Spurious Factors in Linear Asset Pricing Models

SVETLANA BRYZGALOVA[*]

2nd July, 2015

ABSTRACT

When a risk factor has small covariance with asset returns, risk premia in the linear asset pricing models are no longer identified. *Weak factors*, similar to weak instruments, make the usual estimation techniques unreliable. When included in the model, they generate spuriously high significance levels of their own risk premia estimates, overall measures of fit and may crowd out the impact of the true sources of risk. I develop a new approach to the estimation of cross-sectional asset pricing models that: a) provides simultaneous model diagnostics and parameter estimates; b) automatically removes the effect of spurious factors; c) restores consistency and asymptotic normality of the parameter estimates, as well as the accuracy of standard measures of fit; d) performs well in both small and large samples. I provide new insights on the pricing ability of various factors proposed in the literature. In particular, I identify a set of robust factors (e.g. Fama-French ones, but not only), and those that suffer from severe identification problems that render the standard assessment of their pricing performance unreliable (e.g. consumption growth, human capital proxies and others).

JEL classification: G12, C13, C58.

---

[*]London School of Economics and Political Science, Department of Economics and FMG, s.bryzgalova1@lse.ac.uk.

Sharpe (1964) and Lintner (1965) CAPM pioneered the class of linear factor models in asset pricing. Now, decades later, what started as an elegant framework has turned into a well-established and successful tradition in finance. Linear models, thanks to their inherent simplicity and ease of interpretation, are widely used as a reference point in much of the empirical work, having been applied to nearly all kinds of financial assets[1]. In retrospect, however, such heavy use produced a rather puzzling outcome: Harvey, Liu, and Zhu (2013) list over 300 factors proposed in the literature, all of which have been claimed as important (and significant) drivers of the cross-sectional variation in stock returns[2].

One of the reasons for such a wide range of apparently significant risk factors is perhaps a simple lack of model identification, and consequently, an invalid inference about risk premia parameters. As pointed out in a growing number of papers (see e.g. Jagannathan and Wang (1998), Kan and Zhang (1999b), Kleibergen (2009), Kleibergen and Zhan (2013), ?, Gospodinov, Kan, and Robotti (2014a)), in the presence of factors that only weakly correlate with assets (or do not correlate at all), all the risk premia parameters are no longer strongly identified and standard estimation and inference techniques become unreliable. As a result, identification failure often leads to the erroneous conclusion that such factors are important, although they are totally *spurious* by nature. The impact of the true factors could, in turn, be crowded out from the model.

The shrinkage-based estimators that I propose (Pen-FM and Pen-GMM, from the *penalised* version of the Fama-MacBeth procedure or GMM, accordingly), not only allow to detect the overall problem of rank-deficiency caused by irrelevant factors, but also indicate which particular variables are causing it, and recover the impact of strong risk factors without compromising any of its properties (e.g. consistency, asymptotic normality, etc).

My estimator can bypass the identification problem because, in the case of useless (or weak) factors, we know that it stems from the low correlation between these variables and asset returns. This, in turn, is reflected in the regression-based estimates of *betas*, asset exposures to the corresponding sources of risk. Therefore, one can use the L_1-norm of the vector of $\hat{\beta}$'s (or related quantities) to assess the overall factor strength for a given cross-section of returns, and successfully isolate the cases when it is close to zero. Consequently, I modify the second stage of the Fama-MacBeth procedure[3]

---

[1]Notable examples are the 3-factor model of Fama and French (1992), Fama and French (1993); the conditional CAPM of Jagannathan and Wang (1996); the conditional CCAPM of Lettau and Ludvigson (2001b), the Q1-Q4 consumption growth of Jagannathan and Wang (2007), the durable/nondurable consumption CAPM of Yogo (2006); the ultimate consumption risk of Parker and Julliard (2005); the pricing of currency portfolios in Lustig and Verdelhan (2007) and Lustig, Roussanov, and Verdelhan (2011); and the regression-based approach to the term structure of interest rates in Adrian, Crump, and Moench (2013)

[2]In the context of predictive regressions, Novy-Marx (2014) recently demonstrated that many unconventional factors, such as the party of the U.S. President, sunspots, the weather in Manhattan, plannet location and the El-Nino phenomenon have a statistically significant power for the performance of many popular trading strategies, such as those based on market capitalisation, momentum, gross profitability, earnings surprises and others.

[3]The problem of identification is not a consequence of having several stages in the estimation. It is well known that the two-pass procedure gives exactly the same point estimates as GMM with the identity weight matrix under a

(or the GMM objective function) to include a penalty that is inversely proportional to the factor strength, measured by the L_1-norm of the vector $\hat{\beta}$.

One of the main advantages of this penalty type is its ability to simultaneously recognise the presence of both useless and weak factors[4], allowing Pen-FM(GMM) to detect the problem of both under- and weak identification. On the contrary, the critical values for the tests often used in practice[5] are all derived under the assumption of *strictly zero* correlation between the factor and returns. As a result, faced with a *weak* factor, such tests tend to reject the null hypothesis of betas being jointly zero; however, risk premia parameters still have a nonstandard asymptotic distribution, should the researcher proceed with the standard inference techniques[6].

Combining model selection and estimation in one step is another advantage of Pen-FM(GMM), because it makes the model less prone to the problem of *pretesting*, when the outcome of the initial statistical procedure and decision of whether to keep or exclude some factors from the model further distort parameter estimation and inference[7].

Eliminating the influence of irrelevant factors is one objective of the estimator; however, it should also reflect the pricing ability of other variables in the model. I construct the penalty in such a way that does not prevent recovering the impact of strong factors. In fact, I show that Pen-FM(GMM) provide consistent and asymptotically normal estimates of the strong factors risk premia that have *exactly* the same asymptotic distribution as if the irrelevant factors had been known and excluded from the model *ex ante*. Further, I illustrate, with various simulations, that my estimation approach also demonstrates good finite sample performance even for a relatively small sample of 50-150 observations. It is successful in a) eliminating spurious factors from the model, b) retaining the valid ones, c) estimating their pricing impact, and d) recovering the overall quality of fit.

I revisit some of the widely used linear factor models and confirm that many tradable risk factors seem to have substantial covariance with asset returns. This allows researchers to rely on either standard or shrinkage-based estimation procedures, since both deliver identical point estimates and confidence bounds (e.g. the three-factor model of Fama and French (1992), or a four-factor model that additionally includes the quality-minus-junk factor of Asness, Frazzini, and Pedersen (2014)).

There are cases, however, when some of the factors are particularly weak for a given cross-section

---

particular moment normalisation.

[4]If the time series estimates of beta have the standard asymptotic behaviour, then for both useless ($\beta = \mathbf{0_n}$) and weak ($\beta = \frac{B}{\sqrt{T}}$) factors L_1-norm of $\hat{\beta}$ is of the order $\frac{1}{\sqrt{T}}$.

[5]Wald test for the joint spread of betas or more general rank deficiency tests, such as Cragg and Donald (1997), Kleibergen and Paap (2006)

[6]A proper test for the strength of the factor should be derived under the null of weak identification, similar to the critical value of 10 for the first stage F-statistics in the case of a single endogenous variable and 1 instrument in the IV estimation, or more generally the critical values suggested in Stock and Yogo (2005)

[7]See, e.g. simulation designs in Breiman (1996) highlighting the model selection problem in the context of linear regressions and the choice of variables, Guggenberger (2010) for the impact of Hausman pretest in the context of panel data, and Berk, Brown, Buja, Zhang, and Zhao (2013) for recent advances in constructing confidence bounds, robust to prior model selection

of assets, and their presence in the model only masks the impact of the true sources of risk. The new estimator proposed in this paper allows then to uncover this relationship and identify the actual pricing impact of the strong factors. This is the case, for example, of the q-factor model of Hou, Xue, and Zhang (2014) and the otherwise 'hidden' impact of the profitability factor, which I find to be a major driving force behind the cross-sectional variation in momentum-sorted portfolios.

Several papers have recently proposed[8] asset pricing models that highlight, among other things, the role of investment and profitability factors, and argue that these variables should be important drivers of the cross-sectional variation in returns, explaining a large number of asset pricing puzzles[9]. However, when I apply the q-factor model (Hou et al. (2014)) to the momentum-sorted cross-section of portfolios using the Fama-MacBeth procedure, none of the variables seem to command a significant risk premium, although the model produces an impressive R^2 of 93%. Using Pen-FM on the same dataset eliminates the impact of two out of four potential risk drivers, and highlights a significant pricing ability of the profitability factor (measured by ROE), largely responsible for 90% of the cross-sectional variation in portfolio returns. Point estimates of the risk premia (for both market return and ROE), produced by Pen-FM in this case are also closer to the average return generated by a tradable factor, providing further support for the role of the firm's performance in explaining the momentum effect, as demonstrated in Hou et al. (2014). The importance of this factor in explaining various characteristics of stocks is also consistent with the findings of Novy-Marx (2013), who proposes an alternative proxy for expected profitability and argues that it is crucial in predicting the cross-sectional differences of stock returns.

While specifications with tradable factors seem to be occasionally contaminated by the problem of useless factors, the situation seems to be much worse when a nontradable source of risk enters into the model. For example, I find that specifications including such factors as durable consumption growth or human capital proxies are not strongly identified[10] and Pen-FM shrinks their risk premia towards zero. Since conventional measures of fit, such as the crossectional R^2, are often inflated in the presence of spurious factors (Kleibergen and Zhan (2013), Gospodinov, Kan, and Robotti (2014b)), their high in-sample values only mask a poorly identified model.

It is worth noting, however, that when a particular risk driver is identified as weak (or useless), it does not necessarily render the model containing it invalid. The finding merely highlights the impossibility of assessing the size of the risk premia paremeters, significance of their pricing impact and the resulting quality of fit, based on the standard estimation techniques. The method that I propose allows to recover identification and quality of fit only for strong risk factors (which is

---

[8]E.g. Chen, Novy-Marx, and Zhang (2011), Fama and French (2014) and Hou et al. (2014)

[9]There is vast empirical support for shocks to a firm's profitability and investment to be closely related to the company's stock performance, e.g.Ball and Brown (1968), Bernand and Thomas (1990), Chan, Jegadeesh, and Lakonishok (1996), Haugen and Baker (1996), Fairfield, Whisenant, and Yohn (2003), Titman, Wei, and Xie (2004), Fama and French (2006), Cooper, Gulen, and Schill (2008), Xing (2008), Polk and Sapienza (2009), **?**

[10]This finding is consistent with the results of identification tests in Zhiang and Zhan (2013) and **?**

contaminated otherwise), but stays silent regarding the impact of the weak ones. Furthermore, since I focus on the multiple-beta representation, the risk premia reflect the *partial* pricing impact of a factor. Therefore, it is also plausible to have a model with a factor being *priced* within a linear SDF setting, but not contributing anything on its own, that is *conditional on other factors in the model.* When estimated by the Fama-MacBeth procedure, its risk premium is no longer identified. Although the focus of my paper is on the models that admit multivariate beta-representation, nothing precludes extending shrinkage-based estimators to a linear SDF setting to assess the *aggregate* factor impact as well.

Why does identification have such a profound impact on parameter estimates? The reason is simple: virtually any estimation technique relies on the existence of a unique set of true parameter values that satisfies the model's moment conditions or minimises a loss function. Therefore, violations of this requirement in general deliver estimates that are inconsistent, have non-standard distribution, and require (when available) specifically tuned inference techniques for hypothesis testing. Since the true, population values of the β's on an irrelevant factor are zero for all the assets, the risk premia in the second stage are no longer identified, and the entire inference is distorted. Kan and Zhang (1999b) show that even a small degree of model misspecification would be enough to inflate the useless factor t-statistic, creating an illusion of its pricing importance. Kleibergen (2009) further demonstrates that the presence of such factors has a drastic impact on the consistency and asymptotic distribution of the estimates even if the model is correctly specified and the true β's are zero only asymptotically $(\beta = \frac{B}{\sqrt{T}})$.

When the model is not identified, obtaining consistent parameter estimates is generally hard, if not impossible. There is, however, an extensive literature on inference, originating from the problem of weak instruments (see, e.g. Stock, Watson, and Yogo (2002)). Kleibergen (2009) develops identification-robust tests for the two-step procedure of Fama and MacBeth, and demonstrates how to build confidence bounds for the risk premia and test hypotheses of interest in the presence of spurious or weak factors. Unfortunately, the more severe is the identification problem, the less information can be extracted from the data. Therefore, it comes as no surprise that in many empirical applications robust confidence bounds can be unbounded at least from one side, and sometimes even coincide with the whole real line (as in the case of conditional Consumption-CAPM of Lettau and Ludvigson (2001b)), making it impossible to draw any conclusions either in favour of or against a particular hypothesis. In contrast, my approach consists in recovering a subset of parameters that are strongly identified from the data, resulting in their consistent, asymptotically normal estimates and usual confidence bounds. I prove that when the model is estimated by Pen-FM, standard bootstrap techniques can be used to construct valid confidence bounds for the strong factors risk premia even in the presence of useless factors. This is due to the fact that my penalty depends the nature of the second stage regressor (strong or useless), which remaines the same in bootstrap and allows

the shrinkage term to eliminate the impact of the useless factors. As a result, bootstrap remains consistent and does not require additional modifications (e.g. Andrews and Guggenberger (2009), Chatterjee and Lahiri (2011)).

Using various types of penalty to modify the properties of the original estimation procedure has a long and celebrated history in econometrics, with my estimator belonging to the class of Least Absolute Selection and Shrinkage Operator (i.e. lasso, Tibshirani (1996))[11]. The structure of the penalty, however, is new, for it is designed not to choose significant parameters in the otherwise fully identified model, but rather select a subset of parameters that can be strongly identified and recovered from the data. The difference is subtle, but empirically rather striking. Simulations confirm that whereas Pen-FM successfully captures the distinction between strong and weak factors even for a very small sample size, the estimates produced, for instance, by the adaptive lasso (Zou (2006)), display an erratic behaviour[12].

The paper also conributes to a recent strand of literature that examines the properties of conventional asset pricing estimation techniques. Lewellen, Nagel, and Shanken (2010) demonstrate that when a set of assets exhibits a strong factor structure, any variable correlated with those unobserved risk drivers may be indentified as a significant determinant of the cross-section of returns. They assume that model parameters are identified, and propose a number of remedies to the problem, such as increasing the asset span by including portfolios, constructed on other sorting mechanisms, or reporting alternative measures of fit and confidence bounds for them. These remedies, however, do not necessarily lead to better identification.

? highlights the importance of using different SDF normalisations, their effect on the resulting identification conditions and their relation to the useless factor problem. He further suggests using the Kleibergen and Paap (2006) test for rank deficiency as a model selection tool. Gospodinov et al. (2014a) also consider the SDF-based estimation of a potentially misspecified asset pricing model, contaminated by the presence of irrelevant factors. They propose a sequential elimination procedure that successfully identifies spurious factors and those that are not priced in the cross-section of returns, and eliminates them from the candidate model. In contrast, the focus of my paper is on the models with β-representation, which reflect the *partial pricing impact* of different risk factors[13].

---

[11]Various versions of shrinkage techniques have been applied to a very wide class of models, related to variable selection, e.g. adaptive lasso (Zou (2006)) for variable selection in a linear model, bridge estimator for GMM (Caner (2009)), adaptive shrinkage for parameter and moment selection (Liao (2013)), or instrument selection (Caner and Fan (2014))

[12]This finding is expected, since the adaptive lasso, like all other similar estimators, requires identification of the original model parameters used either as part of the usual loss function, or the penalty imposed on it. Should this condition fail, the properties of the estimator will be substantially affected. This does not, however, undermine any results for the correctly identified model

[13]In addition, the two-step procedure could also be used in the applications that rely on the separate datasets used in the estimation of betas and risk premia. For example, Bandi and Tamoni (2014) and Boons and Tamoni (2014) estimate betas from long-horizon regressions and use them to price the cross-section of returns observed at a higher frequency, which would be impossible to do using a standard linear SDF-based approach

Further, I use the simulation design from Gospodinov et al. (2014a) to compare and contrast the finite sample performance of two approaches when the useless factors are assumed to have zero true covariance with asset returns. While both methods can successfully identify and exclude useless factors, my estimator seems to be less conservative, and does not result in the accidental elimination of strong factors from the model with a relatively small sample size, when it is hard to reliably assess the pricing impact of the factor. The difference could be particularly important for empirical applications that use quarterly or yearly data, where the available sample is naturally quite small.

The rest of the paper is organised as follows. I first discuss the structure of a linear factor model and summarise the consequences of identification failure established in the prior literature. Section III introduces Pen-FM and Pen-GMM estimators. I then discuss their asymptotic properties (Section IV) and simulation results (Section V). Section VI presents empirical applications, and Section VII concludes.

I. Linear factor model

I consider a standard linear factor framework for the cross-section of asset returns, where the risk premia for n portfolios are explained through their exposure to k factors, that is

$$
\begin{aligned}
E(R_t^e) &= i_n\lambda_{0,c} + \beta_f\lambda_{0,f}, \\
cov(R_t^e, F_t) &= \beta_f var(F_t), \\
E(F_t) &= \mu_f,
\end{aligned}
\tag{1}
$$

where $t = 1...T$ is the time index of the observations, R_t^e is the $n \times 1$ vector of excess portfolios returns, F_t is the $k \times 1$ vector of factors, $\lambda_{0,c}$ is the intercept (zero-beta excess return), $\lambda_{0,f}$ is the $k \times 1$ vector of the risk premia on the factors, β_f is the $n \times k$ matrix of portfolio betas with respect to the factors, and μ_f is the $k \times 1$ vector of the factors means. Although many theoretical models imply that the common intercept should be equal to 0, it is often included in empirical applications to proxy the imperfect measurement of the risk-free rate, and hence is a common level factor in excess returns.

Model (1) can also be written equivalently as follows

$$
\begin{aligned}
R_t^e &= i_n\lambda_{0,c} + \beta_f\lambda_{0,f} + \beta_f v_t + u_t, \\
F_t &= \mu_f + v_t,
\end{aligned}
\tag{2}
$$

where u_t and v_t are $n \times 1$ and $k \times 1$ vectors of disturbances.

After demeaning the variables and eliminating μ_f, the model becomes:

$$R_t^e = i_n \lambda_{0,c} + \beta_f(\overline{F}_t + \lambda_{0,f}) + \epsilon_t = \beta \lambda_0 + \beta_f \overline{F}_t + \epsilon_t, \tag{3}$$

$$F_t = \mu_f + v_t,$$

where $\epsilon_t = u_t + \beta_f \bar{v}$, $\bar{v} = \frac{1}{T} \sum_{t=1}^{T} v_t$, $\overline{F}_t = F_t - \overline{F}$, $\overline{F} = \frac{1}{T} \sum_{t=1}^{T} F_t$, $\beta = (i_n \ \beta_F)$ is a $n \times (k+1)$ matrix, stacking both the $n \times 1$ unit vector and asset betas, and $\lambda_0 = (\lambda_{0,c}, \lambda'_{0,f})'$ is a $(k+1) \times 1$ vector of the common intercept and risk premia parameters.

Assuming ϵ_t and v_t are asymptotically uncorrelated, our main focus is on estimating the parameters from the first equation in (3). A typical approach would be to use the Fama-MacBeth procedure, which decomposes the parameter estimation in two steps, focusing separately on time series and cross-sectional dimensions.

The first stage consists in time series regressions of excess returns on factors, to get the estimates of β_f:

$$\hat{\beta}_f = \sum_{t=1}^{T} \bar{R}_t^e \bar{F}_t' \left[\sum_{j=1}^{T} \bar{F}_j \bar{F}_j' \right]^{-1},$$

where $\hat{\beta}_f$ is an $n \times k$ matrix, $\overline{R}_t^e$ is a $n \times 1$ vector of demeaned asset returns, $\overline{R}_t^e = R_t^e - \frac{1}{T} \sum_{t=1}^{T} R_t^e$.

While the time series beta reveals how a particular factor correlates with the asset excess returns over time, it does not indicate whether this correlation is *priced* and could be used to explain the differences between required rates of return on various securities. The second stage of the Fama-MacBeth procedure aims to check whether asset holders demand a premium for being exposed to this source of risk (β_j, $j = 1..k$), and consists in using a single OLS or GLS cross-sectional regression of the average excess returns on their risk exposures.

$$\hat{\lambda}_{OLS} = \left[\hat{\beta}' \hat{\beta} \right]^{-1} \hat{\beta}' \bar{R}^e, \tag{4}$$

$$\hat{\lambda}_{GLS} = \left[\hat{\beta}' \hat{\Omega}^{-1} \hat{\beta} \right]^{-1} \hat{\beta}' \hat{\Omega}^{-1} \bar{R}^e,$$

where $\hat{\beta} = [i_n \ \hat{\beta}_f]$ is the extended $n \times (k+1)$ matrix of $\hat{\beta}$'s, $\hat{\lambda} = [\hat{\lambda}_c \ \hat{\lambda}'_f]'$ is a $(k+1) \times 1$ vector of the risk premia estimates, $\bar{R}^e = \frac{1}{T} \sum_{t=1}^{T} R_t^e$ is a $n \times 1$ vector of the average cross-sectional excess returns, and $\hat{\Omega}$ is a consistent estimate of the disturbance variance-covariance matrix, e.g. $\hat{\Omega} = \frac{1}{T-k-1} \sum_{t=1}^{T} (\bar{R}_t^e - \hat{\beta}_f \bar{F}_t)(\bar{R}_t^e - \hat{\beta}_f \bar{F}_t)'$.

If the model is identified, that is, if the matrix of β has full rank, the Fama-MacBeth procedure delivers risk premia estimates that are consistent and asymptotically normal, allowing one to construct confidence bounds and test hypotheses of interest in the usual way (e.g. using t-statistics). In the presence of a useless or weak factor ($\beta_j = \mathbf{0}_n$ or more generally $\beta_j = \frac{B}{\sqrt{T}}$, where B is an $n \times 1$ vector), however, this condition is violated, thus leading to substantial distortions in parameter

inference.

Although the problem of risk premia identification in the cross-section of assets is particularly clear when considering the case of the two-stage procedure, the same issue arises when trying to jointly estimate time series and cross-sectional parameters by GMM, using the following set of moment conditions:

$$E\left[R_t^e - i_n\lambda_{0,c} - \beta_f(\lambda_{0,f} - \mu_f + F_t)\right] = \mathbf{0}_n,$$

$$E\left[(R_t^e - i_n\lambda_{0,c} - \beta_f(\lambda_{0,f} - \mu_f + F_t))F_t'\right] = \mathbf{0}_{n\times k}, \tag{5}$$

$$E[F_t - \mu_f] = \mathbf{0}_k.$$

Assuming the true values of model parameters $\theta_0 = \{vec(\beta_f); \lambda_{0,c}; \lambda_{0,f}; \mu_f\}$ belong to the interior of a compact set $S \in \mathbb{R}^{nk+k+k+1}$, one could then proceed to estimate them jointly by minimizing the following objective function:

$$\hat{\theta} = \arg\min_{\theta\in S} \left[\frac{1}{T}\sum_{t=1}^T g_t(\theta)\right]' W_T(\theta)\left[\frac{1}{T}\sum_{t=1}^T g_t(\theta)\right], \tag{6}$$

where $W_T(\theta)$ is a positive definite weight $(n + nk + k) \times (n + nk + k)$ matrix, and

$$g_t(\theta) = \begin{bmatrix} R_t^e - i_n\lambda_c - \beta_f(\lambda_f - \mu + F_t) \\ vec\left([R_t^e - i_n\lambda_c - \beta_f(\lambda_f - \mu + F_t)] F_t'\right) \\ F_t - \mu \end{bmatrix} \tag{7}$$

is a sample moment of dimension $(n + nk + k) \times 1$.

In the presence of a useless factor the model is no longer identified, since the matrix of first derivatives $G(\theta_0) = E[G_t(\theta_0)] = E\left[\frac{dg_t(\theta_0)}{d\theta}\right]$ has a reduced column rank if at least one of the vectors in β_f is $\mathbf{0}_{n\times 1}$ or $\frac{B}{\sqrt{T}}$, making the estimates from eq.(6) generally inconsistent and having a nonstandard asymptotic distribution, since

$$\frac{dg_t(\theta_0)}{d\theta'} = \begin{bmatrix} [\lambda_{0,f} - \mu_f + F_t]' \otimes I_n & -i_n & \beta_f & \beta_f \\ (F_t \otimes I_n)\left[(\lambda_{0,f} - \mu_f + F_t)' \otimes I_n\right] & -vec(i_nF_t') & -(F_t \otimes I_n)\beta_f & (F_t \otimes I_n)\beta_f \\ \mathbf{0}_{k\times nk} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k} & -I_k \end{bmatrix}, \tag{8}$$

where $\otimes$ denotes the Kronecker product and I_n is the identity matrix of size n. Note that the presence of useless factors affects only the risk premia parameters, since as long as the mean and the variance-covariance matrix of the factors are well-defined, the first moment conditions in eq. (5) would be satisfied for any λ_f as long as $\beta_f(\lambda_f - \lambda_{0,f}) = 0$. Therefore, identification problem relates only to the risk premia, but not the factor exposures, betas.

Throughout the paper, I consider the linear asset pricing framework, potentially contaminated

by the presence of useless/weak factors, whether correctly specified or not. I call the model *correctly specified* if it includes all the true risk factors and eq.(3) holds. The model under estimation, however, could also include a useless/weak risk driver that is not priced in the cross-section of asset returns.

The model is called *misspecified* if eq.(3) does not hold. This could be caused by either omitting some of the risk factors necessary for explaining the cross-section of asset returns, or if the model is actually a non-linear one. The easiest way to model a misspecification would be to assume the true data-generating process including individual fixed effects for the securities in the cross-sectional equation:

$$E(R_t^e) = \lambda_{0,i} + \beta_f \lambda_{0,f}$$

where $\lambda_{0,i}$ is a $n \times 1$ vector of individual intercepts. In the simulations I consider the case of a misspecified model, where the source of misspecification comes from the omitted risk factors. Therefore, it contaminates the estimation of both betas and risk premia.

II. Identification and what if it's not there

Depending on the nature of the particular identification failure and the rest of the model features, conventional risk premia estimators generally lose most of their properties: consistency, asymptotic normality, not to mention the validity of standard errors and confidence interval coverage for *all* the factors in the model. Further, numerical optimisation techniques may have convergence issues, faced with a relatively flat region of the objective function, leading to unstable point estimates.

Kan and Zhang (1999a) are the first to notice the problem generated by including a factor uncorrelated with asset returns in the GMM estimation framework of a linear stochastic discount factor model. They show that if the initial model is misspecified, the Wald test for the risk premia overrejects the null hypothesis of a factor having zero risk premium, and hence a researcher will probably conclude that it indeed explains the systematic differences in portfolio returns. The likelihood of finding significance in the impact of a useless factor increases with the number of test assets; hence, expanding the set of assets (e.g. combining 25 Fama-French with 19 industry portfolios) may even exacerbate the issue (Gospodinov et al. (2014a)). Further, if the model is not identified, tests for model misspecification have relatively low power, thus making it even more difficult to detect the problem.

Gospodinov et al. (2014a) consider a linear SDF model that includes both strong and useless factors, and the effect of misspecificaion-robust standard errors. Their estimator is based on minimizing the Hansen-Jagannathan distance (Hansen and Jagannathan (1997)) between the set of SDF pricing the cross-section of asset returns, and the ones implied by a given linear factor structure. This setting allows to construct misspecification-robust standard errors, because the value of the

objective function can be used to assess the degree of model misspecification. They demonstrate that the risk premia estimates of the useless factors converge to a bounded random variable, and are inconsistent. Under correct model specificatation, strong factors risk premia estimates are consistent; however, they are no longer asymptotically normal. Further, if the model is misspecified, risk premia estimates for the strong factors are inconsistent and their pricing impact could be crowded out by the influence of the useless ones. Useless factors t-statistics, in turn, are inflated and asymptotically tend to infinity.

Kan and Zhang (1999b) study the properties of the Fama-MacBeth two-pass procedure with a single useless risk factor ($\beta = \mathbf{0}_n$), and demonstrate the same outcome. Thus, faced with a finite sample, a researcher is likely to conclude that such a factor explains the cross-sectional differences in asset returns. Kleibergen (2009) also considers the properties of the OLS/GLS two-pass procedure, if the model if weakly identified ($\beta = \frac{B}{\sqrt{T}}$). The paper proposes several statistics that are robust to identification failure and thus could be used to construct confidence sets for the risk premia parameters without pretesting.

Cross-sectional measures of fit are also influenced by the presence of irrelevant factors. Kan and Zhang (1999b) conjecture that in this case cross-sectional OLS-based R^2 tends to be substantially inflated, while its GLS counterpart appears to be less affected. This was later proved by Kleibergen and Zhan (2013), who derive the asymptotic distribution of R^2 and GLS-R^2 statistics and confirm that, although both are affected by the presence of useless factors, the OLS-based measure suffers substantially more. Gospodinov et al. (2014b) consider cross-sectional measures of fit for the families of invariant (i.e. MLE, CUE-GMM, GLS) and non-invariant estimators in both SDF and beta-based frameworks and show that the invariant estimators and their fit are particularly affected by the presence of useless factors and model misspecification.

III.   Pen-FM estimator

Assuming the true values of risk premia parameters $\lambda_0 = (\lambda_{0,c}, \lambda_{0,F})$ lie in the interior of the compact parameter space $\Theta \in \mathbb{R}^k$, consider the following penalised version of the second stage in the Fama-MacBeth procedure:

$$\hat{\lambda}_{pen} = \underset{\lambda \in \Theta}{\arg\min} \left[\bar{R}^e - \hat{\beta}\lambda \right]' W_T \left[\bar{R}^e - \hat{\beta}\lambda \right] + \eta_T \sum_{j=1}^{k} \frac{1}{||\hat{\beta}_j||_1^d} |\lambda_j|, \qquad (9)$$

where $d > 0$ and $\eta_T > 0$ are tuning parameters, and $||\cdot||_1$ stands for the L_1 norm of the vector, $||\hat{\beta}_j||_1 = \sum_{i=1}^{n} |\hat{\beta}_{i,j}|$.

The objective function in eq. 9 is composed of two parts: the first term is the usual loss function, that typically delivers the OLS or GLS estimates of the risk premia parameters in the cross-sectional

regression, depending on the type of the weight matrix, W_T. The second term introduces the penalty that is inversely proportional to the strength of the factors, and is used to eliminate the irrelevant ones from the model.

Eq. 9 defines an estimator in the spirit of the lasso, Least Absolute Selection and Shrinkage Estimator of Tibshirani (1996) or the adaptive lasso of Zou (2006)[14]. The modification here, however, ensures that the driving force for the shrinkage term is not the value of the risk premium or its prior regression-based estimates (which are contaminated by the identification failure), but the nature of the betas. In particular, in the case of the adaptive lasso, the second stage estimates for the risk premia would have the penalty weights inversely proportional to their prior estimates:

$$\hat{\lambda}_{A.Lasso} = \underset{\lambda \in \Theta}{\arg \min} \left[\bar{R}^e - \hat{\beta}\lambda \right]' W_T \left[\bar{R}^e - \hat{\beta}\lambda \right] + \eta_T \sum_{j=1}^{k} \frac{1}{|\hat{\lambda}_{j,ols}|^d} |\lambda_j|, \qquad (10)$$

where $\hat{\lambda}_j$ is the OLS-based estimate of the factor j risk premium. Since these weights are derived from inconsistent estimates, with those for useless factors likely to be inflated under model misspecification, the adaptive lasso will no longer be able to correctly identify strong risk factors in the model. Simulations in Section V further confirm this distinction.

The reason for using the L_1 norm of the vector $\hat{\beta}_j$, however, is clear from the asymptotic behaviour of the latter:

$$vec(\hat{\beta}_j) = vec(\beta_j) + \frac{1}{\sqrt{T}} N\left(0, \Sigma_{\beta_j}\right) + o_p\left(\frac{1}{\sqrt{T}}\right),$$

where $vec(\cdot)$ is the vectorisation operator, stacking the columns of a matrix into a single vector, $N\left(0, \Sigma_{\beta_j}\right)$ is the asymptotic distribution of the estimates of betas, a normal vector with mean 0 and variance-covariance matrix Σ_{β_j}, and $o_p\left(\frac{1}{\sqrt{T}}\right)$ contains the higher-order terms from the asymptotic expansion that do not influence the estimates $\sqrt{T}$ asymptotics. If a factor is strong, there is at least one portfolio that has true non-zero exposure to it; hence the L_1 norm of $\hat{\beta}$ converges to a positive number, different from 0 ($\left\|\hat{\beta}_j\right\|_1 = O_p(1)$). However, if a factor is useless and does not correlate with any of the portfolios in the cross-section, $\beta_j = \mathbf{0_{n \times 1}}$, therefore the L_1 norm of $\hat{\beta}$ converges to $\left\|\hat{\beta}_j\right\|_1 = O_p(\frac{1}{\sqrt{T}})$. This allows to clearly distinguish the estimation of their corresponding risk premia, imposing a higher penalty on the risk premium for a factor that has small absolute betas. Note that in the case of local-to-zero asymptotics in weak identification ($\beta_{sp} = \frac{1}{\sqrt{T}} B_{sp}$), again $\left\|\hat{\beta}_j\right\|_1 = O_p(\frac{1}{\sqrt{T}})$, the same penalty would be able to pick up its scale and shrink the risk premium at the second pass, eliminating its effect.

What is the driving mechanism for such an estimator? It is instructive to show its main features with an example of a single risk factor and no intercept at the second stage.

---

[14]Similar shrinkage-based estimators were later employed in various contexts of parameter estimation and variable selection. For a recent survey of the shrinkage-related techniques, see, e.g. Liao (2013).

$$\hat{\lambda}_{pen} = \arg\min_{\lambda \in \Theta} \left[\bar{R}^e - \hat{\beta}\lambda \right]' W_T \left[\bar{R}^e - \hat{\beta}\lambda \right] + \eta_T \frac{1}{||\hat{\beta}||_1^d} |\lambda|$$

$$= \arg\min_{\lambda \in \Theta} \left[\lambda - \hat{\lambda}_{WLS} \right]' \hat{\beta}' W_T \hat{\beta} (\lambda - \hat{\lambda}_{WLS}) + \eta_T \frac{1}{||\hat{\beta}||_1^d} |\lambda|,$$

where $\lambda_{WLS} = \left[\hat{\beta}' W_T \hat{\beta} \right]^{-1} \hat{\beta}' W_T \bar{R}^e$ is the weighted least squares estimate of the risk premium (which corresponds to either the OLS or GLS cross-sectional regressions).

