_i)$ against x_i. If the distributions of the tests J_{Wald} and J_{GRS} used to calculate p-values p_j^* are correct; then, each p_j^* must be distributed as a Uniform $(0,1)$. This implies that the $\hat{F}(x_i)$ chart against x_i should be as close as possible to a 45° line. Therefore, one can say that with *P-value plot* it is possible to quickly evaluate statistical tests that systematically over-reject, under-reject or those that reject about the right proportion of the time.

For situations where the tests statistics being studied behave close to the expected behavior, i.e., with graphs being close to the 45° line, the authors proposed the *P-value* *discrepancy plot*. This chart plots $\hat{F}(x_i) - x_i$ against x_i[6]. For the *P-value discrepancy plot*, if the distribution is correct, then each p_j^* must be distributed as a Uniform $(0, 1)$ and the graph of $\hat{F}(x_i) - x_i$ against x_i should be near the horizontal axis.

The results for the first simulated scenario derived from a Gaussian disturbance is shown in Figure 1. The left column represent the Wald test, while the right column refer to the GRS test. The graphs in the first row are the EDF of the tests p-values, the graphs in the second row are the *P-value plot*, while those in the third row are the *P-value discrepancy plot*. In the first two types of charts, the straight 45° line is represented by a dashed line. For *P-value discrepancy plot* the dashed line represents the abscissa axis.

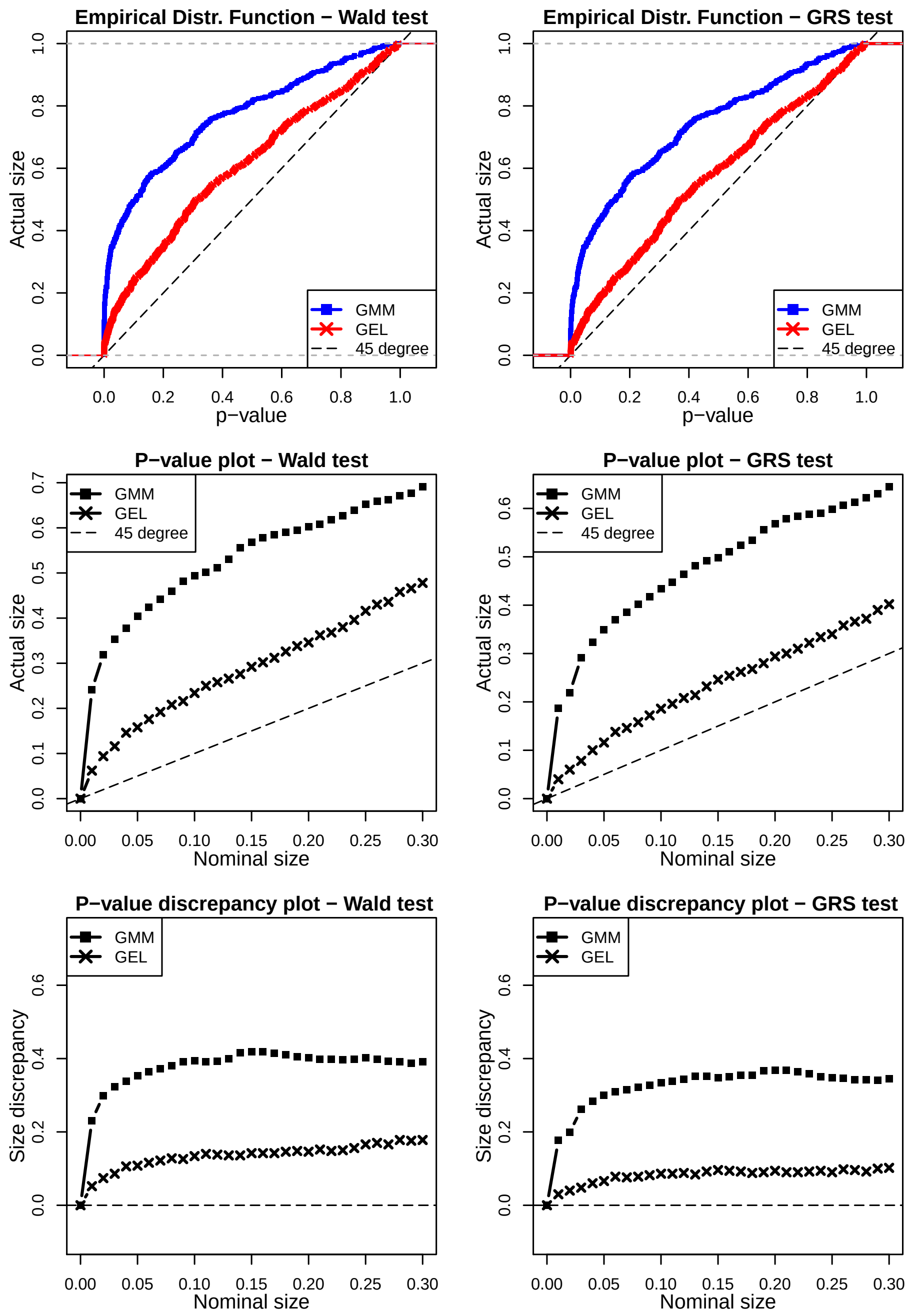
Analyzing the *P-value plot* we note that GEL provides better p-values than GMM, for both, Wald and GRS tests under the null hypothesis. We see that both, GEL and GMM over-reject for any nominal size. For instance, taking 5% nominal size, for Wald test, GMM shows an actual size test of 40.36%, while GEL has less than half of that (15.8%). For the same 5% nominal size, the GRS test derived for finite samples, indeed has better performance for both GMM and GEL. However, GEL still has a better performance. Regarding the *P-value discrepancy plot*, we see similar results. Based on these graphs, it is possible to notice a superiority by GEL compared to GMM for estimating parameters for J_{Wald} and J_{GRS} tests when Gaussian shocks exists. Table 8 in Appendix 5 presents tabulated rejection proportions for the most common nominal sizes, both tests and all four Monte Carlo scenarios.

The results for the second scenario with shocks from a t distribution are presented in Figure 2. The structure of the graphs are the same. In this scenario, by adding a shock from a t distribution, we investigate the tests robustness for data with heavy tail distributions. Clearly, the tests based on GMM perform badly in finite samples for distributions with long tail. For 5% nominal size, the Wald test using the GMM has an actual size of 43.68%, while GEL has slightly more than half of it (23.2%). For the GRS test, the performance of both estimators improve. For the same 5% nominal size, GMM has actual size of 36.47 and GEL 17.8. However, while one can say that GMM has poor performance in finite samples with heavy tails compared to GEL, these results can not hide the fact that both estimators generally over-reject under these circumstances. Even if we consider that GEL performs better, having an actual size of nearly 5 times the 5% nominal size for the Wald test, and an actual size of more than 3 times the 5% nominal size for GRS test, we can not necessarily conclude that their performance are satisfactory.

Figure 3 shows the results for the third scenario, with great magnitudes shocks in the middle of the sample. The goal is to check robustness in the presence of outliers. Here, again the evidence are in the same direction, indicating that GMM has worse performance compared to GEL under the null hypothesis. Note that both estimators always over-reject

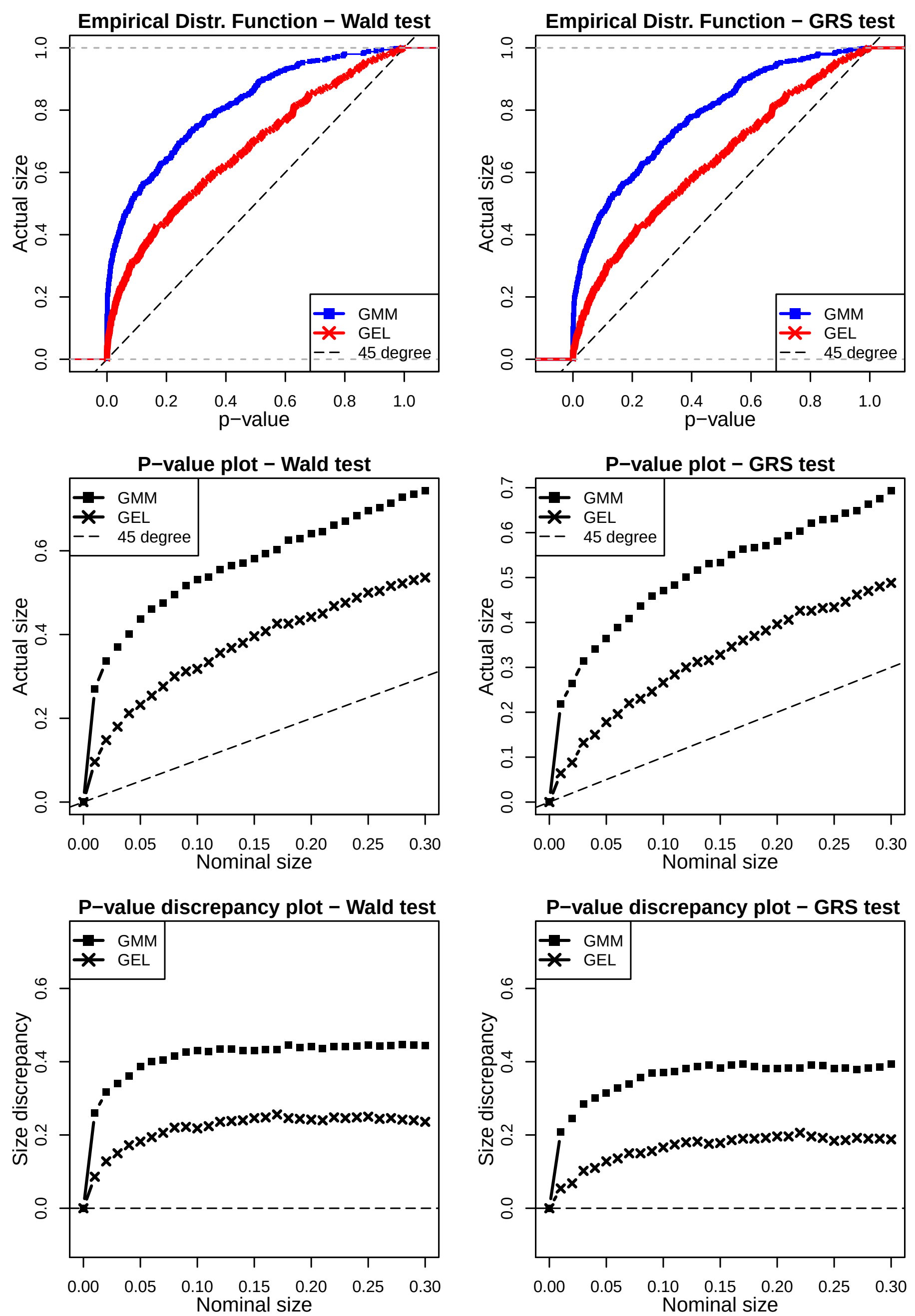
[6]According to the authors, there are advantages and disadvantages in this representation. Among the advantages of this chart, we have that it presents more information than *P-value plot* when statistics of the tests are well behaved. However, this information may be spurious, being just a result of the experiment randomness conducted. Furthermore, there is no natural scale for the vertical axis, which could cause some difficulty in interpretation.