The solution to this problem can easily be seen as a soft-thresholding function:

$$\hat{\lambda}_{pen} = sign\left(\hat{\lambda}_{WLS}\right) \left(|\hat{\lambda}_{WLS}| - \eta_T \frac{1}{2\hat{\beta}' W_T \hat{\beta}||\hat{\beta}||_1^d} \right)_+ \tag{11}$$

$$= \begin{cases} \hat{\lambda}_{WLS} - \eta_T \frac{1}{2\hat{\beta}' W_T \hat{\beta}||\hat{\beta}||_1^d} & \text{if } \hat{\lambda}_{WLS} \geq 0 \text{ and } \eta_T \frac{1}{2\hat{\beta}' W_T \hat{\beta}||\hat{\beta}||_1^d} < |\hat{\lambda}_{WLS}| \\ \hat{\lambda}_{WLS} + \eta_T \frac{1}{2\hat{\beta}' W_T \hat{\beta}||\hat{\beta}||_1^d} & \text{if } \hat{\lambda}_{WLS} < 0 \text{ and } \eta_T \frac{1}{2\hat{\beta}' W_T \hat{\beta}||\hat{\beta}||_1^d} < |\hat{\lambda}_{WLS}| \\ 0 & \text{if } \eta_T \frac{1}{2\hat{\beta}' W_T \hat{\beta}||\hat{\beta}||_1^d} \geq |\hat{\lambda}_{WLS}| \end{cases}$$

Eq. 11 illustrates the whole idea behind the modified lasso technique: if the penalty associated with the factor betas is high enough, the weight of the shrinkage term will asymptotically tend to infinity, setting the estimate directly to 0. At the same time, I set the tuning parameters (d and η_T) to such value that the threshold component does not affect either consistency or the asymptotic distribution for the strong factors (for more details, see Section IV).

If there is more than one regressor at the second stage, there is no analytical solution to the minimization problem of Pen-FM; however, it can be easily derived numerically through a sequence of 1-dimensional optimizations on the partial residuals, which are easy to solve. This is the so-called pathwise coordinate descent algorithm, where, at each point in time only one parameter estimate is updated. The algorithm goes as follows:

Step 1. Pick a factor $i \in [1..k]$ and write the overall objective function as

$$L = \left[\bar{R}^e - \hat{\beta}_i \lambda_i - \hat{\beta}_j \tilde{\lambda}_j \right]' W_T \left[\bar{R}^e - \hat{\beta}_i \lambda_i - \hat{\beta}_j \tilde{\lambda}_j \right] + \eta_T \left(\sum_{j=1, j \neq i}^{k} \frac{1}{\left\| \hat{\beta}_j \right\|_1^d} \left| \tilde{\lambda}_j \right| + \frac{1}{\left\| \hat{\beta}_i \right\|_1^d} |\lambda_i| \right)$$

where all the values of λ_j, except for the one related to factor i, are fixed at certain levels $\tilde{\lambda}_{j, j \neq i}$.

Step 2. Optimise L w.r.t λ_i. Note that this is a univariate lasso-style problem, where the residual pricing errors are explained only by the chosen factor i.

Step 3. Repeat the coordinate update for all the other components of λ.

Step 4. Repeat the procedure in Steps 1-3 until convergence is reached.

The convergence of the algorithm above to the actual solution of Pen-FM estimator problem follows from the general results of Tseng (1988, 2001), who studies the coordinate descent in a general framework. The only requirement for the algorithm to work is that the penalty function is

convex and additively separable in the parameters, which is clearly satisfied in the case of Pen-FM. Pathwise-coordinate descent has the same level of computational complexity as OLS (or GLS), and therefore works very fast. It has been applied before to various types of shrinkage estimators, as in Friedman, Hastie, Hoffling, and Tibshirani (2007), and has been shown to be very efficient and numerically stable. It is also robust to potentially high correlations between the vectors of beta, since each iteration relies only on the residuals from the pricing errors.

As in the two-stage procedure, I define the shrinkage-based estimator for GMM (Pen-GMM) as follows:

$$\hat{\theta}_{pen} = \underset{\theta \in S}{\arg\min} \left[\frac{1}{T} \sum_{t=1}^{T} g_t(\theta) \right]' W_T(\theta) \left[\frac{1}{T} \sum_{t=1}^{T} g_t(\theta) \right] + \eta_T \sum_{j=1}^{k} \frac{1}{||\hat{\beta}||_1^d} |\lambda_j|, \tag{12}$$

where S is a compact set in $\mathbb{R}^{nk+k+k+1}$.

The rationale for constructing such a penalty is the same as before, since one can use the properties of the $\hat{\beta}$s to automatically distinguish the strong factors from the weak ones on the basis of some prior estimates of the latter (OLS or GMM based).

It is important to note that the penalty proposed in this paper does not necessarily need to be based on $||\hat{\beta}_j||_1$. In fact, the proofs can easily be modified to rely on any other variable that has the same asymptotic properties, i.e. being $O_p\left(\frac{1}{\sqrt{T}}\right)$ for the useless factors and $O_p(1)$ for the strong ones. Different scaled versions of the estimates of β, such as partial correlations or their Fischer transformation all share this property. Partial correlations, unlike betas, are invariant to linear transformation of the data, while Fisher transformation ($f(\hat{\rho}) = \frac{1}{2} ln \left(\frac{1+\hat{\rho}}{1-\hat{\rho}} \right)$) provides a map of partial correlations from $[-1, 1]$ to $\mathbb{R}$.

IV. Asymptotic results

Similar to most of the related literature, the present paper relies on the following high-level assumptions regarding the behaviour of the disturbance term ϵ_t:

ASSUMPTION 1: *(Kleibergen (2009)). As $T \to \infty$,*

(a)

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\left(\begin{bmatrix} 1 \\ F_t \end{bmatrix} \otimes \left(R_t - i_n \lambda_{0,c} - \beta_f(\overline{f}_t + \lambda_{0,f}) \right) \right) \right] \xrightarrow{d} \begin{bmatrix} \varphi_R \\ \varphi_\beta \end{bmatrix}$$

where φ_R is $n \times 1$, φ_β is $nk \times 1$, where n is the number of portfolios and k is the number of factors. Further, $\left(\varphi_R', \varphi_\beta' \right)' \sim N(0, V)$, where $V = Q \otimes \Omega$, and

$$\underset{(k+1)\times(k+1)}{Q} = \begin{pmatrix} 1 & \mu_f' \\ \mu_f & V_{ff} + \mu_f \mu_f' \end{pmatrix} = E \left[\begin{pmatrix} 1 \\ F_t \end{pmatrix} \begin{pmatrix} 1 \\ F_t \end{pmatrix}' \right], \quad \underset{n \times n}{\Omega} = var(\varepsilon_t), \quad \underset{k \times k}{V_{ff}} = var(F_t)$$

14

(b)

$$plim_{T\to\infty} \frac{1}{T} \sum_{j=1}^{T} \bar{F}_j \bar{F}'_j = Q_{ff}, \quad plim_{T\to\infty} \bar{F} = \mu_f,$$

where Q_{ff} has full rank.

Assumption 1 provides the conditions required for the regression-based estimates of β_f to be easily computed using conventional methods, i.e. the data should conform to certain CLT and LLN, resulting in the standard $\sqrt{T}$ convergence. This assumption is not at all restrictive, and can be derived from various sets of low-level conditions, depending on the data generating process in mind for the behaviour of the disturbance term and its interaction with the factors, e.g. as in Shanken (1992) or Jagannathan and Wang (1998)[15].

LEMMA 1: *Under Assumption 1, average cross-sectional returns and OLS estimator $\hat{\beta}$ have a joint large sample distribution:*

$$\sqrt{T} \begin{pmatrix} \bar{R} - \beta\lambda_f \\ vec(\hat{\beta} - \beta) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_R \\ \psi_\beta \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega & 0 \\ 0 & V_{ff}^{-1} \otimes \Omega \end{pmatrix}, \right]$$

where $\psi_R = \varphi_R$ is independent of $\psi_\beta = (V_{ff}^{-1} \otimes I_n)(\varphi_\beta - (\mu_f \otimes I_n)\varphi_R)$

Proof. See Kleibergen (2009), Lemma 1. □

A.   Correctly specified model

Having intuitively discussed the driving force behind the proposed shrinkage-based approach, I now turn to its asymptotic properties. The following propositions describe the estimator's behaviour in the presence of irrelevant factors: $\beta = (\beta_{ns}, \beta_{sp})$, where β_{ns} is an $n \times k_1$ matrix of the set of betas associated with k_1 non-spurious factors (including a unit vector) and β_{sp} denotes the matrix of the true value of betas for useless $(\beta_{sp} = \mathbf{0}_{n \times (k+1-k_1)})$ or weak $(\beta_{sp} = \frac{B_{sp}}{\sqrt{T}})$ factors.

PROPOSITION 1: *Under Assumption 1, if $W_T \xrightarrow{p} W$, W is a positive definite $n \times n$ matrix, $\eta_T = \eta T^{-d/2}$ with a finite constant $\eta > 0$, $d > 0$ and $\beta'_{ns}\beta_{ns}$ having full rank, $\hat{\lambda}_{ns} \xrightarrow{p} \lambda_{0,ns}$ and $\hat{\lambda}_{sp} \xrightarrow{p} 0$*

Further, if $d > 2$

$$\sqrt{T} \begin{pmatrix} \hat{\lambda}_{ns} - \lambda_{0,ns} \\ \hat{\lambda}_{sp} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} [\beta'_{ns}W\beta_{ns}]^{-1} \beta'_{ns}W \Psi_{\beta,ns}\lambda_{0,ns} + (\beta'_{ns}W\beta_{ns})^{-1} \beta'_{ns}W\psi_R \\ 0 \end{pmatrix}$$

---

[15]For example, Shanken (1992) uses the following assumptions, which easily result in Assumption 1:
1. The vector ϵ_t is independently and identically distributed over time, conditional on (the time series values for) F, with $E[\epsilon_t|F] = 0$ and $Var(\epsilon_t|F) = \Omega$ (rank N)
2. F_t is generated by a stationary process such that the first and second sample moments converge in probability, as $T \to \infty$ to the true moments which are finite. Also, $\bar{F}$ is asymptotically normally distributed.

Jagannathan and Wang (1998) provide low level conditions for a process with conditional heteroscedasticity.

Proof. See Appendix B.1. □

The intuition behind the proof for consistency is clear: the tuning parameter η_T is set in such a way that the overall effect of the penalty, η_T, disappears with the sample size, and therefore does not affect the consistency of the parameter estimation, unless some of its shrinkage components are inflated by the presence of irrelevant factors. If a factor is useless, the L_1 norm of $\hat{\beta}_j$ tends to 0 at the $\sqrt{T}$ rate, and the penalty converges to a positive constant in front of the corresponding $|\lambda_j|$. Further, since $\hat{\beta}_j \to \mathbf{0}_{\mathbf{n \times 1}}$, λ_j disappears from the usual loss function, $\left[\bar{R}^e - \hat{\beta}\lambda \right]' W_T \left[\bar{R}^e - \hat{\beta}\lambda \right]$, and it is the penalty component that determines its asymptotic behaviour, shrinking the estimate towards 0. At the same time, other parameter estimates are not affected, and their behaviour is fully described by standard arguments.

The shrinkage-based second pass estimator has the so-called *oracle* property for the non-spurious factors: the estimates of their risk premia have the same asymptotic distribution as if we had not included the useless factors in advance. Risk premia estimates are asymptotically normal, with two driving sources of the error component: estimation error from the first pass β's (and the resulting error-in-variables problem), and the disturbance term effect from the second pass.

The risk premia for the useless factors are driven towards 0 even at the level of the asymptotic distribution to ensure that they do not affect the estimation of other parameters. It should be emphasized, that the effect of the penalty does not depend on the actual value of the risk premium. Unlike the usual lasso or related procedures, the mechanism of the shrinkage here is driven by the strength of $\hat{\beta}$, regressors in the second pass. Therefore, there is no parameter discontinuity in the vicinity of 0, and bootstrap methods can be applied to approximate the distribution and build the confidence bounds.

One could argue that the assumption of $\beta = 0$ is quite restrictive, and a more realistic approximation of local-to-zero asymptotics should be used. Following the literature on weak instruments, I model this situation by assuming that $\beta_{sp} = \frac{B_{sp}}{\sqrt{T}}$. This situation could arise when a factor has some finite-sample correlation with the assets that eventually disappears asymptotically. As with the case of useless factors, I present the asymptotic results properties of the Pen-FM estimator, when there are weak factors in the model.

PROPOSITION 2: *Under Assumption 1, if $\beta_{sp} = \frac{B_{sp}}{\sqrt{T}}$, $W_T \xrightarrow{p} W$, W is a positive definite $n \times n$ matrix, $\eta_T = \eta T^{-d/2}$ with a finite constant $\eta > 0$, $d > 0$ and $\beta'_{ns}\beta_{ns}$ having full rank, $\hat{\lambda}_{ns} \xrightarrow{p} \lambda_{0,ns}$ and $\hat{\lambda}_{sp} \xrightarrow{p} 0$*

Further, if $d > 2$

$$\sqrt{T} \begin{pmatrix} \hat{\lambda}_{ns} - \lambda_{0,ns} \\ \hat{\lambda}_{sp} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \left(\beta'_{ns} W^{-1} \beta_{ns} \right]^{-1} \beta'_{ns} W^{-1} B_{sp} \lambda_{0,sp} + \left[\beta'_{ns} W^{-1} \beta_{ns} \right)^{-1} \beta'_{ns} W^{-1} (\psi_R + \Psi_{\beta,ns} \lambda_{0,ns}) \\ 0 \end{pmatrix}$$

16

Proof. See Appendix B.2. □

The logic behind the proof is exactly the same as in the previous case. Recall that even in the case of weak identification again $\left\|\hat{\beta}_j\right\|_1 = O_p(\frac{1}{\sqrt{T}})$. Therefore, the penalty function recognises its impact, shrinking the corresponding risk premia towards 0, while leaving the other parameters intact.

The situation with weak factors is slightly different from that with purely irrelevant ones. While excluding such factors does not influence consistency of the strong factors risk premia estimates, it affects their asymptotic distribution, as their influence does not disappear fast enough (it is of the rate $\frac{1}{\sqrt{T}}$, the same as the asymptotic convergence rate), and hence we get an asymptotic bias apart from the usual components of the distribution. Note, that any procedure eliminating the impact of weak factors from the model (e.g. Gospodinov et al. (2014a), **?**), results in the same effect. In small sample it could influence the risk premia estimates; however, the size of this effect depends on several factors, and in general is likely to be quite small, especially compared to the usual error component.

Note that the $\frac{1}{\sqrt{T}}$ bias arises only if the omitted risk premium is non-zero. This requires a factor that asymptotically is not related to the cross-section of returns, but is nevertheless priced. Though unlikely, one cannot rule out such a case ex ante. If the factor is tradable, the risk premium on it should be equal to the corresponding excess return; hence one can use this property to recover a reliable estimate of the risk premium, and argue about the possible size of the bias or try to correct for it.

B.   Misspecified model

Model misspecification severely exacerbates many consequences of the identification failure[16]; however, its particular influence depends on the degree and nature of such misspecification.

The easiest case to consider is mean-misspecification, when factor betas are properly estimated, but the residual average returns on the second stage are non-zero. One might draw an analogy here with panel data, where the presence of individual fixed effects would imply that the pooled OLS regression is no longer applicable. The case of mean-misspecification is also easy to analyse, because it allows us to isolate the issue of the correct estimation of β from the one of recovering the factor risk premia. For example, one can model the return generation process as follows:

$$\bar{R} = c + \beta\lambda_0 + \frac{1}{\sqrt{T}}\psi_R + o_p\left(\frac{1}{\sqrt{T}}\right),$$

$$vec(\hat{\beta}) = vec(\beta) + \frac{1}{\sqrt{T}}\psi_\beta + o_p\left(\frac{1}{\sqrt{T}}\right),$$

where c is a $n \times 1$ vector of the constants. It is well known that both OLS and GLS, applied to the second pass, result in diverging estimates for the spurious factors risk premia and t-statistics

---

[16]See, e.g. Kan and Zhang (1999a), Jagannathan and Wang (1998), Kleibergen (2009) and Gospodinov et al. (2014a)

asymptotically tending to infinity. Simulations confirm the poor coverage of the standard confidence intervals and the fact that the spurious factor is often found to be significant even in relatively small samples. However, the shrinkage-based second pass I propose successfully recognises the spurious nature of the factor. Since the first-pass estimates of β's are consistent and asymptotically normal, the penalty term behaves in the same way as in the correctly specified model, shrinking the risk premia for spurious factors to 0 and estimating the remaining parameters as if the spurious factor had been omitted from the model. Of course, since the initial model is misspecified to begin with, risk premia estimates would suffer from inconsistency, but it would not stem from the lack of model identification.

A more general case of model misspecification would involve an omitted variable bias (or the nonlinear nature of the factor effects). This would in general lead to the inconsistent estimates of betas (e.g. if the included factors are correlated with the omitted ones), invalidating the inference in both stages of the estimation. However, as long as the problem of rank deficiency caused by the useless factors remains, the asymptotic distribution of Pen-FM estimator will continue to share that of the standard Fama-MacBeth regressions without the impact of spurious factors. A similar result can easily be demonstrated for Pen-GMM.

C. Bootstrap

While the asymptotic distribution gives a valid description of the pointwise convergence, a different procedure is required to construct valid confidence bounds. Although traditional shrinkage-based estimators are often used in conjunction with bootstrap techniques, it has been demonstrated that even in the simplest case of a linear regression with independent factors and i.i.d. disturbances, such inferences will be invalid (Chatterjee and Lahiri (2010)). Intuitively this happens because the classical lasso-related estimators incorporate the penalty function, which behaviour depends on the true parameter values (in particular, whether they are 0 or not). This in turn requires the bootstrap analogue to correctly identify the sign of parameters in the ε-neighborhood of zero, which is quite difficult. Some modifications to the residual bootstrap scheme have been proposed to deal with this feature of the lasso estimator (Chatterjee and Lahiri (2011, 2013)).

Fortunately, the problem explained above is not relevant for the estimator that I propose, because the driving force of the penalty function comes only from the nature of the regressors, and hence there is no discontinuity, depending on the true value of the risk premium. Further, in the baseline scenario I work with a 2-step procedure, where shrinkage is used only in the second stage, leaving the time series estimates of betas and average returns unchanged. All of the asymptotic properties discussed in the previous section result from the first order asymptotic expansions of the time series regressions. Therefore, it can be demonstrated that once a consistent bootstrap procedure for time series regressions is established (be it pairwise bootstrap, blocked or any other technique appropriate

to the data generating process in mind), one can easily modify the second stage so that the bootstrap risk premia have proper asymptotic distributions.

Consider any bootstrap procedure (pairwise, residual or block bootstrap) that remains consistent for the first stage estimates, that is

$$\hat{\beta}^* = \hat{\beta} + \frac{1}{\sqrt{T}}\Psi_\beta + o_p\left(\frac{1}{\sqrt{T}}\right)$$

$$\bar{R}^{e*} = \bar{R}^e + \frac{1}{\sqrt{T}}\Psi_R + o_p\left(\frac{1}{\sqrt{T}}\right),$$

where $\hat{\beta}^*$ and $\bar{R}^*$ are the the bootstrap analogues of $\hat{\beta}$ and $\bar{R}$.

Then

$$\hat{\lambda}^*_{pen} = \arg\min_{\lambda \in \Theta} \left[\bar{R}^* - \hat{\beta}^*\lambda\right]' W_T \left[\bar{R}^* - \hat{\beta}^*\lambda\right] + \eta_T \sum_{j=1}^k \frac{1}{\left\|\hat{\beta}^*_j\right\|_1^d} |\lambda_j| \tag{13}$$

is the bootstrap analogue of $\hat{\lambda}_{pen}$.

Let $\hat{H}_n(\cdot)$ denote the conditional cdf of the bootstrap version $B_T^* = \sqrt{T}(\hat{\lambda}^*_{pen} - \hat{\lambda}_{pen})$ of the centred and scaled Pen-FM estimator of the risk premia $B_T = \sqrt{T}(\hat{\lambda}_{pen} - \lambda_0)$.

PROPOSITION 3: *Under conditions of Theorem 1,*

$$\rho(\hat{H}_T^*, \hat{H}_T) \to 0, \ \ as \ T \to \infty,$$

where $\hat{H}_T = P(B_T \leq x), x \in R$ and ρ denotes weak convergence in distribution on the set of all probability measures on $(\mathbb{R}^{(k+1)}, B(\mathbb{R}^{(k+1)}))$

Proof. See Appendix B.3 □

Proposition 3 implies that the bootstrap analogue of Pen-FM can be used as an approximation for the distribution of the risk premia estimates. This result is similar to the properties of the adaptive lasso, that naturally incorporates soft thresholding with regard to the optimisation solution, and unlike the usual lasso of Tibshirani (1996), does not require aditional corrections (e.g. Chatterjee and Lahiri (2010)).

Let $b_T(\alpha)$ denote the α-quantile of $||B_T||$, $\alpha \in (0,1)$. Define

$$I_{T,\alpha} = b \in R^k : ||b - \hat{\lambda}_{pen}|| \leq T^{-1/2} b_T(\alpha)$$

the level-α confidence set for λ.

PROPOSITION 4: *Let $\alpha \in (0,1)$ be such that $P(||B|| \leq t(\alpha) + \nu) > \alpha$ for all $\nu > 0$. Then under*

the conditions of Proposition 1

$$P(\lambda_0 \in I_{T,\alpha}) \to \alpha \quad as\ T \to \infty$$

This holds if there is at least 1 non-spurious factor, or an intercept in the second stage.

Proof. See Appendix B.4 □

In other words, the above proposition states that having a sample of bootstrap analogues for $\hat{\lambda}_{pen}$, one can construct valid percentile-based confidence bounds for strongly identified parameters.

D.   Generalised Method of Moments

One can modify the objective function in eq. (6) to include a penalty based on the initial OLS estimates of the β_F parameters. Similar to the two-step procedure, this would shrink the risk premia coefficients for the spurious factors to 0, while providing consistent estimates for all the other parameters in the model.

The following set of assumptions provides quite general high level conditions for deriving the asymptotic properties of the estimator in the GMM case.

ASSUMPTION 2:   *1. For all $1 \leq t \leq T$, $T \geq 1$ and $\theta \in S$*

 a) $g_t(\theta)$ is m-dependent

 b) $|g_t(\theta_1) - g_t(\theta_2)| \leq M_t|\theta_1 - \theta_2|$,
 with $lim_{T\to\infty} \sum_{t-1}^{T} EM_t^p < \infty$, for some $p > 2$;
 c) $\sup_{\theta \in S} E|g_t(\theta)|^p < \infty$, for some $p > 2$

 2. Define $E\frac{1}{T}\sum_{t=1}^{T} g_t(\theta) = g_{1T}(\theta)$
 a) Assume that $g_{1T}(\theta) \to g_1(\theta)$ uniformly over S, and $g_{1T}(\theta)$ is continuously differentiable in θ;
 b) $g_1(\theta_{0,ns}, \lambda_{sp} = \mathbf{0}_{k_2}) = 0$, and $g_1(\theta_{ns}, \lambda_{sp} = \mathbf{0}_{k_2}) \neq 0$ for $\theta_{ns} \neq \theta_{0,ns}$, where
 $\theta_{ns} = \{\mu, vec(\beta), \lambda_{f,ns}, \lambda_c\}$

 3. Define the following $(n + nk + k) \times (nk + k + 1 + k)$ matrix: $G_T(\theta) = \frac{dg_{1T}(\theta)}{d\theta'}$. Assume that $G_T(\theta) \xrightarrow{p} G(\theta)$ uniformly in a neighbourhood N of $(\theta_{0,ns}, \lambda_{sp} = \mathbf{0}_{k_2})$, $G(\theta)$ is continuous in theta. $G_{ns}(\theta_{ns,0}, \lambda_{sp} = \mathbf{0}_{k_2})$ is an $(n + nk + k) \times (nk + k_1 + k)$ submatrix of $G(\theta_0)$ and has full column rank.

 4. $W_T(\theta)$ is a positive definite matrix, $W_T(\theta) \xrightarrow{p} W(\theta)$ uniformly in $\theta \in S$, where $W(\theta)$ is an $(n + nk + k) \times (n + nk + k)$ symmetric nonrandom matrix, which is continuous in θ and is positive definite for all $\theta \in S$.

The set of assumptions is fairly standard for the GMM literature and stems from the reliance on the empirical process theory, often used to establish the behaviour of the shrinkage-based GMM estimators (e.g. Caner (2009), Liao (2013)). Most of these assumptions could be further substantially simplified (or trivially established) following the structure of the linear factor model and the moment function for the estimation. However, it is instructive to present a fairly general case. Several comments are in order, however.

Assumption 2.1 presents a widespread sufficient condition for using empirical process arguments, and is very easy to establish for a linear class of models (it also encompasses a relatively large class of processes, including the weak time dependence of the time series and potential heteroscedasticity). For instance, the primary conditions for the two-stage estimation procedure in Shanken (1992) easily satisfy these requirements.

Assumptions 2.2 and 2.3, among other things, provide the identification condition used for the moment function and its parameters. I require the presence of k_2 irrelevant/spurious factors to be the only source for the identification failure, which, once eliminated, should not affect any other parameter estimation. One of the direct consequences is that the first-stage OLS estimates of the betas ($\hat{\beta}$) have a standard asymtotic normal distribution and basically follow the same speed of convergence as in the Fama-McBeth procedure, allowing us to rely on them in formulating the appropriate penalty function.

The following proposition establishes the consistency and asymptotic normality of Pen-GMM:

PROPOSITION 5: *Under Assumption 2, if $\beta_{sp} = \mathbf{0}_{n \times k_2}$, $\eta_T = \eta T^{-d/2}$ with a finite constant $\eta > 0$, and $d > 2$, then*

$$\hat{\lambda}_{sp} \xrightarrow{p} \mathbf{0}_{k_2} \ and \ \hat{\theta}_{ns} \xrightarrow{p} \theta_{0,ns}$$

Further, if $d > 2$

$$\sqrt{T}(\hat{\lambda}_{pen,sp}) \xrightarrow{d} \mathbf{0}_{k_2}$$

$$\sqrt{T}(\hat{\theta}_{pen,ns} - \theta_{0,ns}) \xrightarrow{d} [G_{ns}(\theta_0)'W(\theta_0)G_{ns}(\theta_0)]^{-1}G_{ns}(\theta_0)W(\theta_0)Z(\theta_0)$$

where $\theta_{ns} = \{\mu, vec(\beta,), \lambda_{f,ns}, \lambda_c\}$, $Z(\theta_0) \equiv N(0, \Gamma(\theta_0))$, and
$\Gamma(\theta_0) = lim_{T \to \infty} E\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\theta_0)\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\theta_0)\right]'$

Proof. See Appendix B.5. □

The intuition behind these results is similar to the two-pass procedure of Fama-MacBeth: the penalty function is formulated in such a way as to capture the effect of factors with extremely weak correlation with asset returns. Not only does the resulting estimator retain consistency, but it also has an asymptotically normal distribution. Bootstrap consistency for constructing confidence

bounds could be proved using an argument, similar to the one outlined for the Pen-FM estimator in Propositions 3 and 4.

V.   Simulations

Since many empirical applications are characterised by a rather small time sample of available data (e.g. when using yearly observations), it is particularly important to assess the finite sample performance of the estimator I propose. In this section I discuss the small-sample behaviour of the Pen-FM estimator, based on the simulations for the following sample sizes: $T = 30, 50, 100, 250, 500, 1000$.

For a *correctly specified model* I generate normally distributed returns for 25 portfolios from a one-factor model, CAPM. In order to get factor loadings and other parameters for the data-generating process, I estimate the CAPM on the cross-section of excess returns on 25 Fama-French portfolios sorted on size and book-to-market, using quarterly data from 1947Q2 to 2014Q2 and market excess return, measured by the value-weight return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ. The data is taken from Kenneth French website. I then run separate time series regressions of these portfolios excess returns on R_{mkt}^e to get the estimates of market betas, $\hat{\beta}$ (25 × 1), and the variance-covariance matrix of residuals, $\hat{\Sigma}$ (25 × 25). I then run a cross-sectional regression of the average excess returns on the factor loadings to get $\hat{\lambda}_0$ and $\hat{\lambda}_1$.

The true factor is simulated from a normal distribution with the empirical mean and variance of the market excess return. A spurious factor is simulated from a normal distribution with the mean and variance of the real per capita nondurable consumption growth, constructed for the same time period using data from NIPA Table 7.1 and the corresponding PCE deflator. It is independent of all the other innovations in the model. Finally, returns are generated from the following equation:

$$R_t^e = \hat{\lambda}_0 + \hat{\beta}'\hat{\lambda}_1 + \beta'\bar{R}_{t,mkt}^e + \epsilon_t$$

where ϵ_t is generated from a multivariate normal distribution $N\left(0, \hat{\Sigma}\right)$.

I then compare the performance of 3 estimators: (a) Fama-MacBeth, using the simulated market return as the only factor (I call this the *oracle* estimator, since it includes only the true risk factor *ex ante*), (b) Fama-MacBeth, using the simulated market return and the irrelevant factor, (c) Pen-FM estimator, using the simulated market return and the irrelevant factor.

For a *misspecified model* the data is generated from a 3 factor model, based on 3 canonical Fama-French factors (with parameters obtained and data generated as in the procedure outlined above). However, in the simulations I consider estimating a 1 factor model (thus, the source of misspecification is omitting the SMB and HML factors). Again, I compare the performance of 3

estimators: (a) Fama-MacBeth, using the simulated market return as the only factor, (b) Fama-MacBeth, using the simulated market return and the irrelevant factor, (c) Pen-FM estimator, using the simulated market return and the irrelevant factor.

For each of the simulations, I also compute conventional measures of fit:

$$R_{ols}^2 = 1 - \frac{var\left(\bar{R}^e - \hat{\lambda}_{ols}\hat{\beta}\right)}{var(\bar{R}^e)}$$

$$R_{gls,1}^2 = 1 - \frac{var\left(\hat{\Omega}^{-1/2}(\bar{R}^e - \hat{\lambda}_{ols}\hat{\beta})\right)}{var(\hat{\Omega}^{-1/2}\bar{R}^e)}$$

$$R_{gls,2}^2 = 1 - \frac{var\left(\hat{\Omega}^{-1/2}(\bar{R}^e - \hat{\lambda}_{gls}\hat{\beta})\right)}{var(\hat{\Omega}^{-1/2}\bar{R}^e)}$$

$$APE = \frac{1}{n}\sum_{i=1}^{n}|\alpha_i|$$

$$HJ = \sqrt{\hat{\lambda}_{ols}'\left(\sum_{t=1}^{T}R_tR_t'\right)\hat{\lambda}_{ols}}$$

$$q = \alpha'(y\hat{\Omega}y')^+\alpha$$

$$T^2 = \alpha'\left(\frac{(1+\hat{\lambda}_f'\Sigma_f\hat{\lambda}_f)q}{T}\right)^+\alpha$$

where R_{ols}^2 is the cross-sectional OLS-based R^2, $R_{gls,1}^2$ is the GLS-R^2, based on the OLS-type estimates of the risk premia, $\hat{\Omega}$ is the sample variance-covariance matrix of returns, $R_{gls,2}^2$ is the GLS-R^2, based on the GLS-type estimates of the risk premia $\alpha_i = \bar{R}_i^e - \hat{\lambda}_{ols}\hat{\beta}_i$ is the average time series pricing error for portfolio i, HJ is the Hansen-Jagannathan distance, $^+$ stands for the pseudo-inverse of a matrix, $y = I - \hat{\beta}(\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}$, T^2 is the cross-sectional test of Shanken (1985), Σ_f is the variance-covariance matrix of the factors, and $\hat{\lambda}_f$ is a $k \times 1$ vector of the factors risk premia (excluding the common intercept).

For the Pen approach, I use the penalty defined through Fischer transformation of the partial correlations of the factors and returns (since they are invariant to the data scale). I set the level tuning parameter, η to $\bar{\sigma}$, the average standard deviation of the residials from the first stage, and the curvature parameter, d, to 4. In Section V.C, I investigate the impact of tuning parameters on the estimator performance, and show that changing values of the tuning parameters has only little effect on the estimator's ability to eliminate or retain strong/weak factors.

A. Correctly specified model

Table I demonstrates the performance of the three estimation techniques in terms of their point estimates: the Fama-MacBeth two-pass procedure without the useless factor (denoted as the *oracle* estimator), the Fama-MacBeth estimator, which includes both useful and useless factors in the model and the Pen-FM estimator. All three us an identity weight matrix at the second stage. For each of the estimators the table reports the mean point estimate of the risk premia and the intercept, their bias and mean squared error. I also report in the last column the average factor shrinkage rates for the Pen-FM estimator, produced using 10,000 simulations (i.e. how often the corresponding risk premia estimate is set exactly to 0).

The results are striking. The useless factor is correctly identified in the model with the correponding risk premia shrunk to 0 with 100% accuracy even for such a small sample size as 30 observations. At the same time, the useful factor (market excess return) is correctly preserved in the specification, with the shrinkage rate below 1% for all the sample sizes. Starting from $T = 50$, the finite sample bias of the parameter estimates produced by the Pen-FM estimator is much closer to that of the oracle Fama-MacBeth cross-sectional regression, which exludes the useless factor ex ante. For example, when $T = 50$, the average finite sample bias of the useful factor risk premium, produced by the oracle Fama-MacBeth estimator is 0.093 %, 0.114 % for the two-step procedure which includes the useless factor, and 0.091% for the estimates produced by Pen-FM.

The mean squared errors of the estimates demonstrate a similar pattern: for $T \geq 50$ the MSE for Pen-FM is virtually identical to that of the Fama-MacBeth without the useless factor in the model. At the same time, the mean squared error for the standard Fama-MacBeth estimator stays at the same level of about 0.32% regardless of sample size, illustrating the fact that the risk premia estimate of the useless factor is inconsistent, converging to a bounded random variable, centred at 0.

[TABLE I ABOUT HERE]

The size of the confidence intervals constructed by bootstrap is slighly conservative (see Table A1). However, it is not a feature particular to the Pen-FM estimator. Even without the presence of useless factors in the model, bootstrapping risk premia parameters seems to produce similar slightly conservative confidence bounds, as illustrated in Table A1, Panel A.

Fig. 1-5 also illustrate the ability of Pen-FM estimator to restore the original quality of fit for the model. Fig. 1 shows the distribution of the cross-sectional R^2 for the various sample size. The measures of fit, produced by the model in the absence of the useless factor and with it, when estimated by Pen-FM, are virtually identical. At the same time, $\bar{R}^2$, produced by the conventional Fama-MacBeth approach seems to be inflated by the presence of a useless factor, consistent with the theoretical findings in Kleibergen and Zhan (2013). The distribution of the in-sample measure of fit

seems to be quite wide (e.g. for T=100 it fluctuates a good deal from 0 to 80%), again highlighting the inaccuracy of a single point estimate and a need to construct confidence bounds for the measures of fit (e.g. as suggested in Lewellen et al. (2010). Even if we estimate the true model specification, empirically the data contains quite a lot of noise (which was also captured in the simulation design, calibrating data generating parameters to their sample analogues). Thus it is not surprising to find that the probability of getting a rather low value of the R^2 is still high for a moderate sample size. Only when the number of observations is high (e.g. T=1000), does the peak of the probability density function seem to approach 80%; however, even then the domain remains quite wide.