Figure 1: Simulated scenario 1 with Gaussian innovations ($\hat{\varepsilon}_{i,t}^{\text{Sim}*} = \hat{\xi}_{i,t}^{\text{Sim1}}$) in Wald and GRS tests (Model=Fama-French, N=6, T=120, 500 simulations)



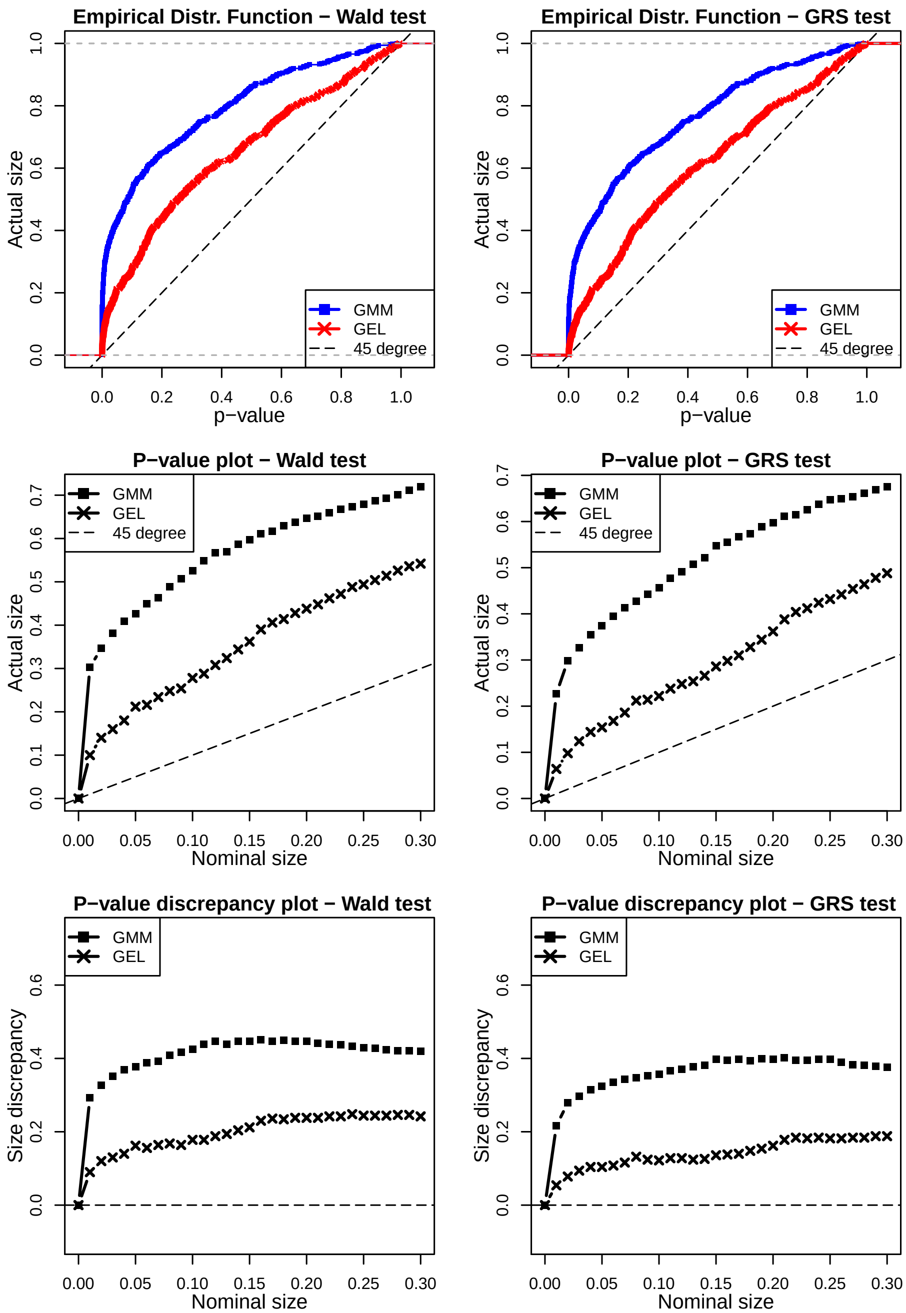
The left column shows simulations for J_{Wald} test, while the right column shows simulations for J_{GRS} test. The top two graphs are the EDF of the p-values obtained via GMM and GEL for both tests. The two graphs in the middle are the *P-value plot*, while the bottom two are the *P-value discrepancy plot*. In order to facilitate visualization, we show in the EDF and *P-value plot* charts a dashed line representing a 45° line. For the *P-value discrepancy plot* the dashed line represents the x-axis.

Figure 2: Simulated scenario 2 with shocks from a t distribution ($\hat{\varepsilon}_{i,t}^{\text{Sim}^*} = \hat{\nu}_{i,t}^{\text{Sim2}}$) in Wald and GRS tests (Model=Fama-French, N=6, T=120, 500 simulations)



The left column shows simulations for J_{Wald} test, while the right column shows simulations for J_{GRS} test. The top two graphs are the EDF of the p-values obtained via GMM and GEL for both tests. The two graphs in the middle are the *P-value plot*, while the bottom two are the *P-value discrepancy plot*. In order to facilitate visualization, we show in the EDF and *P-value plot* charts a dashed line representing a 45° line. For the *P-value discrepancy plot* the dashed line represents the x-axis.

Figure 3: Simulated scenario 3 with shocks at t=T/2 defined by $\hat{\varepsilon}^{\text{Sim}^*}_{i,t} = \mathbb{1}_{t=T/2}(\hat{\kappa}^{\text{Sim3}}_{i,t})$ in Wald and GRS tests (Model=Fama-French, N=6, T=120, 500 simulations)



The left column shows simulations for J_{Wald} test, while the right column shows simulations for J_{GRS} test. The top two graphs are the EDF of the p-values obtained via GMM and GEL for both tests. The two graphs in the middle are the *P-value plot*, while the bottom two are the *P-value discrepancy plot*. In order to facilitate visualization, we show in the EDF and *P-value plot* charts a dashed line representing a 45° line. For the *P-value discrepancy plot* the dashed line represents the x-axis.

when we add a random shock with 5 standard deviations in middle of the sample.

Finally, in Figure 4 we have the results for the fourth scenario, in which we also seek to evaluate robustness to outliers. In this scenario, the innovation comes from a Gaussian distribution, being added a shock with 5 standard deviations from a Normal distribution, when we drawn from a Uniform $(0, 1)$ a value lower than 0.05. Here we have interesting results that differ from earlier ones. The J_{Wald} and J_{GRS} tests based on GMM estimations show better results than via GEL for any nominal size we take. However, note that this superiority is tenuous, being more discernible for nominal values below 10%. Taking 5% nominal size, the Wald test with GMM has an actual size of 90.34%, while GEL has 95.6%. For GRS test, assuming the same 5% nominal size, GMM has 86.5% and GEL 93%. Analyzing the *P-value discrepancy plot* we can note similar pattern with an important feature: for both tests, both GMM and GEL estimations tend to consistently improve performance after reaching a peak of discrepancy around nominal size of 5%.

In summary, analyzing all the results presented in this section, it is possible to observe evidence that efficiency tests in finite samples with GEL estimations tend to have better performance compared to estimations via GMM. Furthermore, tests using GEL are more robust to the presence of heavy tails. To assess the robustness for outliers, depending on the generating process assumed, both GMM and GEL can be advantageously compared with each other. However, these results also demonstrate that, whichever the estimator and the test we evaluate, in general, the Wald and GRS tests have a tendency to over-reject.

4. Empirical Analysis

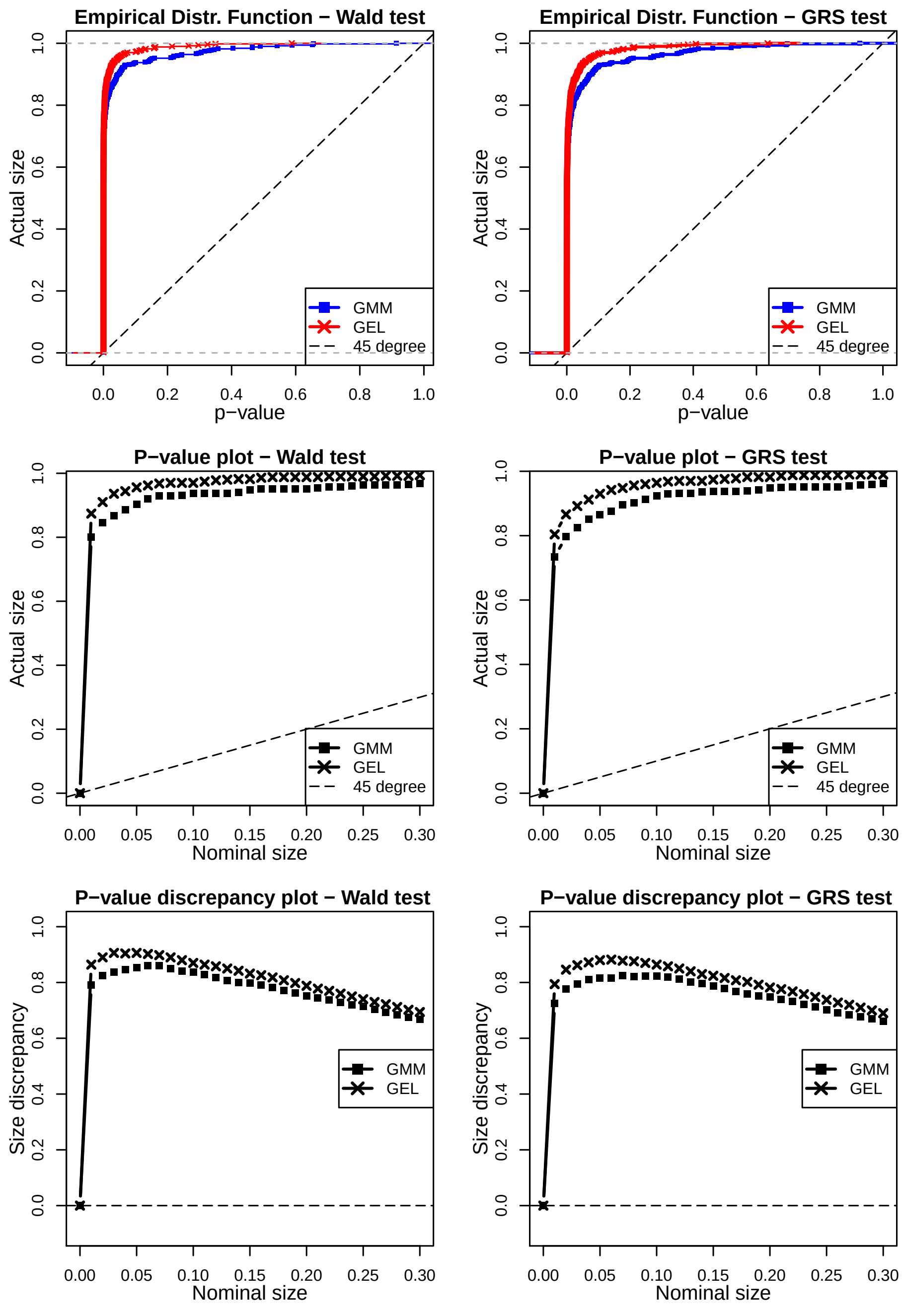
In this session, we evaluate how efficiency tests based on GEL and GMM estimations can lead to different decisions using real datasets. In order to better analyze this issue, we compare the test results for different sample sizes and portfolios types, for two asset pricing models, as well as including or not conditional information. Additional analysis can be found on Appendix 5.

4.1. Instruments and Factors

Following previous studies, we selected a limited number of instruments from those commonly used to measure the state of the economy. One can say that the lagged variables chosen are part of a standard set of instruments for this purpose.

The set of lagged variables consists of 5 instruments. The first is the lagged value of a 3-month Treasury-bill yield (Ferson and Qian, 2004). The second is the spread between corporate bond yields with different ratings. This spread is derived from the difference between the Moody's Baa and Aaa corporate bond yields (Keim and Stambaugh, 1986; Ferson and Siegel, 2009). Another instrument is the spread between the 10-year and 1-year Treasury-bill yield with constant maturity (Fama and French, 1989; Ferson and Siegel, 2009). Following Ferson and Qian (2004), we included the percentage change in the U.S. inflation, measured by the *Consumer Price Index* (CPI). Lastly, the monthly growth rate of the seasonally adjusted industrial production is also used, measured by the *Industrial*

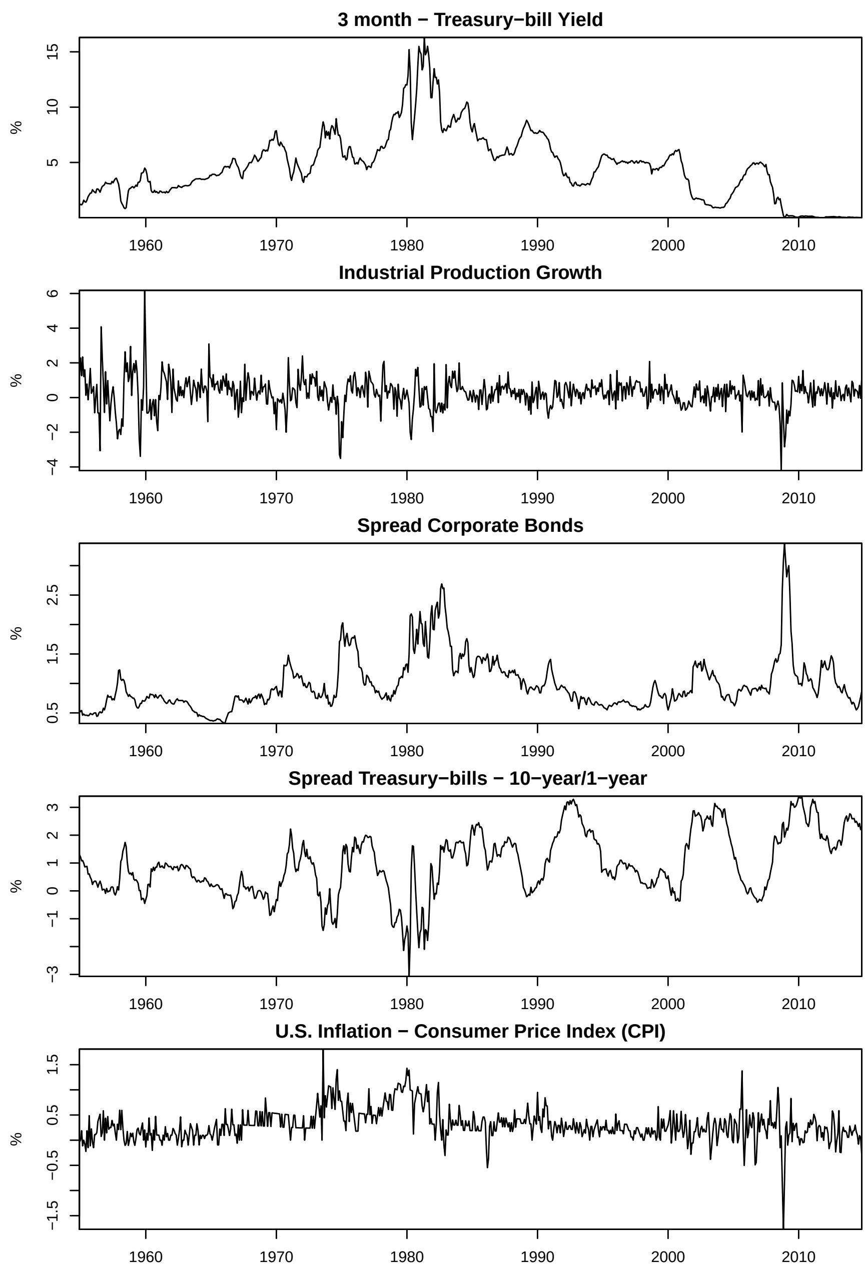
Figure 4: Simulated scenario 4 with shocks defined by $\hat{\varepsilon}_{i,t}^{\mathrm{Sim}*} = \hat{\xi}_{i,t}^{\mathrm{Sim4}} - \mathbb{1}_{\hat{p}_{i,t}<0.05}(\hat{\kappa}_{i,t}^{\mathrm{Sim4}})$ in Wald and GRS tests (Model=Fama-French, N=6, T=120, 500 simulations)



All three left panels are simulations for J_{Wald} test, while the three right panels are simulations for J_{GRS} test. The two top panels are the EDF graphics of the p-values obtained via GMM and GEL for both tests. The two central panels are the *P-value plot*, while the two bottom panels are the *P-value discrepancy plot*. In order to facilitate visualization, it is included in the EDF and *P-value plot* charts a dashed line representing a 45° line. For the *P-value discrepancy plot* the dashed line represents the x-axis.

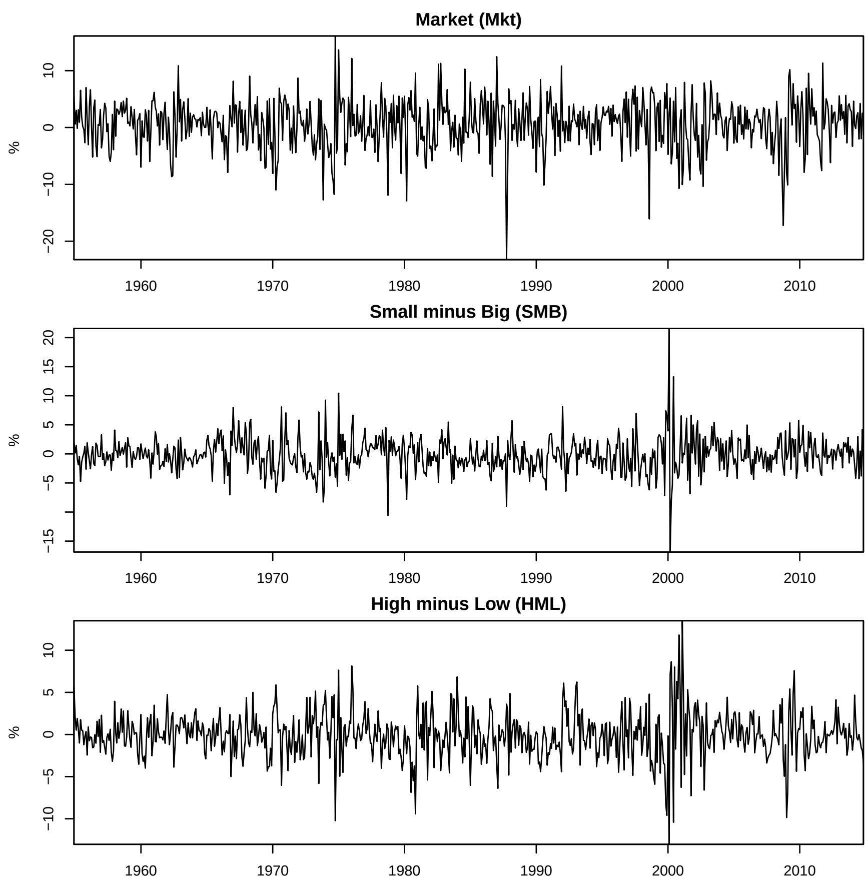
Production Index (Ferson and Qian, 2004). All data were extracted from the historical time series provided by the Federal Reserve.

Figure 5: Historical series of the instruments for 720 months (from jan-1955 to dec-2014)



In view that we focus on the CAPM and Fama-French three-factor model, we extracted

Figure 6: Historical series of the factors for 720 months (from jan-1955 to dec-2014)



the factors for both approaches from Kenneth R. French website[7]. The market portfolio[8] consists of the weighted return of the value of all companies listed on the NYSE, AMEX and NASDAQ[9]. The SMB and HML factors are computed in accordance with Fama and French (1993)[10].

Table 1: Descriptive statistics of the lagged variables and factors for a period of 720 months (60 years) from jan-1955 to dec-2014

	Mean	Std. Dev.	Min	Max	ρ_1
Lagged Variables					
3 month - Treasury-bill Yield	0.047	0.030	0.000	0.163	0.99
Industrial Production Growth	0.002	0.009	-0.042	0.062	0.37
Spread Corporate Bonds	0.010	0.004	0.003	0.034	0.97
Spread Treasury-bills - 10-year/1-year	0.010	0.011	-0.031	0.034	0.97
U.S. Inflation - Consumer Price Index (CPI)	0.003	0.003	-0.018	0.018	0.61
Factors					
Market (Mkt)	0.005	0.044	-0.232	0.161	0.08
Small minus Big (SMB)	-0.002	0.030	-0.169	0.216	0.06
High minus Low (HML)	-0.000	0.027	-0.130	0.135	0.16

Monthly returns of the 5 lagged variables and the 3 factors from asset pricing models. First column presents the sample mean, the second shows the sample standard deviation, the third and fourth column present minimum and maximum returns, and the last column presents the first-order autocorrelation. The sample period is January 1955 through December 2014 (720 observations).

Figures 5 and 6 respectively present the complete historical series of the lagged state variables and factors used. The common maximum time span for all instruments is 720 months (60 years) prior to December 2014. From the graphs of the five instruments, important events in this 60-year range are easily seen through peaks and valleys. The oil crisis and the 2008 financial crisis are examples that impacted the lagged variables of the economy.

Table 1 remarks some descriptive statistics of instruments and factors for the maximum period of 720 months. Observing the first order autocorrelation, we can see that the instruments are highly persistent, while this is not observed for the factors. Note that for most of the five instruments, the first order autocorrelation is 97% or higher. The only

[7] $http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html$

[8] Also known as the *wealth portfolio* from CAPM and Fama-French

[9] Precisely, the market portfolio consists of the value-weight return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11 at the beginning of month t, good shares and price data at the beginning of t, and good return data for t.

[10] The first factor is the average return of 3 smaller portfolios subtracted by the average return of the three largest portfolios; while the second one is the average return of the two portfolios with high book-to-market subtracted from the average return of the two portfolios with low book-to-market.

instrument that can not be considered persistent is the *Industrial Production Index*, which has the first order autocorrelation of 37%. The three factors have first order autocorrelation lower than 20%.

4.2. Portfolios

The portfolio chosen is formed by six portfolios selected with equal weights by size and book-to-market[11]. In Appendix 5 we also performed the same tests for different types and sizes of portfolios[12]. Table 2 reports the descriptive statistics of the monthly returns of the 6 portfolios. The table presents the 720 months period prior to December 2014, since this is the longest period of data for all instruments. The lagged variables are used to compute the R^2 statistic. Note that the mean ranges from 0.5% to 1.2% and the standard deviation from 4.7% to 7.2%. The table also presents the first order autocorrelation, which is generally low and between 12% and 26%; as well as the R^2 from the regression of the 5 instruments on the returns. Note that the adjustment coefficient is very low for all 6 assets, being of the order of 2%.

Table 2: Descriptive statistics of the monthly returns for the portfolio with 6 assets for a period of 720 months (60 years) from jan-1955 to dec-2014

	Mean	Std. Dev.	Min	Max	ρ_1	R^2
Small - 1 (Low)	0.005	0.072	-0.329	0.434	0.19	0.02
Small - 2	0.010	0.056	-0.286	0.316	0.22	0.02
Small - 3 (High)	0.012	0.057	-0.273	0.359	0.26	0.02
Big - 1 (Low)	0.006	0.053	-0.265	0.205	0.12	0.02
Big - 2	0.007	0.047	-0.242	0.218	0.13	0.02
Big - 3 (High)	0.009	0.049	-0.209	0.264	0.12	0.02

Monthly returns of the 6 portfolios formed on size and book-to market (2x3). The portfolios are based on equal-weighted returns extracted from the database provided by Kenneth R. French. First column presents the sample mean, the second shows the sample standard deviation, the third and fourth column present minimum and maximum returns, ρ_1 is the first-order autocorrelation, and the R^2 is the adjusted coefficient of determination from the regression of the returns on the lagged instruments. The sample period is January 1955 through December 2014 (720 observations).

The evaluated periods begin with 60 months prior to December 2014. Then, we extend the window adding more months prior to the fixed last date in our sample (Dec/2014).

[11] *6 Portfolios Formed on Size and Book-to-Market (2 x 3).* Following the methodology from Fama and French (1993), this portfolio is constructed from the intersection of two portfolios composed by market cap with portfolios composed by the ratio of book equity with market cap. Again, data for these portfolios were extracted from Kenneth R. French website.

[12] We also considered *25 Portfolios Formed on Size and Book-to-Market (5x5)* and the *49 Industry Portfolios* from Kenneth R. French website.

4.3. Tests of Efficiency Using GEL and GMM for N=6

In short, in this section we are interested to know how efficiency tests based on GEL and GMM estimations can lead to different decisions. We compare both methods when (i) no conditional information is used, and when (ii) *managed portfolios* structure is used under the *unconditional mean variance efficiency* with respect to information. To do so, the analysis is done comparing the test results for different sample sizes, as well as for two asset pricing models (CAPM and the Fama-French three-factor).

We start analyzing the portfolio consisting of the six assets ($N = 6$), and Appendix 5 extend for portfolios with higher number of assets (e.g., $N = 25$ and $N = 49$). For all portfolios, we seek to test whether the factors from each of the asset pricing models explain the portfolios average returns. For CAPM, we are interested to evaluate the efficiency of the market portfolio discussed in the previous section. For the Fama-French model, we want to assess the efficiency based on the three factors, namely, Mkt, SMB and HML.

4.3.1. No Conditional Information

Table 3 presents the estimation results by GMM and GEL for an increasing sequence of months, starting with the last 60 months and extending the window up to the limit of sample, i.e., 1020 months (85 years). Each sample begins in January of a given year, and ends in December 2014. The table also presents for each time interval, the estimation of both asset pricing models of interest, CAPM and Fama-French.

Initially examining the tests results when using either GMM or GEL, we note that for all periods above 180 months, both CAPM and Fama-French models show strong evidence to reject the hypothesis of efficiency for each model. This is due to the p-values of Wald and GRS tests being practically zero.

However, for short T we see strong disagreement between both methodologies. While for $T = 60$ (i.e., 5 years), we see no evidence to reject the efficiency using either GMM or GEL for both model and tests, for $T = 90$, $T = 120$ and $T = 150$, GMM and GEL point to opposite directions.

For 90 months, GMM rejects efficiency at 5% significance level for the CAPM using either Wald or GRS. We do not see the same conclusion using GEL for the same sample size. For Fama-French we do not see such a strong disagreement between them.

For 120 months (i.e., 10 years) we see similar results. With GEL, the p-values for Wald and GRS tests are 0.30 and 0.35 respectively for the CAPM model. With GMM these p-values are much smaller and provide evidence against the null hypothesis that the alphas are jointly equal to zero at a standard 5% significance level. For the Fama-French model, the p-values generated by GMM and GEL are very similar: 0.02 and 0.05 (Wald) and 0.04 and 0.08 (GRS) respectively.