The GLS R^2, based on either OLS or GLS second stage estimates (Fig. 2 and 3), seem to have a much tighter spread (in particular, if one relies on the OLS second stage). As the sample size increases, the measures of fit seem to better indicate the pricing ability of the true factor. The GLS R^2 is less affected by the problem of the useless factor (as demonstrated in Kleibergen and Zhan (2013)), but there is still a difference between the estimates, and if the model is not identified, R^2 seems to be slightly higher, as in the OLS case. This effect, however, is much less pronounced. Once again, the distribution of GLS R^2 for Pen-FM is virtually identical to that of the conventional Fama-MacBeth estimator without the useless factor in the model. A similar spurious increase in the quality of fit may be noted, considering the distribution of the average pricing errors (Fig. 5), which is shifted to the left in the presence of a useless factor. The Hansen-Jagannathan distance is also affected by the presence of the useless factor (as demonstrated in Gospodinov et al. (2014a)); however, not as much (Fig. 4). In contrast to the standard Fama-McBeth estimator, even for a very small sample size the average pricing error and the Hansen-Jagannathan distance produced by Pen-FM are virtually identical to those of the model that does not include the spurious factor ex ante.

Figs. A1 and A2 demonstrate the impact of the useless factors on the distribution of the T^2 and q statistics respectively. I compute their values, based on the risk premia estimates produced by Fama-MacBeth approach with or without the useless factor, but not Pen-FM, since that would require an assumption on the dimension of the model, and the shrinkage-based estimation is generally silent about testing the size of the model (as opposed to identifying its parameters). The distribution of q is extremely wide and when the model is contaminated by the useless factors is naturally inflated. The impact on the distribution of T^2 is naturally a combination of the impact coming from the Shanken correction term (which is affected by the identification failure through the risk premia estimates), and q quadratics. As a result, the distribution is much closer to that of the oracle estimator; however, it is still characterised by an appreciably heavy right tail, and is generally slightly inflated.

B. Misspecified model

The second simulation design that I consider corresponds to the case of a misspecified model, where the cause of misspecification is the omitted variable bias. The data is generated from a 3-factor model, based on 3 canonical Fama-French factors (with data generating parameters obtained from the in-sample model estimation similar to the previous case). However, in the simulations I consider estimating a one factor model (thus, the source of misspecification is omitting the SMB and HML factors). Again, I compare the performance of 3 estimators: (a) Fama-MacBeth, using the simulated market return as the only factor, (b) Fama-MacBeth, using the simulated market return and the irrelevant factor, (c) Pen-FM estimator, using the simulated market return and the irrelevant factor.

Table II describes the pointwise distribution of the oracle estimator (Fama-MacBeth with an identity weight matrix, applied using only the market excess return as a risk factor), Fama-MacBeth and Pen-FM estimators, when the model includes both true and useless factors.

The results are similar to the case of the correctly specified model. Pen-FM successfully identifies both strong and useless factors with very high accuracy (the useless one is always eliminated from the model by shrinking its premium to 0 even when $T = 30$). The mean squared error and omitted variable bias for all the parameters are close to those of the oracle estimator. At the same time, column 9 demonstrates that the risk premium for the spurious factor, produced by conventional Fama-MacBeth procedure diverges as the sample size increases (its mean squared error increases from 0.445 for T=50 to 1.979 for T=1000). However, the risk premia estimates remain within a reasonable range of parameters, so even if the Fama-MacBeth estimates diverge, it may be difficult to detect it in practice.

Confidence intervals based on t-statistics for the Fama-MacBeth estimator overreject the null hypohesis of no impact of the useless factors (see Tables A4 and A6), and should a researcher rely on them, she would be likely to identify a useless factor as priced in the cross-section of stock returns.

[TABLE II ABOUT HERE]

Fig. 6-10 present the quality of fit measures in the misspecified model contaminated by the presence of a useless factor and the ability of Pen-FM to restore them. Fig. 6 shows the distribution of the cross-sectional R^2 for various sample sizes. The similarity between the measures of fit, produced by the model in the absence of the useless factor and with it, but estimated by Pen-FM, is striking: even for such a small sample size as 50 time series observations, the distributions of the R^2 produced by the Fama-MacBeth estimates in the absence of a useless factor, and Pen-FM in a nonidentified model, are virtually identical. This is expected, since, as indicated in Table II, once the useless factor is eliminated from the model, the parameter estimates produced by Pen-FM are nearly identical to those of the one-factor version of Fama-MacBeth. As the sample size increases, the true sample

26

distribution of R^2 becomes much tighter, and peaks around 10-15%, illustrating the model's failure to capture all the variation in the asset returns, while omitting two out of three risk factors.

The cross-sectional R^2 produced by the conventional Fama-MacBeth method is severely inflated by the presence of a useless factor, and its distribution is so wide that it looks almost uniform on $[0, 1]$. This illustration is consistent with the theoretical findings of Kleibergen and Zhan (2013) and Gospodinov et al. (2014b), who demonstrate that under misspecification, the cross-sectional R^2 seems to be particularly affected by the identification failure.

Fig. 7 describes the distribution of GLS R^2, when the second stage estimates are produced using the identity weight matrix. Interestingly, when the model is no longer identified, GLS R^2 tends to be lower than its true in-sample value, produced by Pen-FM or the Fama-MacBeth estimator without the impact of the useless factor. This implies that if a researcher were to rely on this measure of fit, she would be likely to underestimate the pricing ability of the model. Fig. 8 presents similar graphs for the distribution of the GLS R^2, when the risk premia parameters are estimated by GLS in the second stage. The difference between various methods of estimation is much less pronounced, although Fama-MacBeth tends to somewhat overestimate the quality of fit produced by the model.

The average pricing errors displayed in Fig. 10 also indicate a substantial impact of the useless factor in the model. When such a factor is included, and risk premia parameters are estimated using the conventional Fama-MacBeth approach, the APE seem to be smaller than they actually are, resulting in s spurious improvement in the model's ability to explain the difference in asset returns. Again, this is nearly perfectly restored once the model is estimated by Pen-FM.

The Hansen-Jagannathan distance (Fig. 9) is often used to assess model misspecification, since the greater is the distance between the set of SDFs that price a given set of portfolios and the one suggested by a particular specification, the higher is the degree of mispricing. When a useless factor is included, HJ in the Fama-MacBeth estimation has a much wider support than it normally does; and, on average, it tends to be higher.

Fig. A1 and A2 demonstrate the impact of the useless factors on the distribution of T^2 and q statistics in a misspecified model. Again, I compute their values on the basis of the risk premia estimates produced by the Fama-MacBeth approach with or without the useless factor, but not Pen-FM, since computing these statistics requires using the matrices with the dimension, depending on the number of factors in the model (and not just their risk premia values). When the model contains a spurious factor, the distribution of q becomes extremely wide and skewed to the right. The effect of spurious factors on the distribution of T^2 is naturally a combination of the influence coming from the Shanken correction term (which is affected by the identification failure through the risk premia estimates), and q. T^2 is generally biased towards 0, making it harder to detect the model misspecification in the presence of a useless factor.

C.   Robustness check

In order to assess the numerical stability and finite sample properties of the Pen-FM estimator, I study how the survival rates of useful and useless factors depend on the tuning parameters within the same simulation design of either the correct or the misspecified model descibed in the earlier sections.

Table III summarises the survival rates for the useful and useless factors as a function of the tuning parameter d, which defines the curvature of the penalty. In Proposition 1 I proved the Pen-FM estimator to be consistent and asymptotically normal for all values of $d > 2$. In this simulation I fix the other tuning parameter value, $\eta = \bar{\sigma}$, and vary the value of d from 3 to 10. Each simulation design is once again repeated 10,000 times, and the average shrinkage rates of the factors are reported. Intuitively, the higher the curvature parameter, the harsher is the estimated difference between a strong and a weak factor, and hence, one would also expect a slighly more pronounced difference between their shrinkage rates.

It can be clearly seen that the behaviour of the estimates is nearly identical for different values of the curvature parameter and within 1% difference from each other. The only case that stands out, is when the sample is very small (30-50 observations) and $d = 3$. In this case the useful factor has been mistakenly identified as the spurious one in 1-2.5% of the simulations, but these types of fluctuations are fully expected when dealing with such a small sample with a relatively low signal-to-noise ratio. A similar pattern characterises the shrinkage rates for the useless factors, which are extremely close to 1.

[TABLES III, IV ABOUT HERE]

Table IV shows how the shrinkage rates of Pen-FM depend on the value of the other tuning parameter, η, which is responsible for the overall weight on the penalty compared with the standard component of the loss function (see eq. 9) and could be thought of as the level parameter. Once again, I conduct 10,000 simulations of the correctly or incorrectly specified model for the various sample size, and compute the shrinkage rates for both useful and useless factors. I fix the curvature tuning parameter, d, at $d = 4$, and vary η.

I consider the following range of parameters:

1. $\eta = \bar{R}^e$, the average excess return on the portfolio;

2. $\eta = ln(\bar{\sigma}^2)$, log of the average volatility of the residuals from the first stage;

3. $\eta = \bar{\sigma}$, the average standard deviation of the first stage residuals;

4. the value of η is chosen by fivefold cross-validation;

5. the value of η is chosen by leave-one-out cross-validation.

I have chosen the values of the tuning parameter η that either capture the scale of the data (for example, whether excess returns are displayed in percentages or not), or are suggested by some of the data-driven techniques[17]. Cross-validation (CV) is intuitively appealing, because it is a data-driven method and it naturally allows one to assess the out-of sample performance of the model, treating every observation as part of the validation set only once. CV-based methods have been extensively used in many different applications, and have proved to be extremely useful[18]. Here I briefly describe the so-called k-fold cross-validation.

The original sample is divided into k equal size subsamples, followed by the following algorithm.

- Pick a subsample and call it a validation set; all the other subsamples form a training set.

- Pick a point on the grid for the tuning parameters. For the chosen values of the tuning parameters estimate the model on the training set and assess its performnce on the validation set by the corresponding loss function $(L_T(\hat{\lambda}))$.

- Repeat the procedure for all the other subsamples.

- Compute the average of the loss function (CV criterion).

- Repeat the calculations for all the other values of the tuning parameters. Since the location of the minimum CV value is a random variable, it is often suggested that the one to pick the one that gives the largest CV criterion within 1 standard deviation of its absolute minimum on the grid, to ensure the robustness of the result (Friedman, Hastie, and Tibshirani (2010)).

Table IV summarises the shrinkage rates of the useful and useless factors for different values of the level tuning parameter, η. Similar to the findings in Table III, the tuning parameter impact is virtually negligible. The useless factor is successfully identified and eliminated from the model in nearly 100% of the simulations, even for a very small sample size, regardless of whether the model is correctly or incorrectly specified, while the strong factor is successfully retained with an equally high probability. The only setting where it causes some discrepancy (within 2-3% confidence bounds) is the case of a misspecified model and a very small sample size ($T = 30$ or 50); but it is again entirely expected for the samples of such size, and therefore does not raise any concerns.

---

[17]Although the table presents the results for the tuning parameteres selected by cross-validation, I have also considered such alternative procedures as BIC, Generalised BIC and the *pass* selection stability criterion. The outcomes are similar both quantitively and qualitatively, and are available upon request.

[18]For an excellent overview see, e.g. Hastie, Tibshirani, and Friedman (2011)

D.   Comparing Pen-FM with alternatives

In this section I compare the finite sample performance of the sequential elimination procedure proposed in Gospodinov et al. (2014a) and that of Pen-FM with regard to identifying the strong and useless factors.

I replicate the simulation designs used in Table 4 of Gospodinov et al. (2014a)[19], to reflect various combinations of the risk drivers in a potential four-factor model: strong factors that are either priced in the cross-section of asset returns or not, and irrelevant factors. For each of the variables I compute the frequency with which it is identified as a strong risk factor in the cross-section of asset returns and consequently retained in the model. Each simulation design is repeated 10,000 times.

Panel A in Table V summarises the factor survival rates for a correctly specified model. The top panel focuses on the case of 2 priced strong factors, 1 strong factor that is correlated with returns, but not priced, and 1 purely irrelevant factor, which does not correlate with asset returns[20]. For each of the variables I present its survival rate, based on the misspecification-robust t_m- statistic of Gospodinov et al. (2014a)[21] for a linear SDF model, the frequency with which the corresponding risk premium estimate was *not* set exactly to 0 by the Pen-FM estimator and one minus the average shrinkage rate from the 10,000 bootstrap replica. The latter also provides an additional comparison of the performance of the pointwise estimator with its bootstrap analogue. A good procedure should be able to recognise the presence of a strong factor and leave it in the model with probability close to 1. At the same time, faced with the useless factor, one needs to recognise it and eliminate from the model, forcing the survival rate to be close to 0.

Consider the case of a correctly specified model, with 2 useful factors that are priced in the cross-section of asset returns, 1 useful, but unpriced factor (with a risk premium equal to zero), and a useless factor, presented in the top panel of Table V. The useless factor is correctly identified and effectively eliminated from the model by both the misspecification-robust $t-$test and the Pen-FM estimator even for a very small sample size (e.g. for a time series of 100 observations, the useless factor is retained in the model in no more than 1% of the simulations. For the smallest sample size of 50 observations, Pen-FM seems also to outperform the sequential elimination procedure, since it retained the useless factor in less than 1.5% of the models only, while the latter was keeping it as part of the specification in roughly 15% of the simulations.

The t_m-test is designed to eliminate not only the useless factors from the linear model, but also those factors that are strongly correlated with asset returns, but not priced in the cross-section of returns. As a result, in 95-99% of cases the useful factor with $\lambda = 0$ is also eliminated from the model.

---

[19] I am very grateful to Cesare Robotti for sharing the corresponding routines.

[20] The setting proxies the estimation of a 4-factor model on the set of portfolios similar to 25 size and book-to-market and 17 industry portfolios. For a full description of the simulation design, please refer to Gospodinov et al. (2014a)

[21] The t_c-statistic for a correctly specified model performs very similar to t_m in terms of the factor survival rates. Since it is not known *ex ante*, whether the model is correctly specified or not, I focus on the outcome of the t_m test.

However, the Pen-FM estimator eliminates only the impact of useless factors, and thus retains its presence in the model in 92-98% of the simulations, depending on the sample size.

Small sample simulations also seem to highlight an unfortunate propensity of the t_m-test to eliminate even strong factors from the model, when the span of the data is not sufficiently large. For example, when the sample size is only about 200 observations, the strong factor is mistakenly identified as a useless one in 40-50% of the simulations. When $T = 50$, the survival rates for the strong factors are accordingly only 6 and 11%. The inference is restored once the sample size is increased to about 600 observations (which would correspond empirically to roughly 50 years of monthly observations). However, the small sample performance of the t_m-test raises some caution with regard to its application, and interpretation of the results for models that rely on quarterly or yearly data, where the sample size is rather small. At the same time, the Pen-FM estimator seems to be quite promising in this regard, because it retains strong factors in the model with a very high probability (the first strong factor is retained in 99.9% of the cases for all the sample sizes, while the second one is retained in 92-98% of the simulations). It also worth highlghting that the pointwise and bootstrap shrinkage rates of Pen-FM are very close to each other, with the difference within 2%, supporting the notion that bootstrap replicas approximate the pointwise distribution of the estimates rather well, even with a very small sample size.

The second panel presents similar findings for a correctly specified model with 2 useful (and priced) and two useless factors. The results are quite similar - both approaches are able to identify the presence of irrelevant factors starting from a very small sample size (again, for $T = 50$, Pen-FM seems to have a little advantage). At the same time, for a small to moderate sample, the t_m-statistic tends to eliminate strong factors, as well as the weak ones, while Pen-FM remains consistent in keeping those variables in the model.

Panel B in Table V presents the case of a misspecified model. The results are quite similar, for both the t_m-test and the Pen-FM estimator correctly identify the presence of a useless factor and either shrink it to 0 or eliminate from the model. As before, however, in the small sample the t_m-statistic tends to mistakenly identify a strong factor as a weak one with a rather high probability; however, correct factor classification is restored at a sample size of at least 600 observations. Again, Pen-FM estimator seems to outperform the sequential elimination procedure in a finite sample. The only difference arises for $T = 50$, when the Pen-FM retains the second strong factor in only 77-78% of the simulations compared with the usual 92-95% observed for this sample size in other simulations designs. However, this is still substantially better than the 8% survival rate produced by the sequential elimination procedure; for $T = 100$ the strong factor is retained already in 91-92%

of the simulations.

These simulation results suggest that the t_m-test is rather conservative with regard to the factor elimination, and in a small sample could result in mistakenly excluding a true risk driver from the model. Pen-FM seems to be better at deciphering the strength of a factor even for such a small sample size as 50 time series observations, therefore, making it more reliable when working with quarterly or yearly data, where the sample size is naturally small.

Table VI summarises the factor survival rates produced by the adaptive lasso in the same simulation design of Gospodinov et al. (2014a). As discussed in Section III, when the model is no longer identified, the adaptive lasso is not expected to correctly identify the factors that are priced in the cross-section of asset returns.

$$\hat{\lambda}_{AdL} = \underset{\lambda \in \Theta}{\arg\min} \left[\bar{R}^e - \hat{\beta}\lambda \right]' W_T \left[\bar{R}^e - \hat{\beta}\lambda \right] + \eta_T \sum_{j=1}^{k} \frac{1}{|\hat{\lambda}_{j,ols}|^d} |\lambda_j|,$$

When the model includes useless factors, prior OLS-based estimates of the risk premia that define the individual weights in the penalty no longer have the desired properties, since weak identification contaminates their estimation. As a result, adaptive lasso produces erratic behaviour for the second stage estimates, potentially shrinking true risk drivers and/or retaining the useless ones. Particular shrinkage rates will depend on the strength of the factor, its relation to the other variables, and the prior estimates of the risk premia.

Table VI summarises the average factor survival rates produced by the Pen-FM estimator with $d = 4$ and $\eta = \bar{\sigma}$ (the baseline scenario) with those of the adaptive lasso, when the tuning parameter is chosen via the BIC[22].

[TABLE VI ABOUT HERE]

For a correctly specified model (Panel A), the adaptive lasso nearly always retains the second useful factor, but not the first, which is often eliminated from the model for a relatively moderate sample size (e.g. when $T = 250$, it is retained in only 62.6% of the simulations). Furthermore, unlike the Pen-FM, the adaptive lasso estmator is not able to recognise the presence of a useless factor, and it is never eliminated.

If the model is misspecified, the impact of the identification failure on the original penalty weights is particularly severe, which results in worse factor survival rates for the adaptive lasso. The first of the useful factors is eliminated from the model with a high probability (e.g. for $T = 250$, it is retained only in 45.66% and 34.31%of the simulations, respectively, depending on whether the simulation design includes 1 or 2 useless factors). The second useless factor is always retained in the

---

[22]I am grateful to Dennis D. Boos for sharing his R routine, which is available at his webpage, http://www4.stat.ncsu.edu/ boos/var.select/lasso.adaptive.html

model, and the first one increasingly so (e.g. for a sample of 50 observations it is a part of the model in 56.54% of the simulations, while for $T = 1000$ already in 96.18%). This finding is expected, since as the sample size increases, the risk premia for the useless factors in the misspecified models tend to grow larger (along with their t-statistic) and the adaptive lasso penalty becomes automatically smaller, suggesting that it would be useful to preserve such factors in the model. The simulations confirm the different nature of the estimators and a quite drastic difference in the estimation of risk premia parameters in the presence of useless factors.

VI. Empirical applications

A. Data description

I apply the Pen-FM estimator to a large set of models that have been proposed in the empirical literature, and study how using different estimation techniques may alter parameter estimates and the assessment of model model pricing ability[23]. I focus on the following list of models/factors for the cross-section of stock returns.

CAPM. The model is estimated using monthly excess returns on a cross-section of 25 Fama-French portfolios, sorted by size and book-to-market ratio. I use 1-month Treasury rate as a proxy for the risk-free rate of return. The market portfolio is the value-weighted return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ. Data is taken from Kenneth French website. To be consistent with other applications, relying on tradable factors, I consider the period of January 1972 - December, 2013[24].

Fama-French 3 factor model. The model is estimated using monthly excess returns on a cross-section of 25 Fama-French portfolios, sorted by size and book-to-market ratio. I use 1-month Treasury rate as a proxy for the risk-free rate of return. Following Fama and French (1992), I use market excess return, SMB and HML as the risk factors. SMB is a zero-investment portfolio formed by a long position on the stocks with small capitalisation (cap), and a short position on big cap stocks. HML is constructed in a similar way, going long on high book-to-market (B/M) stocks and short on low B/M stocks.

Carhart 4 factor model. I consider two cross-sections of asset returns to test the Carhart (1997) model: 25 Fama-French portfolios, sorted by size and book-to-market, and 25 Fama-French portfolios, sorted by value and momentum. In addition to the 3 Fama-French factors, the model includes the

---

[23]I have applied the new estimator to a wide set of models; however, for reasons of brevity, in this paper I focus on a particular subset. Additional empirical results are available upon request.

[24]I have also estimated the models, using other time samples, e.g. the largest currently available, 1947-2013, 1961-2013, or the samples used at the time of the papers publication. There was no qualitative difference between the relative performance of Pen-FM and the Fama-MacBeth estimator (i.e. if the factor has been identified as a strong/weak one, it continues to be so when a different time span is used to estimate the model). Additional empirical results are available upon request.

momentum factor (UMD), a zero-cost portfolio constructed by going long the previous 12-month return winners and short the previous 12-month loser stocks.

"Quality-minus-junk". A quality-minus-junk factor (QMJ), suggested in Asness et al. (2014), is constructed by forming a long/short portfolio of stocks sorted by their quality (which is measured by profitability, growth, safety and payout). I use the set of excess returns on Fama-French 25 portfolios, sorted by size and book-to-market as the test assets, and consider a 4 factor model, which includes market excess return, SMB, HML and QMJ.

q-factor model. I consider the so-called q-factor model, various specifications of which have been suggested in the prior literature linking stock performance to investment-related factors (e.g. Liu, Whited, and Zhang (2009), Hou et al. (2014), Li and Zhang (2010)). I consider the 4 factor specification adopted in Hou et al. (2014), and that includes market excess return, the size factor (ME), reflecting the difference between the portfolios of large and small stocks, the investment factor (I/A), reflecting the difference in returns on stocks with high/low investment-to-assets ratio, and the profitability factor, built in a similar way from sorting stocks on their return-on-equity (ROE)[25]. I apply the model to several collections of test assets: excess returns on 25 Fama-French portfolios sorted by size and book-to-market, 25 Fama-French portfolios sorted by value and momentum, 10 portfolios sorted on momentum, and 25 portfolios sorted on price/earnings ratio.

cay-CAPM. This is the version of scaled CAPM suggested by Lettau and Ludvigson (2001b); it uses the long-run consumption-wealth cointegration relationship in addition to the market factor and their interaction term. I replicate their results for exactly the same time sample and a cross-section of the portfolios that were used in the original paper. The data is quarterly, 1963Q3-1998Q3.

cay-CCAPM. Similar to cay-CAPM, the model relies on nondurable consumption growth, *cay*, and their interaction term.

Human Capital CAPM. Jagannathan and Wang (1996) suggested using return on human capital (proxied by after-tax-labour income), as an additional factor for the cross-section of stock returns. I estimate the model on the same dataset, as in Lettau and Ludvigson (2001b).

Durable consumption model. Yogo (2006) suggested a model of the representative agent, deriving utility from the flow of nondurable goods, and the stock of durables. In the linearised version, the model includes three factors: market excess returns and nondurable/durable consumption growth. I estimate the model using several cross-sections of the asset returns that were used in the original paper: 25 portfolios sorted by size and book-to-market, 24 portfolios sorted by book-to-market within industry, and 24 portfolios sorted by market and HML betas. The data is quarterly, 1951Q3-2001Q4.

---

[25]I am very grateful to Lu Zhang and Chen Xue for sharing the factors data

B. *Tradable factors and the cross-section of stock returns*

Panel A in Table VII below summarises the estimation of the linear factor models that rely on tradable factors. For each of the specifications, I provide the p-value of the Wald test[26] for the corresponding factor betas to be jointly equal to 0. I also apply the sequential elimination procedure of Gospodinov et al. (2014a), based on the t_m test statistic[27] and indicate whether a particular factor survives it. I then proceed to estimate the models using the standard Fama-MacBeth approach and Pen-FM, using the identity weight matrix. For the estimates produced by the Fama-MacBeth cross-sectional regression, I provide standard errors and p-values, based on t-statistics with and without Shanken correction, and the p-values based on 10,000 replicas of the stationary bootstrap of Politis and Romano (1994), and cross-sectional R^2 of the model fit. For the Pen-FM estimator, I provide the point estimates of risk premia, their average bootstrap shrinkage rates, bootstrap-based p-values and cross-sectional R^2. To be consistent, when discussing the statistical significance of the parameters, I refer to bootstrap-based p-values for both estimators. Greyshading indicates the factors that are identified as weak (or irrelevant) and eliminated from the model by Pen-FM.

[TABLE VII ABOUT HERE]

There is no difference whether CAPM parameters are estimated by the Fama-MacBeth or the Pen-FM estimator. Both methods deliver identical risk premia (-0.558% per month for market excess return), bootstrap-based p-values and R^2 (13%). A similar result is obtained when I estimate the Fama-French 3 factor model, where both methods deliver identical pricing performance. Market premium is significant at 10%, but negative. This is consistent with other empirical estimates of the market risk premium (e.g. Lettau and Ludvigson (2001b) also report a negative, but insignificant market premium for the cross-section of quarterly returns). HML, however, is significant and seems to be a strong factor. Overall, the model captures a large share of the cross-sectional variation, as indicated by the in-sample value of R^2 at 71%. The common intercept, however, is still quite large, at about 1.3%. There is no significant shrinkage for any of the factors in bootstrap, either, and the parameter estimates are nearly identical.

Including the quality-minus-junk factor improves the fit of the model, as R^2 increases from 71 to 83-84%. The QMJ factor risk premium is set exactly to 0 in 8.4% of bootstrap replicas; however, its impact remains significant at 10%, providing further evidence that including this factor improves the pricing ability of the model. In the Fama-MacBeth estimation, the common intercept was weakly significant at 10%, however, in the case of Pen-FM, it is no longer significant, decreasing from 0.7 to 0.57% (which is partly due to a slightly larger risk premium for HML).

---

[26]I use heteroscedasticity and autocorrelation-robust standard errors, based on the lag truncation rule in Andrews (1991)

[27]Since it is not known ex ante, whether the model is correctly specified or not, I use the misspecification-robust test

The Carhart (1997) 4-factor model is estimated on two cross-sections of porfolios, highlighting a rather interesting, but at the same time expected, finding, that the sorting mechanism used in portfolio construction affects the pricing ability of the factors. When I estimate the 4-factor model on the cross-section of 25 portfolios, sorted by size and book-to-market ratio, momentum factor is identified by the Pen-FM estimator as the irrelevant one, since the correponsing risk premia is shrunk exactly to 0 in 99.6% of the bootstrap replicas. As a result of this elimination, cross-sectional R^2 in the model estimated by Pen-FM is the same as for the 3-factor Fama-French model, 71%.

On the other hand, when portfolios are sorted on value and momentum, HML is indicated as the irrelevant one, while momentum clearly drives most of the cross-sectional variation. Both models exhibit the same R^2, 90%. Interestingly, once HML is eliminated by Pen-FM from the model, the risk premium on SMB becomes weakly significant at 10%, recovering the true impact of the size factor. This illustration of different pricing ability of the risk factors, when facing different cross-sections of asset returns, is not new, but it is interesting to note that the impact can be so strong as to affect the model identification.

Hou et al. (2014) suggest a 4 factor model that, the authors claim, manages to explain most of the puzzles in empirical finance literature, with the main contribution coming from investment and profitability factors. Their specification outperforms Fama-French and Carhart models with regards to many anomalies, including operating accrual, R&D-to-market and momentum. Therefore, it seems to be particularly interesting to assess model performance on various test assets. For 25 Fama-French portfolios, the profitability factor impact is not strongly identified, as it is eliminated from the model in 82.2% of the bootstrap replica. At the same time, investment remains a significant determinant of the cross-sectional variation, commanding a premium of 0.36%. A different outcome is observed when using the cross-section of stocks sorted by value and momentum. In this case the profitability factor is removed from the model as the weak one. Size and ROE factors are identified as strong determinants of the cross-sectional variation of returns, with risk premia estimates of 0.484% and 0.63% accordingly. It is interesting to note that, although the I/A factor is eliminated from the model, the cross-sectional R^2 remains at the same high level of 88%.

A particular strength of the profitability factor becomes apparent when evaluating its performance on the cross-section of stocks sorted on momentum. When the conventional Fama-MacBeth estimator is applied to the data, none of the factors command a significant risk premium, although the model explains 93% of the cross-sectional dispersion in portfolio excess returns. Looking at the estimates produced by Pen-FM, one can easily account for this finding: it seems that size and investment factors are only weakly related to momentum-sorted portfolio returns, while it is the profitability factor that drives nearly all of their variation. The model delivers a positive (but highly insignificant) market risk premium, and a large and positive risk premium for ROE (0.742%). Although both M/E and I/A are eliminated from the model, the cross-sectional R^2 is at an impressive level of 90%. This

may be due to an identification failure, caused by the presence of useless (or weak) factors, which was masking the impact of the true risk drivers.

When stocks are sorted in portfolios based on their price/earnings ratio, the Fama-MacBeth estimator results in high cross-sectional R^2 (81%), but insignificant risk premia for all the four factors, and a rather large average mispricing at 2.71%. In contrast, the Pen-FM estimator shrinks the impact of the size and profitability factors (which are elimininated in 96.8% and 84.5% of the bootstrap replicas, respectively). As a result, investment becomes weakly significant, commanding a premium of 0.44%, the market premium is also positive (but insignificant) at 0.27%, while the common intercept, which is often viewed as the sign of model misspecification, is only 0.25% (and insignificant). The model again highlights the ability of the Pen-FM estimator to identify and eliminate weak factors from the cross-section of returns, while maintaining the impact of the strong ones. In particular, investment and market factors alone explain 76% of the cross-sectional variation in portfolios, sorted by the P/E ratio.

C. Nontradable factors and the cross-section of stock returns

Standard consumption-based asset pricing models feature a representative agent who trades in financial securities in order to optimize her consumption flow (e.g. Lucas (1976), Breeden (1979)). In this framework the only source of risk is related to the fluctuations in consumption, and hence, all the assets are priced in accordance with their ability to hedge against it. In the simplest version of the CCAPM, the risk premium associated with a particular security is proportional to its covariance with the consumption growth:

$$E[R_i^e] \approx \lambda \ cov\left(R_{t,i}^e, \Delta c\right)$$

If the agent has the CRRA utility function, λ is directly related to the relative risk aversion, γ, and hence, one of the natural tests of the model consists in estimating this parameter and comparing it with the plausible values for the risk aversion (i.e. < 10). Mehra and Prescott (1985) and Weil (1989) show that in order to match historical data, one would need to have a coefficient of risk aversion much larger than any plausible empirically supported value, thus leading to the so-called equity premium and risk-free rate puzzles. The model was strongly rejected on US data (Hansen and Singleton (1982), Hansen and Singleton (1983), Mankiw and Shapiro (1986)), but led to a tremendous growth in the consumption-based asset pricing literature, which largely developed in two main directions: modifying the model framework in terms of preferences, production sector and various frictions related to decision-making, or highlighting the impact of the data used to validate the model[28].

---

[28]The literature on consumption-based asset pricing is vast; for an overview see Campbell (2003) and Ludvigson (2013)

Not only the estimates of the risk aversion parameter turn out to be unrealistically large, but they are also characterised by extremely wide confidence bounds (e.g. Yogo (2006) reports $\hat{\gamma} = 142$ with the standard errors of 25 when estimating the CCAPM using the Fama-French 25 portfolios). The impact of low covariance between consumption and asset returns could not merely explain a high estimate of the risk aversion, but also lead to the models being weakly identified, implying a potential loss of consistency, nonstandard asymptotic distribution for the conventional OLS or GMM estimators, and the need to rely on identification-robust inference procedures.

Panel B in Table VII reports estimation of some widely used empirical models, relying on non-tradable factors, such as consumption. The scaled version of CAPM, motivated by the long-run relationship between conumption and wealth dynamics in Lettau and Ludvigson (2001a), seems to be rather weakly identified, as both *cay* and its product with the market return are eliminated from the model by the Pen-FM estimator in 97.6% and 85.1% of the bootstrap replicas, respectively. The resulting specification includes only the market excess return as the only factor for the cross-section of quarterly stock returns, which leads to the well-known illustration of the inability of the classical CAPM to explain any cross-sectional variation, delivering the R^2 of only 1%. The scaled version of Consumption-CAPM also seems to be contaminated by identification failure. Not only the estimates of the risk preia of all three factors are shrunk to 0 with a very high frequency, but even the Wald test for the vector of betas indicates nondurable consumption growth as a rather weak risk factor.

This finding provides a new aspect to the well-known failure of the CCAPM and similar specifications to both match the equity premium and explain the cross-sectional variation in returns.

One of the natural solutions to the problem could lie in using alternative measures for consumption and investment horizons. Kroencke (2014) explicitly models the filtering process used to construct NIPA time series, and finds that the unfiltered flow consumption produces a much better fit of the basic consumption-based asset pricing model and substantially lowers the required level of risk aversion. Daniel and Marshall (1997) show that while the contemporaneous correlation of consumption growth and returns is quite low for the quarterly data, it is substantially increased at lower frequency. This finding would be consistent with investors' rebalancing their portfolios over longer periods of time, either due to transaction costs (market frictions or the costs of information processing), or due to external constraints (e.g. some of the calendar effects). Lynch (1996) further studies the effect of decision frequency and its synchronisation between agents, demonstrating that it could naturally result in a lower contemporaneous correlation between consumption risk and returns. Jagannathan and Wang (2007) state that investors are more likely to make decisions at the end of the year, and, hence, consumption growth, if evaluated then, would be a more likely determinant of the asset returns. These papers could also be viewed as a means to improve model identification.

Jagannathan and Wang (1996) and Santos and Veronesi (2006) argue that human capital (HC) should be an important risk driver for financial securities. I estimate their HC-CAPM on the dataset

used in Lettau and Ludvigson (2001b), and find that this model is also contaminated by the identification problem. While the true risk factor may command a significant premium, the model is still poorly identified, as indicated by Table VII, and after-tax labour income, as a proxy for human capital, is eliminated by Pen-FM from the model for stock returns. The scaled version of the HC-CAPM also seems to be weakly identified, since the only robust risk factor seems to be market excess return.

Unlike the baseline models that mainly focus on nondurable consumption goods and services, Yogo (2006) argues that the stock of durables is an important driver of financial returns, and taking it into account substantially improves the ability of the model to match not only the level of aggregate variables (e.g. the equity premium, or the risk-free rate), but also the cross-sectional spread in portfolios, sorted on various characteristics. Table VII illustrates the estimation of durable consumption CAPM, that includes market returns, as well as durable and nondurable consumption growth as factors on several cross-sections of portfolios. Both consumption-related factors seem to be rather weak drivers for the cross-section of stocks, and are eliminated in roughly 99% of the bootstrap replicas. This finding is also robust across the different sets of portfolios. Once the weak factors are eliminated from the model, only the market excess return remains; however, its price of risk is negative and insignificant, while the resulting R^2 is rather low at only 1-11%.