Finally, for $T = 150$ months the same pattern repeats. For the CAPM, the p-value under GEL of the Wald statistic is 0.29, while the p-value of the F distribution under the assumption of normality given by the GRS test is 0.33. GMM provides much smaller p-values, both of them showing evidence to reject the efficiency hypothesis for a significance level of 5%. For the Fama-French model, the difference between the p-values using either GEL or GMM is smaller. Thus, the divergence between them is is more tenuous. Overall,

Table 3: Tests of portfolio efficiency using 6 portfolios formed on size and book-to-market for selected periods of time

	Months	Wald Test		GRS Test		Wald Test		GRS Test	
		Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
		CAPM				**FF**			
GMM	60	10.5	0.104	1.5	0.180	9.8	0.132	1.4	0.236
	90	14.7	0.022	2.3	0.045	12.4	0.054	1.9	0.098
	120	15.0	0.020	2.4	0.035	14.6	0.023	2.3	0.043
	150	13.7	0.033	2.2	0.048	17.6	0.007	2.8	0.015
	180	20.3	0.002	3.3	0.005	31.2	0.000	4.9	0.000
	240	28.4	0.000	4.6	0.000	64.4	0.000	10.3	0.000
	360	60.6	0.000	9.9	0.000	137.7	0.000	22.4	0.000
	480	73.0	0.000	12.0	0.000	219.3	0.000	35.9	0.000
	600	74.2	0.000	12.2	0.000	274.8	0.000	45.1	0.000
	720	78.5	0.000	13.0	0.000	285.8	0.000	47.0	0.000
	840	81.6	0.000	13.5	0.000	277.6	0.000	45.8	0.000
	960	83.2	0.000	13.8	0.000	268.5	0.000	44.3	0.000
	1020	70.2	0.000	11.6	0.000	242.7	0.000	40.1	0.000
GEL	60	7.5	0.279	1.1	0.374	8.3	0.218	1.2	0.335
	90	8.0	0.239	1.2	0.300	11.6	0.071	1.7	0.122
	120	7.2	0.304	1.1	0.351	12.5	0.052	1.9	0.083
	150	7.3	0.291	1.2	0.329	17.0	0.009	2.7	0.018
	180	17.3	0.008	2.8	0.013	27.1	0.000	4.3	0.000
	240	26.0	0.000	4.2	0.000	51.6	0.000	8.3	0.000
	360	54.0	0.000	8.8	0.000	109.6	0.000	17.8	0.000
	480	61.0	0.000	10.0	0.000	185.6	0.000	30.4	0.000
	600	51.5	0.000	8.5	0.000	221.6	0.000	36.4	0.000
	720	56.8	0.000	9.4	0.000	236.6	0.000	38.9	0.000
	840	56.6	0.000	9.4	0.000	227.4	0.000	37.5	0.000
	960	72.7	0.000	12.0	0.000	176.6	0.000	29.2	0.000
	1020	47.6	0.000	7.9	0.000	171.9	0.000	28.4	0.000

Tests of portfolio efficiency using 6 portfolios formed on size and book-to-market (2x3) for 9 selected periods of time: $T = 60$ (5 years), $T = 90$ (7.5 years), $T = 120$ (10 years), $T = 150$ (12.5 years), $T = 180$ (15 years), $T = 240$ (20 years), $T = 360$ (30 years), $T = 480$ (40 years), $T = 600$ (50 years), $T = 720$ (60 years), $T = 840$ (70 years), $T = 960$ (80 years) and $T = 1020$ (85 years). The tests are conducted based on both estimations methodology: via GMM are on top, while via GEL are on the bottom. Tests of efficiency under the CAPM asset pricing model are on the left, while tests under Fama-French three-factor model (represented as "FF") are on the right. Table presents the statistic and the p-values of the Wald and the GRS tests for each case.

in Table 3 we see evidence to endorse the simulations results presented in Section 3, as GMM over-rejects the null hypothesis when compared to tests done via GEL, especially in a finite sample context.

4.3.2. Managed Portfolios

Table 4: Tests of portfolio efficiency using 6 portfolios formed on size and book-to-market under *scaled returns* for selected periods of time

	Months	Wald Test		GRS Test		Wald Test		GRS Test	
		Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
		CAPM				**FF**			
GMM	60	NA	NA	NA	NA	10.3	0.113	1.5	0.211
	90	NA	NA	NA	NA	NA	NA	NA	NA
	120	128.0	0.000	20.1	0.000	15.7	0.015	2.4	0.031
	150	39.3	0.000	6.2	0.000	57.2	0.000	9.0	0.000
	180	63.2	0.000	10.1	0.000	45.3	0.000	7.2	0.000
	240	32.1	0.000	5.2	0.000	97.5	0.000	15.6	0.000
	360	79.0	0.000	12.9	0.000	209.7	0.000	34.1	0.000
	480	111.5	0.000	18.3	0.000	310.4	0.000	50.8	0.000
	600	111.1	0.000	18.3	0.000	398.1	0.000	65.3	0.000
	720	102.4	0.000	16.9	0.000	380.0	0.000	62.5	0.000
GEL	60	45.2	0.000	6.7	0.000	65.6	0.000	9.3	0.000
	90	38.4	0.000	5.9	0.000	17.3	0.008	2.6	0.024
	120	5.6	0.466	0.9	0.509	16.3	0.012	2.5	0.025
	150	7.6	0.271	1.2	0.308	22.8	0.001	3.6	0.003
	180	32.2	0.000	5.2	0.000	46.3	0.000	7.3	0.000
	240	31.6	0.000	5.1	0.000	82.3	0.000	13.2	0.000
	360	63.3	0.000	10.3	0.000	174.9	0.000	28.4	0.000
	480	78.9	0.000	13.0	0.000	275.2	0.000	45.0	0.000
	600	72.2	0.000	11.9	0.000	241.8	0.000	39.7	0.000
	720	76.0	0.000	12.5	0.000	332.5	0.000	54.7	0.000

Tests of portfolio efficiency using 6 portfolios formed on size and book-to-market (2x3) for 6 selected periods of time: $T = 60$ (5 years), $T = 90$ (7.5 years), $T = 120$ (10 years), $T = 150$ (12.5 years), $T = 180$ (15 years), $T = 240$ (20 years), $T = 360$ (30 years), $T = 480$ (40 years), $T = 600$ (50 years) and $T = 720$ (60 years). The tests are evaluated using conditioning information when instruments are incorporated to the pricing equation. The lagged variables consisting the conditioning information are: (i) 3 month Treasury-bill yield, (ii) industrial production growth, (iii) yield spreads of low-grade over high-grade corporate bonds, (iv) yield spreads of long-term over short-term Treasury-bills (10-year/1-year) and (v) U.S. inflation (CPI). The tests are conducted based on both estimations methodology: via GMM are on top, while via GEL are on the bottom. Tests of efficiency under the CAPM asset pricing model are on the left, while tests under Fama-French three-factor model (represented as "FF") are on the right. Table presents the statistic and the p-values of the Wald and the GRS tests for each case."NA" denotes not applicable, situations in which singularity problems occur impeding the inversion of the covariance matrix. For both models when $T = 90$, even though we obtained estimates for the coefficients through GMM, we were not able to invert the covariance matrix and perform the tests.

Table 4 presents the results of the market portfolio efficiency tests for the *scaled returns* approach. Here, we use the *managed portfolios* where the lagged variables are used to provide signals regarding the state of the economy. To this end, the five instruments previously described are used.

As before, we have compelling evidence to reject efficiency for all intervals of 180 months and above, for all tests and models based on estimations from either GMM or GEL. Note that the longest period for which we can perform the tests is 720 months due to the data restriction of all five instruments.

While for longer periods the p-values are virtually zero, for $T = 60$, $T = 120$ and $T = 150$ months[13], tests inference via GMM and GEL are conflicting. For CAPM, GEL again shows no indication to reject efficiency (for $T = 120$ and $T = 150$), while GMM does (p-values are practically zero for Wald and GRS tests). When we used fixed weights in the Table 3, the results were similar.

With the use of instruments, the tests os efficiency for the Fama-French do not necessarily provide different inference regarding the rejection of the null hypothesis. But we still see that GMM generates smaller p-values (i.e., higher tests statistics) for both tests when compared to GEL. However, for $T = 60$, GMM and GEL strongly disagrees again. While GMM generates p-values higher than 10%, GEL has p-values practically equal to zero.

Figure 7 summarizes the estimated coefficients of α and β_{Mkt} with the use of instruments for CAPM and Fama-French models. As for every T and each model there are 6 estimated alphas and betas (Mkt), we present the estimated coefficients for each methodology (GMM and GEL) using boxplots. Figure 16 in Appendix 5 plots the $\hat{\beta}_{SMB}$ and $\hat{\beta}_{HML}$ for Fama-French model.

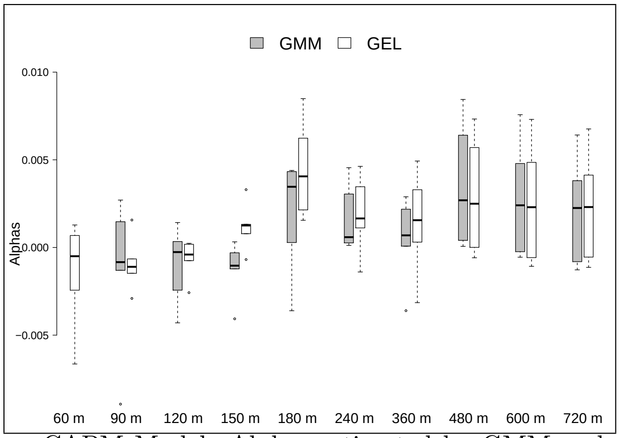
Notice that as the boxplots distributions are not identical. GMM and GEL generate different coefficients estimates, not only different covariance matrix from these estimates. For CAPM, which is located in the top two panels, note that for shorter periods differences are evident between both methodologies. Regarding the $\hat{\alpha}$, the biggest differences are for T equal to 90, 120, 240 and 360. For $\hat{\beta}_{Mkt}$, this is clear for sample sizes smaller than 240 months. For Fama-French model, differences between estimates by GMM and GEL are more subtle. However, the $\hat{\alpha}$ for very short samples, as for 60 months, differences in estimates are not negligible.

From asset pricing theory one knowns that the returns of any asset should be higher if the asset has higher betas. Figure 8 plots for CAPM the estimated betas ($\hat{\beta}_{Mkt}$) against the sample mean of monthly excess returns ($\hat{E}(R_i)$) of each of the six assets. The model states that the average returns should be proportional to betas. Each panel shows one of the time periods evaluated. The GMM estimates are located on the left, while GEL estimates are on the right.

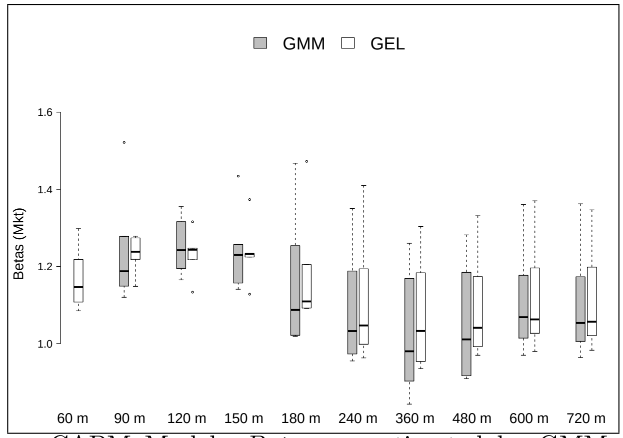
The distance between the points and the straight line must represent the pricing errors, i.e., estimated alphas. We see a clear difference in point estimations depending whether

[13] Singularity problems may occur during the estimations impeding the inversion of the covariance matrix. These cases are shown as "NA". For $T = 90$ we could not perform the tests for both models using GMM. Even though we obtained estimates for the CAPM coefficients through GMM, we were not able to invert the covariance matrix and perform the tests.

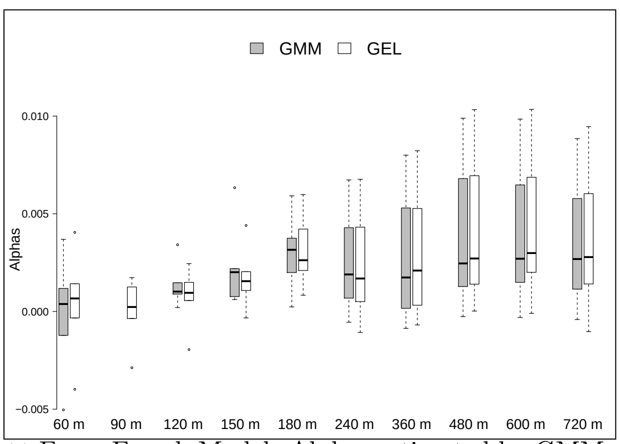
Figure 7: Boxplots for comparison of estimations by GMM and GEL using 6 portfolios formed on size and book-to-market under *scaled returns* for selected periods of time



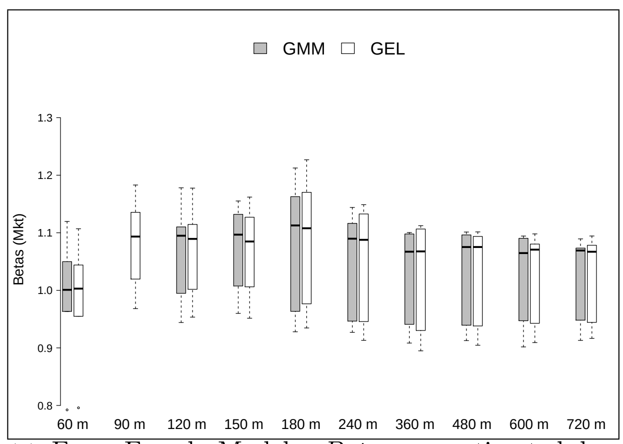
(a) CAPM Model: Alphas estimated by GMM and GEL



(b) CAPM Model: $Betas_{Mkt}$ estimated by GMM and GEL



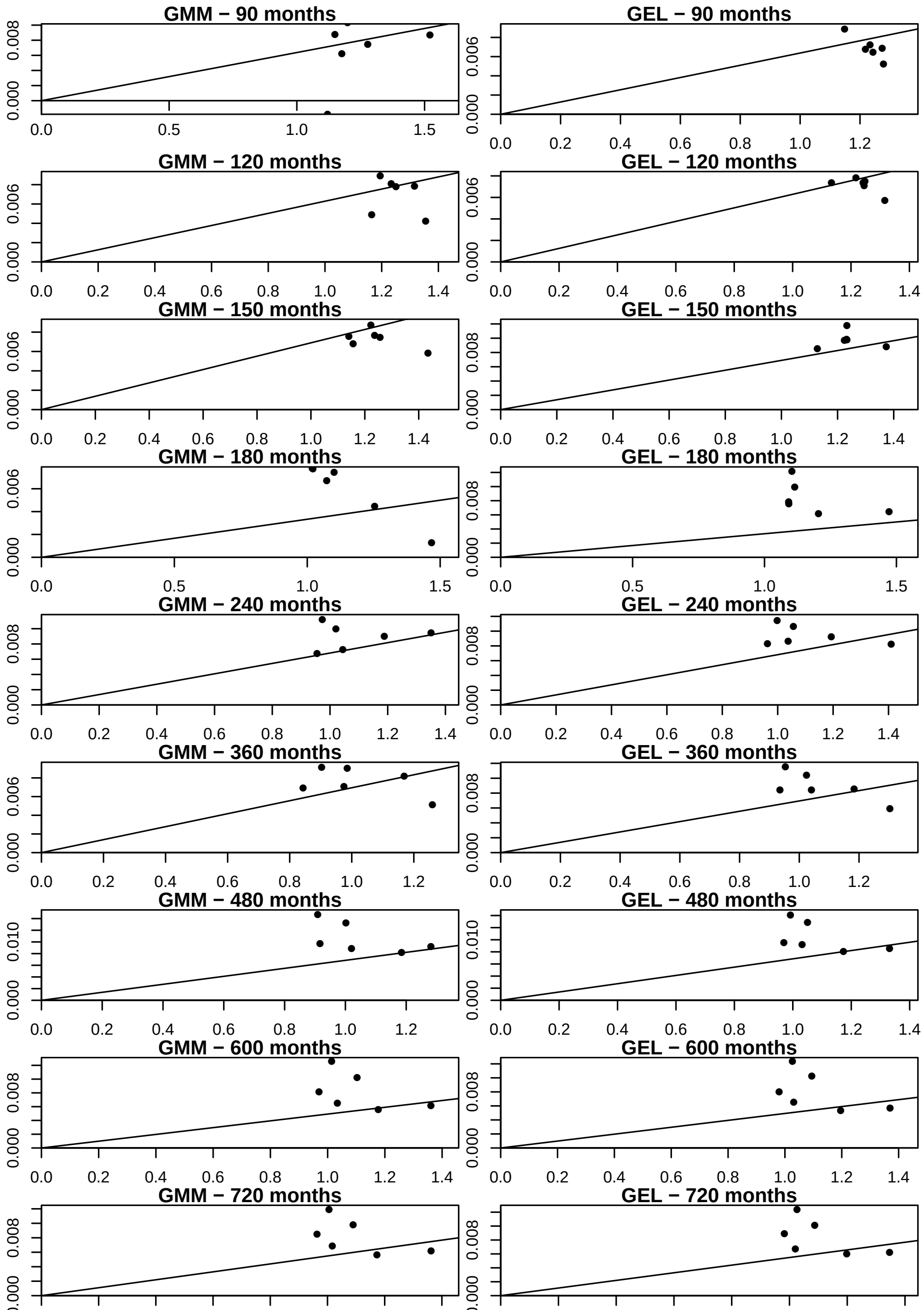
(c) Fama-French Model: Alphas estimated by GMM and GEL



(d) Fama-French Model: $Betas_{Mkt}$ estimated by GMM and GEL

In all 4 panels, the x-axis represents the time intervals, starting with 120 months before December 2014 to 720 months (60 years) prior to this date. The y-axis are the estimated coefficients values by GMM and GEL. Estimation by GMM is represented by gray boxplots, while GEL estimation is represented by white boxplots. For $T = 60$ (CAPM) and $T = 90$ (Fama-French), we were not able to estimate the coefficients through GMM.

Figure 8: CAPM Model - Comparison of GMM and GEL estimated betas under *scaled returns* against the sample mean of monthly excess returns for the portfolio with 6 assets



In all panels, the estimated betas ($\hat{\beta}_{Mkt}$) are in the x-axis, and the sample mean of the monthly excess returns for each of the $n = 6$ assets in the portfolio are in the y-axis. Estimations via GMM are on the left, while via GEL are on the right. Each panel represents one of the time intervals, starting with 120 months before December 2014 to 720 months (60 years) prior to this date.

GMM or GEL is used, being more evident the divergence for short T. Note that in the panels with T lower than 180 months, while the estimates by GMM are more dispersed, those with GEL are more grouped and closer with slope line equal $\hat{E}(R_{Mkt})$.

5. Conclusions

We evaluate the behavior of the GMM and GEL estimators in tests of portfolio efficiency. We argue that both estimators have different statistical features, and therefore, tests of portfolio efficiency based on them may reflect these differences.

First, we assess the robustness of the tests with the use of GMM and GEL estimators in a finite sample context. Defining different DGPs to incorporate different specifications, we perform several Monte Carlo experiments to examine the effects of distortions in the data can cause on tests of efficiency, and consequently, in decisions based on these results. In general, we see evidence that GEL estimators have better performance when heavy tails are present. Depending on the characteristics of the DGP choosen, both GMM and GEL may have better robustness to outliers compared among them. However, under the null hypothesis, for both estimators, the Wald and GRS tests have a tendency to over-reject the hypothesis of efficiency in finite samples.

In the second part, we use returns from real datasets in our analysis. Comparing the results for (a) different portfolios sizes (see Appendix 5 for more details), (b) different factor pricing models, as well as (c) increasing periods of time, we see that (i) in general, efficiency tests using GEL generate lower estimates (i.e., higher p-values) compared to tests using the standard approach with GMM; and (ii) when the sample size has finite characteristics, with low N and T, we note that the results are conflicting among the methodologies. These results are significant when the hypothesis of efficiency is evaluated for both models, the CAPM and the Fama-French three-factor model. Under the unconditional efficiency structure these findings also apply for either fixed weights, or conditional information with instruments in the multiplicative approach.

Finally, these results may be an evidence that estimators from GEL class really performs differently in small samples. In addition, they may show that tests based on GMM, or even by maximum likelihood estimation, have a tendency to over-reject the null hypothesis of efficiency.

References

Almeida, C., Garcia, R., 2008. Empirical likelihood estimators for stochastic discount factors. In: EFA 2008 Athens Meetings Paper.

Almeida, C., Garcia, R., 2012. Assessing misspecified asset pricing models with empirical likelihood estimators. Journal of Econometrics 170 (2), 519–537.

Anatolyev, S., 2005. GMM, GEL, serial correlation, and asymptotic bias. Econometrica 73 (3), 983–1002.

Anatolyev, S., Gospodinov, N., 2011. Methods for Estimation and Inference in Modern Econometrics. CRC Press.

Andrews, D. W., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica, 817–858.

Campbell, J. Y., Lo, A. W.-C., MacKinlay, A. C., et al., 1997. The Econometrics of Financial Markets. Vol. 2. Princeton University Press.

Chaussé, P., 2010. Computing generalized method of moments and generalized empirical likelihood with R. Journal of Statistical Software 34 (11), 1–35.

Cochrane, J. H., 2005. Asset Pricing. Princeton University Press.

Davidson, R., MacKinnon, J., 1998. Graphical methods for investigating the size and power of hypothesis tests. The Manchester School of Economic & Social Studies 66 (1), 1–26.

Fama, E. F., French, K. R., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25 (1), 23–49.

Fama, E. F., French, K. R., 1992. The cross-section of expected stock returns. The Journal of Finance 47 (2), 427–465.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33 (1), 3–56.

Ferson, W. E., Qian, M., 2004. Conditional performance evaluation, revisited. Research Foundation Monograph of the CFA Institute.

Ferson, W. E., Siegel, A. F., Jul 2009. Testing portfolio efficiency with conditioning information. Review of Financial Studies 22 (7), 2735–2758.

Gibbons, M. R., Ross, S. A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica, 1121–1152.

Hansen, L. P., Heaton, J., Yaron, A., 1996. Finite-sample properties of some alternative GMM estimators. Journal of Business & Economic Statistics 14 (3), 262–280.

Hansen, L. P., Singleton, K. J., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica, 1269–1286.

Jensen, M. C., 1968. The performance of mutual funds in the period 1945–1964. The Journal of Finance 23 (2), 389–416.

Keim, D. B., Stambaugh, R. F., 1986. Predicting returns in the stock and bond markets. Journal of Financial Economics 17 (2), 357–390.

Kitamura, Y., 1997. Empirical likelihood methods with weakly dependent processes. The Annals of Statistics 25 (5), 2084–2102.

Kitamura, Y., Stutzer, M., 1997. An information-theoretic alternative to generalized method of moments estimation. Econometrica, 861–874.

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. The Review of Economics and Statistics, 13–37.

Markowitz, H., 1952. Portfolio selection. The Journal of Finance 7 (1), 77–91.

Newey, W. K., Smith, R. J., 2004. Higher order properties of GMM and generalized empirical likelihood estimators. Econometrica 72 (1), 219–255.

Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.

Owen, A. B., 2001. Empirical Likelihood. Chapman and Hall.

Peñaranda, F., Sentana, E., 2012. Spanning tests in return and stochastic discount factor mean–variance frontiers: A unifying approach. Journal of Econometrics 170 (2), 303–324.

Sharpe, W. F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance 19 (3), 425–442.

Smith, R. J., 1997. Alternative semi-parametric likelihood approaches to generalised method of moments estimation. The Economic Journal, 503–519.

Smith, R. J., 2000. Empirical likelihood estimation and inference. Applications of Differential Geometry to Econometrics, 119.

Appendix

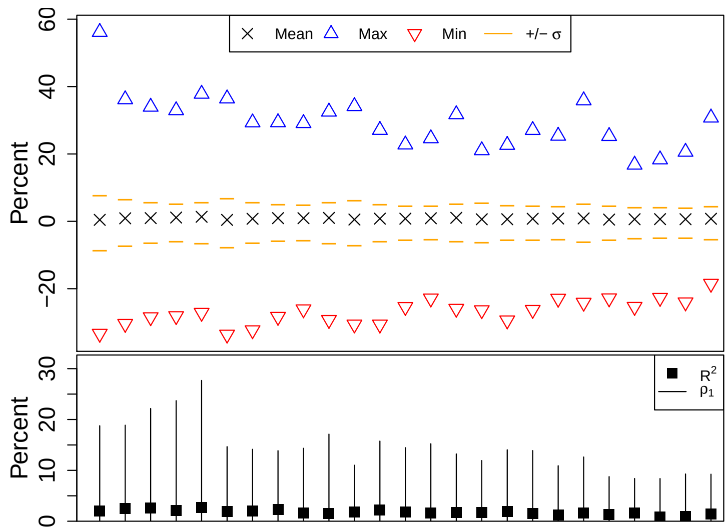
Results for Different Types and Sizes of Portfolios

As in Section 4, here we use different types and sizes of portfolio to evaluate how efficiency tests using either GEL or GMM can cause different inference conclusions. Again, we make use of (i) fixed weights portfolios (no conditional information), as well as (ii) *managed portfolios* under the *unconditional mean variance efficiency* with respect to information approach. We consider CAPM and the Fama-French three-factor models and different sample sizes (T).

Data - Portfolios with 25 and 49 assets

In order to examine the estimators behavior for different types of portfolios and higher number of assets, we selected other two portfolios. To avoid using a single portfolio composition methodology, the first one is based on size and book-to-market, while the second one is composed with categories derived from industries classification according to the business segment[14]. The chosen portfolios are: the (i) 25 assets selected with equal weights by size and book-to-market[15], and the (ii) 49 Industry Portfolios.