One of the potential explanations behind such a subpar performance of the nontradable risk factors consists in the measurement error problem. Indeed, if the nondurable consumption growth (or any other variable) is observed with a measurement error, it causes an attenuation bias in the estimates of betas, which could in turn lead to a weak factor problem in small sample[29]. I address this issue by constructing mimicking portfolios of the nontradable factors using a simple linear projection on the cross-section of the corresponding stock returns. By construction, the resulting projection preserves the pricing impact of the original variable, however, it does not have the same measurement error component, as before.

Table VIII illustrates the use of mimicking portfolios for some of the models with nontradable factors. While there is considerable improvement in the performance of the nondurable consumption (unless the market return is also included into the model), the main finding remains unchanged: the model still suffer from the identification failures. Cross-products of the consumption-to-wealth ratio and consumption, durable consumption growth, labour and its cross-product still do not generate enough asset exposure to the risk factors to identify the associated risk premia, even when used as mimicking portfolios.

---

[29]Note, that the classical measurement error leads to a a multiplicative attenuation bias, and therefore can be the sole reason for the lack of identification. In finite sample, however, its presence makes the inference unreliable and, if large enough, could substantially exacerbate the underlying problem

VII.   Conclusion

Identification conditions play a major role in model estimation, and one must be very cautious when trying to draw quantitaive results from the data without considering this property first. While in some cases this requirement is fairly easy to test, the use of more complicated techniques sometimes makes it more difficult to analyze. This paper deals with one particular case of underidentification: the presence of useless factors in the linear asset pricing models. I proposed a new estimator that can be used simulatenously as a model diagnostic and estimation technique for the risk premia parameters. While automatically eliminating the impact of the factors that are either weakly correlated with asset returns (or do not correlate at all), the method restores the identification of the strong factors in the model, their estimation accuracy, and quality of fit.

Applying this new technique to real data, I find support for the pricing ability of several tradable factors (e.g. the three Fama-French factors or the 'quality-minus-junk' factor). I further demonstrate that the profitability factor largely drives the cross-section of momentum-sorted portfolios, contrary to the outcome of the standard Fama-MacBeth estimation.

It seems that much of the cross-sectional research with nontradable factors, however, should also be considered through the prism of model identification, as nearly all the specifications cosidered are contaminated by the problem of rank deficiency. How and whether the situation is improved in nonlinear models are undoubtedly very important questions, and form an interesting agenda for future research.

REFERENCES

ADRIAN, T., R. K. CRUMP, AND E. MOENCH (2013): "Pricing the Term Structure with Linear Regressions," *Journal of Financial Economics*, 110, 103–27.

ANDREWS, D. AND P. GUGGENBERGER (2009): "Incorrect Asymptotic Size of Subsampling Procedures Based on Post-consistent Model Selection Estimators," *Journal of Econometrics*, 152, 19–27.

ANDREWS, D. W. K. (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica*, 59, 307–345.

——— (1994): "Empirical Process Methods in Econometrics," *Econometrica*, 4, 2247–2294.

ASNESS, C., A. FRAZZINI, AND L. H. PEDERSEN (2014): "Quality Minus Junk," Working Paper.

BALL, R. AND P. BROWN (1968): "An Empirical Evaluation of Accounting Income Numbers," *Journal of Accounting Research*, 6, 159–78.

BANDI, F. AND A. TAMONI (2014): "Business-Cycle Consumption Risk and Asset Prices," Working Paper.

BERK, R., L. BROWN, A. BUJA, K. ZHANG, AND L. ZHAO (2013): "Valid Post-selection Inference," *Annals of Statistics*, 41, 802–37.

BERNAND, V. AND J. THOMAS (1990): "Evidence that Stock Prices do not Fully Reflect the Implications of Current Earnings for Future Earnings," *Journal of Accounting and Economics*, 13, 305–40.

BOONS, M. AND A. TAMONI (2014): "Horizon-Specific Macroeconomic Risks and the Cross-Section of Expected Returns," Working Paper.

BREEDEN, D. T. (1979): "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265–96.

BREIMAN, L. (1996): "Heuristics of instability and stabilization in model selection," *Annals of Statistics*, 24, 97–126.

CAMPBELL, J. Y. (2003): "Consumption-Based Asset Pricing," in *Handbook of the Economics of Finance*, ed. by G. Constantinides, M. Harris, and R. Stulz, Amsterdam: North-Holland, vol. 1, chap. 13.

CANER, M. (2009): "LASSO Type GMM Estimator," *Econometric Theory*, 25, 1–23.

CANER, M. AND M. FAN (2014): "Hybrid GEL Estimators: Instrument Selection with Adaptive Lasso," Working Paper.

CARHART, M. M. (1997): "On Persistence in Mutual Fund Performance," *Journal of Finance*, 52, 57–82.

CHAN, L., N. JEGADEESH, AND J. LAKONISHOK (1996): "Momentum Strategies," *Journal of Finance*, 51, 1681–713.

CHATTERJEE, A. AND S. N. LAHIRI (2010): "Asymptotic Properties of the Residual Bootstrap for Lasso Estimators," *Proceedings of the American Mathematical Society*, 138, 4497–4509.

——— (2011): "Bootstrapping Lasso Estimators," *Journal of the American Statistical Association*, 106, 608–25.

——— (2013): "Rates of Convergence of the Adaptive Lasso Estimators to the Oracle Distribution and Higher Order Refinements by the Bootstrap," *Annals of Statistics*, 41, 1232–59.

CHEN, L., R. NOVY-MARX, AND L. ZHANG (2011): "An Alternative Three-Factor Model," Working Paper.

COOPER, M., H. GULEN, AND M. SCHILL (2008): "Asset Growth and the Cross-section of Stock Returns," *Journal of Finance*, 63, 1609–52.

CRAGG, J. G. AND S. G. DONALD (1997): "Inferring the Rank of a Matrix," *Journal of Econometrics*, 76, 223–50.

DANIEL, K. D. AND D. MARSHALL (1997): "The Equity Premium Puzzle and the Risk-Free Rate Puzzle at Long Horizons," *Macroeconomic Dynamics*, 1, 452–84.

FAIRFIELD, P., S. WHISENANT, AND T. YOHN (2003): "Accrued Earnings and Growth: Implications for Future Profitability and Market Mispricing," *The Accounting Review*, 78, 353–71.

FAMA, E. F. AND K. R. FRENCH (1992): "The Cross-Section of Expected Stock Returns," *The Journal of Finance*, 47, 427–465.

——— (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *The Journal of Financial Economics*, 33, 3–56.

——— (2006): "Profitability, Investment, and Average returns," *Journal of Financial Economics*, 82, 491–518.

——— (2014): "A Five-Factor Asset Pricing Model," Fama-Miller Working Paper.

FRIEDMAN, J., T. HASTIE, H. HOFFLING, AND R. TIBSHIRANI (2007): "Pathwise Coordinate Descent," *The Annals of Applied Statistics*, 1, 302–32.

FRIEDMAN, J., T. HASTIE, AND R. TIBSHIRANI (2010): "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, 33, 1–22.

GOSPODINOV, N., R. M. KAN, AND C. ROBOTTI (2014a): "Misspecification-Robust Inference in Linear Asset Pricing Models with Irrelevant Risk Factors," *Review of Financial Studies*, 27, 2139–2170.

——— (2014b): "Spurious Inference in Unidentified Asset-Pricing Models," Working Paper.

GUGGENBERGER, P. (2010): "The Impact of a Hausman Pretest on the Size of a Hypothesis Test: the Panel Data Case," *Journal of Econometrics*, 156, 337–43.

HANSEN, L. P. AND R. JAGANNATHAN (1997): "Asssessing Specification Errors in Stochastic Discount Factor Models," *The Journal of Finance*, 52, 557–590.

HANSEN, L. P. AND K. J. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–86.

——— (1983): "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 91, 249–68.

HARVEY, C., Y. LIU, AND H. ZHU (2013): "...and the cross-section of expected returns," Working Paper, Duke University.

HASTIE, T., R. TIBSHIRANI, AND J. FRIEDMAN (2011): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd Edition)*, Springer Series in Statistics, New York: Springer-Verlag.

HAUGEN, R. AND N. BAKER (1996): "Commonality in the Determinants of Expected Stock Returns," *Journal of Financial Economics*, 41, 401–439.

HOU, K., C. XUE, AND L. ZHANG (2014): "Digesting Anomalies: An Investment Approach," *Review of Financial Studies*, forthcoming.

JAGANNATHAN, R. AND Z. WANG (1996): "The Conditional CAPM and the Cross-Section of Expected Returns," *The Journal of Finance*, 51, 3–53.

——— (1998): "An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regression," *The Journal of Finance*, 53, 1285–1309.

——— (2007): "Lazy Investors, Discretionary Consumption, and the Cross-Section of Stock Returns," *The Journal of Finance*, 62, 1623–61.

KAN, R. M. AND C. ZHANG (1999a): "GMM Tests of Stochastic Discount Factor Models with Useless Factors," *Journal of Financial Economics*, 54, 103–27.

——— (1999b): "Two-pass Tests of Asset Pricing Models with Useless Factors," *Journal of Finance*, 54, 204–35.

KLEIBERGEN, F. (2009): "Tests of Risk Premia in Linear Factor Models," *Journal of Econometrics*, 149, 149–73.

KLEIBERGEN, F. AND R. PAAP (2006): "Generalised Reduced Rank Test Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97–126.

KLEIBERGEN, F. AND Z. ZHAN (2013): "Unexplained Factors and their Effects on Second Pass R-squareds and t-tests," Working Paper.

KNIGHT, K. AND W. FU (2000): "Asymptotics for Lasso-type Estimators," *The Annals of Statistics*, 28, 1356–78.

KROENCKE, T. A. (2014): "Asset Pricing Without Garbage," Annual Conference 2014 (Hamburg): Evidence-based Economic Policy 100476, Verein fr Socialpolitik / German Economic Association.

LETTAU, M. AND S. LUDVIGSON (2001a): "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance*, 56, 815–49.

——— (2001b): "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying," *Journal of Political Economy*, 109, 1238–1286.

LEWELLEN, J., S. NAGEL, AND J. SHANKEN (2010): "A skeptical appraisal of asset pricing tests," *Journal of Financial Economics*, 96, 175–194.

LI, D. AND L. ZHANG (2010): "Does q-theory with Investment Frictions Explain Anomalies in the Cross-section of Returns?" *Journal of Financial Economics*, 210, 297–314.

LIAO, Z. (2013): "Adaptive GMM Shrinkage Estimation with Consistent Moment Selection," *Econometric Theory*, 29, 1–48.

LINTNER, J. (1965): "Security Prices, Risk, and Maximal Gains from Diversification," *Journal of Finance*, 20, 587–615.

LIU, L. X., T. M. WHITED, AND L. ZHANG (2009): "Investment-based Expected Stock Returns," *Journal of Political Economy*, 112, 1105–39.

LUCAS, JR., R. E. (1976): "Econometric Policy Evaluation: A Critique," in *The Phillips Curve and Labor Markets*, Amsterdam: North-Holland, vol. 1 of *Carnegie Rochester Conference Series on Public Policy*, 19–46.

LUDVIGSON, S. (2013): "Advances in Consumption-Based Asset Pricing: Empirical Tests," in *Handbook of the Economics of Finance*, ed. by G. Constantinides, M. Harris, and R. Stulz, Amsterdam: North-Holland, vol. 2, chap. 12.

LUSTIG, H., N. ROUSSANOV, AND A. VERDELHAN (2011): "Common Risk Factors in Currency Markets," *Review of Financial Studies*, 3731–77.

LUSTIG, H. AND A. VERDELHAN (2007): "The Cross-Section of Foreign Currency Risk Premia and US Consumption Growth Risk," *American Economic Review*, 89–117.

Lynch, A. W. (1996): "Decision Frequency and Synchronization Across Agents: Implications for Aggregate Consumption and Equity Return," *Journal of Finance*, 51, 1479–97.

Mankiw, N. G. and M. D. Shapiro (1986): "Risk and Return: Consumption Beta Versus Market Beta," *Review of Economics and Statistics*, 68, 452–59.

Mehra, R. and E. C. Prescott (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–61.

Novy-Marx, R. (2013): "The Other Side of Value: The Gross Profitability Premium," *Journal of Financial Economics*, 108, 1–28.

——— (2014): "Predicting Anomaly Performance with Politics, the Weather, Global Warming, Sunspots, and the Stars," *Journal of Financial Economics*, forthcoming.

Parker, J. A. and C. Julliard (2005): "Consumption Risk and the Cross-Section of Expected Returns," *Journal of Political Economy*, 113.

Politis, D. N. and J. P. Romano (1994): "The Stationary Bootstrap," *The Journal of American Statistical Association*, 89, 1303–1313.

Polk, C. and P. Sapienza (2009): "The Stock Market and Corporate Investment: A Test of Catering Theory," *Review of Financial Studies*, 22, 187–217.

Pollard, D. (1991): "Asymptotics for Least Absolute Deviations Regression Estimators," *Econometric Theory*, 7, 186–99.

——— (1994): "On the Asymptotics of Constrained M-estimation," *Annals of Statistics*, 22, 1993–2010.

Santos, T. and P. Veronesi (2006): "Labor Income and Predictable Stock Returns," *Review of Financial Studies*, 19, 1–44.

Shanken, J. (1985): "Multivariate Tests of the Zero-beta CAPM," *Journal of Financial Economics*, 14, 327–47.

——— (1992): "On the Estimation of Beta-Pricing Models," *The Review of Financial Studies*, 5, 1–33.

Sharpe, W. F. (1964): "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 425–42.

Stock, J., M. Watson, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business and Economic Statistics*, 20, 518 29.

Stock, J. H. and M. Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Economic Models: Essays in Honour of Thomas Rothenberg*, ed. by D. Andrews and J. Stock, Cambridge, UK: Cambridge University Press, 80–108.

Tibshirani, R. (1996): "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society: Series B*, 58, 267–88.

Titman, S., K. Wei, and F. Xie (2004): "Capital Investments and Stock Returns," *Journal of Financial and Quantitative Analysis*, 39, 677–700.

Tseng, P. (1988): "Coordinate Ascent for Maximizing Nondifferentiable Concave Functions," *Technical Report LIDS-P 1840*.

——— (2001): "Convergence of Block Coordinate Descent Method for Nondifferentiable Maximization," *Journal of Optimization Theory Applications*, 109, 474–94.

Weil, P. (1989): "The Equity Premium Puzzle and the Risk-Free Rate Puzzle," *Journal of Monetary Economics*, 24, 401–421.

Xing, Y. (2008): "Interpreting the Value Effect through the q-theory: An Empirical Investigation," *Review of Financial Studies*, 21, 1767–95.

Yogo, M. (2006): "A Consumption-Based Explanation of Expected Stock Returns," *Journal of Finance*, 61, 539–580.

Zhiang, L. and Z. Zhan (2013): "An Empirical Comparison of Non-traded and Traded Factors in Asset Pricing," Working paper.

Zou, H. (2006): "The Adaptive Lasso and Its Oracle Properties," *Journal of Americal Statistical Association*, 101, 1418–29.

Table I. Estimates of risk premia in a correctly specified model

	True parameter	Mean Estimate			Bias			MSE			Mean shrinkage
	value	Oracle	FM	Pen-FM	Oracle	FM	Pen-FM	Oracle	FM	Pen-FM	rate
Panel A: T=30											
Intercept	3.277	3.265	3.273	3.25	-0.012	-0.004	-0.027	2.259	2.26	2.203	0
Useful factor	-0.647	-0.673	-0.672	-0.659	-0.026	-0.026	-0.012	2.25	2.247	2.196	0.007
Useless factor	-	-	0.002	0	-	0.002	0	-	0.317	0	1
Panel B: T=50											
Intercept	3.277	3.17	3.149	3.173	-0.107	-0.128	-0.104	1.877	1.837	1.848	0
Useful factor	-0.647	-0.553	-0.533	-0.556	0.093	0.114	0.091	1.875	1.827	1.848	0.009
Useless factor	-	-	0.01	0	-	0.01	0	-	0.314	0	1
Panel C: T=100											
Intercept	3.277	3.213	3.195	3.21	-0.064	-0.082	-0.067	1.449	1.447	1.444	0
Useful factor	-0.647	-0.593	-0.575	-0.591	0.054	0.072	0.056	1.421	1.427	1.417	0.003
Useless factor	-	-	0.01	0	-	0.01	0	-	0.318	0	1
Panel D: T=250											
Intercept	3.277	3.267	3.271	3.266	-0.011	-0.007	-0.011	0.902	0.894	0.901	0
Useful factor	-0.647	-0.642	-0.648	-0.642	0.005	-0.001	0.005	0.887	0.885	0.886	0
Useless factor	-	-	-0.002	0	-	-0.002	0	-	0.325	0	1
Panel E: T=500											
Intercept	3.277	3.277	3.281	3.276	-0.001	0.004	-0.001	0.628	0.647	0.628	0
Useful factor	-0.647	-0.645	-0.65	-0.645	0.001	-0.003	0.002	0.627	0.646	0.627	0
Useless factor	-	-	0.007	0	-	0.007	0	-	0.31	0	1
Panel F: T=1000											
Intercept	3.277	3.286	3.278	3.286	0.009	0	0.009	0.435	0.441	0.435	0
Useful factor	-0.647	-0.655	-0.646	-0.654	-0.008	0.001	-0.008	0.421	0.431	0.421	0
Useless factor	-	-	-0.012	0	-	-0.012	0	-	0.321	0	1

Note. The table summarises the properties of the Fama-MacBeth and Pen-FM estimators with an identity weight matrix in a model for 25 portfolios with a common intercept and one true factor driving the returns. λ_0 is the value of the intercept, λ_1 and λ_2 are the corresponding risk premia of the true risk factor and the useless one. The model is simulated 10 000 times for different values of the sample size (T). The "Oracle" estimator corresponds to the Fama-MacBeth procedure omitting the useless factor, "FM" and "Pen-FM" stand for the Fama-MacBeth and Pen-FM estimators in the model with a useful and a useless factor. The table presents the mean point estimates of the parameters, their bias, and the mean squared error (MSE). The mean shrinkage rate corresponds to the average percentage of times the corresponding coefficient was set to exactly 0 during 10,000 simulations.

Returns are generated from the multivariate normal distribution with the mean and variance-covariance matrix equal to those of the nominal quarterly excess returns on 25 Fama-French portfolios sorted by size and book-to-market ratio during the period 1962Q2 : 2014Q2. The useful factor drives the cross-section of asset returns, and is calibrated to have the same mean and variance as the quarterly excess return on the market. The useless factor is generated from a multivariate normal distribution with the mean and variance equal to their sample analogues of nondurable consumption growth for the same time period. Betas, common intercept and risk premium for the useful factor come from the Fama-MacBeth estimates of a one factor model with market excess return estimated on the cross-section of the 25 Fama-French portfolios.

Table II. Estimates of risk premia in a missspecified model

	True parameter value (λ)	Mean Estimate			Bias			MSE			Mean shrinkage rate
		Oracle	FM	Pen-FM	Oracle	FM	Pen-FM	Oracle	FM	Pen-FM	
Panel A: T=30											
Intercept	3.315	3.192	3.041	3.149	-0.123	-0.274	-0.166	1.287	1.514	1.253	0
Useful factor	-1.316	-0.619	-0.629	-0.578	0.698	0.687	0.739	1.392	1.58	1.378	0.022
Useless factor		-	-0.019	0	-	-0.019	0	-	0.34	0	1
Panel B: T=50											
Intercept	3.315	3.184	3.053	3.177	-0.132	-0.262	-0.138	1.105	1.456	1.097	0
Useful factor	-1.316	-0.592	-0.621	-0.587	0.724	0.696	0.729	1.252	1.542	1.25	0.014
Useless factor		-	-0.021	0	-	-0.021	0	-	0.445	0	1
Panel C: T=100											
Intercept	3.315	3.253	3.142	3.247	-0.062	-0.173	-0.068	0.781	1.318	0.78	0
Useful factor	-1.316	-0.639	-0.692	-0.634	0.677	0.624	0.682	0.986	1.407	0.989	0.003
Useless factor		-	0.021	0	-	0.021	0	-	0.605	0	1
Panel D: T=250											
Intercept	3.315	3.261	3.159	3.259	-0.054	-0.156	-0.057	0.488	1.138	0.488	0
Useful factor	-1.316	-0.637	-0.708	-0.635	0.679	0.609	0.681	0.814	1.255	0.816	0
Useless factor		-	-0.004	0	-	-0.004	0	-	0.979	0	1
Panel E: T=500											
Intercept	3.315	3.276	3.246	3.275	-0.04	-0.069	-0.04	0.363	1.117	0.363	0
Useful factor	-1.316	-0.649	-0.794	-0.649	0.667	0.522	0.667	0.745	1.212	0.745	0
Useless factor		-	-0.008	0	-	-0.008	0	-	1.374	0	1
Panel F: T=1000											
Intercept	3.315	3.262	3.157	3.262	-0.053	-0.158	-0.053	0.255	1.053	0.255	0
Useful factor	-1.316	-0.634	-0.703	-0.634	0.682	0.614	0.682	0.72	1.197	0.72	0
Useless factor		-	0.049	0	-	0.049	0	-	1.979	0	1

Note. The table summarises the properties of the Fama-MacBeth and Pen-FM estimators with an identity weight matrix in a model for 25 portfolios with a common intercept and 3 factors driving the returns, but with only the first and a useless one considered in the estimation. λ_0 is the value of the intercept; λ_1 and λ_2 are the corresponding risk premia of the first useful factor and the useless one. The model is simulated 10,000 times for different values of the sample size (T). The "Oracle" estimator corresponds to the Fama-MacBeth procedure omitting the useless factor, "FM" and "Pen-FM" stand for the Fama-MacBeth and Pen-FM estimators in the model with a useful and a useless factor. The table summarises the mean point estimates of the parameters, their bias and the mean squared error. The mean shrinkage rate corresponds to the percentage of times the corresponding coefficient was set to exactly 0 during 10 000 simulations.

Returns are generated from the multivariate normal distribution with the mean and variance-covariance matrix equal to those of the quarterly nominal excess returns on 25 Fama-French portfolios sorted on size and book-to-market ratio during the period 1962Q2 : 2014Q2. Returns are simulated from a 3-factor model, the latter calibrated to have the same mean and variance as the three Fama-French factors (market excess return, SMB and HML portfolios). The useless factor is generated from a multivariate normal distribution with the mean and variance equal to their sample analogues of nondurable consumption per capita growth rate during the same time period. Betas, common intercept and risk premium for the useful factor come from the Fama-MacBeth estimates of a 3-factor model on the cross-section of 25 Fama-French portfolios. In the estimation, however, only the market return and the irrelevant factor are used; thus the source of misspecification is the omitted factors.

Table III. Shrinkage rate dependence on the value of the tuning parameter d

	Shrinkage rates for the									
	useful factor, $\lambda_1 \neq 0$					useless factor				
T	$d=3$	$d=4$	$d=5$	$d=7$	$d=10$	$d=3$	$d=4$	$d=5$	$d=7$	$d=10$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Correctly specified model										
30	0.0137	0.0032	0.0031	0.0014	0.0000	1.0000	0.9947	0.9915	0.9957	0.9981
50	0.0126	0.0011	0.0013	0.0012	0.0011	1.0000	0.9936	0.9926	0.9968	0.9992
100	0.0095	0.0010	0.0001	0.0002	0.0001	1.0000	0.9989	0.9987	1.0000	1.0000
250	0.0011	0.0001	0.0000	0.0001	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000
500	0.0001	0.0001	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1000	0.000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Panel B: Misspecified model										
30	0.0284	0.0063	0.0001	0.0001	0.0000	0.9968	0.9905	0.9821	0.9810	0.9989
50	0.0252	0.0011	0.0000	0.0000	0.0000	0.9976	0.9947	0.9768	0.9922	1.0000
100	0.0063	0.0001	0.0000	0.0000	0.0000	0.9978	0.9968	0.9905	1.0000	1.0000
250	0.0000	0.0000	0.0000	0.0000	0.0000	0.9998	0.9947	1.0000	1.0000	1.0000
500	0.0000	0.0000	0.0000	0.0000	0.0000	0.998	0.9999	1.0000	1.0000	1.0000
1000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Note. The table summarises the shrinkage rates for the useful/useless factor produced by the Pen-FM estimator for various sample sizes (T) and a range of parameters, $d = 2, 3, 5, 7, 10$, when η_0 is set at the average standard deviation of the residuals from the first stage. Simulation designs for the correctly specified and misspecified models correspond to those described in Tables I and II. Each sample is repeated 10,000 times.

Table IV. Shrinkage rate dependence on the value of the tuning parameter η_0

	Shrinkage rates for the									
	useful factor, $\lambda_1 \neq 0$					useless factor				
T	$\eta_0 = \bar{R}^e$	$\eta_0 = ln(\sigma^2)$	$\eta_0 = \bar{\sigma}$	$CV(5)$	$CV(n-1)$	$\eta_0 = \bar{R}^e$	$\eta_0 = ln(\sigma^2)$	$\eta_0 = \bar{\sigma}$	$CV(5)$	$CV(n-1)$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Correctly specified model										
30	0.0008	0.0008	0.0031	0.0014	0.0000	0.9888	0.9873	0.9947	0.9957	0.9981
50	0.0016	0.0001	0.0010	0.0012	0.0011	0.9857	0.9944	0.9936	0.9968	0.9992
100	0.0000	0.0000	0.0007	0.0002	0.0001	0.9976	0.9960	0.9989	1.0000	1.0000
250	0.0000	0.0000	0.0001	0.0001	0.0001	0.9992	0.9992	1.0000	1.0000	1.0000
500	0.0000	0.0000	0.0001	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Panel B: Misspecified model										
30	0.0284	0.0063	0.0001	0.0001	0.0000	0.9968	0.9905	0.9640	0.9637	0.9989
50	0.0252	0.0011	0.0000	0.0000	0.0000	0.9976	0.9947	0.9749	0.9912	1.0000
100	0.0063	0.0001	0.0000	0.0000	0.0000	0.9978	0.9968	0.9975	0.9971	1.0000
250	0.0000	0.0000	0.0000	0.0000	0.0000	0.9998	0.9947	1.0000	1.0000	1.0000
500	0.0000	0.0000	0.0000	0.0000	0.0000	0.998	0.9999	1.0000	1.0000	1.0000
1000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Note. The table illustrates the shrinkage rates for the useful/useless factor produced by the Pen-FM estimator for various sample sizes (T) and a range of parameters η_0, while $d = 4$. Simulation designs for the correctly specified and misspecified models correspond to those described in Tables I and II. Tuning parameter η_0 is set to be equal to 1) average excess return on the portfolio, 2)logarithm of average variance of the residuals from the first stage, 3) average standard deviation of the residuals from the first stage, 4) the average value of the tuning parameter chosen by 5-fold cross-validation, 5) the average value of the tuning parameter chosen by leave-one-out cross-validation. Each sample is repeated 10,000 times.

Table V. Survival rates of useful and irrelevant factors

Panel A: Correctly specified model

	Useful			Useful			Useful			Useless		
	$(\lambda_1 \neq 0)$			$(\lambda_2 \neq 0)$			$(\lambda_3 = 0)$					
T	$t_m(\lambda_1)$	Pen-FM	Bootstrap	$t_m(\lambda_2)$	Pen-FM	Bootstrap	$t_m(\lambda_3)$	Pen-FM	Bootstrap	$t_m(\lambda_4)$	Pen-FM	Bootstrap
50	0.0628	1	0.9995	0.1120	0.9166	0.9245	0.0588	0.9233	0.9426	0.1541	0	0.0126
100	0.1760	1	0.9997	0.2431	0.9403	0.9439	0.0207	0.9539	0.9606	0.0072	0	0.0079
150	0.3444	1	0.9998	0.4623	0.9652	0.9598	0.0232	0.9622	0.9695	0.0031	0	0.0053
200	0.5142	1	0.9998	0.6599	0.9787	0.9686	0.0231	0.9788	0.9749	0.0023	0	0.0040
250	0.6614	1	0.9998	0.8035	0.9761	0.9742	0.0231	0.9746	0.9786	0.0017	0	0.0032
600	0.9864	1	0.9998	0.9987	0.9802	0.9784	0.0141	0.9777	0.9813	0.0006	0	0.0027
1000	0.9999	1	0.9999	1.0000	0.9828	0.9815	0.0117	0.9833	0.9829	0.0003	0	0.0023

	Useful			Useful			Useless			Useless		
	$(\lambda_1 \neq 0)$			$(\lambda_2 \neq 0)$								
T	$t_m(\lambda_1)$	Pen-FM	Bootstrap	$t_m(\lambda_2)$	Pen-FM	Bootstrap	$t_m(\lambda_3)$	Pen-FM	Bootstrap	$t_m(\lambda_4)$	Pen-FM	Bootstrap
50	0.0573	1	0.9999	0.0984	1	1	0.1537	0	0.0164	0.1485	0	0.0232
100	0.1739	1	0.9999	0.2351	1	1	0.0068	0	0.0065	0.0085	0	0.0119
150	0.2020	1	1	0.2290	1	1	0.0080	0	0.0059	0.0032	0	0.0079
200	0.5265	1	1	0.6582	1	1	0.0017	0	0.0044	0.0025	0	0.0059
250	0.6742	1	1	0.8080	1	1	0.0015	0	0.0035	0.0015	0	0.0040
600	0.9880	1	1	0.9985	1	1	0.0007	0	0.0029	0.0003	0	0.0034
1000	1	1	1	0.9900	1	1	0.0000	0	0.0025	0.0002	0	0.0029

Panel B: Misspecified model

	Useful			Useful			Useful			Useless		
	$(\lambda_1 \neq 0)$			$(\lambda_2 \neq 0)$			$(\lambda_3 = 0)$					
T	$t_m(\lambda_1)$	Pen-FM	Bootstrap	$t_m(\lambda_2)$	Pen-FM	Bootstrap	$t_m(\lambda_3)$	Pen-FM	Bootstrap	$t_m(\lambda_4)$	Pen-FM	Bootstrap
50	0.0640	1	0.9995	0.1167	0.9453	0.9122	0.0676	0.9132	0.9437	0.1790	0	0.0133
100	0.1696	1	0.9996	0.2353	0.9617	0.9415	0.0224	0.9425	0.9614	0.0142	0	0.0075
150	0.3343	1	0.9997	0.4389	0.9733	0.9566	0.0221	0.9566	0.9699	0.0088	0	0.0052
200	0.5016	1	0.9998	0.6298	0.9787	0.9653	0.0240	0.9652	0.9751	0.0080	0	0.0039
250	0.6526	1	0.9998	0.7750	0.9826	0.9713	0.0238	0.9775	0.9786	0.0079	0	0.0031
600	0.9806	1	0.9998	0.9963	0.9850	0.9758	0.0138	0.9751	0.9812	0.0073	0	0.0026
1000	0.9972	1	0.9998	0.9989	0.9871	0.9792	0.0121	0.9764	0.9830	0.0088	0	0.0022

	Useful			Useful			Useless			Useless		
	$(\lambda_1 \neq 0)$			$(\lambda_2 \neq 0)$								
T	$t_m(\lambda_1)$	Pen-FM	Bootstrap	$t_m(\lambda_2)$	Pen-FM	Bootstrap	$t_m(\lambda_3)$	Pen-FM	Bootstrap	$t_m(\lambda_4)$	Pen-FM	Bootstrap
50	0.0406	1	0.9986	0.0815	0.7782	0.7805	0.1669	0	0.0228	0.1660	0	0.0279
100	0.0985	1	0.9992	0.1310	0.9246	0.9181	0.0138	0	0.0150	0.0141	0	0.0184
150	0.1928	1	0.9994	0.2493	0.9781	0.9744	0.0083	0	0.0123	0.0093	0	0.0134
200	0.3058	1	0.9996	0.3840	0.9793	0.9717	0.0074	0	0.0101	0.0081	0	0.0103
250	0.4221	1	0.9997	0.5180	0.9921	0.9927	0.0073	0	0.0082	0.0078	0	0.0087
600	0.9026	1	0.9997	0.9516	1	0.9980	0.0097	0	0.0069	0.0086	0	0.0073
1000	0.9822	1	0.9997	0.9922	1	1	0.0102	0	0.0059	0.0096	0	0.0063

Note. The table summarises the survival rates for the useful/useless factors in the simulations of a 4-factor model (correctly or incorrectly specified) for different sample sizes. For each of the factors, I compute its survival rate from 10,000 simulations, based on the t_m statistic from Gospodinov et al. (2014a)(Table 4), the pointwise estimates produced by the Pen-FM estimator (e.g. the frequency with which the risk premia estimate was *not* set exactly to 0), and one minus the average shrinkage rate from the Pen-FM estimator in 10,000 bootstrap replicas. Panel A presents the survival rates for the correctly specified model, when it is generated with 2 useful and 2 useless factors, or a combination of 2 useful (and priced) factors, 1 useful (but not priced) and 1 useless factor. Panel B presents similar results for a misspecified model. For a complete description of the simulation design, please refer to Gospodinov et al. (2014a)

Table VI. Comparison of the Pen-FM estimator with the adaptive lasso, based on the survival rates of useful and useless factors.

Panel A: Correctly specified model

	Useful $(\lambda_1 \neq 0)$		Useful $(\lambda_2 \neq 0)$		Useful $(\lambda_3 = 0)$		Useless	
T	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)
50	1	0.4172	0.9166	1	0.9233	0.7340	0	1
100	1	0.4745	0.9403	1	0.9539	0.8392	0	1
150	1	0.5173	0.9652	1	0.9622	0.9262	0	1
200	1	0.5743	0.9787	1	0.9748	0.9431	0	1
250	1	0.6260	0.9761	1	0.9746	0.9694	0	1
600	1	0.8132	0.9802	1	0.9777	1	0	1
1000	1	0.9099	0.9828	1	0.9833	1	0	1

	Useful $(\lambda_1 \neq 0)$		Useful $(\lambda_2 \neq 0)$		Useless		Useless	
T	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)
50	1	1	1	0.9322	0	1	0	1
100	1	1	1	0.9851	0	1	0	1
150	1	1	1	0.9955	0	1	0	1
200	1	1	1	1	0	1	0	1
250	1	1	1	1	0	1	0	1
600	1	1	1	1	0	1	0	1
1000	1	1	1	1	0	1	0	1

Panel B: Misspecified model

	Useful $(\lambda_1 \neq 0)$		Useful $(\lambda_2 \neq 0)$		Useful $(\lambda_3 = 0)$		Useless	
T	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)
50	1	0.4691	0.9453	1	0.9132	0.6133	0	1
100	1	0.4782	0.9617	1	0.9424	0.7134	0	1
150	1	0.4784	0.9733	1	0.9566	0.7650	0	1
200	1	0.4870	0.9787	1	0.9652	0.7612	0	1
250	1	0.4566	0.9826	1	0.9775	0.8377	0	1
600	1	0.5179	0.9850	1	0.9751	0.9810	0	1
1000	1	0.6433	0.9989	1	0.9764	0.9959	0	1

	Useful $(\lambda_1 \neq 0)$		Useful $(\lambda_2 \neq 0)$		Useless		Useless	
T	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)	Pen-FM	AdaLasso (BIC)
50	1	0.5352	0.7782	1	0	0.5654	0	1
100	1	0.4833	0.9246	1	0	0.6132	0	1
150	1	0.5090	0.9781	1	0	0.6911	0	1
200	1	0.4566	0.9793	1	0	0.7177	0	1
250	1	0.3431	0.9921	1	0	0.7432	0	1
600	1	0.3217	1	1	0	0.9210	0	1
1000	1	0.2918	1	1	0	0.9618	0	1

Note. The table summarises the survival rates for the useful/useless factors in the simulations of a 4-factor model (correctly or incorrectly specified) for different sample sizes. For each of the factors, I compute its survival rate from 10,000 simulations, based on the shrinkage rate of Pen-FM estimator ($d = 4$ and $\nu_0 = \bar{\sigma}$) in 10,000 bootstrap replicas. I then compute the corresponding factor survival rates of the adaptive lasso with the tuning parameter chosen by BIC. Panel A presents the survival rates for the correctly specified model when it is generated with 2 useful and 2 useless factors, or a combination of 2 useful (and priced), 1 useful (but not priced) factors, and 1 useless factor. Panel B presents similar results for a misspecified model.

For a complete description of the simulation designs, please refer to Gospodinov et al. (2014a)

Table VII. Models for the cross-section of stock returns

Model (1)	Factors (2)	p-value (Wald) (3)	GKR (2014) (4)	λ_j (5)	st.error (OLS) (6)	p-value (OLS) (7)	st.error (Shanken) (8)	p-value (Shanken) (9)	p-value (Bootstrap) (10)	R^2 (%) (11)	λ_j (12)	Shrinkage rate (Bootstrap) (13)	p-value (Bootstrap) (14)	R^2 (%) (15)
					Fama-MacBeth estimator							Pen-FM estimator		
				Panel A: tradable factors										
				25 portfolios, sorted by size and book-to-market										
CAPM	Intercept	-	-	1.431***	0.4282	0.0008	0.4325	0.0009	0.002	19	1.4307***	0	0.002	19
	MKT	0	yes	-0.658	0.4256	0.1222	0.4764	0.1674	0.184		-0.658	0	0.184	
				25 portfolios, sorted by size and book-to-market										
Fama and French (1992)	Intercept	-	-	1.252	0.2987	0	0.3054	0	0	70	1.2533	0	0	70
	MKT	0	yes	-0.703*	0.3035	0.0205	0.3721	0.0587	0.06		-0.704*	0	0.06	
	SMB	0	yes	0.145	0.0291	0	0.1424	0.3083	0.376		0.145	0	0.376	
	HML	0	no	0.43***	0.031	0	0.1376	0.0018	0.008		0.429***	0	0.008	
				25 portfolios, sorted by size and book-to-market										
"Quality-minus-junk" Asness, Frazzini Pedersen (2014)	Intercept	-	-	0.7*	0.3257	0.0317	0.3422	0.0409	0.092	84	0.576	0	0.212	83
	MKT	0	yes	-0.327	0.327	0.3177	0.4013	0.4155	0.412		-0.206	0	0.684	
	SMB	0	yes	0.174	0.0217	0	0.1459	0.2325	0.288		0.172	0	0.292	
	HML	0	no	0.398**	0.0265	0	0.1387	0.0041	0.016		0.416***	0	0.008	
	QMJ	0	no	0.44**	0.1102	0.0001	0.1599	0.006	0.016		0.324*	0.084	0.084	
				25 portfolios, sorted by size and book-to-market										
Carhart (1997)	Intercept	-	-	0.684**	0.3199	0.0325	0.381	0.0726	0.032	76	1.289***	0	0	71
	MKT	0	no	-0.001	0.3208	0.998	0.411	0.9984	0.546		-0.648	0.001	0.085	
	SMB	0	no	0.111	0.0263	0	0.1044	0.2869	0.276		0.106	0.001	0.301	
	HML	0	yes	0.408***	0.024	0	0.0997	0	0		0.385***	0	0	
	UMD	0	no	2.257	0.5999	0.0002	0.7276	0.0019	0.29		0	0.996	0.996	
				25 portfolios sorted by value and momentum										
	Intercept	-	-	0.898	0.396	0.0233	0.4076	0.0275	0.12	90	1.074*	0	0.052	90
	MKT	0	no	-0.146	0.3765	0.6989	0.416	0.7263	0.724		-0.304	0.002	0.434	
	SMB	0	no	0.201	0.0696	0.0039	0.1227	0.1013	0.196		0.224*	0	0.1	
	HML	0	no	0.122	0.253	0.6305	0.2773	0.6608	0.832		0	0.374	0.87	
	UMD	0	yes	0.811***	0.0258	0	0.1397	0	0		0.804***	0	0	
				25 portfolios, sorted by size and book-to-market										
q-factor model Hou, Xue and Zhang (2014)	Intercept	-	-	1.045***	0.3164	0.001	0.3354	0.0018	0.004	77	1.034***	0	0	70
	MKT	0	no	-0.553	0.3168	0.0807	0.3937	0.16	0.166		-0.505	0	0.184	
	M/E	0	yes	0.363**	0.0542	0	0.1513	0.0165	0.05		0.255	0.002	0.158	
	I/A	0	yes	0.407***	0.0976	0	0.1329	0.0022	0.004		0.363**	0.004	0.012	
	ROE	0	no	0.494**	0.2029	0.0148	0.2446	0.0432	0.042		0	0.822	0.822	
				25 portfolios sorted by value and momentum										
	Intercept	-	-	0.256	0.5046	0.6115	0.5381	0.6339	0.66	88	0.454	0	0.218	88
	MKT	0	no	0.285	0.4758	0.5489	0.5474	0.6024	0.604		0.105	0.001	0.921	
	M/E	0	yes	0.5***	0.0658	0	0.1566	0.0014	0.004		0.482***	0	0.006	
	I/A	0	no	0.063	0.2438	0.796	0.273	0.8174	0.788		0	0.759	0.979	
	ROE	0	yes	0.665***	0.1467	0	0.1951	0.0006	0.006		0.63***	0	0.004	

Table VII. Models for the cross-section of stock returns

				Fama-MacBeth estimator							Pen-FM estimator			
Model (1)	Factors (2)	p-value (Wald) (3)	GKR (2014) (4)	λ_j (5)	st.error (OLS) (6)	p-value (OLS) (7)	st.error (Shanken) (8)	p-value (Shanken) (9)	p-value (Bootstrap) (10)	R^2 (%) (11)	λ_j (12)	Shrinkage rate (Bootstrap) (13)	p-value (Bootstrap) (14)	R^2 (%) (15)
				10 portfolios sorted on momentum										
	Intercept	-	-	1.164	0.7529	0.1222	0.8086	0.1502	0.432	93	-0.064	0	0.582	90
	MKT	0	no	-0.631	0.7234	0.3834	0.8037	0.4327	0.73		0.578	0.001	0.951	
	M/E	0	yes	0.73	0.7362	0.3213	0.803	0.3632	0.614		0	0.968	0.968	
	I/A	0	no	0.02	0.5142	0.9685	0.5585	0.971	0.91		0	0.582	0.6	
	ROE	0	yes	0.468	0.3192	0.1425	0.3621	0.1961	0.206		0.742**	0.005	0.033	
				19 portfolios sorted by P/E ratio										
	Intercept	-	-	2.71	1.447	0.0611	1.7586	0.1233	0.504	81	0.2578	0	0.544	76
	MKT	0	yes	-2.124	1.4002	0.1293	1.714	0.2153	0.7		0.272	0	0.968	
	M/E	0	yes	1.132	0.5639	0.0447	0.6995	0.1056	0.54		0	0.957	0.967	
	I/A	0	no	0.056	0.2395	0.8144	0.3027	0.8527	0.374		0.443*	0.051	0.095	
	ROE	0	no	0.072	0.281	0.798	0.3609	0.842	0.946		0	0.669	0.845	

Panel B: nontradable factors

25 portfolios sorted by size and book-to-market

Model	Factors	p-value (Wald)	GKR (2014)	λ_j	st.error (OLS)	p-value (OLS)	st.error (Shanken)	p-value (Shanken)	p-value (Bootstrap)	R^2 (%)	λ_j	Shrinkage rate (Bootstrap)	p-value (Bootstrap)	R^2 (%)
Lettau and Ludvigson (2001)	Intercept	-	-	3.24	0.6601	0	0.7261	0	0	16	3.8128	0.0000	0	0
	Δc	0.0061	no	0.216	0.1653	0.1922	0.1861	0.2468	0.398		0	0.889	0.975	
	Intercept	-	-	3.681**	0.9501	0.0001	1.4126	0.0092	0.012	31	4.106***	0	0.004	1
	cay	0.2006	no	-0.516	2.1813	0.813	3.3786	0.8786	0.518		0	0.958	0.976	
(scaled CAPM)	MKT	0	no	-0.056	0.9751	0.9545	1.6072	0.9724	0.72		-0.251	0	0.902	
	$cay\times$ MKT	0	no	1.136	0.3035	0.0002	0.4602	0.0136	0.398		0	0.839	0.851	
(scaled CCAPM)	Intercept	-	-	4.281***	0.7043	0	1.0093	0	0	70	4.958***	0	0	25
	cay	0.001	no	-0.125	0.2784	0.6526	0.41	0.7599	0.584		-0.385	0.319	0.513	
	Δc	0.7378	no	0.024	0.1095	0.8292	0.1619	0.884	0.654		0	0.689	0.721	
	$cay \times \Delta c$	0.0002	no	0.057	0.0171	0.0009	0.0254	0.0246	0.298		0	0.989	0.989	
(HC-CAPM)	Intercept	-	-	4.467***	0.9389	0	1.684	0.008	0.008	58	4.160***	0.0000	0	1
	MKT	0	no	-1.097	0.9322	0.2392	1.81	0.5445	0.798		-0.296	0	0.918	
	Δy	0.6566	no	1.259	0.3641	0.0005	0.6569	0.0552	0.1		0	0.815	0.815	
(scaled HC-CAPM)	Intercept	-	-	5.184***	0.9293	0	1.5628	0.0009	0.016	77	4.268***	0	0.002	7
	cay	0.0534	no	-0.445	0.2629	0.0908	0.4521	0.3254	0.502		0	0.987	0.987	
	MKT	0	no	-1.987	0.9226	0.0313	1.6995	0.2424	0.564		-0.438	0	0.8	
	Δy	0.3859	no	0.557	0.254	0.0282	0.4331	0.1982	0.23		0	0.792	0.792	
	$cay\times$ MKT	0	no	0.341	0.1841	0.0643	0.3225	0.2908	0.422		0	0.858	0.87	
	$cay \times \Delta y$	0.5135	no	-0.167	0.0678	0.014	0.1153	0.1485	0.422		-0.009	0.809	0.809	

Table VII. Models for the cross-section of stock returns

Model (1)	Factors (2)	p-value (Wald) (3)	GKR (2014) (4)	λ_j (5)	st.error (OLS) (6)	p-value (OLS) (7)	st.error (Shanken) (8)	p-value (Shanken) (9)	p-value (Bootstrap) (10)	R^2 (%) (11)	λ_j (12)	Shrinkage rate (Bootstrap) (13)	p-value (Bootstrap) (14)	R^2 (%) (15)
				25 portfolios, sorted by size and book-to-market										
Durable	Intercept	-	-	2.335**	0.9331	0.0123	1.5056	0.1209	0.03	55	3.445***	0	0.002	11
consumption CAPM	Δc_{ndur}	0.1116	no	0.641	0.2197	0.0035	0.3565	0.0721	0.126		0	0.974	0.974	
	Δc_{dur}	0.6711	no	0.013	0.1305	0.9215	0.2139	0.952	0.884		0	0.99	0.996	
	MKT	0	no	-0.152	0.9662	0.8754	1.6615	0.9273	0.592		-1.03	0.001	0.359	
				24 portfolios, sorted by book-to-market within industry										
	Intercept	-	-	1.767	0.862	0.0404	0.9431	0.061	0.414	11	1.317	0	0.344	3
	Δc_{ndur}	0.1513	no	0.232	0.1061	0.029	0.1221	0.0579	0.526		0	0.993	0.995	
	Δc_{dur}	0.6878	no	-0.002	0.1836	0.9891	0.2043	0.9902	0.738		0	0.976	0.998	
	MKT	0	no	0.44	0.8346	0.5977	1.0789	0.6831	0.46		0.89	0.002	0.488	
				24 portfolios, sorted by market and HML betas										
	Intercept	-	-	1.558**	0.6859	0.0231	0.9654	0.1066	0.014	44	2.185***	0	0.004	1
	Δc_{ndur}	0.9222	no	0.522	0.1578	9e-04	0.2253	0.0206	0.272		0	0.999	0.999	
	Δc_{dur}	0.021	no	0.112	0.1277	0.3823	0.1836	0.5434	0.456		0	0.996	0.998	
	MKT	0	no	0.338	0.6665	0.6122	1.1002	0.7587	0.842		-0.169	0	0.942	

Note. The table presents the risk premia estimates and fit for different models of the cross-section of stocks. Panel A summarises results for the models that rely on tradable risk factors, while Panel B demonstrated similar results for the models, relying on nontradable factors. First column describes the estimated model, or refers to the paper where the original factor was first proposed. Column 2 presents the list of the risk factors used in the corresponding speicification. Column 3 presents the p-value of the Wald test for the factor being a useless one, based on the first stage estimates of betas and heteroscedasticity and autocorrelation-robust standard errors, based on the lag truncation rule suggested in Andrews (1991). Column 4 indicates whether a particular risk factor has survived the sequential elimination procedure based on the misspecification-robust t_m-statistic of Gospodinov et al. (2014a). Columns 5-11 present the results of the model estimation based on the Fama-MacBeth procedure with an identity weight matrix ($W = I_n$), and include point estimates of the risk premia, OLS and Shanken standard errors, the corresponding p-values, and the p-value based on 10,000 pairwise block stationary bootstrap of Politis and Romano (1994). Column 11 presents the cross-sectional R^2 of the model estimated by the Fama-MacBeth procedure. Columns 12-15 describe Pen-FM estimation of the model, and summarise the point estimates of the risk premia, their shrinkage rate in the 10,000 bootstrap samples, the corresponding p-value of the parameter, and the cross-sectional R^2. Grey areas highlight the factors that are identified as useless/weak by the Pen-FM estimator (and, hence, experience a substantial shrinkage rate)

Table VIII. Mimicking portfolios of the nontradeable factor and the cross-section of stock returns

Model (1)	Factors (2)	p-value (Wald) (3)	GKR (2014) (4)	Fama-MacBeth estimator							Pen-FM estimator			
				λ_j (5)	st.error (OLS) (6)	p-value (OLS) (7)	st.error (Shanken) (8)	p-value (Shanken) (9)	p-value (Bootstrap) (10)	R^2 (%) (11)	λ_j (12)	Shrinkage rate (Bootstrap) (13)	p-value (Bootstrap) (14)	R^2 (%) (15)
				25 portfolios sorted by size and book-to-market										
Scaled CAPM	Intercept	-	-	2.604	1.0306	0.0115	1.2044	0.0306	0.01	27	3.5769	0	0.008	8
	cay	0	no	-2.46	0.5413	0	0.7895	0.0018	0.105		-0.382	0.9725	0.9735	
	MKT	0	no	1.249	1.0998	0.2563	1.4605	0.3926	0.765		0.232	0.0005	0.8125	
	cay× MKT	0	no	-0.114	0.0959	0.2344	0.1211	0.3463	0.93		0	0.9955	0.9955	
Scaled CCAPM	Intercept	-	-	2.253	0.9141	0.0137	1.0663	0.0346	0.019	69	2.4462	0	0.023	68
	cay	0	no	-0.111	0.0731	0.128	0.0974	0.2538	0.321		-0.099	0.071	0.311	
	Δc	0	no	0.175	0.0567	0.002	0.0721	0.015	0.024		0.164	0.006	0.015	
	cay $\times \Delta c$	0	no	-0.005	0.0049	0.3124	0.0061	0.4124	0.806		0	0.9805	0.9845	
HC-CAPM	Intercept	-	-	0.704	1.3241	0.595	1.6608	0.6716	0.759	93	4.7047	0	0.111	9
	MKT	0	no	2.472	1.3211	0.0613	1.7963	0.1688	0.078		-0.774	0	0.811	
	Δy	0	no	0.641	0.1689	1e-04	0.2237	0.0042	0.001		0	0.812	0.812	
Scaled HC-CAPM	Intercept	-	-	5.184	0.9293	0	1.5628	9e-04	0.002	77	4.1585	0	0.003	1
	cay	0	no	-4.446	2.6289	0.0908	4.5212	0.3254	0.535		0	0.997	0.997	
	MKT	0	no	-1.987	0.9226	0.0313	1.6995	0.2424	0.547		-0.298	0	0.945	
	Δy	0	no	0.557	0.254	0.0282	0.4331	0.1982	0.223		0	0.981	0.981	
	cay× MKT	0	no	0.341	0.1841	0.0643	0.3225	0.2908	0.457		0	0.9895	0.9915	
	cay $\times \Delta y$	0	no	-0.167	0.0678	0.014	0.1153	0.1485	0.394		0	0.986	0.986	
Durable consumption model	Intercept	-	-	2.333	0.9333	0.0124	1.0883	0.0321	0.027	55	3.6587	0.0000	0	21
	Δc_{ndur}	0	no	0.136	0.0525	0.0096	0.0643	0.0346	0.073		0	0.976	0.976	
	Δc_{dur}	0	no	-0.019	0.0284	0.5011	0.0381	0.6161	0.987		0	0.998	1	
	MKT	0	no	-0.19	0.9624	0.8433	1.2608	0.88	0.71		-1.252	0	0.22	

Note. The table presents the risk premia estimates and fit for different models of the cross-section of stocks using mimicking portfolios for the nontradable factors. First column describes the estimated model, or refers to the paper where the original factor was first proposed. Column 2 presents the list of the risk factors used in the corresponding specification. Column 3 presents the p-value of the Wald test for the factor being a useless one, based on the first stage estimates of betas and heteroscedasticity and autocorrelation-robust standard errors, based on the lag truncation rule suggested in Andrews (1991). Column 4 indicates whether a particular risk factor has survived the sequential elimination procedure based on the misspecification-robust t_m-statistic of Gospodinov et al. (2014a). Columns 5-11 present the results of the model estimation based on the Fama-MacBeth procedure with an identity weight matrix ($W = I_n$), and include point estimates of the risk premia, OLS and Shanken standard errors, the corresponding p-values, and the p-value based on 10,000 pairwise block stationary bootstrap of Politis and Romano (1994). Column 11 presents the cross-sectional R^2 of the model estimated by the Fama-MacBeth procedure. Columns 12-15 describe Pen-FM estimation of the model, and summarise the point estimates of the risk premia, their shrinkage rate in the 10,000 bootstrap samples, the corresponding p-value of the parameter, and the cross-sectional R^2. Grey areas highlight the factors that are identified as useless/weak by the Pen-FM estimator (and, hence, experience a substantial shrinkage rate)

Figure 1 Distribution of the cross-sectional R^2 in a correctly specified model
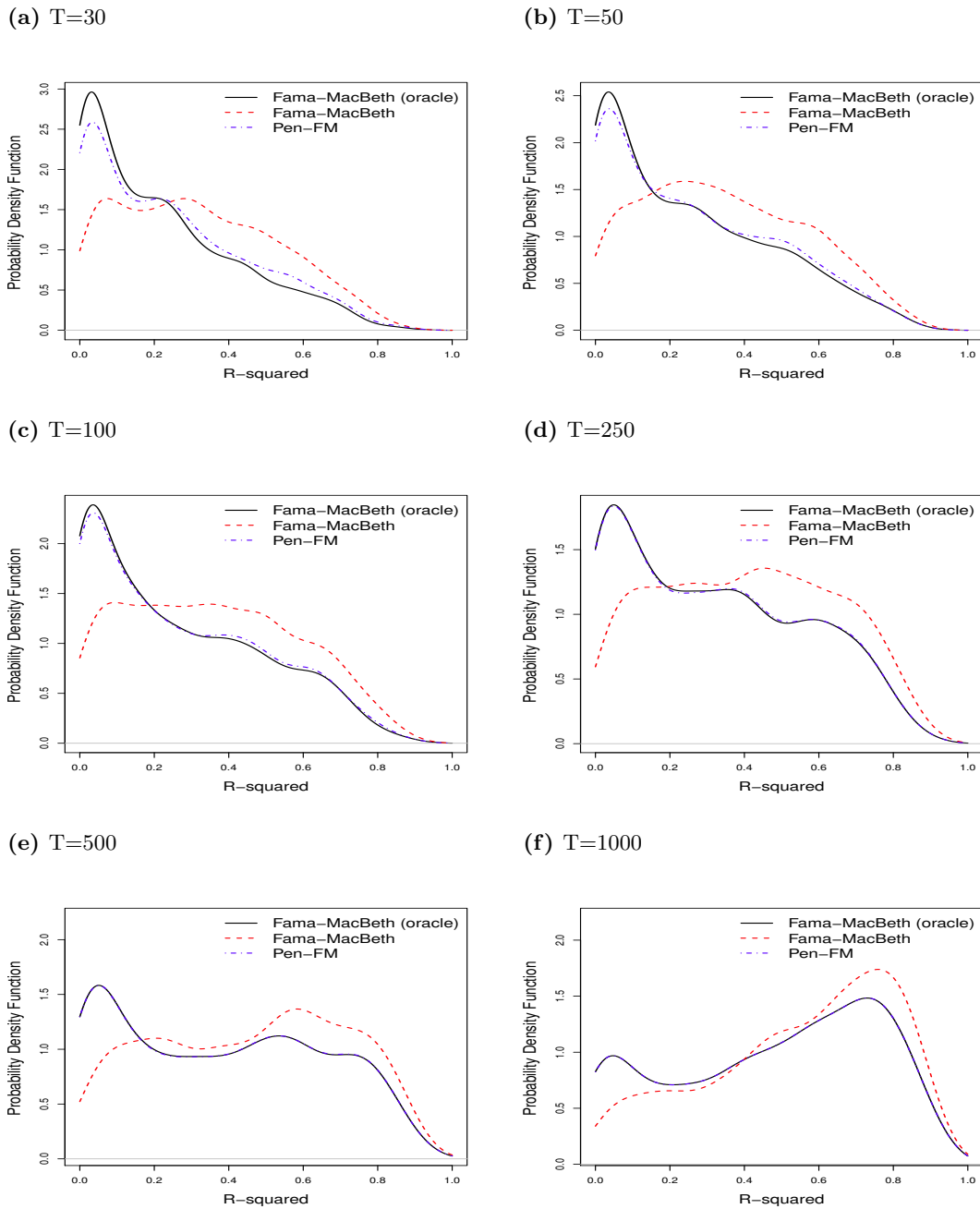
(a) T=30

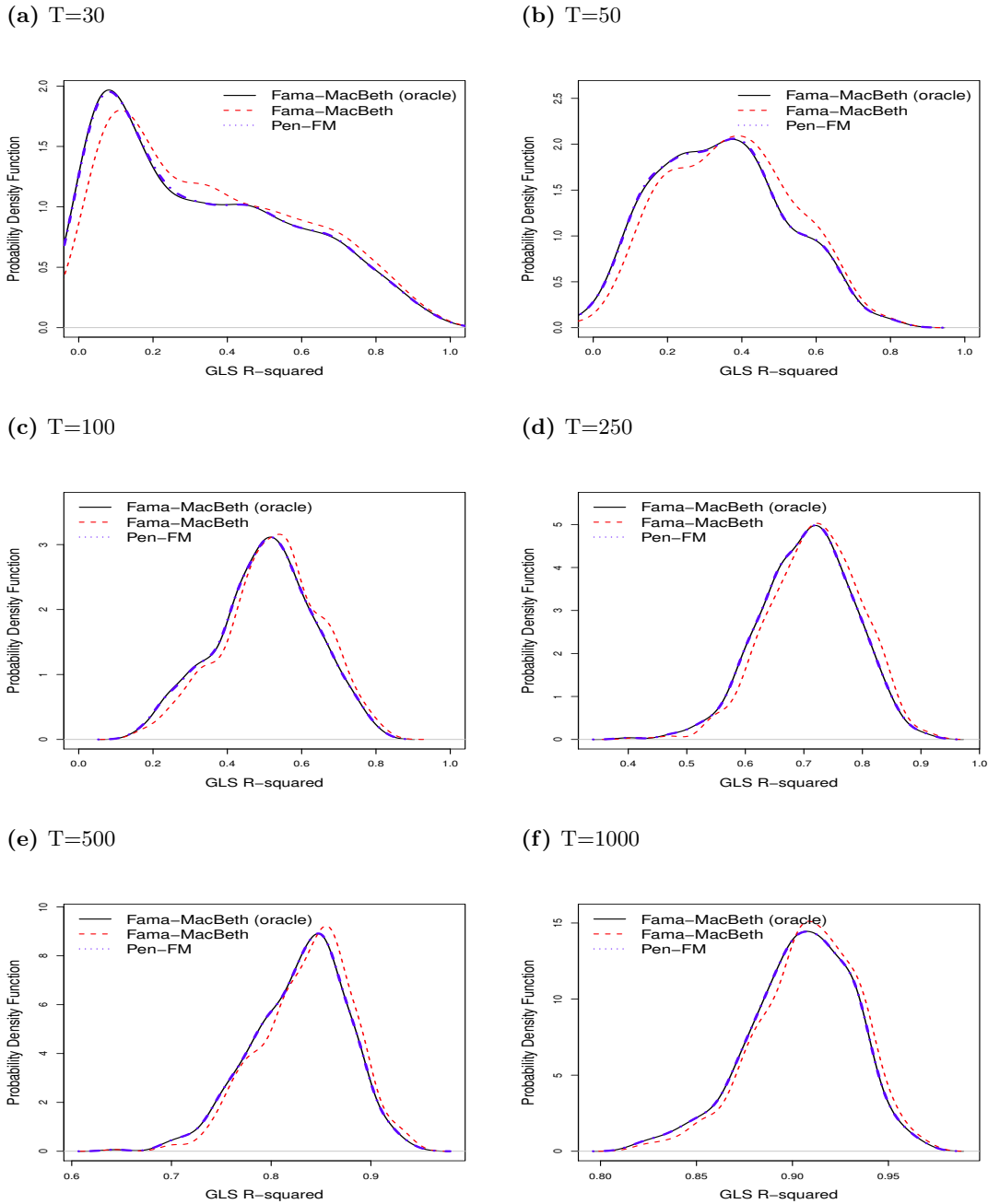**(b)** T=50

(c) T=100
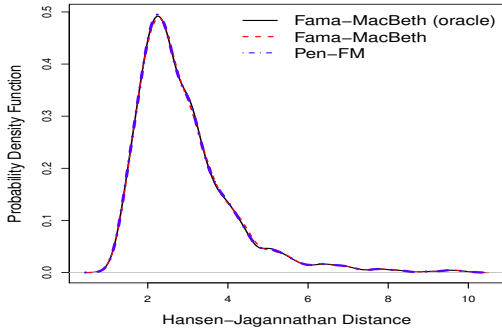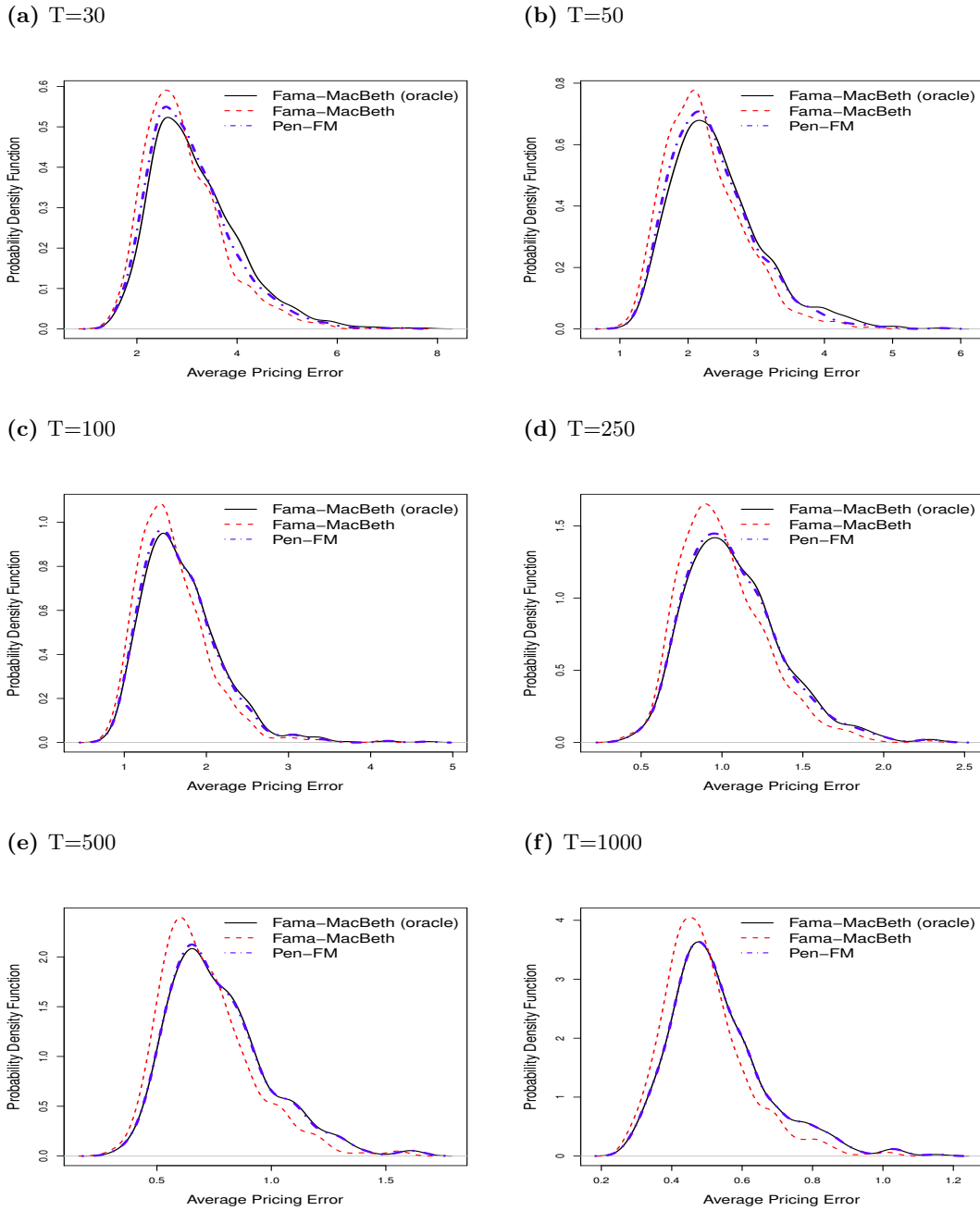
(d) T=250

(e) T=500

(f) T=1000

Note. The graphs present the probability density function for the cross-sectional R-squared in a simulation of a correctly specified model, potentially contaminated by the presence of an irrelevant factor for various sample sizes (T=30, 50, 100, 250, 500, 1000). For each of the sample sizes, the solid line represents the p.d.f. of the R-squared in the model without a useless factor, when the risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of the cross-sectional R-squared when the model is estimated by the Fama-MacBeth procedure, and a useless factor is included, while the dash-dotted line stands for the R^2 when the Pen-FM estimator is employed in the same scenario of the contaminated model.

For a detailed description of the simulation design, please refer to Table I.

**Figure 2** Distribution of the cross-sectional $GLS\ R^2$ in a correctly specified model based on the OLS risk premia estimates in the second stage

(a) T=30
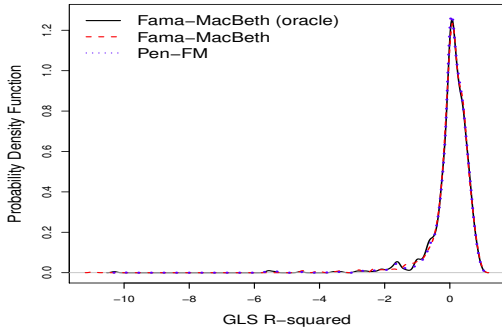
(b) T=50
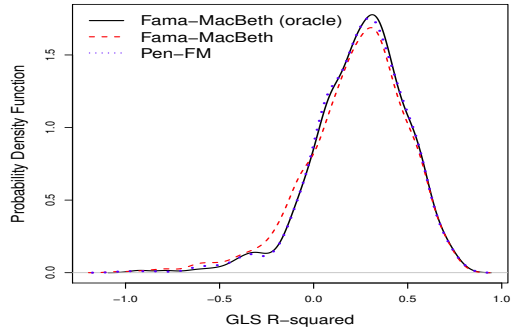
(c) T=100
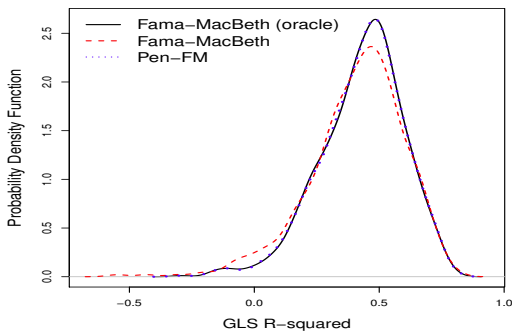
(d) T=250
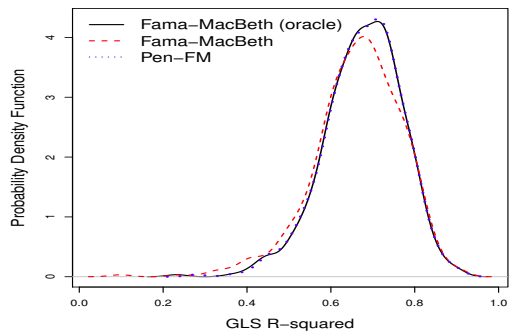
(e) T=500
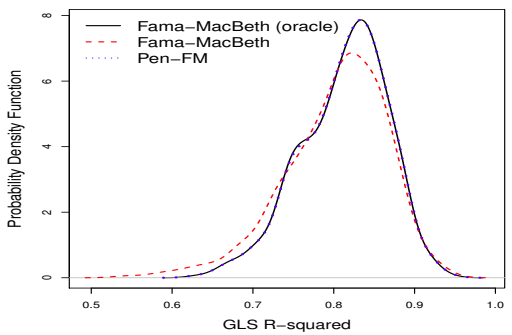
(f) T=1000
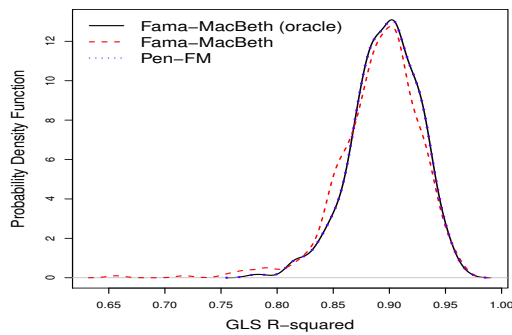
Note. The graphs demonstrate the probability density function for the cross-sectional $GLS\ R^2$ in a simulation of a correctly specified model, potentially contaminated by the presence of an irrelevant factor, and estimated using an identity weight matrix on the second stage ($W = I_n$). For each of the sample sizes (T=30, 50, 100, 250, 500, 1000), the solid line represents p.d.f. of the $GLS\ R^2$ in the model without a useless factor, when risk premia are estimated by Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of the cross-sectional $GLS\ R^2$ when the model is estimated by Fama-MacBeth procedure, and a useless factor is included, while the dash-dotted line stands for the $GLS\ R^2$ when Pen-FM estimator is employed in the same scenario of the contaminated model.