Figure 9: Descriptive statistics of the monthly returns for the portfolio with 25 assets for a period of 720 months (60 years) from jan-1955 to dec-2014



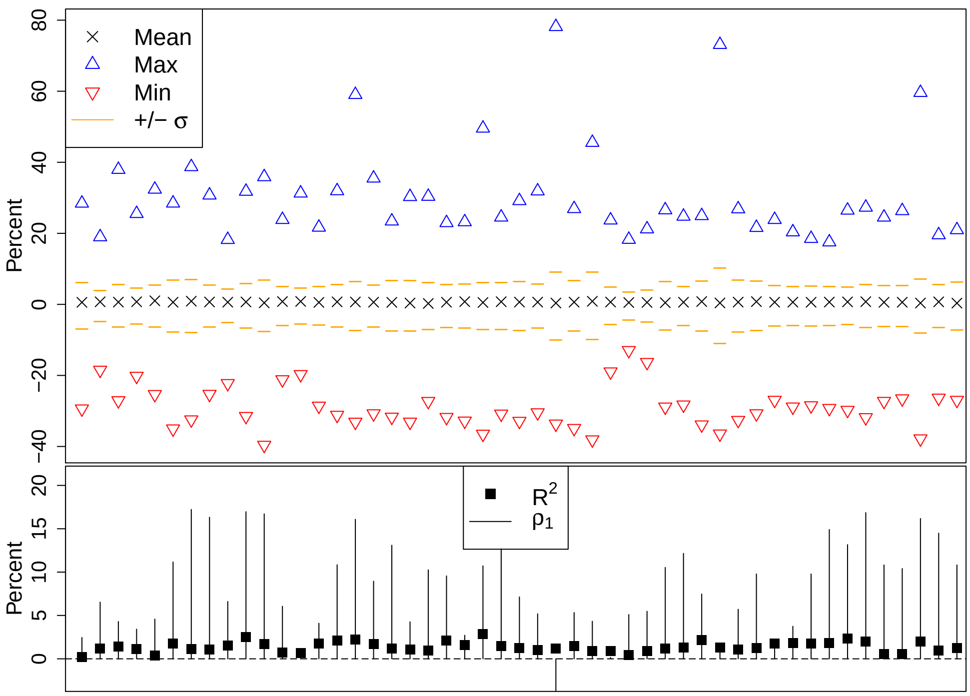
The top panel shows the sample mean statistics (represented by "X"), Max (represented by a blue triangle), Min (represented by an upside-down red triangle), and the distance between the two horizontal bars represent the range of $\pm\sigma$ for a 720 months period. The bottom panel shows the first-order autocorrelation ρ_1 (bar) and R^2 (square point), which is the adjusted coefficient of determination in percent from the regression of the returns on the 5 instruments. In both panels, the x-axis represents the 25 assets, and the y-axis is expressed in percentage.

Figure 9 presents the main descriptive statistics of the 25 portfolios formed on size and book-to-market. Compared to the 6 portfolio presented in Section 4.2, the statistics are very similar, except that we can see some portfolios with first order autocorrelation less than 10%. Figure 10 remarks the descriptive statistics of the 49 Industry Portfolios. Notice that both mean and standard deviation are similar to the previous portfolio, while the maximum and minimum returns are magnified. Most of the first order autocorrelation are lower than 20% and there is only one asset with negative value. The R^2 keeps low values, with adjusted coefficients of determination no higher than 5%.

[14] The data for these portfolios were extracted from Kenneth R. French website.

[15] *25 Portfolios Formed on Size and Book-to-Market (5 x 5)*

Figure 10: Descriptive statistics of the monthly returns for the portfolio with 49 assets for a period of 720 months (60 years) from jan-1955 to dec-2014



The top panel shows the sample mean statistics (represented by "X"), Max (represented by a blue triangle), Min (represented by an upside-down red triangle), and the distance between the two horizontal bars represent the range of $\pm\sigma$ for a 720 months period. The bottom panel shows the first-order autocorrelation ρ_1 (bar) and R^2 (square point), which is the adjusted coefficient of determination in percent from the regression of the returns on the 5 instruments. In both panels, the x-axis represents the 49 assets, and the y-axis is expressed in percentage.

Results - Tests of Efficiency Using GEL and GMM for Portfolios with 25 and 49 assets

25 Portfolios Formed on Size and Book-to-Market

No Conditional Information

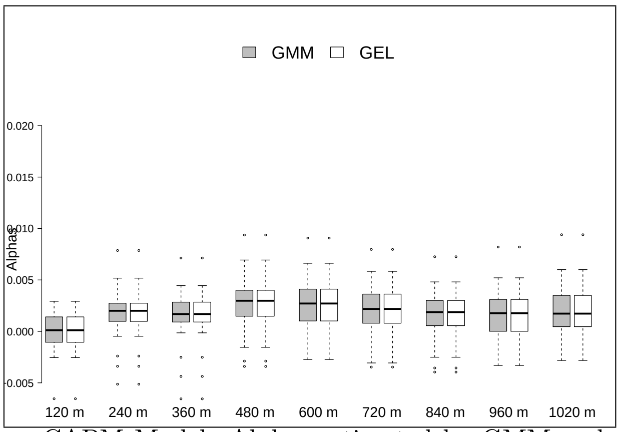
Table 5: Tests of portfolio efficiency using 25 portfolios formed on size and book-to-market for selected periods of time

	Months	Wald Test		GRS Test		Wald Test		GRS Test	
		Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
		CAPM				**FF**			
GMM	120	98.5	0.000	3.1	0.000	NA	NA	NA	NA
	240	93.0	0.000	3.3	0.000	114.9	0.000	4.1	0.000
	360	168.7	0.000	6.3	0.000	257.5	0.000	9.5	0.000
	480	123.4	0.000	4.7	0.000	335.7	0.000	12.6	0.000
	600	133.4	0.000	5.1	0.000	465.2	0.000	17.7	0.000
	720	125.5	0.000	4.8	0.000	459.7	0.000	17.7	0.000
	840	122.5	0.000	4.8	0.000	413.2	0.000	16.0	0.000
	960	131.9	0.000	5.1	0.000	398.2	0.000	15.5	0.000
	1020	129.2	0.000	5.0	0.000	403.8	0.000	15.7	0.000
GEL	120	53.6	0.001	1.7	0.039	92.1	0.000	2.8	0.000
	240	85.4	0.000	3.0	0.000	138.9	0.000	4.9	0.000
	360	135.0	0.000	5.0	0.000	253.8	0.000	9.4	0.000
	480	90.6	0.000	3.4	0.000	329.0	0.000	12.4	0.000
	600	85.1	0.000	3.3	0.000	362.4	0.000	13.8	0.000
	720	83.2	0.000	3.2	0.000	374.1	0.000	14.4	0.000
	840	80.6	0.000	3.1	0.000	300.4	0.000	11.6	0.000
	960	90.3	0.000	3.5	0.000	300.5	0.000	11.7	0.000
	1020	84.1	0.000	3.3	0.000	310.5	0.000	12.1	0.000

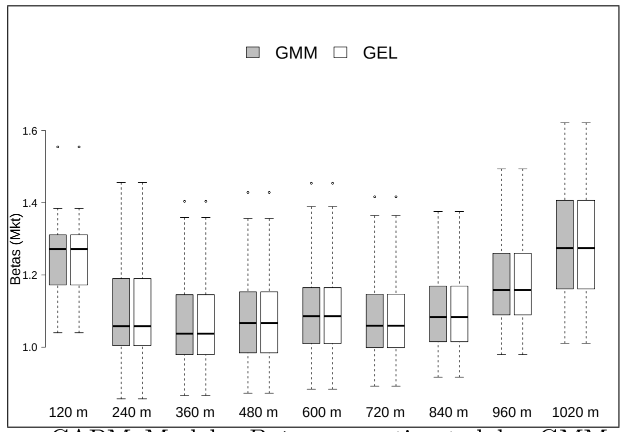
Tests of portfolio efficiency using 25 portfolios formed on size and book-to-market (5x5) for 9 selected periods of time: T=120 (10 years), T=240 (20 years), T=360 (30 years), T=480 (40 years), T=600 (50 years), T=720 (60 years), T=840 (70 years), T=960 (80 years) and T=1020 (85 years). The tests are conducted based on both estimations methodology: via GMM are on top, while via GEL are on the bottom. Tests of efficiency under the CAPM asset pricing model are on the left, while tests under Fama-French three-factor model (represented as "FF") are on the right. Table presents the statistic and the p-values of the Wald and the GRS tests for each case. "NA" denotes not applicable, situations in which singularity problems occur impeding the inversion of the covariance matrix. "NA" denotes not applicable, situations in which singularity problems occur impeding the inversion of the covariance matrix. For $T = 120$ and Fama-French three-factor model, even though we obtained estimates for the coefficients through GMM, we were not able to invert the covariance matrix and perform the tests.

Table 5 presents the results of the estimations by GMM and GEL for an increasing sequence of months. Comparing GMM and GEL for CAPM, both estimators provide compelling evidence to reject the market portfolio efficiency hypothesis. The p-values for

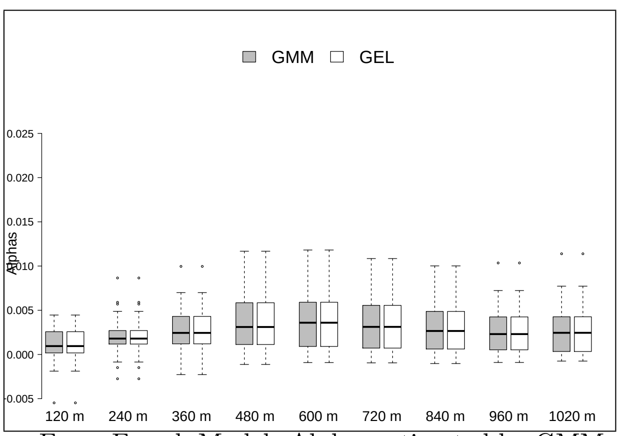
Figure 11: Boxplots for comparison of estimations by GMM and GEL using 25 portfolios formed on size and book-to-market for selected periods of time



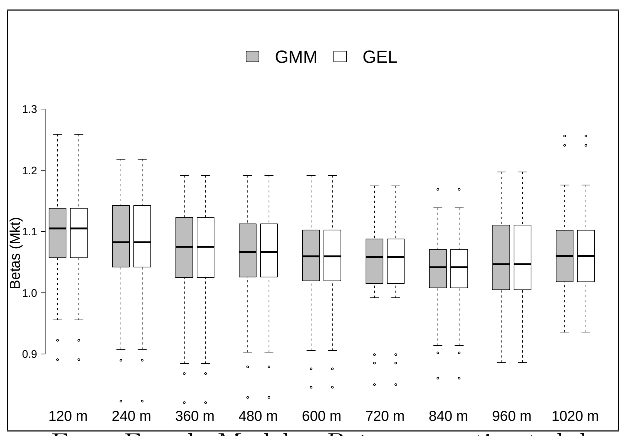
(a) CAPM Model: Alphas estimated by GMM and GEL



(b) CAPM Model: $Betas_{Mkt}$ estimated by GMM and GEL



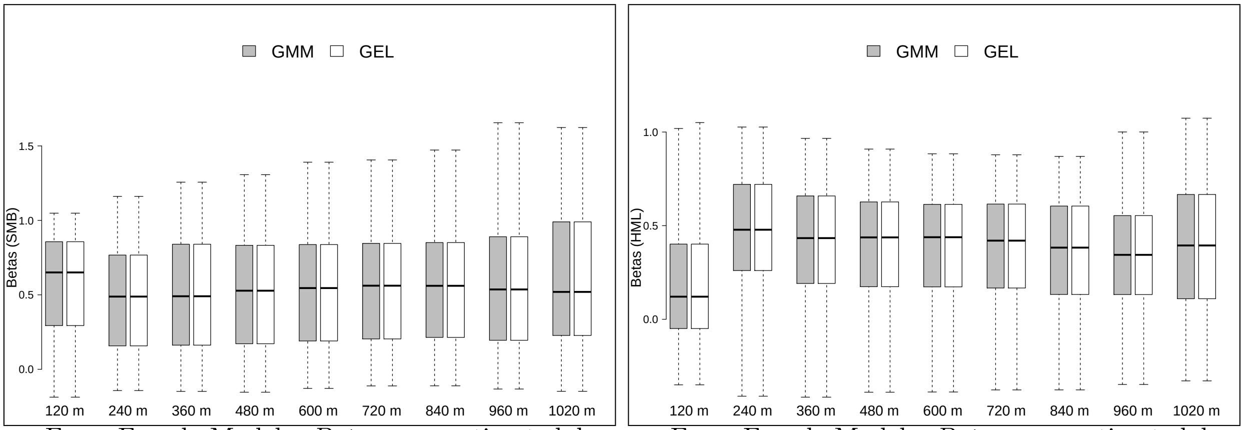
(c) Fama-French Model: Alphas estimated by GMM and GEL



(d) Fama-French Model: $Betas_{Mkt}$ estimated by GMM and GEL

In all 4 panels, the x-axis represents the time intervals, starting with 120 months before December 2014 to 1020 months prior to this date. The y-axis are the estimated coefficients values by GMM and GEL. Estimation by GMM is represented by gray boxplots, while GEL estimation is represented by white boxplots.

Figure 12: Boxplots for comparison of estimations by GMM and GEL for Fama-French model using 25 portfolios formed on size and book-to-market for selected periods of time



(a) Fama-French Model: $Betas_{SMB}$ estimated by GMM and GEL

(b) Fama-French Model: $Betas_{HML}$ estimated by GMM and GEL

In both panels, the x-axis represents the time intervals, starting with 120 months before December 2014 to 1020 months prior to this date. The y-axis are the estimated coefficients values by GMM and GEL. Estimation by GMM is represented by gray boxplots, while GEL estimation is represented by white boxplots.

both approaches are practically zero for every T, except for 120 months using GEL.[16]. However, notice that GEL consistently generates lower test statistics when compared to GMM (Wald and GRS).