For a detailed description of the simulation design, please refer to Table I.

Figure 3 Distribution of the cross-sectional GLS R^2 in a correctly specified model based on the GLS risk premia estimates in the second stage

(a) T=30

(b) T=50

(c) T=100

(d) T=250

(e) T=500

(f) T=1000

Note. The graphs present the probability density function for the cross-sectional $GLS\ R^2$ in a simulation of a correctly specified model, potentially contaminated by the presence of an irrelevant factor, and estimated using the FGLS weight matrix on the second stage ($W = \hat{\Omega}^{-1}$). For each of the sample sizes (T=30, 50, 100, 250, 500, 1000), the solid line represents the p.d.f. of the $GLS\ R^2$ in the model without a useless factor, when risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of the cross-sectional $GLS\ R^2$ when the model is estimated by Fama-MacBeth procedure, and a useless factor is included, while the dash-dotted line stands for the $GLS\ R^2$ when the Pen-FM estimator is employed in the same scenario of the contaminated model.

For a detailed description of the simulation design, please refer to Table I.

Figure 4 Distribution of the Hansen-Jagannathan distance in a correctly specified model
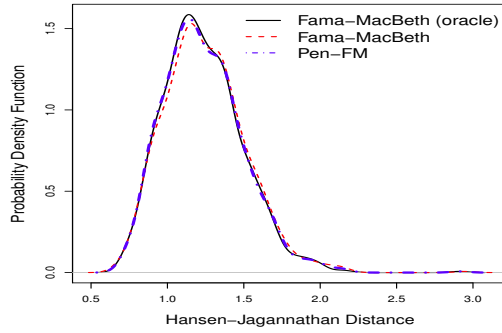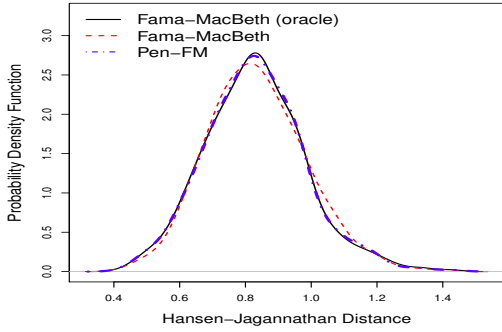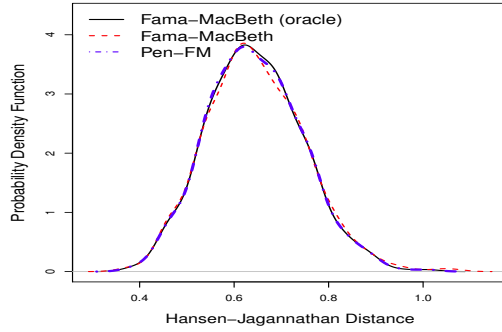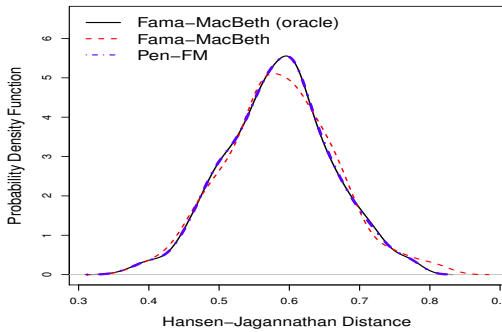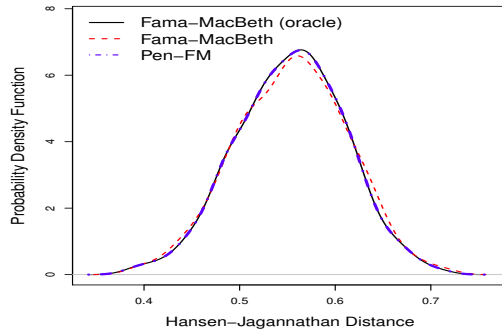
(a) T=30

(b) T=50

(c) T=100

(d) T=250

(e) T=500

(f) T=1000

Note. The graphs present the probability density function for the Hansen-Jagannathan distance in the simulations of a correctly specified model, potentially contaminated by the presence of an irrelevant factor, and the risk premia estimated using an identity weight matrix on the second stage ($W = I_n$). For each of the sample sizes (T=30, 50, 100, 250, 500, 1000), the solid line represents the p.d.f. of HJ in the model without a useless factor, when the risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of HJ when the model is estimated by the Fama-MacBeth procedure, and a useless factor is included, while the dash-dotted line stands for HJ when the Pen-FM estimator is employed in the same scenario of the contaminated model.

For a detailed description of the simulation design, please refer to Table I.

Figure 5 Distribution of the average pricing error in a correctly specified model

(a) T=30

(b) T=50

(c) T=100

(d) T=250

(e) T=500

(f) T=1000

Note. The graphs present the probability density function for the average pricing error (APE) in the simulations of a correctly specified model, potentially contaminated by the presence of an irrelevant factor, and the risk premia estimated using an identity weight matrix on the second stage ($W = I_n$). For each of the sample sizes (T=30, 50, 100, 250, 500, 1000), the solid line represents the p.d.f. of the APE in the model without a useless factor, when the risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of APE when the model is estimated by Fama-MacBeth procedure, and a useless factor is included as well, while the dash-dotted line stands for the APE when the Pen-FM estimator is employed in the same scenario of the contaminated model.

For a detailed description of the simulation design, please refer to Table I.

Figure 6 Distribution of the cross-sectional R^2 in a misspecified model

(a) T=30

(b) T=50

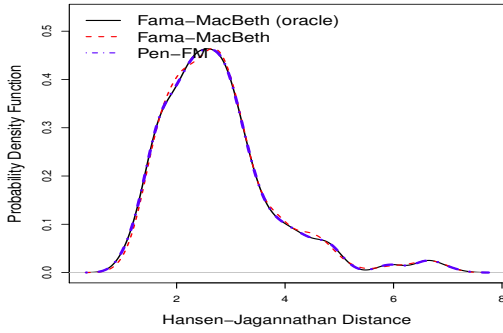**(c)** T=100

(d) T=250

(e) T=500

(f) T=1000
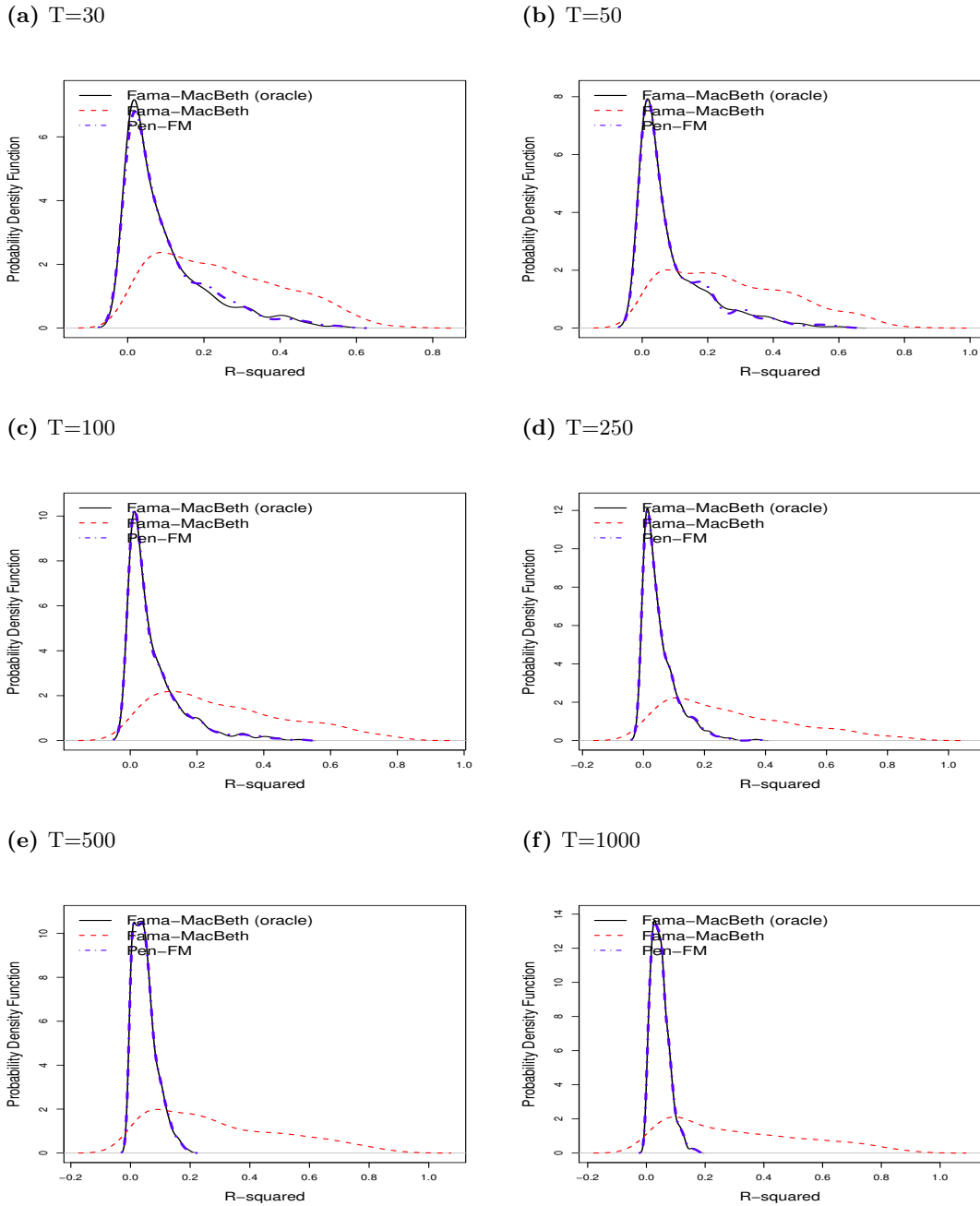
Note. The graphs present the probability density function for the cross-sectional R^2 in a simulation of a misspecified model with omitted variable bias and further potentially contaminated by the presence of an irrelevant factor for various sample sizes (T=30, 50, 100, 250, 500, 1000). The second stage estimates are produced using an identity weight matrix. For each of the sample sizes, the solid line represents p.d.f. of the R^2 statistic in the model without a useless factor, when the risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of the cross-sectional R^2 when the model is estimated by the Fama-MacBeth procedure, including both the useful and the useless factor, while the dash-dotted line stands for R^2 when the Pen-FM estimator is employed in the same scenario of the contaminated model.

For a detailed description of the simulation design for the misspecified model, please refer to Table II.

**Figure 7** Distribution of the $GLS\ R^2$ in a misspecified model based on the OLS estimates of the risk premia in the second stage

(a) T=30
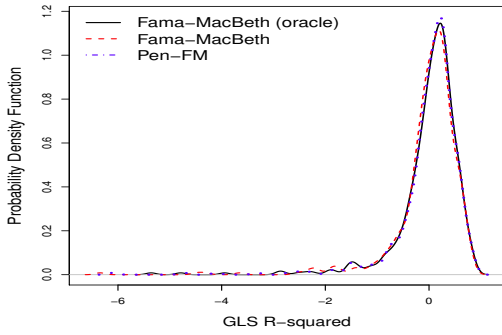
(b) T=50
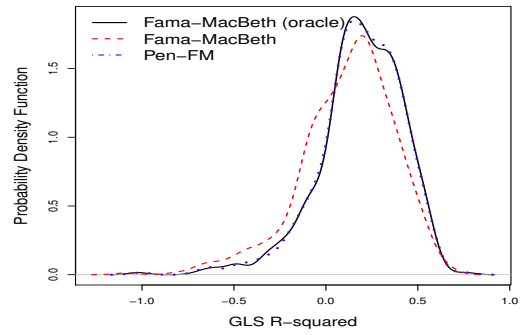
(c) T=100
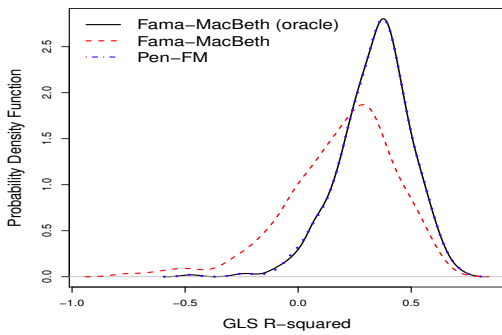
(d) T=250
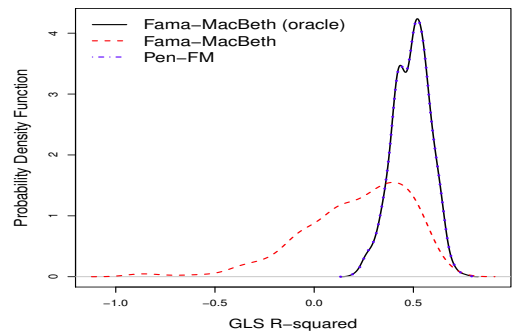
(e) T=500
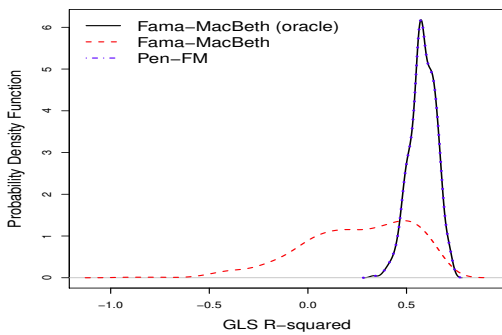
(f) T=1000
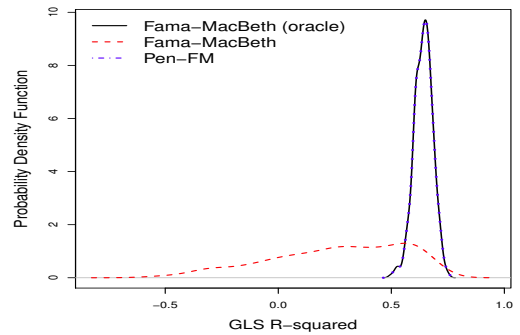
Note. The graphs illustrate the probability density function for the cross-sectional $GLS\ R^2$ in a simulation of a misspecified model with omitted variable bias and further potentially contaminated by the presence of an irrelevant factor for various sample sizes (T=30, 50, 100, 250, 500, 1000). The second stage estimates are produced using an identity weight matrix. For each of the sample sizes, the solid line represents the p.d.f. of the $GLS\ R^2$ statistic in the model without a useless factor, when the risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of the cross-sectional $GLS\ R^2$ when the model is estimated by the Fama-MacBeth procedure, including both the useful and the useless factor, while the dash-dotted line stands for R^2 when the Pen-FM estimator is employed in the same scenario of the contaminated model.

For a detailed description of the simulation design for the misspecified model, please refer to Table II.

Figure 8 Distribution of the cross-sectional GLS R^2 in a misspecified model with risk premia estimates based on the GLS second stage
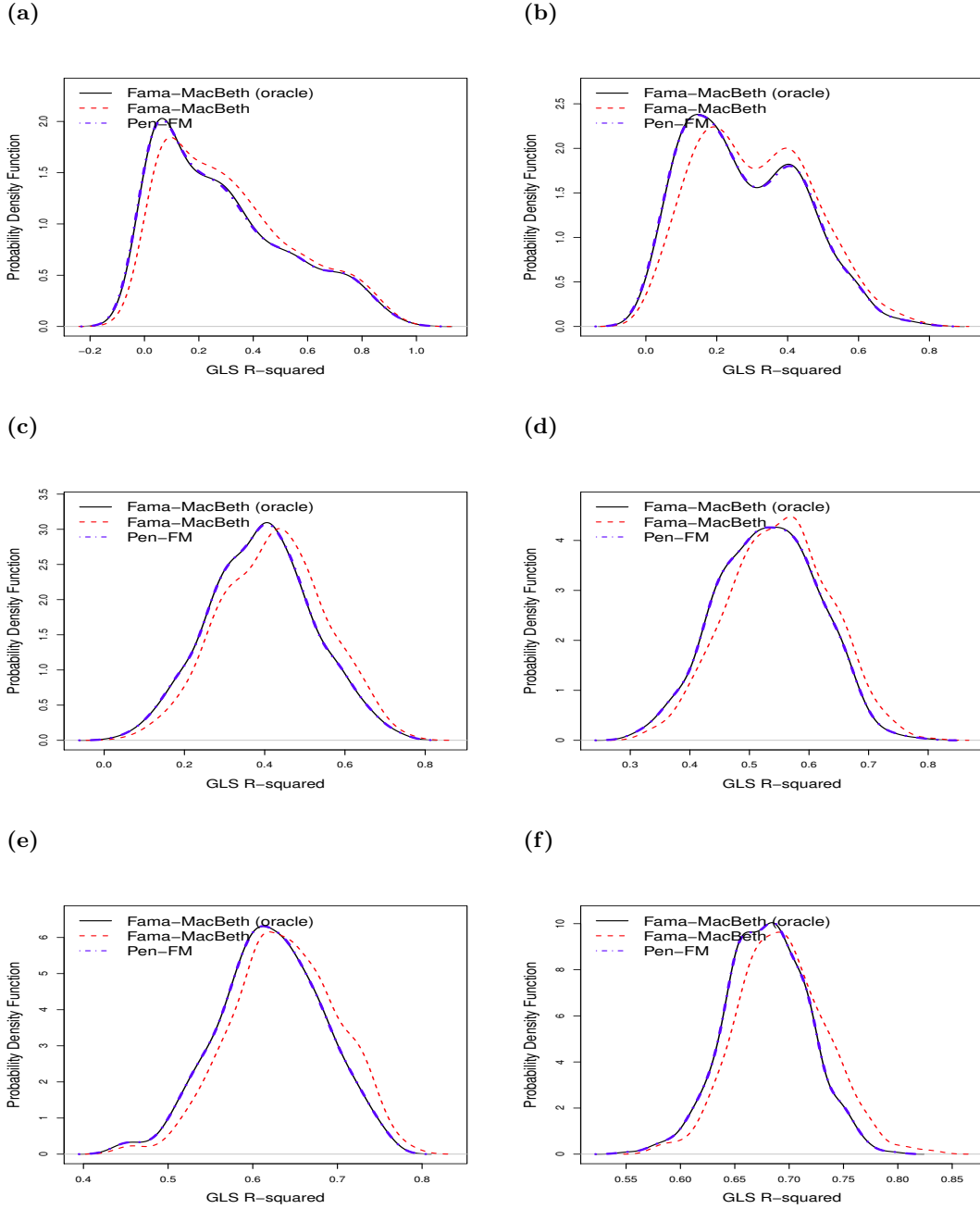
(a)

(b)

(c)

(d)

(e)

(f)

Note. The graphs present the probability density function for the cross-sectional GLS R^2 in a simulation of a misspecified model with omitted variable bias and further potentially contaminated by the presence of an irrelevant factor for various sample sizes (T=30, 50, 100, 250, 500, 1000). The second stage estimates are produced using the FGLS weight matrix ($W = \hat{\Omega}^{-1}$). For each of the sample sizes, the solid line represents the p.d.f. of the GLS R^2 statistic in the model without a useless factor, when the risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of the cross-sectional GLS R^2 when the model is estimated by the Fama-MacBeth procedure, including both the useful and the useless factor, while the dash-dotted line corresponds to the case of the Pen-FM estimator employed in the same scenario of the contaminated model.

For a detailed description of the simulation design for the misspecified model, please refer to Table II.

Figure 9 Distribution of the Hansen-Jagannathan distance in a misspecified model
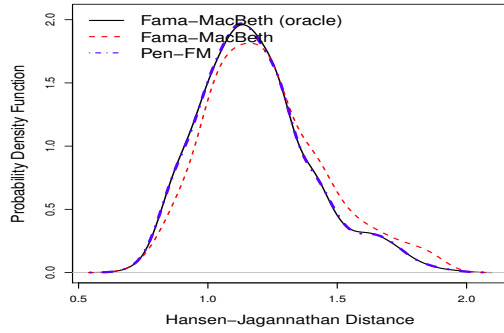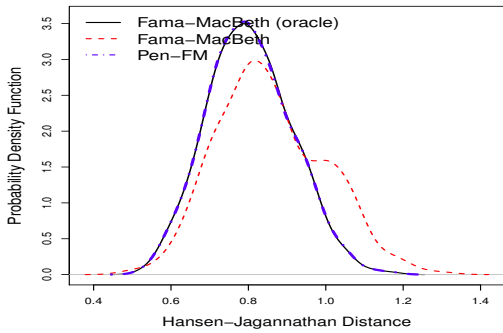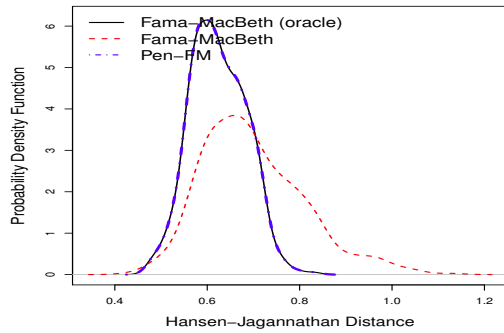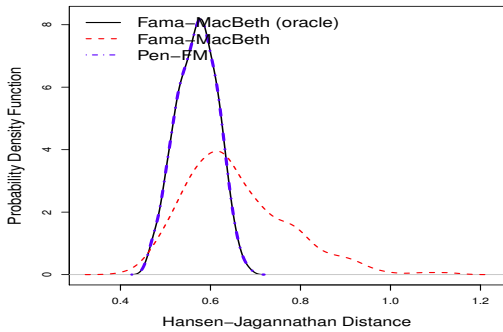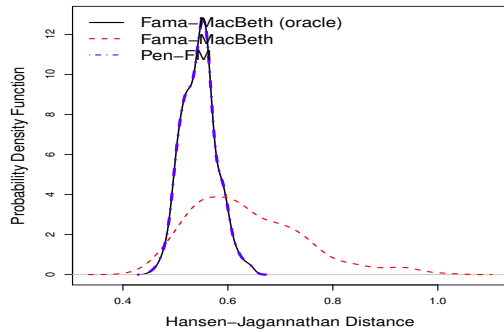
(a) T=30

(b) T=50

(c) T=100

(d) T=250

(e) T=500

(f) T=1000

Note. The graphs present the probability density function for the Hansen-Jagannathan distance (HJ) in a simulation of a misspecified model with omitted variable bias and further potentially contaminated by the presence of an irrelevant factor for various sample sizes (T=30, 50, 100, 250, 500, 1000). The second stage estimates are produced using an identity weight matrix. For each of the sample sizes, the solid line represents the p.d.f. of HJ in the model without a useless factor, when risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of HJ when the model is estimated by the Fama-MacBeth procedure, including both the useful and the useless factor, while the dash-dotted line corresponds to the case of the Pen-FM estimator employed in the same scenario of the contaminated model.

For a detailed description of the simulation design for the misspecified model, please refer to Table II.

Figure 10 Distribution of the average pricing error in a misspecified model
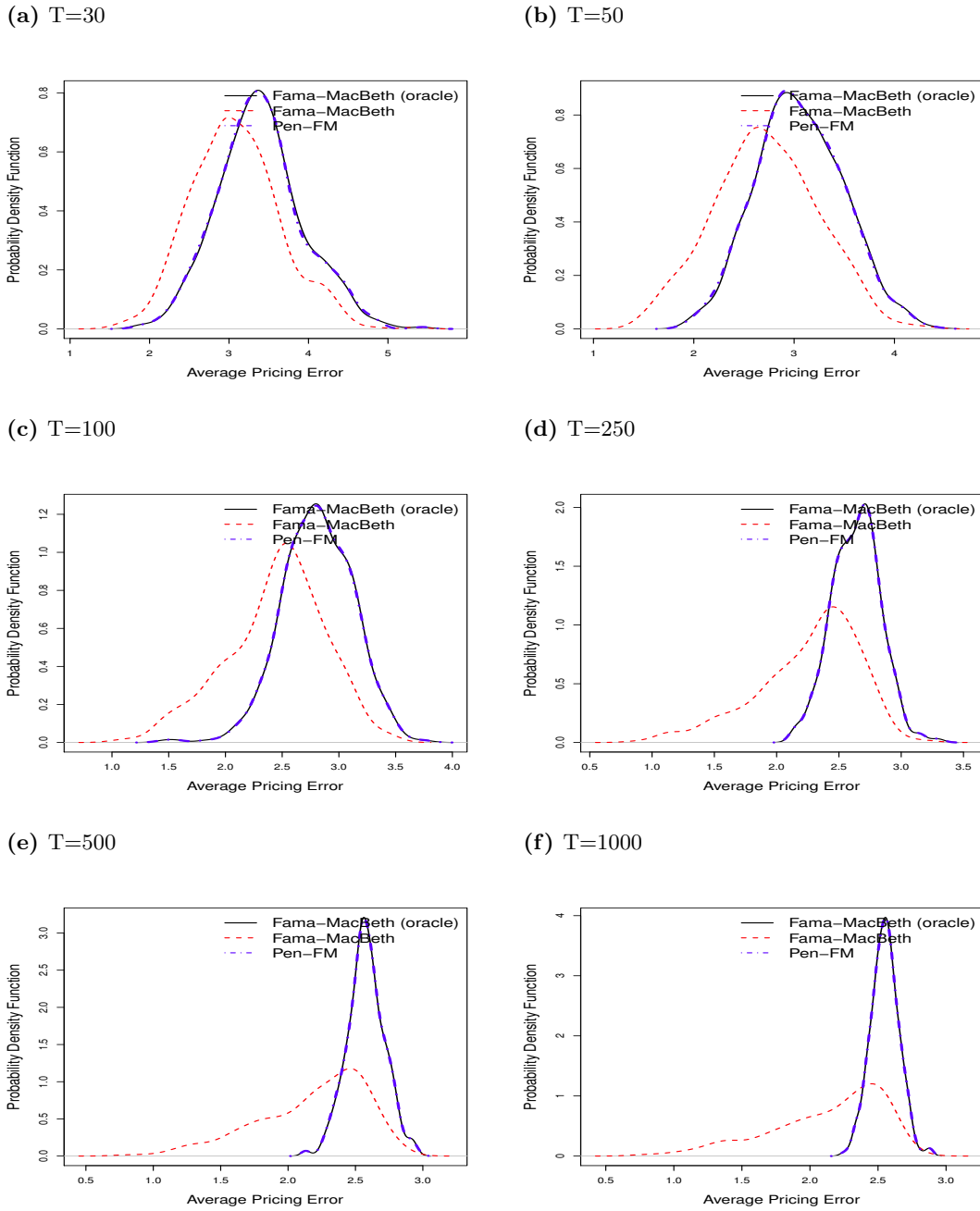
(a) T=30

(b) T=50

(c) T=100

(d) T=250

(e) T=500

(f) T=1000

Note. The graphs present the probability density function for the average pricing error (APE) in a simulation of a misspecified model with omitted variable bias and further potentially contaminated by the presence of an irrelevant factor for various sample sizes (T=30, 50, 100, 250, 500, 1000). The second stage estimates are produced using an identity weight matrix. For each of the sample sizes, the solid line represents the p.d.f. of APE in the model without a useless factor, when risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of the APE when the model is estimated by the Fama-MacBeth procedure, including both the useful and the useless factor, while the dash-dotted line corresponds to the case of the Pen-FM estimator employed in the same scenario of the contaminated model.

For a detailed description of the simulation design for the misspecified model, please refer to Table II.

Appendix A.

Table A1. Empirical size of the bootstrap-based confidence bounds in a correctly specified model

T	λ_0 10%	5%	1%	Useful factor, $\lambda_1 \neq 0$ 10%	5%	1%	Useless factor, $\lambda_2 = 0$ 10%	5%	1%
Panel A: Fama-MacBeth estimator in a model with only a useful factor									
30	0.065	0.029	0.003	0.033	0.009	0.000	-	-	-
50	0.075	0.037	0.009	0.047	0.016	0.001	-	-	-
100	0.096	0.055	0.015	0.062	0.026	0.005	-	-	-
250	0.103	0.049	0.009	0.054	0.023	0.003	-	-	-
500	0.106	0.057	0.008	0.059	0.027	0.003	-	-	-
1000	0.101	0.043	0.008	0.048	0.018	0.002	-	-	-
Panel B: Fama-MacBeth estimator in a model with a useful and a useless factor									
25	0.045	0.023	0.003	0.024	0.006	0.001	0.008	0.001	0
50	0.061	0.032	0.002	0.029	0.011	0.002	0.007	0.002	0
100	0.068	0.029	0.004	0.033	0.009	0.000	0.009	0.002	0
250	0.069	0.027	0.003	0.034	0.011	0.001	0.006	0.002	0
500	0.071	0.036	0.008	0.043	0.018	0.001	0.005	0.003	0.001
1000	0.063	0.028	0.007	0.037	0.012	0.006	0.005	0	0
Panel C: Pen-FM estimator in a model with a useful and a useless factor									
25	0.051	0.027	0.006	0.027	0.007	0.001	0.002	0	0
50	0.081	0.038	0.005	0.036	0.016	0.002	0	0	0
100	0.09	0.041	0.005	0.047	0.014	0.001	0	0	0
250	0.093	0.05	0.008	0.048	0.025	0.001	0	0	0
500	0.095	0.054	0.013	0.055	0.026	0.004	0	0	0
1000	0.097	0.042	0.01	0.061	0.021	0.007	0	0	0

Note. The table summarises the empirical size of the bootstrap-based confidence bounds for the Fama-MacBeth and Pen-FM estimators with the identity weight matrix in the second stage and at various significance levels (α=10%, 5%, 1%). The model includes a true risk factor and a useless one. λ_0 stands for the value of the intercept, λ_1 and λ_2 are the corresponding risk premia of the factors. Panel A corresponds to the case of the Fama-MacBeth estimator with an identity weight matrix, when the model includes only the useful factor. Panels B and C present the empirical size of the confidence bounds of the risk premia when the model includes both a useful and a useless factor, and the parameters are estimated by Fama-MacBeth or Pen-FM estimator accordingly. The model is simulated 10,000 times for different values of the sample size (T). The confidence bounds are constructed from 10,000 pairwise bootstrap replicas.

For a detailed description of the simulation design, please refer to Table I.

Table A2. Empirical size of the confidence bounds, based on the t-statistic in a correctly specified model

T	Intercept, λ_0			Useful factor, $\lambda_1 \neq 0$			Useless factor, $\lambda_2 = 0$		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Fama-MacBeth estimator in a model with only a useful factor,									
without Shanken correction									
25	0.1265	0.073	0.0345	0.127	0.07	0.0325	-	-	-
50	0.114	0.064	0.025	0.1155	0.057	0.0255	-	-	-
100	0.1115	0.058	0.0275	0.105	0.055	0.0285	-	-	-
250	0.107	0.051	0.019	0.1065	0.0575	0.0175	-	-	-
500	0.096	0.0465	0.0195	0.1025	0.052	0.021	-	-	-
1000	0.09	0.047	0.018	0.095	0.043	0.0175	-	-	-
Panel B: Fama-MacBeth estimator in a model with only a useful factor,									
with Shanken correction									
30	0.095	0.0435	0.011	0.035	0.0085	0.001	-	-	-
50	0.092	0.04	0.014	0.04	0.0115	0.003	-	-	-
100	0.097	0.0445	0.0185	0.051	0.02	0.003	-	-	-
250	0.1	0.0445	0.0145	0.0585	0.016	0.003	-	-	-
500	0.0925	0.045	0.0175	0.056	0.023	0.0055	-	-	-
1000	0.089	0.046	0.017	0.0495	0.019	0.007	-	-	-
Panel C: Fama-MacBeth estimator in a model with a useless factor,									
without Shanken correction									
30	0.1325	0.076	0.035	0.1305	0.0725	0.037	0.123	0.0695	0.034
50	0.117	0.0625	0.0255	0.1225	0.062	0.028	0.1115	0.0595	0.029
100	0.1115	0.0565	0.0245	0.1065	0.053	0.025	0.106	0.0505	0.0205
250	0.101	0.05	0.0195	0.099	0.051	0.02	0.11	0.049	0.0195
500	0.1075	0.0495	0.021	0.111	0.0515	0.0225	0.0935	0.048	0.023
1000	0.089	0.0485	0.0145	0.09	0.0485	0.0175	0.113	0.058	0.021
Panel D: Fama-MacBeth estimator in a model with a useless factor,									
with Shanken correction									
30	0.0875	0.0355	0.011	0.031	0.0055	0.001	0.0285	0.0055	0
50	0.074	0.0365	0.0085	0.033	0.007	0.002	0.033	0.0085	0.0015
100	0.0765	0.033	0.0125	0.037	0.015	0.0035	0.024	0.0065	0.0005
250	0.0755	0.0305	0.0085	0.0435	0.0145	0.0025	0.034	0.005	0
500	0.0835	0.0345	0.01	0.0445	0.0165	0.0025	0.0305	0.0075	0.0005
1000	0.0675	0.0335	0.0095	0.041	0.013	0.005	0.0365	0.0075	0

Note. The table presents the empirical size of the t-statistic-based confidence bounds for the Fama-MacBeth estimator with an identity weight matrix in a model with a common intercept for 25 portfolios and a single risk factor, with or without a useless one. λ_0 is the value of the intercept; λ_1 and λ_2 are the corresponding risk premia of the factors. The model is simulated 10,000 times for different values of the sample size (T). Panels A and C present the size of the t-statistic, computed using OLS-based heteroscedasticity-robust standard errors. Panels B and D present results based on Shanken correction.

For a detailed description of the simulation design, please refer to Table I.

Table A3. Empirical size of the bootstrap-based confidence bounds for true values in a misspecified model

T	λ_0 10%	5%	1%	Useful factor, $\lambda_1 \neq 0$ 10%	5%	1%	Useless factor, $\lambda_2 = 0$ 10%	5%	1%
Panel A: Fama-MacBeth estimator in a model with only a useful factor								
25	0.004	0	0	0.003	0	0	-	-	-
50	0.005	0	0	0.003	0.001	0	-	-	-
100	0.004	0.001	0	0.005	0.002	0	-	-	-
250	0.001	0	0	0.009	0.001	0	-	-	-
500	0	0	0	0.05	0.005	0	-	-	-
1000	0.002	0.001	0	0.183	0.049	0.005	-	-	-
Panel B: Fama-MacBeth estimator in a model with a useful and a useless factor								
25	0.01	0.003	0	0.002	0.001	0	0.002	0	0
50	0.005	0	0	0.002	0	0	0.011	0.002	0
100	0.003	0.001	0	0.002	0.001	0	0.055	0.02	0.001
250	0.004	0.002	0	0.005	0.001	0	0.093	0.052	0.014
500	0.002	0	0	0.01	0.001	0	0.088	0.05	0.01
1000	0.003	0.001	0	0.019	0.005	0	0.122	0.066	0.019
Panel C: Pen-FM estimator in a model with a useful and a useless factor								
25	0.007	0.001	0	0.004	0.001	0	0	0	0
50	0.002	0	0	0.002	0	0	0	0	0
100	0.003	0	0	0.003	0	0	0	0	0
250	0.002	0.001	0	0.009	0	0	0	0	0
500	0.004	0.001	0	0.055	0.012	0	0	0	0
1000	0.002	0	0	0.161	0.046	0.002	0	0	0

Note. The table summarises the empirical size of the bootstrap-based confidence bounds for the Fama-MacBeth and Pen-FM estimators with an identity weight matrix in the second stage and at various significance levels (α=10%, 5%, 1%). The misspecified model includes only 1 out of 3 true risk factors, and is further contaminated by the presence of a useless one. λ_0 stands for the value of the intercept; λ_1 and λ_2 are the corresponding risk premia of the factors. Panel A corresponds to the case of the Fama-MacBeth estimator with an identity weight matrix, when the model includes only one useful factor. Panels B and C present empirical size of the confidence bounds of the risk premia when the model includes both a useful and a useless factor, and their parameters are estimated by the Fama-MacBeth or Pen-FM procedures accordingly. The model is simulated 10 000 times for different values of the sample size (T). The confidence bounds are constructed from 10 000 pairwise bootstrap replicas.

For a detailed description of the simulation design for the misspecified model, please refer to Table II.

Table A4. Empirical size of the confidence bounds for the true values of the risk premia, based on the t-statistic in a mispecified model

T	Intercept, λ_0			Useful factor, $\lambda_1 \neq 0$			Useless factor, $\lambda_2 = 0$		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Fama-MacBeth estimator in a model with only a useful factor,									
without Shanken correction									
30	0.0145	0.004	0.001	0.02	0.006	0.002	-	-	-
50	0.01	0.003	0.0015	0.015	0.004	0.001	-	-	-
100	0.0035	0	0	0.015	0.002	0	-	-	-
250	0.0025	0.001	0	0.0555	0.014	0.002	-	-	-
500	0.004	0.0015	0	0.1535	0.051	0.009	-	-	-
1000	0.0035	0	0	0.408	0.206	0.0785	-	-	-
Panel B: Fama-MacBeth estimator in a model with only a useful factor,									
with Shanken correction									
30	0.011	0.0015	5e-04	0.002	0	0	-	-	-
50	0.0065	0.002	0.001	0.003	0.001	0	-	-	-
100	0.0015	0	0	0.001	0	0	-	-	-
250	0.0025	0.001	0	0.014	0.0015	0	-	-	-
500	0.004	0.0015	0	0.0585	0.0115	0.002	-	-	-
1000	0.003	0	0	0.238	0.086	0.0115	-	-	-
Panel C: Fama-MacBeth estimator in a model with a useless factor,									
without Shanken correction									
25	0.0355	0.015	0.002	0.0435	0.016	0.003	0.135	0.055	0.016
50	0.0435	0.0175	0.0055	0.0555	0.022	0.007	0.2885	0.139	0.0465
100	0.0745	0.0375	0.015	0.0935	0.0465	0.02	0.5945	0.441	0.2375
250	0.178	0.1115	0.0605	0.2325	0.154	0.084	0.805	0.7595	0.696
500	0.301	0.229	0.162	0.3845	0.2935	0.202	0.872	0.8425	0.806
1000	0.4095	0.347	0.2815	0.5355	0.451	0.3625	0.932	0.915	0.8885
Panel D: Fama-MacBeth estimator in a model with a useless factor,									
with Shanken correction									
25	0.018	0.004	0.0015	0.003	0.001	0	0.0185	0.003	0.001
50	0.0165	0.005	0.0015	0.007	0.001	0	0.054	0.0065	5e-04
100	0.0185	0.0065	0.0015	0.0155	0.003	0.0005	0.249	0.044	0.002
250	0.0305	0.008	0.0015	0.0375	0.011	0.0015	0.672	0.3305	0.051
500	0.0445	0.018	0.006	0.0725	0.027	0.0095	0.8105	0.5845	0.2095
1000	0.056	0.0235	0.011	0.124	0.0595	0.025	0.8895	0.731	0.3965

Note. The table presents the empirical size of the t-statistic-based confidence bounds for the true risk premia values for the Fama-MacBeth estimator with the identity weight matrix in a model with a common intercept for 25 portfolios and a single risk factor, with or without a useless one. λ_0 is the value of the intercept, λ_1 and λ_2 are the corresponding risk premia of the factors. The model is simulated 10,000 times for different values of the sample size (T). Panels A and C present the size of the t-statistic computed using heteroscedasticity-robust standard errors. Panels B and D present the results based on Shanken correction.

For a detailed description of the simulation design for the misspecified model, please refer to Table II.

Table A5. Empirical size of the bootstrap-based confidence bounds for the pseudo-true values in a misspecified model

T	λ_0 10%	5%	1%	Useful factor, $\lambda_1 \neq 0$ 10%	5%	1%	Useless factor, $\lambda_2 = 0$ 10%	5%	1%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Fama-MacBeth estimator in a model with only a useful factor									
25	0.004	0	0	0	0	0	-	-	-
50	0.005	0	0	0	0	0	-	-	-
100	0.004	0.001	0	0.001	0	0	-	-	-
250	0	0	0	0	0	0	-	-	-
500	0	0	0	0	0	0	-	-	-
1000	0.001	0	0	0	0	0			
Panel B: Fama-MacBeth estimator in a model with a useful and a useless factor									
25	0.01	0.003	0	0.001	0	0	0.002	0	0
50	0.001	0	0	0	0	0	0.011	0.002	0
100	0.003	0.001	0	0	0	0	0.055	0.02	0.001
250	0.004	0	0	0.001	0	0	0.093	0.052	0.014
500	0.002	0	0	0.001	0	0	0.088	0.05	0.01
1000	0.004	0.001	0	0.003	0	0	0.122	0.066	0.019
Panel C: Pen-FM estimator in a model with a useful and a useless factor									
25	0.007	0.001	0	0.001	0	0	0	0	0
50	0.002	0	0	0	0	0	0	0	0
100	0.003	0	0	0	0	0	0	0	0
250	0.002	0.001	0	0.001	0	0	0	0	0
500	0.003	0.001	0	0	0	0	0	0	0
1000	0.002	0	0	0	0	0	0	0	0

Note. The table summarises the empirical size of the bootstrap-based confidence bounds for the Fama-MacBeth and Pen-FM estimators with an identity weight matrix at the second stage and various significance levels (α=10%, 5%, 1%). The misspecified model includes only 1 out of 3 true risk factors, and is further contaminated by the presence of a useless one. λ_0 stands for the value of the intercept; λ_1 and λ_2 are the corresponding risk premia of the factors. The pseudo-true values of the risk premia are defined as the limit of the risk premia estimates in a misspecified model without the influence of the useless factor. Panel A corresponds to the case of the Fama-MacBeth estimator with an identity weight matrix, when the model includes only one useful factor. Panels B and C present the empirical size of the confidence bounds of risk premia when the model includes both a useful and a useless factor, and their parameters are estimated by the Fama-MacBeth or Pen-FM procedures accordingly. The model is simulated 10 000 times for different values of the sample size (T). The confidence bounds are constructed from 10 000 pairwise bootstrap replicas.

For a detailed description of the simulation design for the misspecified model, please refer to Table II.

Table A6. Empirical size of the confidence bounds for the pseudo-true values of risk premia, based on the t-statistic in a mispecified model

T	Intercept, λ_0			Useful factor, $\lambda_1 \neq 0$			Useless factor, $\lambda_2 = 0$		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Fama-MacBeth estimator in a model with only a useful factor,									
without Shanken correction									
30	0.015	0.004	0.001	0.0085	0.002	0	-	-	-
50	0.0085	0.003	0.0015	0.007	0.002	0.0005	-	-	-
100	0.003	0	0	0.0015	0	0	-	-	-
250	0.0025	0.001	0	0.002	0	0	-	-	-
500	0.0045	0.0015	0	0.0035	5e-04	0	-	-	-
1000	0.003	0	0	0.0005	0	0	-	-	-
Panel B: Fama-MacBeth estimator in a model with only a useful factor,									
with Shanken correction									
30	0.01	0.0015	0.0005	0	0	0	-	-	-
50	0.0055	0.0015	0.001	0.0015	0	0	-	-	-
100	0.0015	0	0	0	0	0	-	-	-
250	0.002	0.001	0	0	0	0	-	-	-
500	0.004	0.0015	0	0.001	0	0	-	-	-
1000	0.002	0	0	0	0	0	-	-	-
Panel C: Fama-MacBeth estimator in a model with a useless factor,									
without Shanken correction									
30	0.0345	0.015	0.002	0.0225	0.0055	0.0015	0.135	0.055	0.016
50	0.04	0.017	0.0045	0.0315	0.01	0.004	0.2885	0.139	0.0465
100	0.0715	0.036	0.0145	0.06	0.027	0.01	0.5945	0.441	0.2375
250	0.18	0.11	0.0595	0.168	0.1015	0.0515	0.805	0.7595	0.696
500	0.305	0.2245	0.1605	0.297	0.2275	0.158	0.872	0.8425	0.806
1000	0.405	0.342	0.283	0.4145	0.347	0.28	0.932	0.915	0.8885
Panel D: Fama-MacBeth estimator in a model with a useless factor,									
with Shanken correction									
30	0.0175	0.0035	0.0005	0.001	0	0	0.0185	0.003	0.001
50	0.015	0.0055	0.0015	0.005	0.001	0	0.054	0.0065	0.0005
100	0.0175	0.0065	0.001	0.0055	0.001	0	0.249	0.044	0.002
250	0.028	0.0065	0.0015	0.011	0.002	0	0.672	0.3305	0.051
500	0.0445	0.018	0.006	0.0305	0.007	0.0015	0.8105	0.5845	0.2095
1000	0.055	0.022	0.009	0.0445	0.0155	0.0015	0.8895	0.731	0.3965

Note. The table summarises the empirical size of the *t*-statistic-based confidence bounds for the Fama-MacBeth and Pen-FM estimators with an identity weight matrix at the second stage and at various significance levels (α=10%, 5%, 1%). The misspecified model includes only 1 out of 3 true risk factors, and is further contaminated by the presence of a useless factor. λ_0 stands for the value of the common intercept; λ_1 and λ_2 are the corresponding risk premia of the factors. The pseudo-true values of the risk premia are defined as the limit of the risk premia estimates in a misspecified model without the influence of the useless factor. Panels A and C present the size of the t-statistic confidence bounds, computed using OLS-based heteroscedasticity-robust standard errors that do not take into account the error-in-variables problem of the second stage. The model is estimated with/without the useless factor. Panels B and D present similar results for the case of Shanken correction. The model is simulated 10,000 times for different values of the sample size (T).

For a detailed description of the simulation design for the misspecified model, please refer to Table II.

Figure A1 Distribution of the T^2 statistic in a correctly specified model
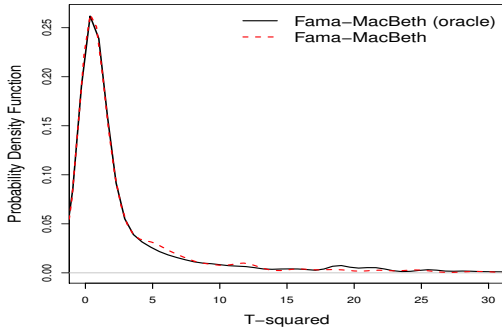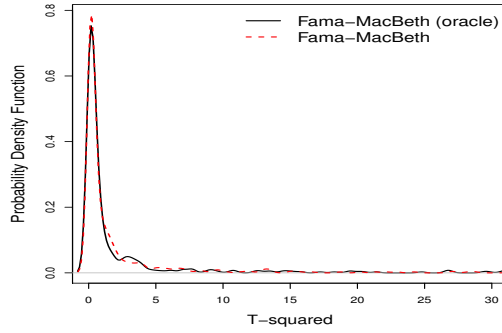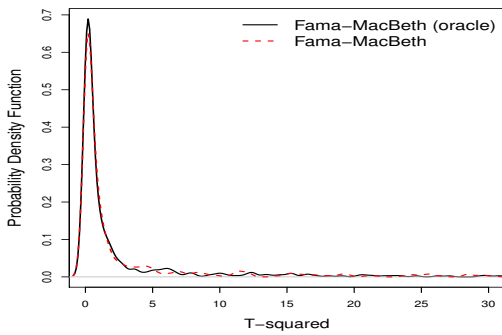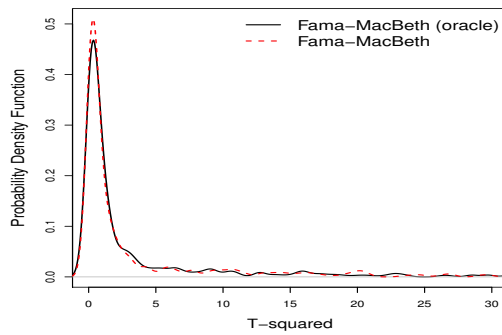
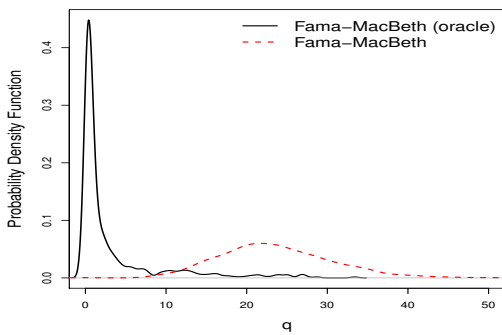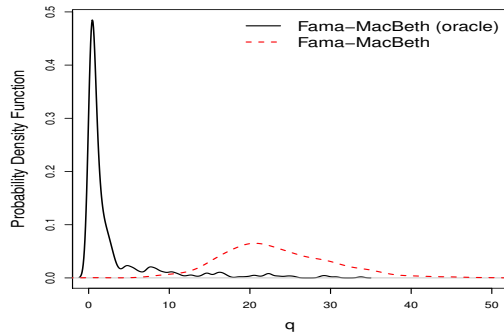**(a)** T=30

(b) T=50

(c) T=100

(d) T=250

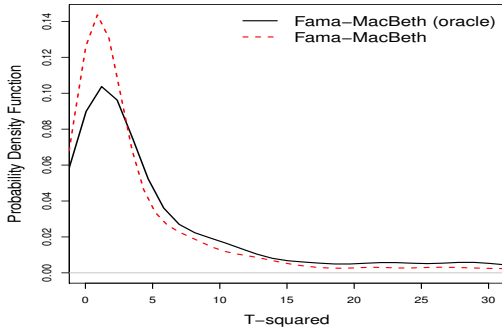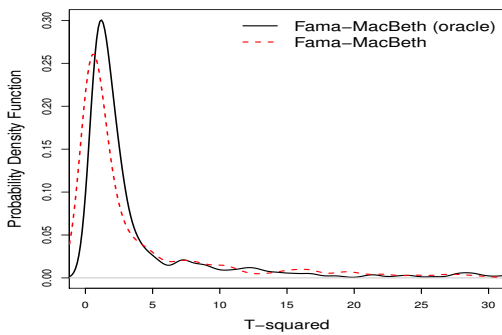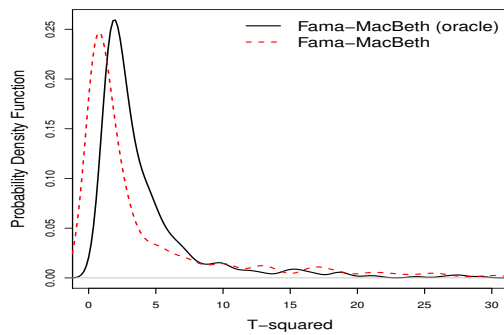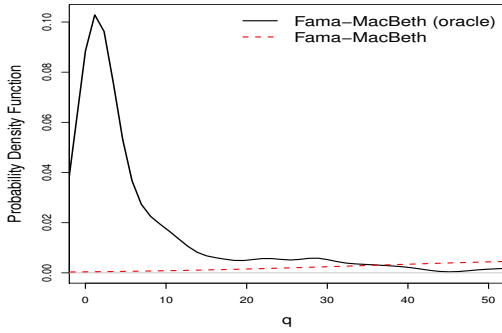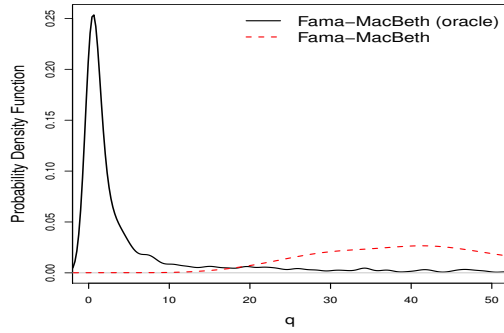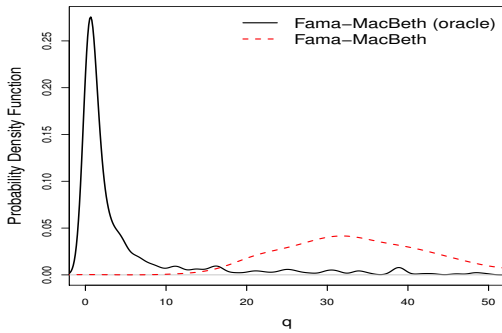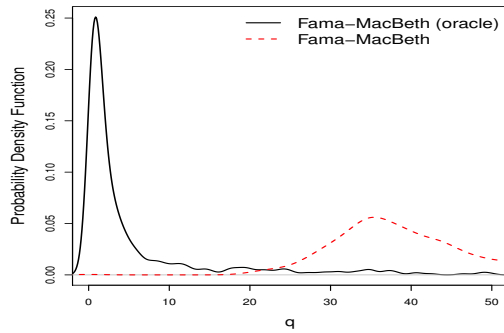**(e)** T=500
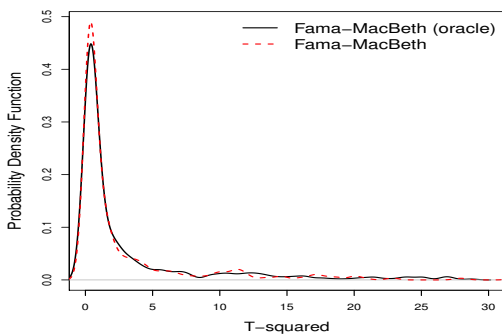
(f) T=1000
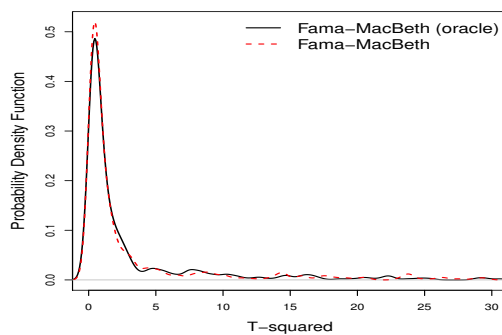
Note. The graphs present the probability density function for the T^2 statistic in the simulations of a correctly specified model, potentially contaminated by the presence of an irrelevant factor, and the risk premia estimated using an identity weight matrix in the second stage ($W = I_n$). For each of the sample sizes (T=30, 50, 100, 250, 500, 1000), the solid line represents the p.d.f. of the T^2 in the model without a useless factor, when risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of T^2 when the model is estimated by the Fama-MacBeth procedure in the presence of a useless factor.

For a detailed description of the simulation design, please refer to Table I.

Figure A2 Distribution of q-statistic in a correctly specified model

(a) T=30
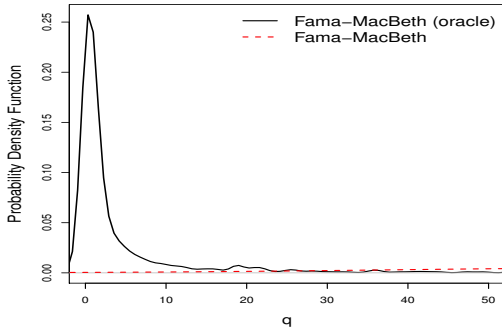
(b) T=50
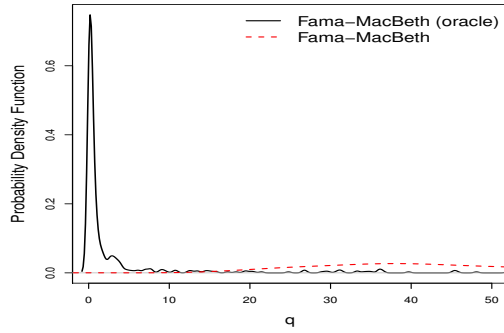
(c) T=100
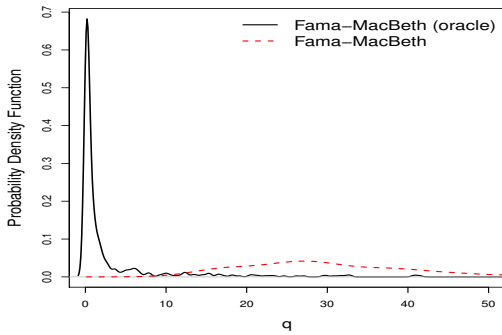
(d) T=250
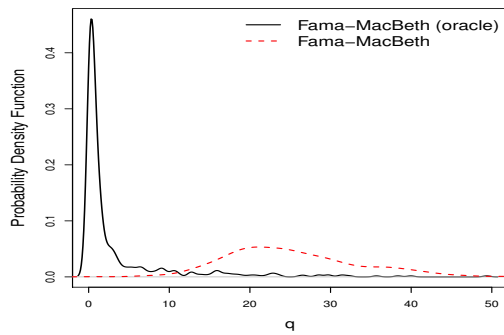
(e) T=500

(f) T=1000

Note. The graphs present the probability density function of the q-statistic in the simulations of a correctly specified model, potentially contaminated by the presence of an irrelevant factor, and the risk premia estimated using an identity weight matrix in the second stage ($W = I_n$). For each of the sample sizes (T=30, 50, 100, 250, 500, 1000), the solid line represents the p.d.f. of q in the model without a useless factor, when risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), and the dashed line depicts the distribution of q when the model is estimated by the Fama-MacBeth procedure under the presence of a useless factor.

For a detailed description of the simulation design, please refer to Table I.

Figure A3 Distribution of the T^2-statistic in a misspecified model
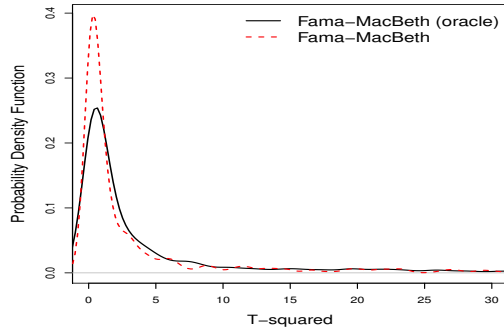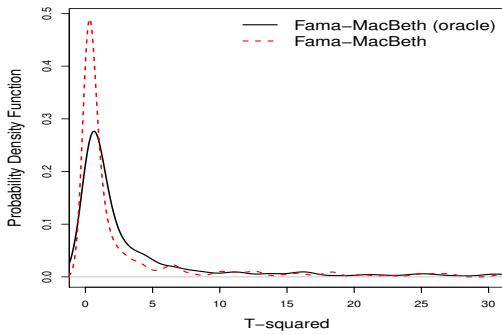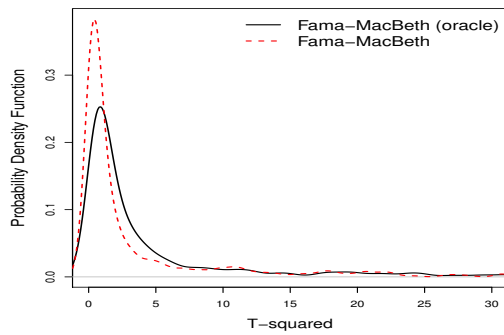
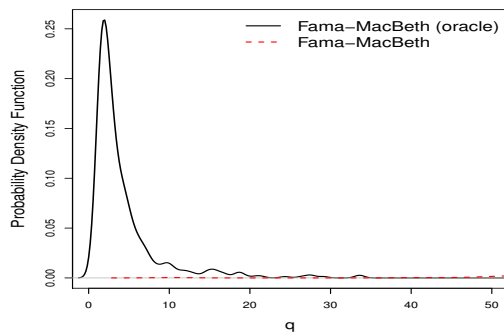**(a)** T=30

(b) T=50

(c) T=100

(d) T=250

(e) T=500

(f) T=1000

Note. The graphs present the probability density function for the T^2-statistic in a simulation of a misspecified model, potentially contaminated by the presence of an irrelevant factor for various sample sizes (T=30, 50, 100, 250, 500, 1000). The second stage estimates are produced using an identity weight matrix. For each of the sample sizes, the solid line represents p.d.f. of T^2 in the model without a useless factor, when risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of T^2 when the model is estimated by the Fama-MacBeth procedure, including both the useful and the useless factor.

For a detailed description of the simulation design for the misspecified model, please refer to Table II.

Figure A4 Distribution of the q-statistic in a misspecified model

(a) T=30

(b) T=50

(c) T=100

(d) T=250

(e) T=500
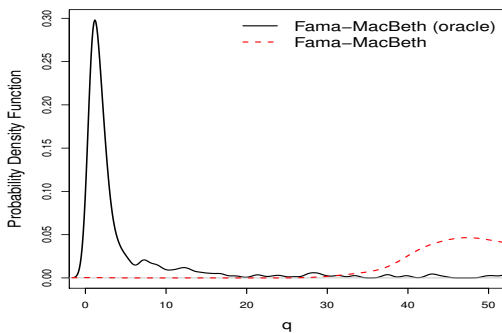
(f) T=1000

Note. The graphs present the probability density function for q-statistic in a simulation of a misspecified model, potentially contaminated by the presence of an irrelevant factor for various sample sizes (T=30, 50, 100, 250, 500, 1000). The second stage estimates are produced using an identity weight matrix. For each of the sample sizes, the solid line represents the p.d.f. of q in the model without a useless factor, when risk premia are estimated by the Fama-MacBeth estimator (the *oracle* case), the dashed line depicts the distribution of q when the model is estimated by the Fama-MacBeth procedure, including both the useful and the useless factor.

For a detailed description of the simulation design for the misspecified model, please refer to Table II.

Appendix B

B.1. Proof of Proposition 1

Proof. Consider the quadratics in the objective function.

$$\left[\bar{R} - \hat{\beta}\lambda\right]' W_T \left[\bar{R} - \hat{\beta}\lambda\right] \xrightarrow{p} [E[R] - \beta_{ns}\lambda_{ns}]'W[E[R] - \beta_{ns}\lambda_{ns}]$$

For the strong factors that have substantial covariance with asset returns (whether their risk is priced or not), $\eta_T \frac{1}{\|\hat{\beta}_j\|_1^d} \overset{a}{\sim} \eta T^{-d/2} O_p(1) \xrightarrow{d} 0$, where $\overset{a}{\sim}$ denotes equivalence of the asymptotic expansion up to $o_p\left(\frac{1}{\sqrt{T}}\right)$. For the useless factors we have $\eta_T \frac{1}{\|\hat{\beta}_j\|_1^d} \overset{a}{\sim} \eta T^{-d/2} c_j T^{d/2} \xrightarrow{d} \tilde{c}_j > 0$. Therefore, in the limit the objective function becomes the following convex function of λ:

$$[E[R] - \beta_{ns}\lambda_{ns}]'W[E[R] - \beta_{ns}\lambda_{ns}] + \sum_{j=1}^{k} \tilde{c}_j |\lambda_j| \mathbb{1}\{\beta_j = 0\}$$

Since c_j are some positive constants,

$$\mathbf{0} = \underset{\lambda_{sp} \in \Theta_{sp}}{\arg\min}[E[R] - \beta_{ns}\lambda_{ns}]'W[E[R] - \beta_{ns}\lambda_{ns}] + \sum_{j=1}^{k} \tilde{c}_j |\lambda_j| \mathbb{1}\{\beta_j = 0\}$$

The risk premia for the strong factors are still identified, as

$$\lambda_{0,ns} = \underset{\lambda_{ns} \in \Theta_{ns}}{\arg\min} \ [E[R] - \beta_{ns}\lambda_{ns}]'W[E[R] - \beta_{ns}\lambda_{ns}] = (\beta_{ns}'W\beta_{ns})^{-1}\beta_{ns}'WE[R] = (\beta_{ns}'W\beta_{ns})^{-1}\beta_{ns}'W\beta_{ns}\lambda_{0,ns}$$

By the convexity lemma of Pollard (1991), the estimator is consistent.

To establish asymptotic normality, it is first instructive to show the distribution of the usual Fama-McBeth estimator in the absence of identification failure.

Following Lemma 1, the first stage estimates have the following asymptotic representations

$$\hat{\beta}_{ns} = \beta_{ns} + \frac{1}{\sqrt{T}}\Psi_{\beta,ns} + o_p\left(\frac{1}{\sqrt{T}}\right), \quad \bar{R} = \beta_{ns}\lambda_{0,ns} + \frac{B_{sp}}{\sqrt{T}}\lambda_{0,sp} + \frac{1}{\sqrt{T}}\psi_R + o_p\left(\frac{1}{\sqrt{T}}\right)$$

where $\Psi_{\beta,ns} = vecinv(\psi_{\beta,ns})$ and $vecinv$ is the inverse of the vectorisation operator.

Consider the WLS estimator of the cross-section regression:

$$\hat{\lambda}_{ns} = \left(\hat{\beta}_{ns}'W_T\hat{\beta}_{ns}\right)^{-1}\hat{\beta}_{ns}'W_T\bar{R} \overset{a}{=} \left(\hat{\beta}_{ns}'W_T\hat{\beta}_{ns}\right)^{-1}\hat{\beta}_{ns}'\left(\beta_{ns}\lambda_{0,ns} + \frac{1}{\sqrt{T}}\psi_R\right)$$

$$= \left(\hat{\beta}_{ns}'W_T\hat{\beta}_{ns}\right)^{-1}\hat{\beta}_{ns}'W_T\left(\hat{\beta}_{ns}\lambda_{0,ns} + (\beta_{ns} - \hat{\beta}_{ns})\lambda_{0,ns} + \frac{1}{\sqrt{T}}\psi_R\right) =$$

$$= \lambda_{0,ns} + \left(\hat{\beta}_{ns}'W_T\hat{\beta}_{ns}\right)^{-1}\hat{\beta}_{ns}'W_T(\hat{\beta}_{ns} - \beta_{ns})\lambda_{0,ns} + \left(\hat{\beta}_{ns}'W_T\hat{\beta}_{ns}\right)^{-1}\hat{\beta}_{ns}'W_T\frac{1}{\sqrt{T}}\psi_R$$

Finally, since as $T \to \infty$

$$\hat{\beta}_{ns}'W_T\hat{\beta}_{ns} \overset{a}{=} \left[\beta_{ns} + \frac{1}{\sqrt{T}}\Psi_{\beta,ns}\right]' W_T \left[\beta_{ns} + \frac{1}{\sqrt{T}}\Psi_{\beta,ns}\right] \xrightarrow{p} \beta_{ns}'W\beta_{ns}$$

$$\hat{\beta}_{ns} - \beta_{ns} \overset{a}{=} -\frac{1}{\sqrt{T}}\Psi_{\beta,ns} = \frac{1}{\sqrt{T}}\Psi_{\beta,ns}$$

75

it follows that

$$\sqrt{T}(\hat{\lambda}_{ns} - \lambda_{0,ns}) \xrightarrow{d} [\beta'_{ns} W \beta_{ns}]^{-1} \beta'_{ns} W \Psi_{\beta,ns} \lambda_{0,ns} + (\beta'_1 W \beta_1)^{-1} \beta'_{ns} W \psi_R$$

In order to demonstrate the asymptotic distribution of the shrinkage-based estimator, I reformulate the objective function in terms of the centred parameters $u = \frac{\lambda - \lambda_0}{\sqrt{T}}$:

$$L_T(u) = \left[\bar{R} - \hat{\beta}\left(\lambda_0 + \frac{u}{\sqrt{T}}\right)\right]' W_T \left[\bar{R} - \hat{\beta}\left(\lambda_0 + \frac{u}{\sqrt{T}}\right)\right] + \eta_T \sum_{j=1}^{k} \frac{1}{\left\|\hat{\beta}^j\right\|_1^d} \left|\lambda_{0j} + \frac{u}{\sqrt{T}}\right|$$

Solving the original problem in 9 w.r.t. λ is the same as optimizing $\bar{L}(u) = T(L_T(u) - L_T(0))$ w.r.t. u. Since

$$\left[\bar{R} - \hat{\beta}\left(\lambda_0 + \frac{u}{\sqrt{T}}\right)\right]' W_T \left[\bar{R} - \hat{\beta}\left(\lambda_0 + \frac{u}{\sqrt{T}}\right)\right] =$$

$$= \bar{R}' W_T \bar{R} + \left[\lambda_0 + \frac{u}{\sqrt{T}}\right]' \hat{\beta}' W_T \hat{\beta} \left[\lambda_0 + \frac{u}{\sqrt{T}}\right] - 2\left[\lambda_0 + \frac{u}{\sqrt{T}}\right]' \hat{\beta}' W_T \bar{R}$$

$$\lambda_0' \hat{\beta}' W_T \hat{\beta} \lambda_0 + \frac{u'}{\sqrt{T}} \hat{\beta}' W_T \hat{\beta} \frac{u}{\sqrt{T}} + \frac{2}{\sqrt{T}} u' \hat{\beta}' W_T \hat{\beta} \lambda_0 - 2\left[\lambda_0 + \frac{u}{\sqrt{T}}\right]' \hat{\beta}' W_T \bar{R}$$

$$= \lambda_0' \hat{\beta}' W_T \hat{\beta} \lambda_0 + \frac{u'}{\sqrt{T}} \hat{\beta}' W_T \hat{\beta} \frac{u}{\sqrt{T}} - 2\lambda_0' \hat{\beta}' W_T \bar{R} + \frac{2}{\sqrt{T}} u' \hat{\beta}' W_T (\hat{\beta} \lambda_0 - \bar{R})$$

Therefore, in localized parameters u the problem looks as follows:

$$\hat{u} = \arg\min_{u \in K} \ u' \hat{\beta}' W_T \hat{\beta} u + 2\sqrt{T} u' \hat{\beta}' W_T (\hat{\beta} \lambda_0 - \bar{R}) + T\eta_T \sum_{j=1}^{k} \frac{1}{\left\|\hat{\beta}_j\right\|_1^d} \left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right]$$

$$= \arg\min_{u \in K} \ u' \hat{\beta}' W_T \hat{\beta} u + 2\sqrt{T} \hat{\beta}' W_T (\hat{\beta} - \beta) \lambda_0 - 2u' \hat{\beta}' W_T \varphi_R +$$

$$+ T\eta_T \sum_{j=1}^{k} \frac{1}{\left\|\hat{\beta}_j\right\|_1^d} \left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right]$$

where K is a compact set in $\mathbb{R}^k$.