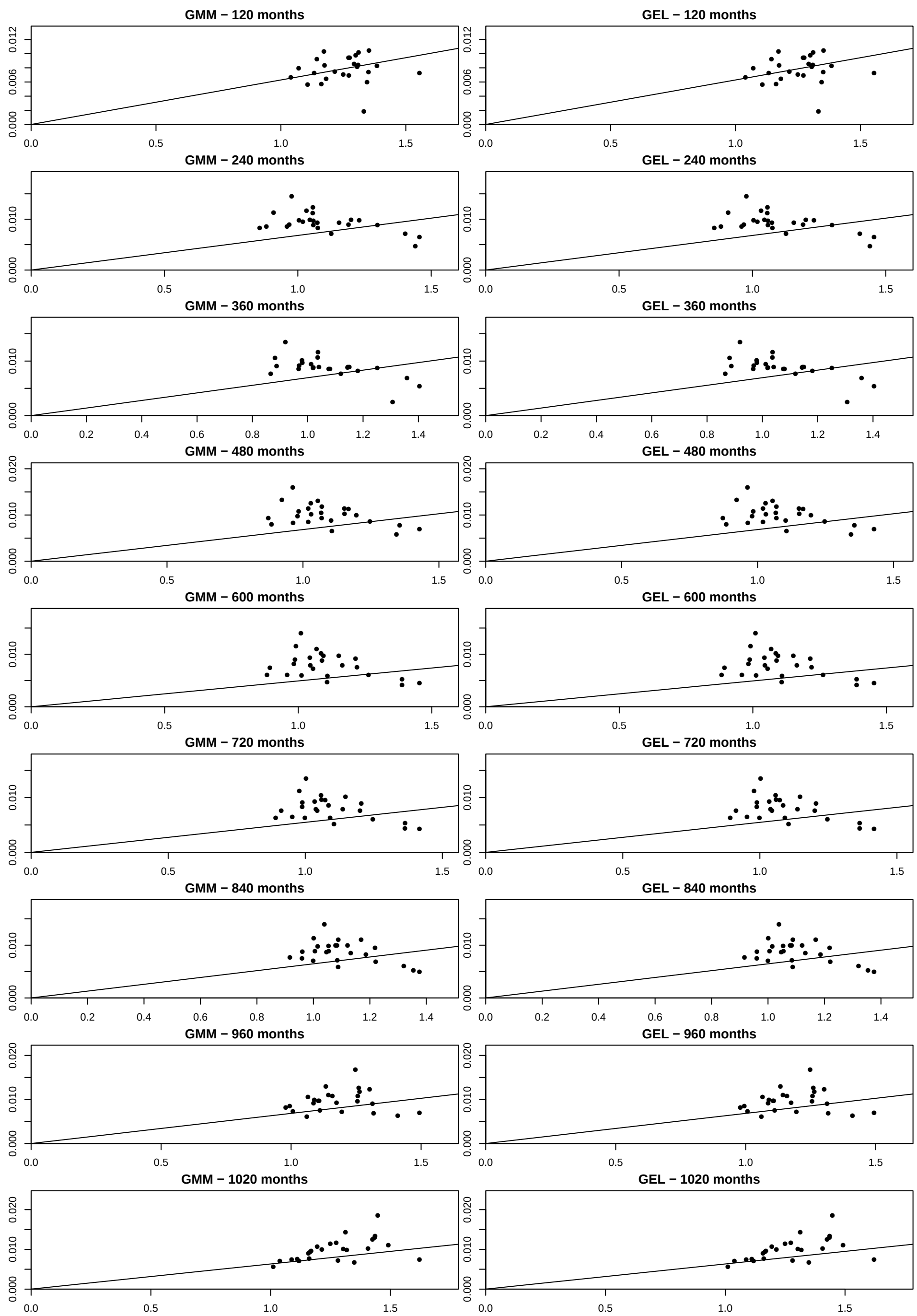
For the Fama-French model, the conclusions to reject the null hypothesis do not change. Here we had a single case (120 months and GMM) where we were not able to invert the covariance matrix of the alphas. Again, in general GEL generates lower statistics for both tests (except for 240 months and Fama-French model). This difference seems to become more relevant as the sample period increases, i.e., we have evidence of an increasing difference relationship when the sample expands.

In Figure 11 we present the boxplots of $\hat{\alpha}$ and $\hat{\beta}_{Mkt}$ (CAPM and Fama-French). The boxplots of $\hat{\beta}_{SMB}$ and $\hat{\beta}_{HML}$ (Fama-French model) are shown in Figure 12. Each boxplot summarizes the 25 estimates using either GMM or GEL for each time interval. Note that for every sample size T we obtained estimates for the coefficients, even though for $T = 120$ and Fama-French we were not able to invert the covariance matrix and perform the tests through GMM.

The boxplots of the alphas and betas show similar distributions for GMM and GEL. The GRS and Wald tests (Table 5) have different values between both estimators mainly

[16]If we compare these results with other studies regarding the efficiency of a market index, many of them come to similar conclusions (Ferson and Siegel, 2009; Fama and French, 1992). Even though the portfolios and the exact time intervals used in these works have a strong resemblance, one should be cautious since they are not exactly the same.

Figure 13: CAPM Model - Comparison of GMM and GEL estimated betas against the sample mean of monthly excess returns for the portfolio with 25 assets



In all panels, the estimated betas ($\hat{\beta}_{Mkt}$) are in the x-axis, and the sample mean of the monthly excess returns for each of the $n = 25$ assets in the portfolio are in the y-axis. Estimations via GMM are on the left, while via GEL are on the right. Each panel represents one of the time intervals, starting with 120 months before December 2014 to 1020 months (85 years) prior to this date.

because of the difference in the estimated covariance matrix. We see that for both models the estimated alphas are low for the shortest period ($T = 120$) with increasing values as t grows; and from 720 months on, the alphas fall. Inversely, the betas have similar characteristics: starting with high estimates, they decrease to finally increase again. For CAPM, Figure 13 plots the sample mean of the monthly excess returns ($\hat{E}(R_i)$) against the estimated betas ($\hat{\beta}_{Mkt}$) for the 25 assets.

Managed Portfolios

Table 6: Tests of portfolio efficiency using 25 portfolios formed on size and book-to-market under *scaled returns* for selected periods of time

	Months	Wald Test		GRS Test		Wald Test		GRS Test	
		Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
		CAPM				**FF**			
GMM	120	NA	NA	NA	NA	NA	NA	NA	NA
	240	NA	NA	NA	NA	NA	NA	NA	NA
	360	NA	NA	NA	NA	NA	NA	NA	NA
	480	802.7	0.000	30.4	0.000	849.3	0.000	32.0	0.000
	600	480.7	0.000	18.4	0.000	967.6	0.000	36.9	0.000
	720	320.5	0.000	12.4	0.000	851.1	0.000	32.7	0.000
GEL	120	NA	NA	NA	NA	NA	NA	NA	NA
	240	NA	NA	NA	NA	NA	NA	NA	NA
	360	NA	NA	NA	NA	NA	NA	NA	NA
	480	NA	NA	NA	NA	NA	NA	NA	NA
	600	NA	NA	NA	NA	NA	NA	NA	NA
	720	NA	NA	NA	NA	NA	NA	NA	NA

Tests of portfolio efficiency using 25 portfolios formed on size and book-to-market (5x5) for 6 selected periods of time: T=120 (10 years), T=240 (20 years), T=360 (30 years), T=480 (40 years), T=600 (50 years) and T=720 (60 years). The tests are evaluated using conditioning information when instruments are incorporated to the pricing equation. The lagged variables consisting the conditioning information are: (i) 3 month Treasury-bill yield, (ii) industrial production growth, (iii) yield spreads of low-grade over high-grade corporate bonds, (iv) yield spreads of long-term over short-term Treasury-bills (10-year/1-year) and (v) U.S. inflation (CPI). The tests are conducted based on both estimations methodology: via GMM are on top, while via GEL are on the bottom. Tests of efficiency under the CAPM asset pricing model are on the left, while tests under Fama-French three-factor model (represented as "FF") are on the right. Table presents the statistic and the p-values of the Wald and the GRS tests for each case. "NA" denotes not applicable, situations in which singularity problems occur impeding the inversion of the covariance matrix.

Table 6 presents the results of the efficiency tests using the multiplicative approach with 5 instruments. A quick inspection shows that in many cases it was not possible to compute the tests[17]. All "NA" occur due to the impossibility of the methods to estimate the parameters of the model. Having 25 assets and five instruments can cause the optimal

[17]From the formulas in (33) and (34), inverting the covariance matrix can become an impediment to the

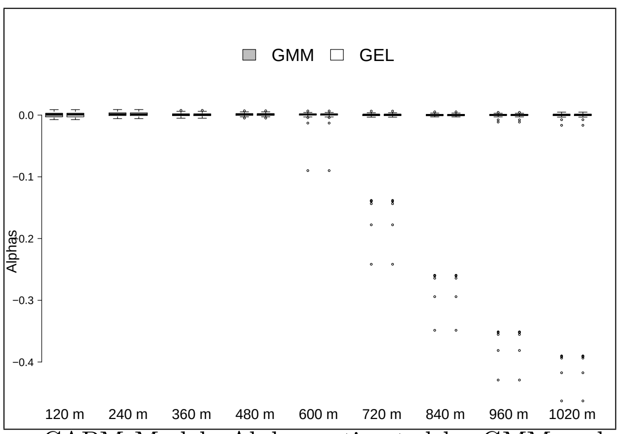
long-run covariance matrix to be singular. Singularity problems is a fairly common issue, especially with portfolios with high number of assets and under the multiplicative approach with instruments. Ferson and Siegel (2009) also had to deal with this issue (see e.g. Peñaranda and Sentana (2012) for advanced treatment).

Note that only for $T = 480$, 600 and 720 it was possible to perform the tests using GMM. Both tests generated considerably higher estimates (especially Wald), when compared with the fixed weights approach, leading to a very strong evidence to reject the null hypothesis. Using GEL we were not able to estimate the coefficients for any T.

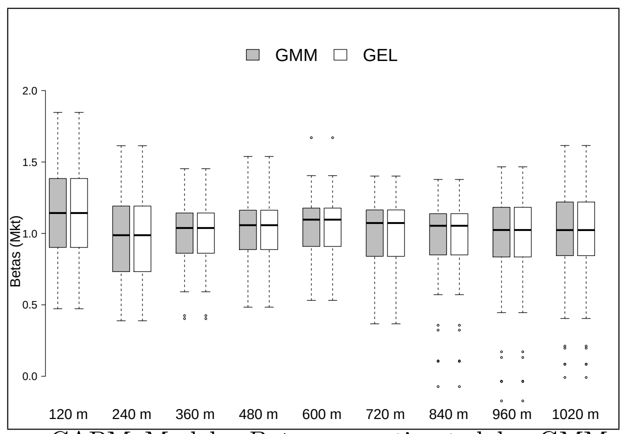
49 Industry Portfolios

No Conditional Information

Figure 14: Boxplots for comparison of estimations by GMM and GEL for CAPM model using 49 industry portfolios for selected periods of time



(a) CAPM Model: Alphas estimated by GMM and GEL



(b) CAPM Model: $Betas_{Mkt}$ estimated by GMM and GEL

In both panels, the x-axis represents the time intervals, starting with 120 months before December 2014 to 1020 months prior to this date. The y-axis are the estimated coefficients values by GMM and GEL. Estimation by GMM is represented by gray boxplots, while GEL estimation is represented by white boxplots.

Table 7 shows the results of the estimations by GMM and GEL for the portfolio formed using 49 industrial categories. With this number of assets, singularities issues are more common. Thus, for many sample sizes we could not perform the tests[18].

Examining the results for the periods in which we obtained results, we see that tests using GEL and GMM generate high p-values (CAPM) and provide no evidence to reject the efficiency hypothesis of the market proxy. Note that for these cases, the p-values

estimation of the tests, given the fact that singular matrices can arise. However, there are cases in which we cannot estimate the parameters of the model. This situation is also represented by "NA" in the table.

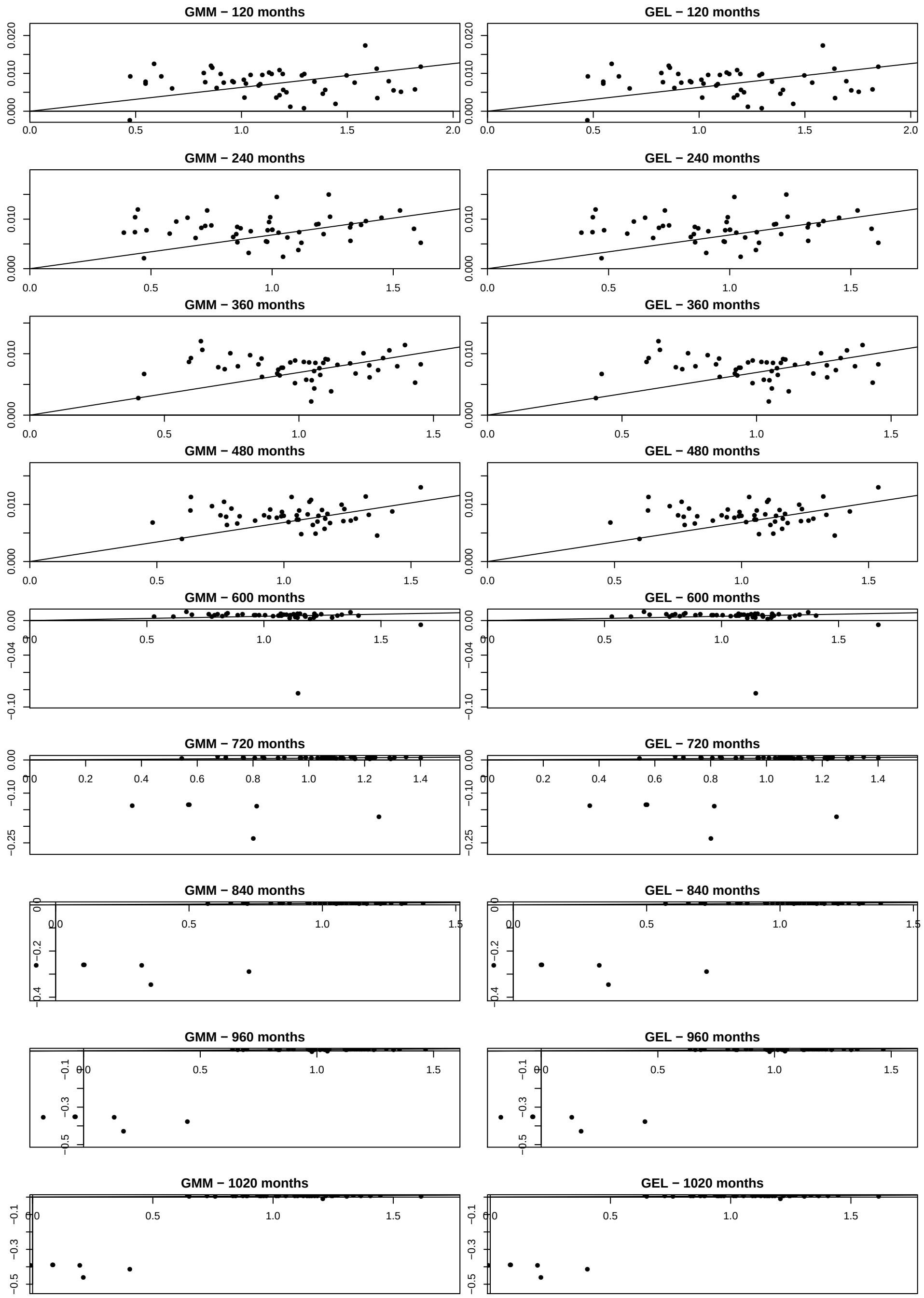
[18]Not surprisingly, we were not able to perform the tests of efficiency by neither estimation method when using the multiplicative approach (*managed portfolios*).

Table 7: Tests of portfolio efficiency using 49 industry portfolios for selected periods of time

	Months	Wald Test		GRS Test		Wald Test		GRS Test	
		Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
		CAPM				**FF**			
GMM	120	NA	NA	NA	NA	NA	NA	NA	NA
	240	41.9	0.753	0.7	0.946	NA	NA	NA	NA
	360	50.6	0.411	0.9	0.685	31.3	0.977	0.5	0.994
	480	60.7	0.123	1.1	0.292	43.7	0.689	0.8	0.838
	600	NA	NA	NA	NA	NA	NA	NA	NA
	720	NA	NA	NA	NA	NA	NA	NA	NA
	840	NA	NA	NA	NA	NA	NA	NA	NA
	960	NA	NA	NA	NA	NA	NA	NA	NA
	1020	NA	NA	NA	NA	NA	NA	NA	NA
GEL	120	108.7	0.000	1.3	0.160	NA	NA	NA	NA
	240	48.0	0.514	0.8	0.852	NA	NA	NA	NA
	360	44.1	0.672	0.8	0.860	NA	NA	NA	NA
	480	52.4	0.344	1.0	0.557	NA	NA	NA	NA
	600	81.9	0.002	1.5	0.014	NA	NA	NA	NA
	720	142.8	0.000	2.7	0.000	NA	NA	NA	NA
	840	206.4	0.000	4.0	0.000	NA	NA	NA	NA
	960	212.8	0.000	4.1	0.000	NA	NA	NA	NA
	1020	231.9	0.000	4.5	0.000	NA	NA	NA	NA

Tests of portfolio efficiency using 49 industry portfolios for 9 selected periods of time: T=120 (10 years), T=240 (20 years), T=360 (30 years), T=480 (40 years), T=600 (50 years), T=720 (60 years), T=840 (70 years), T=960 (80 years) and T=1020 (85 years). The tests are conducted based on both estimations methodology: via GMM are on top, while via GEL are on the bottom. Tests of efficiency under the CAPM asset pricing model are on the left, while tests under Fama-French three-factor model (represented as “FF”) are on the right. Table presents the statistic and the p-values of the Wald and the GRS tests for each case. “NA” denotes not applicable, situations in which singularity problems occur impeding the inversion of the covariance matrix.

Figure 15: CAPM Model - Comparison of GMM and GEL estimated betas against the sample mean of monthly excess returns for the portfolio with 49 assets



In all panels, the estimated betas ($\hat{\beta}_{Mkt}$) are in the x-axis, and the sample mean of the monthly excess returns for each of the $n = 49$ assets in the portfolio are in the y-axis. Estimations via GMM are on the left, while via GEL are on the right. Each panel represents one of the time intervals, starting with 120 months before December 2014 to 1020 months (85 years) prior to this date.

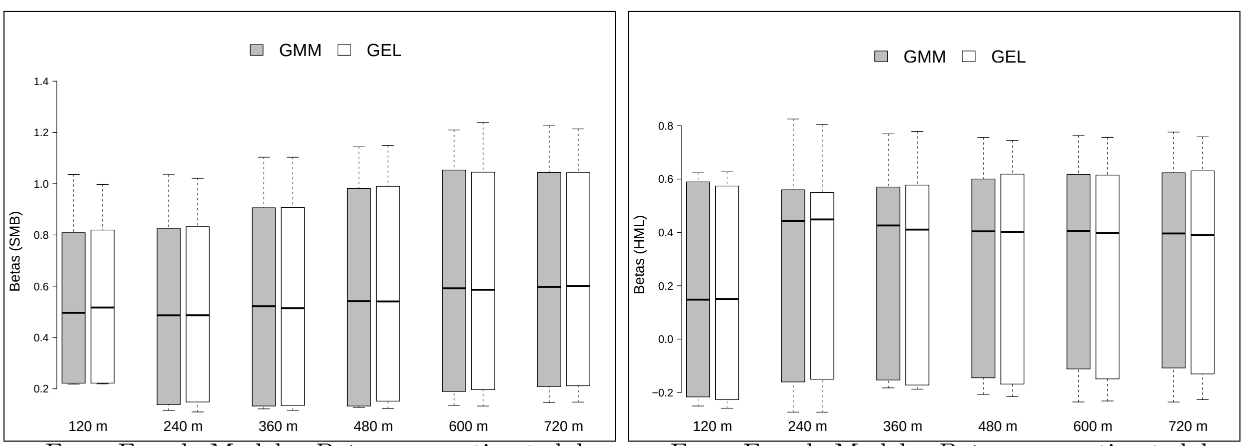
decrease as the sample size increases. For Fama-French, while GEL did not provide any test statistics, GMM results were similar to the CAPM, but with higher p-values. The p-values of the Wald test is 0.98 and 0.69 for 360 and 480 months respectively, while the GRS test p-values are 0.99 and 0.84 for the same periods. On the other hand, for CAPM, the tests of efficiency using GEL have small p-values for $T = 600$ and higher, while GMM had singularities issues for the same intervals.

These results show some evidence that the Wald test, based on large sample distribution, tends to reject the null hypothesis more often compared to tests that rely on finite sample distributions (like the GRS). This characteristic of the tests is in line with the analysis of the size and power of efficiency tests in Campbell et al. (1997).

Figure 14 shows the boxplots of $\hat{\alpha}$ and $\hat{\beta}_{Mkt}$ (CAPM). We do not present the boxplots for the Fama-French model boxplots due to GEL's failure to provide estimates. The boxplots show similar distributions for GMM and GEL[19]. Again, GRS and Wald tests (Table 7) have different values between both estimators mainly because of the difference in the estimated covariance matrix. For CAPM, Figure 15 plots the sample mean of the monthly excess returns ($\hat{E}(R_i)$) against the estimated betas ($\hat{\beta}_{Mkt}$) for the 49 assets.

Additional Figures and Tables

Figure 16: Boxplots for comparison of estimations by GMM and GEL for Fama-French model using 6 portfolios formed on size and book-to-market under *scaled returns* for selected periods of time



(a) Fama-French Model: $Betas_{SMB}$ estimated by GMM and GEL

(b) Fama-French Model: $Betas_{HML}$ estimated by GMM and GEL

In both panels, the x-axis represents the time intervals, starting with 120 months before December 2014 to 720 months prior to this date. The y-axis are the estimated coefficients values by GMM and GEL. Estimation by GMM is represented by gray boxplots, while GEL estimation is represented by white boxplots.

[19]Note that the longer the sample size, more common is the presence of outliers in the alpha estimates.

Table 8: Rejections proportions of J_{Wald} and J_{GRS} tests for nominal sizes of 1%, 2.5%, 5% e 10% from the Monte Carlo experiments for each of the 4 scenarios (Model=Fama-French, N=6, T=120, 500 simulations)

	1%				**2.50%**				**5%**				**10%**			
	Wald Test		GRS Test		Wald Test		GRS Test		Wald Test		GRS Test		Wald Test		GRS Test	
	GMM	GEL	GMM	GEL	GMM	GEL	GMM	GEL	GMM	GEL	GMM	GEL	GMM	GEL	GMM	GEL
$\hat{\varepsilon}_{i,t}^{\text{Sim}^*} = \hat{\xi}_{i,t}^{\text{Sim1}}$	0.2410	0.0620	0.1867	0.0400	0.3454	0.1100	0.2631	0.0720	0.4036	0.1580	0.3494	0.1160	0.4940	0.2340	0.4337	0.1860
$\hat{\varepsilon}_{i,t}^{\text{Sim}^*} = \hat{\nu}_{i,t}^{\text{Sim2}}$	0.2705	0.0960	0.2184	0.0640	0.3547	0.1580	0.2926	0.1140	0.4369	0.2320	0.3647	0.1780	0.5311	0.3180	0.4709	0.2660
$\hat{\varepsilon}_{i,t}^{\text{Sim}^*} = \mathbb{1}_{t=T/2}(\hat{\kappa}_{i,t}^{\text{Sim3}})$	0.3026	0.1000	0.2265	0.0640	0.3667	0.1480	0.3126	0.1080	0.4269	0.2120	0.3747	0.1540	0.5251	0.2780	0.4569	0.2220
$\hat{\varepsilon}_{i,t}^{\text{Sim}^*} = \hat{\xi}_{i,t}^{\text{Sim4}} - \mathbb{1}_{\hat{p}_{i,t}<0.05}(\hat{\kappa}_{i,t}^{\text{Sim4}})$	0.8008	0.8740	0.7344	0.8040	0.8571	0.9260	0.8189	0.8820	0.9034	0.9560	0.8652	0.9300	0.9376	0.9700	0.9235	0.9640

Tabulated rejection proportions from the Monte Carlo experiments for nominal sizes of 1%, 2.5%, 5% e 10%. The sample data used is the 6 portfolios formed on size and book-to-market (2x3) with T=120 observations for each asset. Tests of efficiency are evaluated only for the Fama-French three-factor asset pricing model. Table presents results for 4 different scenarios with 500 simulations each. First case ($\hat{\varepsilon}_{i,t}^{\text{Sim}^*} = \hat{\xi}_{i,t}^{\text{Sim1}}$) assess Wald and GRS tests under Gaussian shocks. For the second scenario ($\hat{\varepsilon}_{i,t}^{\text{Sim}^*} = \hat{\nu}_{i,t}^{\text{Sim2}}$) innovations are drawn from a t-Student distribution. In the third case ($\hat{\varepsilon}_{i,t}^{\text{Sim}^*} = \mathbb{1}_{t=T/2}(\hat{\kappa}_{i,t}^{\text{Sim3}})$) the DGP insert a large magnitude shock on a fixed date. Finally, the fourth scenario ($\hat{\varepsilon}_{i,t}^{\text{Sim}^*} = \hat{\xi}_{i,t}^{\text{Sim4}} - \mathbb{1}_{\hat{p}_{i,t}<0.05}(\hat{\kappa}_{i,t}^{\text{Sim4}})$) the DGP assumes that there is a probability of 5% that an outlier exists.