It is easy to show that since as $t \to \infty$

$$\hat{\beta}' W_T \hat{\beta} \stackrel{a}{=} \begin{bmatrix} \beta'_{ns} + \frac{1}{\sqrt{T}} \Psi'_{\beta,ns} \\ \frac{1}{\sqrt{T}} \Psi'_{\beta,sp} \end{bmatrix} W_T \begin{bmatrix} \beta_{ns} + \frac{1}{\sqrt{T}} \Psi_{\beta,ns} & \frac{1}{\sqrt{T}} \Psi_{\beta,sp} \end{bmatrix} \stackrel{a}{=} \begin{bmatrix} \beta'_{ns} W \beta_{ns} & 0 \\ 0 & 0 \end{bmatrix}$$

the following identities hold:

$$u' \hat{\beta}' W_T \hat{\beta} u \stackrel{a}{=} \begin{bmatrix} u'_{ns} & u'_{sp} \end{bmatrix} \begin{bmatrix} \beta'_{ns} W \beta_{ns} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{ns} \\ u_{sp} \end{bmatrix} = u'_{ns} [\beta'_{ns} W \beta_{ns}] u_{ns} ,$$

$$u' \hat{\beta}' W_T (\hat{\beta} - \beta) \lambda_0 \stackrel{a}{=} \begin{bmatrix} u'_{ns} & u'_{sp} \end{bmatrix} \begin{bmatrix} \beta'_{ns} + \frac{1}{\sqrt{T}} \Psi'_{\beta,ns} \\ \frac{1}{\sqrt{T}} \Psi'_{\beta,sp} \end{bmatrix} W \begin{bmatrix} \frac{1}{\sqrt{T}} \Psi_{\beta,ns} & \frac{1}{\sqrt{T}} \Psi_{\beta,sp} \end{bmatrix} \begin{bmatrix} \lambda_{0,ns} \\ 0 \end{bmatrix} \stackrel{a}{=}$$

$$= \begin{bmatrix} u'_{ns} & u'_{sp} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{T}} \beta'_{ns} W \Psi_{\beta,ns} & \frac{1}{\sqrt{T}} \beta'_{ns} W \Psi_{\beta,sp} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_{0,ns} \\ 0 \end{bmatrix} = \frac{1}{\sqrt{T}} u'_{ns} \beta'_{ns} W \Psi_{\beta,ns} \lambda_{0,ns} ,$$

$$u' \hat{\beta}' W_T \varphi_R \stackrel{a}{=} \begin{bmatrix} u'_{ns} & u'_{sp} \end{bmatrix} \begin{bmatrix} \beta'_{ns} + \frac{1}{\sqrt{T}} \Psi'_{\beta,ns} \\ \frac{1}{\sqrt{T}} \Psi'_{\beta,sp} \end{bmatrix} W \varphi_R .$$

Finally, this implies that the overall objective function asymptotically looks as follows:

$$\bar{L}_T(u) \overset{a}{=} u'_{ns}\left[\beta'_{ns}W\beta_{ns}\right]u_{ns} + 2u'_{ns}\beta'_{ns}W\Psi_{\beta,ns}\lambda_{0,ns} - 2u'_{ns}(\beta_{ns} + \frac{1}{\sqrt{T}}\Psi_{\beta,ns})'W\varphi_R$$

$$-\frac{2}{\sqrt{T}}u'_{sp}\Psi'_{\beta,sp}W\varphi_R + T\eta_T\sum_{j=1}^{k}\frac{1}{\left\|\hat{\beta}^j\right\|_1^d}\left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right] \overset{a}{=}$$

$$= u'_{ns}\beta'_{ns}W\beta_{ns}u_{ns} - 2u'_{ns}\beta'_{ns}W(\varphi_R - \Psi_{\beta,ns}\lambda_{0,ns}) + T\eta_T\sum_{j=1}^{k}\frac{1}{\left\|\hat{\beta}^j\right\|_1^d}\left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right]$$

Now, for a spurious factor: $T\eta_T\frac{1}{\left\|\hat{\beta}_j\right\|_1^d}\left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right] = \sqrt{T}\eta T^{-d/2}c_j T^{d/2}|u_j| = \sqrt{T}\tilde{c}_j|u_j|$,
while for the strong ones: $T\eta_T\frac{1}{\left\|\hat{\beta}_j\right\|_1^d}\left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right] = c_j\sqrt{T}T^{-d/2}u_j sgn(\lambda_{0j}) \to 0$, since $d > 2$.
Therefore, as $T \to \infty$, $\bar{L}_T(u) \overset{d}{\to} \bar{L}_n$ for every u, where

$$\bar{L}(u) = \begin{cases} -u'_{ns}\beta'_{ns}W\beta_{ns}u_{ns} - 2u'_{ns}W\beta_{ns}(\phi_R - \Psi_{\beta,ns}\lambda_{0,ns}) & \text{if } u_{sp} = 0 \\ \infty & \text{otherwise} \end{cases}$$

Note that $\bar{L}_T(u)$ is a convex function with a unique optimum given by

$$\left([\beta'_{ns}W\beta_{ns}]^{-1}\beta'_{ns}W\Psi_{\beta,ns}\lambda_{0,ns} + [\beta'_{ns}W\beta_{ns}]^{-1}\beta'_{ns}W\psi_R, 0\right)'.$$

Therefore, due to the epiconvergence results of Pollard (1994) and Knight and Fu (2000), we have that

$$\hat{u}_{ns} \overset{d}{\to} [\beta'_{ns}W\beta_{ns}]^{-1}\beta'_{ns}W\Psi_{\beta,ns}\lambda_{0,ns} + [\beta'_{ns}W\beta_{ns}]^{-1}\beta'_{ns}W\psi_R,$$

$$\hat{u}_{sp} \overset{d}{\to} 0.$$

Hence, the distribution of the risk premia estimates for the useful factors coincides with the one without the identification problem. Therefore, Pen-FM exhibits the so-called *oracle* property. $\qquad\square$

B.2. Proof of Proposition 2

Proof. I am going to prove consistency first. Consider the objective function. As $T \to \infty$

$$\left[\bar{R} - \hat{\beta}\lambda\right]'W_T\left[\bar{R} - \hat{\beta}\lambda\right] \overset{p}{\to} [E[R] - \beta_{ns}\lambda_{ns}]'W[E[R] - \beta_{ns}\lambda_{ns}]$$

Also note that for the strong factors $\eta_T\frac{1}{\left\|\hat{\beta}_j\right\|_1^d} \sim \eta T^{-d/2}O_p(1) \to 0$, while for the weak ones $\eta_T\frac{1}{\left\|\hat{\beta}_j\right\|_1^d} \sim \eta T^{-d/2}c_j T^{d/2} \to \tilde{c}_j > 0$.
Therefore, the limit objective function becomes

$$[E[R] - \beta_{ns}\lambda_{ns}]'W[E[R] - \beta_{ns}\lambda_{ns}] + \sum_{j=1}^{k}\tilde{c}_j|\lambda_j|\mathbb{1}\left\{\beta_j = O_p\left(\frac{1}{\sqrt{T}}\right)\right\}$$

Since $\tilde{c}_j$ are positive constants,

$$0 = \underset{\lambda_{sp}\in\Theta_{sp}}{\arg\min}[E[R] - \beta_{ns}\lambda_{ns}]'W[E[R] - \beta_{ns}\lambda_{ns}] + \sum_{j=1}^{k}\tilde{c}_j|\lambda_j|\mathbb{1}\left\{\beta_j = O_p\left(\frac{1}{\sqrt{T}}\right)\right\}$$

However, the risk premia for the strong factors are still strongly identified, since

$$\underset{\lambda_j \in \Theta_{ns}}{\arg \min}[E\,[R] - \beta_{ns}\lambda_{ns}]'W[E\,[R] - \beta_{ns}\lambda_{ns}] = \lambda_{0,ns} + \frac{1}{\sqrt{T}}\,(\beta'_{ns}W\beta_{ns})^{-1}\,\beta'_{ns}WB_{sp}\lambda_{0,sp} \rightarrow \lambda_{0,ns}$$

Therefore, once again, due to the convexity lemma of Pollard (1991), the estimator is consistent.

Again, I first demonstrate the asymptotic distribution in the usual Fama-McBeth estimator in the absence of weak factors. Recall that

$$\hat{\beta}_{ns} = \beta_{ns} + \frac{1}{\sqrt{T}}\Psi_{\beta,ns} + o_p\left(\frac{1}{\sqrt{T}}\right), \qquad \bar{R} = \beta_{ns}\lambda_{0,ns} + \frac{B_{sp}}{\sqrt{T}}\lambda_{0,sp} + \frac{1}{\sqrt{T}}\psi_R + o_p\left(\frac{1}{\sqrt{T}}\right)$$

where $\Psi_{\beta,ns} = vecinv(\psi_{\beta,ns})$.

Therefore, the second stage estimates have the following asymptotic expansion

$$\hat{\lambda}_{ns} = \left(\hat{\beta}'_{ns}W_T\hat{\beta}_{ns}\right)^{-1}\hat{\beta}'_{ns}W_T\bar{R} \overset{a}{=} \left(\hat{\beta}'_{ns}W_T\hat{\beta}_{ns}\right)^{-1}\hat{\beta}'_{ns}\left(\beta_{ns}\lambda_{0,ns} + \frac{B_{sp}}{\sqrt{T}}\lambda_{0,sp} + \frac{1}{\sqrt{T}}\psi_R\right) =$$

$$= \left(\hat{\beta}'_{ns}W_T\hat{\beta}_{ns}\right)^{-1}\hat{\beta}'_{ns}W_T\left(\hat{\beta}_{ns}\lambda_{0,ns} + (\beta_{ns} - \hat{\beta}_{ns})\lambda_{0,ns} + \frac{B_{sp}}{\sqrt{T}}\lambda_{0,sp} + \frac{1}{\sqrt{T}}\psi_R\right) =$$

$$= \lambda_{0,ns} + \left(\hat{\beta}'_{ns}W_T\hat{\beta}_{ns}\right)^{-1}\hat{\beta}'_{ns}W_T(\hat{\beta}_{ns} - \beta_{ns})\lambda_{0,ns} + \left(\hat{\beta}'_{ns}W_T\hat{\beta}_{ns}\right)^{-1}\hat{\beta}'_{ns}W_T\frac{B_{sp}}{\sqrt{T}}\lambda_{0,sp}$$

$$+ \left(\hat{\beta}'_{ns}W_T\hat{\beta}_{ns}\right)^{-1}\hat{\beta}'_{ns}W_T\frac{1}{\sqrt{T}}\psi_R$$

Finally, since

$$\hat{\beta}'_{ns}W_T\hat{\beta}_{ns} \overset{a}{=} \left[\beta_{ns} + \frac{1}{\sqrt{T}}\Psi_{\beta,ns}\right]'W_T\left[\beta_{ns} + \frac{1}{\sqrt{T}}\Psi_{\beta,ns}\right] \overset{p}{\rightarrow} \beta'_{ns}W\beta_{ns}$$

$$\hat{\beta}_{ns} - \beta_{ns} \overset{a}{=} -\frac{1}{\sqrt{T}}\Psi_{\beta,ns} = \frac{1}{\sqrt{T}}\Psi_{\beta,ns}$$

we get

$$\sqrt{T}(\hat{\lambda}_{ns} - \lambda_{0,ns}) \overset{d}{\rightarrow}$$

$$[\beta'_{ns}W\beta_{ns}]^{-1}\beta'_{ns}W\Psi_{\beta,ns}\lambda_{0,ns} + (\beta'_1W\beta_1)^{-1}\beta'_{ns}W\psi_R + \left(\hat{\beta}'_{ns}W_T\hat{\beta}_{ns}\right)^{-1}\hat{\beta}'_{ns}W_TB_{sp}\lambda_{0,sp}$$

The asymptotic distribution of risk premia estimates has three components:

- $[\beta'_{ns}W\beta_{ns}]^{-1}\beta'_{ns}W\Psi_{\beta,ns}\lambda_{0,ns}$, which arises due to the error-in-variables problem, since we observe not the true values of betas, but only their estimates, i.e. the origin for Shanken (1992) correction;

- $(\beta'_1W\beta_1)^{-1}\beta'_{ns}W\psi_R$, which corresponds to the usual sampling error, associated with the WLS estimator;

- $\left(\hat{\beta}'_{ns}W_T\hat{\beta}_{ns}\right)^{-1}\hat{\beta}'_{ns}W_TB_{sp}\lambda_{0,sp}$, which is the $\frac{1}{\sqrt{T}}$ omitted variable bias, due to eliminating potentially priced weak factors from the model.

Similar to the previous case, in order show the asymptotic distribution of the Pen-FM estimator, I rewrite the objective function in terms of the localised parameters, $u = \frac{\lambda - \lambda_0}{\sqrt{T}}$, as follows:

$$\hat{u} = \underset{u \in K}{\arg \min}\ u'\hat{\beta}'W_T\hat{\beta}u + 2\sqrt{T}u'\hat{\beta}'W_T(\hat{\beta}\lambda_0 - \bar{R}) + T\eta_T\sum_{j=1}^{k}\frac{1}{\left\|\hat{\beta}_j\right\|_1^d}\left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right],$$

since

$$\left[\bar{R} - \hat{\beta}\left(\lambda_0 + \frac{u}{\sqrt{T}}\right)\right]' W_T \left[\bar{R} - \hat{\beta}\left(\lambda_0 + \frac{u}{\sqrt{T}}\right)\right] =$$

$$= \bar{R}' W_T \bar{R} + \left[\lambda_0 + \frac{u}{\sqrt{T}}\right]' \hat{\beta}' W_T \hat{\beta} \left[\lambda_0 + \frac{u}{\sqrt{T}}\right] - 2\left[\lambda_0 + \frac{u}{\sqrt{T}}\right]' \hat{\beta}' W_T \bar{R} ,$$

$$\left[\lambda_0 + \frac{u}{\sqrt{T}}\right]' \hat{\beta}' W_T \hat{\beta} \left[\lambda_0 + \frac{u}{\sqrt{T}}\right] = \lambda_0' \hat{\beta}' W_T \hat{\beta} \lambda_0 + \frac{u'}{\sqrt{T}} \hat{\beta}' W_T \hat{\beta} \frac{u}{\sqrt{T}} + \frac{2}{\sqrt{T}} u' \hat{\beta}' W_T \hat{\beta} \lambda_0$$

$$-2\left[\lambda_0 + \frac{u}{\sqrt{T}}\right]' \hat{\beta}' W_T \bar{R} = -2\lambda_0' \hat{\beta}' W_T \bar{R} - \frac{2}{\sqrt{T}} u' \hat{\beta}' W_T \bar{R} .$$

Recall that

$$\hat{\beta}' W_T \hat{\beta} \overset{a}{=} \begin{bmatrix} \beta_{ns}' + \frac{1}{\sqrt{T}}\Psi_{\beta,ns}' \\ \frac{1}{\sqrt{T}}(B_{sp}' + \Psi_{\beta,sp}') \end{bmatrix} W_T \begin{bmatrix} \beta_{ns} + \frac{1}{\sqrt{T}}\Psi_{\beta,ns} & \frac{B_{sp}}{\sqrt{T}} + \frac{1}{\sqrt{T}}\Psi_{\beta,sp} \end{bmatrix} \overset{a}{=}$$

$$= \begin{bmatrix} \beta_{ns}' W \beta_{ns} + \frac{2}{\sqrt{T}}\Psi_{\beta,ns}' W \beta_{ns} & \frac{1}{\sqrt{T}}\beta_{ns}' W(B_{sp} + \Psi_{\beta,sp}) \\ \frac{1}{\sqrt{T}}(B_{sp} + \Psi_{\beta,sp})W \beta_{ns} & 0 \end{bmatrix} .$$

Hence,

$$u' \hat{\beta}' W_T \hat{\beta} u \overset{a}{=} \begin{bmatrix} u_{ns}' & u_{sp}' \end{bmatrix} \begin{bmatrix} \beta_{ns}' W \beta_{ns} + \frac{2}{\sqrt{T}}\Psi_{\beta 1}' W \beta_{ns} & \frac{1}{\sqrt{T}}\beta_{ns}' W(B_{sp} + \Psi_{\beta,sp}) \\ \frac{1}{\sqrt{T}}(B_{sp} + \Psi_{\beta,ns})' W \beta_{ns} & 0 \end{bmatrix} \begin{bmatrix} u_{ns} \\ u_{sp} \end{bmatrix} =$$

$$= u_{ns}' \left[\beta_{ns}' W \beta_{ns} + \frac{2}{\sqrt{T}}\Psi_{\beta,ns}' W \beta_{ns}\right] u_{ns} + u_{sp}' \left[\frac{1}{\sqrt{T}}(B_{sp} + \Psi_{\beta,sp})' W \beta_{ns}\right] u_{ns} + u_{ns}' \left[\frac{1}{\sqrt{T}}\beta_{ns}' W(B_{sp} + \Psi_{\beta,sp})\right] u_{sp}$$

$$u' \hat{\beta}' W_T \hat{\beta} \lambda_0 - u' \hat{\beta}' W_T \bar{R} = u' \hat{\beta}' W_T \left[\hat{\beta}_{ns}\lambda_{0,ns} - \beta_{ns}\lambda_{0,ns} - \frac{B_{sp}}{\sqrt{T}}\lambda_{0,sp} - \frac{1}{\sqrt{T}}\varphi_R\right] \overset{a}{=}$$

$$= \begin{bmatrix} u_{ns}' & u_{sp}' \end{bmatrix} \begin{bmatrix} \beta_{ns}' + \frac{1}{\sqrt{T}}\Psi_{\beta,ns}' \\ \frac{1}{\sqrt{T}}(B_{sp}' + \Psi_{\beta,sp}') \end{bmatrix} W_T \left[\frac{1}{\sqrt{T}}\Psi_{\beta,ns}\lambda_{0,ns} - \frac{B_{sp}}{\sqrt{T}}\lambda_{0,sp} - \frac{1}{\sqrt{T}}\varphi_R\right] \overset{a}{=}$$

$$= \begin{bmatrix} u_{ns}' & u_{sp}' \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{T}}\beta_{ns}' W \Psi_{\beta,ns}\lambda_{0,ns} - \frac{1}{\sqrt{T}}\beta_{ns}' W B_{sp}\lambda_{0,sp} - \frac{1}{\sqrt{T}}\beta_{ns}' W \varphi_R \\ 0 \end{bmatrix}$$

Finally, this implies that the overall objective function asymptotically looks as follows:

$$\bar{L}_T(u) \overset{a}{=} u_{ns}' \left[\beta_{ns}' W \beta_{ns} + \frac{2}{\sqrt{T}}\Psi_{\beta,ns}' W \beta_{ns}\right] u_{ns} + u_{sp}' \left[\frac{1}{\sqrt{T}}(B_{sp} + \Psi_{\beta,sp})' W \beta_{ns}\right] u_{ns}$$

$$+u_{ns}' \left[\frac{1}{\sqrt{T}}\beta_{ns}' W(B_{sp} + \Psi_{\beta,sp})\right] u_{sp} + 2u_{ns}' \left[\beta_{ns}' W \Psi_{\beta,ns}\lambda_{0,ns} - \beta_{ns}' W B_{sp}\lambda_{0,sp} - \beta_{ns}' W \varphi_R\right]$$

$$+T\eta_T \sum_{j=1}^{k} \frac{1}{\left\|\hat{\beta}^j\right\|_1^d} \left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right] \overset{a}{=}$$

$$= u_{ns}' \beta_{ns}' W \beta_{ns} u_{ns} + 2u_{ns}' \left[\beta_{ns}' W \Psi_{\beta,ns}\lambda_{0,ns} - \beta_{ns}' W B_{sp}\lambda_{0,sp} - \beta_{ns}' W \varphi_R\right] + T\eta_T \sum_{j=1}^{k} \frac{1}{\left\|\hat{\beta}^j\right\|_1^d} \left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right]$$

Now, for a spurious factor $T\eta_T \frac{1}{\|\hat{\beta}_j\|_1^d} \left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right] = \sqrt{T}\eta T^{-d/2} c_2 T^{d/2} |u_j| = \sqrt{T}\tilde{c}|u_j|$ while for the

strong ones $T\eta_T \frac{1}{||\hat{\beta}_j||_1^d}\left[\left|\lambda_{0j} + \frac{u_j}{\sqrt{T}}\right| - |\lambda_{0j}|\right] = c_2\sqrt{T}T^{-d/2}u_j sgn(\lambda_{0j}) \to 0$, since $d > 2$.

Hence, as $T \to \infty$, $\bar{L}_T(u) \xrightarrow{d} \bar{L}_n$ for every u, where

$$\bar{L}(u) = \begin{cases} -u'_{ns}\beta'_{ns}W\beta_{ns}u_{ns} - 2u'_{ns}W\beta_{ns}(\phi_R + B_{sp}\lambda_{0,ns} - \Psi_{\beta,ns}\lambda_{0,ns}) & \text{if } u_{sp} = 0 \\ \infty & \text{otherwise} \end{cases}$$

Note that $\bar{L}_T(u)$ is a convex function with the unique optimum given by

$$\left([\beta'_{ns}W\beta_{ns}]^{-1}\beta'_{ns}W\left[\Psi_{\beta,ns}\lambda_{0,ns} + B_{sp}\lambda_{0,sp}\right] + [\beta'_{ns}W\beta_{ns}]^{-1}\beta'_{ns}W\psi_R, 0\right)'.$$

Therefore, due to the epiconvergence results of Pollard (1994) and Knight and Fu (2000),

$$\hat{u}_{ns} \xrightarrow{d} [\beta'_{ns}W\beta_{ns}]^{-1}\beta'_{ns}WB_{sp}\lambda_{0,sp} + [\beta'_{ns}W\beta_{ns}]^{-1}\beta'_{ns}W(\psi_R + \Psi_{\beta,ns}\lambda_{0,ns}),$$

$$\hat{u}_{sp} \xrightarrow{d} 0.$$

$\square$

B.3. Proof of Proposition 3

Proof. Consider the bootstrap counterpart of the second stage regression.

$$\hat{\lambda}^* = \arg\min_{\lambda \in \Theta} (\bar{R}^* - \hat{\beta}^*\lambda)'W_T^*(\bar{R}^* - \hat{\beta}^*\lambda) + \mu_T\sum_{j=1}^{k}\frac{1}{||\hat{\beta}_j^*||^d}[|\lambda_j|]$$

Similar to Proposition 1, in terms of localised parameters, $\lambda = \hat{\lambda}_{pen} + \frac{u}{\sqrt{T}}$, the centred problem becomes

$$\hat{u}^* = \arg\min_{u \in K} (\hat{\lambda}_{pen} + \frac{u}{\sqrt{T}})'\hat{\beta}^{*'}W_T^*\hat{\beta}^*(\hat{\lambda}_{pen} + \frac{u}{\sqrt{T}}) - 2(\hat{\lambda}_{pen} + \frac{u}{\sqrt{T}})'\hat{\beta}^{*'}W_T^*\bar{R}^* +$$

$$+\mu_T\sum_{j=1}^{k}\frac{1}{||\hat{\beta}_j^*||^d}[|\hat{\lambda}_{j,pen} + \frac{u}{\sqrt{T}}| - |\hat{\lambda}_{j,pen}|]$$

where K is a compact set on $\mathbb{R}^{k+1}$. Note that the problem is equivalent to the following one

$$\hat{u}^* = \arg\min_{u \in K} u'\hat{\beta}^{*'}W_T^*\hat{\beta}^*u + 2\sqrt{T}u'\hat{\beta}^{*'}W_T^*\hat{\beta}^*\hat{\lambda}_{pen} - 2\sqrt{T}u'\hat{\beta}_{*'}W_T^*\bar{R}^* +$$

$$+\mu_T\sum_{j=1}^{k}\frac{1}{||\hat{\beta}_j^*||^d}[|\hat{\lambda}_{j,pen} + \frac{u}{\sqrt{T}}| - |\hat{\lambda}_{j,pen}|].$$

If $\beta_{sp} = 0$

$$\begin{bmatrix} u''_{ns} & u'_{sp} \end{bmatrix}\begin{bmatrix} \hat{\beta}'_{ns} + \frac{\Psi_{\beta,ns}}{\sqrt{T}}' \\ \hat{\beta}'_{sp} + \frac{\Psi_{\beta,sp}}{\sqrt{T}}' \end{bmatrix}W_T^*\begin{bmatrix} \hat{\beta}_{ns} + \frac{\Psi_{\beta,ns}}{\sqrt{T}} & \hat{\beta}_{sp} + \frac{\Psi_{\beta,ns}}{\sqrt{T}} \end{bmatrix}\begin{bmatrix} u_{ns} \\ u_{sp} \end{bmatrix} \overset{a}{=} u'_{ns}\beta'_{ns}W\beta_{ns}u_{ns}$$

$$2\sqrt{T}u'\hat{\beta}^{*'}W_T^*\hat{\beta}^*\hat{\lambda}_{pen} - 2\sqrt{T}u'\hat{\beta}^{*'}W_T^*\bar{R}^* = 2u'\begin{bmatrix} \hat{\beta}'_{ns} + \frac{\Psi_{\beta,ns}}{\sqrt{T}}' \\ \hat{\beta}'_{sp} + \frac{\Psi_{\beta,sp}}{\sqrt{T}}' \end{bmatrix}W_T^*\sqrt{T}\left[\hat{\beta}^*\hat{\lambda}_{pen} - \hat{\beta}^*\bar{R}^*\right]$$

Further,

$$\hat{\beta}^*\hat{\lambda}_{pen} - \hat{\beta}^*\bar{R}^* = -\frac{1}{\sqrt{T}}\psi_R + \frac{1}{\sqrt{T}}\hat{\beta}\hat{\lambda}_{pen} + \left[\bar{R} - \hat{\beta}\hat{\lambda}_{pen}\right]$$

$$\left[\beta_0 + \frac{1}{\sqrt{T}}\Psi_\beta\right]' W \left[\beta_0\lambda_0 + \frac{1}{\sqrt{T}}\psi_R\right] - \left[\beta_0 + \frac{1}{\sqrt{T}}\Psi_\beta\right]' W \left[\beta_0 + \frac{1}{\sqrt{T}}\Psi_\beta\right]\left[\lambda_0 + \frac{1}{\sqrt{T}}\psi_{pen}\right] = o_p\left(\frac{1}{\sqrt{T}}\right)$$

since $\psi_{pen} = [(\beta'_{ns}\beta_{ns}]^{-1}W\beta'_{ns}[-\psi_R + \beta_{ns}\Psi_R]$

This in turn implies that the bootstrap counterpart of the second stage satisfies

$$\hat{u}^* = \arg\min_{u \in K} u'\beta'_{ns}W\beta_{ns}u + 2u'_{ns}\beta'_{ns}W(-\psi_R + \Psi_{\beta,ns}\lambda_{0,ns}) + \eta_T \sum_{j=1}^k \frac{1}{||\hat{\beta}^*_j||^d}[|\hat{\lambda}_{j,pen} + \frac{u}{\sqrt{T}}| - |\hat{\lambda}_{j,pen}|]$$

The weak convergence of $\sqrt{T}(\hat{\lambda}^*_{pen} - \hat{\lambda}_{pen})$ to $\sqrt{T}(\hat{\lambda}_{pen} - \lambda_0)$ now follows from the argmax theorem of Knight and Fu (2000). $\qquad \square$

B.4. Proof of Proposition 4

Proof. The condition in Proposition 4 requires the strict monotonicity of the cdf to the right of a particular α-quantile. This implies that if $B_T \to B$ weakly, then $B_T^{-1}(\alpha) \to B^{-1}(\alpha)$ as $T \to \infty$. Hence, $P(\lambda_0 \in I_{T,\alpha}) \to \alpha$ as $T \to \infty$.

If there is at least one non-spurious component (e.g. a common intercept for the second stage or any useful factor), the limiting distribution of the estimate will be a continuous random variable, thus implying the monotonicity of its cdf, and again, driving the desired outcome. $\qquad \square$

B.5. Proof of Proposition 5

Proof. The argument for the consistency and asymptotic normality of the Pen-GMM estimator is derived on the basis of the empirical process theory. The structure of the argument is similar to the existing literature on the shrinkage estimators for the GMM class of models, e.g. Caner (2009), Liao (2013), and Caner and Fan (2014). I first demonstrate the consistency of the estimator.

The sample moment function can be decomposed in the following way:

$$\frac{1}{T}\sum_{t=1}^T g_t(\theta) = \frac{1}{T}\sum_{t=1}^T (g_t(\theta) - Eg_t(\theta)) + \frac{1}{T}\sum_{t=1}^T Eg_t(\theta)$$

Under Assumption 2, by the properties of the empirical processes (Andrews (1994))

$$\frac{1}{\sqrt{T}}\sum_{t=1}^T (g_t(\theta) - Eg_t(\theta)) = O_p(1)$$

Further, by Assumption 2.2

$$E\left(\frac{1}{T}\sum_{t=1}^T g_t(\theta)\right) \xrightarrow{p} g_1(\theta)$$

Also note that for the strong factors $\eta_T \frac{1}{\|\hat{\beta}_j\|_1^d} \sim \eta T^{-d/2}O_p(1) \to 0$, while for the spurious ones $\eta_T \frac{1}{\|\hat{\beta}_j\|_1^d} \sim \eta T^{-d/2}c_j T^{d/2} \to \tilde{c}_j > 0$

Therefore, the whole objective function converges uniformly in $\theta \in S$ to the following expression

$$g_1(\theta)'W(\theta)g_1(\theta) + \sum_{j=1}^k \tilde{c}_j |\lambda_j| \mathbb{1}\{\beta_j = 0\}$$

81

Finally, since $g_1(\theta_{0,ns}, \lambda_{sp}) = g_1(\theta_{0,ns}, \mathbf{0}_{k2})$, and $\{\mu_f, vec(\beta_f), \lambda_{0,ns}, \lambda_{0,c}\}$ are identified under Assumption 2.4, $\{\theta_{0,ns}, \mathbf{0}_{k_2}\}$ is the unique minimum of the limit objective function.

Therefore

$$\hat{\theta}_{pen} \xrightarrow{p} \underset{\theta \in S}{\arg\min} \ g_1(\theta)'W(\theta)g_1(\theta) + \sum_{j=1}^{k} \widetilde{c}_j \, |\lambda_j| \, \mathbb{1}\{\beta_j = 0\}$$

and

$$\hat{\theta}_{pen,ns} = \{\hat{\mu}_f, vec(\hat{\beta}), \hat{\lambda}_{ns}, \hat{\lambda}_c\} \to \theta_{0,ns} = \{\mu_f, vec(\beta_f), \lambda_{0,ns}, \lambda_{0,c}\}$$

$$\hat{\lambda}_{sp} \xrightarrow{p} \mathbf{0}_{k2}$$

Similar to the case of the Fama-MacBeth estimator, in order to derive the asymptotic distribution of the Pen-GMM, I rewrite the original optimization problem in the centred parameters $u = \sqrt{T}(\hat{\theta}_{pen} - \theta_0)$:

$$\hat{u} = \underset{u \in K}{\arg\min} \ L_T(u)$$

where

$$L_T(u) = \left[\frac{1}{T}\sum_{t=1}^{T} g_t\left(\theta_0 + \frac{u}{\sqrt{T}}\right)\right]' W_T\left(\theta_0 + \frac{u}{\sqrt{T}}\right) \left[\frac{1}{T}\sum_{t=1}^{T} g_t\left(\theta_0 + \frac{u}{\sqrt{T}}\right)\right]$$

$$- \left[\frac{1}{T}\sum_{t=1}^{T} g_t\left(\theta_0\right)\right]' W_T\left(\theta_0\right) \left[\frac{1}{T}\sum_{t=1}^{T} g_t\left(\theta_0\right)\right] + \eta_T \sum_{j=1}^{k} \frac{1}{||\hat{\beta}||_1^d} \left(\left|\lambda_{j,0} + \frac{u_{\lambda,j}}{\sqrt{T}}\right| - |\lambda_{j,0}|\right)$$

and K is a compact subset in $\mathbb{R}^{nk+2k+1}$.

Using the empirical process results (Andrews (1994)), from Assumption 2.1 it follows that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t\left(\theta_0 + \frac{u}{\sqrt{T}}\right) - Eg_t\left(\theta_0 + \frac{u}{\sqrt{T}}\right) \Rightarrow Z(\theta_0) \equiv N(0, \Gamma)$$

Now, since $Eg_t(\theta_0) = 0$ and by Assumption 2.3,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Eg_t\left(\theta_0 + \frac{u}{\sqrt{T}}\right) \to G(\theta_0)u$$

uniformly in u.

Therefore,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Eg_t\left(\theta_0 + \frac{u}{\sqrt{T}}\right) \Rightarrow Z(\theta_0) + G(\theta_0)u$$

Note that under the presence of useless factors, $G(\theta_0)u = G_{ns}(\theta_0)u_{ns}$ for $u \in K$, where $u_{ns} = \{u_\mu, u_\beta; , u_{\lambda,ns}, u_{\lambda,c}\}$, that is all the localized parameters, except for the those corresponding to the risk premia of the spurious factors.

Therefore, by Assumption 2.4 the first part of the objective function becomes

$$V_T(u) = TL_T(u) \Rightarrow [Z(\theta_0) + G_{ns}(\theta_0)u_{ns}]' W(\theta_0) [Z(\theta_0) + G_{ns}(\theta_0)u_{ns}] - Z(\theta_0)'W(\theta_0)Z(\theta_0)$$
$$= u_{ns}' G_{ns}(\theta_0)'W(\theta_0)G_{ns}(\theta_0)u_{ns} + 2u_{ns}' G_{ns}(\theta_0)'W(\theta_0)Z(\theta_0)$$

Now, for the spurious factors: $T\eta_T \frac{1}{||\hat{\beta}_j||_1^d} \left[\left|\lambda_{0j} + \frac{u_{\lambda,j}}{\sqrt{T}}\right| - |\lambda_{0j}|\right] = \sqrt{T}\eta T^{-d/2}c_2 T^{d/2} |u_{\lambda,j}| = \sqrt{T}\widetilde{c} \, |u_{\lambda,j}|,$

while for the usual ones: $T\eta_T \frac{1}{\|\hat{\beta}_j\|_1^d}\left[\left|\lambda_{0j} + \frac{u_{\lambda,j}}{\sqrt{T}}\right| - |\lambda_{0j}|\right] = c_2\sqrt{T}T^{-d/2}u_{\lambda,j}sgn(\lambda_{0j}) \to 0$, since $d > 2$

Therefore, $\bar{V}_T(u) \xrightarrow{d} \bar{L}_n$ for every u, where

$$\bar{L}(u) = \begin{cases} u'_{ns}G_{ns}(\theta_0)'W(\theta_0)G(\theta_0)u + 2u'_{ns}G(\theta_0)'W(\theta_0)Z(\theta_0) & \text{if } u_{\lambda,sp} = \mathbf{0}_{k_2} \\ \infty & \text{otherwise} \end{cases}$$

Due to the epiconvergence theorem of Knight and Fu (2000),

$$\sqrt{T}(\hat{\lambda}_{pen,sp}) \xrightarrow{d} \mathbf{0}_{k_2}$$
$$\sqrt{T}(\hat{\theta}_{pen,ns} - \theta_{0,ns}) \xrightarrow{d} [G_{ns}(\theta_0)'W(\theta_0)G_{ns}(\theta_0)]^{-1}G_{ns}(\theta_0)W(\theta_0)Z(\theta_0)$$

where $\theta_{0,ns} = \{\mu_f, vec(\beta_f), \lambda_{0,ns}, \lambda_{0,c}\}$

$\square$