Misspecification-Robust Inference in Linear Asset Pricing Models with Irrelevant Risk Factors

Nikolay Gospodinov, Raymond Kan, and Cesare Robotti

Working Paper 2013-9
October 2013

Abstract: We show that in misspecified models with useless factors (for example, factors that are independent of the returns on the test assets), the standard inference procedures tend to erroneously conclude, with high probability, that these irrelevant factors are priced and the restrictions of the model hold. Our proposed model selection procedure, which is robust to useless factors and potential model misspecification, restores the standard inference and proves to be effective in eliminating factors that do not improve the model's pricing ability. The practical relevance of our analysis is illustrated using simulations and empirical applications.

JEL classification: G12, C12, C52

Key words: asset pricing models, lack of identification, model misspecification, GMM estimation

---

The authors thank the editor (Andrew Karolyi), two anonymous referees, Anton Braun, Chu Zhang, Guofu Zhou, and seminar participants at the Federal Reserve Bank of Atlanta, Imperial College London, Queen's University, Singapore Management University, and the University of Georgia. They also thank participants at the third Annual CIRPÉE Applied Financial Time Series Workshop at HEC (Montreal), Mathematical Finance Days 2012, and the 2013 SoFiE Conference for helpful discussions and comments. Gospodinov gratefully acknowledges financial support from Fonds Québécois de recherche sur la société et la culture, Institut de Finance Mathématique de Montréal, and the Social Sciences and Humanities Research Council of Canada. Kan gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada and the National Bank Financial of Canada. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to Nikolay Gospodinov, Federal Reserve Bank of Atlanta, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street NE, Atlanta, GA 30309-4470, nikolay.gospodinov@atl.frb.org; Raymond Kan, University of Toronto, or Cesare Robotti (contact author), Imperial College Business School, Tanaka Building, South Kensington Campus, London SW7 2AZ, United Kingdom, 44 (0)20 7589 5111, c.robotti@imperial.ac.uk.

Misspecification-Robust Inference in Linear Asset Pricing Models with Irrelevant Risk Factors

It seems natural to view all asset pricing models only as approximations of the true data generating process and, hence, potentially misspecified. It is also often the case that these models include nontraded factors which exhibit very low correlations with the returns on the test assets and may jeopardize the identifiability of the model parameters. In the presence of misspecification and lack of identification, the finite-sample distributions of the statistics of interest can depart substantially from the standard asymptotic approximations developed under the assumption of correctly specified and fully identified models. In general, ignoring possible model misspecification and identification failure tends to result in an overly positive assessment of the pricing performance of the asset pricing model and the individual risk factors. This paper attempts to raise the awareness of applied researchers about the pitfalls of incorrectly assuming correct specification and identification of the model and proposes a more conservative (but asymptotically valid) approach to selecting risk factors and determining if they are priced or not.

It is now widely documented that misspecification is an inherent feature of many asset pricing models and reliable statistical inference crucially depends on its robustness to potential model misspecification. Kan and Robotti (2008, 2009), Kan, Robotti, and Shanken (2013), and Gospodinov, Kan, and Robotti (2013) show that by ignoring model misspecification, one can mistakenly conclude that a risk factor is priced when, in fact, it does not contribute to the pricing ability of the model. While these papers provide a general statistical framework for inference, evaluation and comparison of potentially misspecified asset pricing models (see also Ludvigson, 2013), the misspecification-robust inference is developed under the assumption that the covariance matrix of asset returns and risk factors is of full column rank, i.e., the parameters of interest in these models are identified. Importantly, the issues with statistical inference under potential model misspecification become particularly acute when the pricing model includes factors that are only weakly correlated with the returns on the test assets, such as macroeconomic factors.

In this paper, we further generalize the setup in the papers mentioned above to allow for possible identification failure in a stochastic discount factor (SDF) framework. We show that in the extreme case of model misspecification with one or more "useless" factors (i.e., factors that are independent of the asset returns), the identification condition fails and the validity of the statistical inference is compromised. We focus on linear SDFs mainly because the useless factor problem is well-defined for this class of models. In addition, we choose to present our results for the distance metric introduced by Hansen and Jagannathan (HJ, 1997). This measure has gained increased popularity in the empirical asset pricing literature and has been used both as a model diagnostic and as a tool for model selection by many researchers.

The impact of the violation of this identification condition on the asymptotic properties of parameter restriction and specification tests in linear asset pricing models estimated via generalized method of moments (GMM) was first studied by Kan and Zhang (1999b).[1] Kan and Zhang (1999b) analyze the behavior of the standard Wald test (which uses a variance matrix derived under the assumption of correct model specification) for potentially misspecified models. They show that in this setup, using the standard Wald tests would lead to conclude too often that the useless factor is priced. Furthermore, the specification tests have low power in rejecting a misspecified model. An immediate implication of this result is that many poor models and risk factors may have erroneously been deemed empirically successful as our empirical applications illustrate.

We extend the analysis of Kan and Zhang (1999b) along several dimensions. First, unlike Kan and Zhang (1999b), we study the asymptotic and finite-sample properties of misspecification-robust parameter tests and investigate whether the model misspecification adjustment can restore the validity of the standard inference in the presence of useless factors. In particular, we demonstrate that the misspecification-robust Wald test for the significance of the SDF parameter on the useless factor is asymptotically distributed as a chi-squared random variable with one degree of freedom. This result is new to the literature and is somewhat surprising given the identification failure. It stands in sharp contrast with the Wald test constructed under the assumption of correct specification which is shown by Kan and Zhang (1999b) to be asymptotically chi-squared distributed with degrees of freedom given by the difference between the number of assets and the number of factors included in the model. As a consequence, using standard inference will result in a rather extreme over-rejection (with limiting rejection probability equal to one) of the null hypothesis that the risk premium on the useless factor is equal to zero.[2]

Second, we add to the analysis in Kan and Zhang (1999b) by also studying the limiting behavior

[1] Burnside (2010, 2011) discusses analogous identification failures for alternative normalizations of the SDF. Kan and Zhang (1999a) study the consequences of lack of identification for two-pass cross-sectional regressions while Kleibergen (2009, 2010) and Khalaf and Schaller (2011) propose test procedures that exhibit robustness to the degree of correlation between returns and factors in a two-pass cross-sectional regression framework.

[2] Our use of the term "over-rejection" is somewhat non-standard since the true risk premium on a useless factor is not identifiable. Nevertheless, since a useless factor does not improve the pricing performance of the model, testing the null of a zero risk premium is of most practical importance.

of the estimates and Wald tests associated with the useful factors. We show that in misspecified models, the estimator of the coefficient associated with the useless factor diverges with the sample size while the parameters on the useful factors are not consistently estimable. The limiting distributions of the t-statistics corresponding to the useful factors are found to be non-standard and less dispersed when a useless factor is present. Regardless of whether the model is correctly specified or misspecified, the misspecification-robust standard errors ensure asymptotically valid inference and allow us to identify factors that do not contribute to the pricing of the test assets (i.e., useless factors and factors that do not reduce the HJ-distance). To conserve space, we relegate some of the theoretical results on the explicit form of the limiting distributions of the estimators, the t-tests under correct model specification and misspecification as well as the HJ-distance test to an online appendix available on the authors' websites.

Third, we provide a constructive solution to the useless factor problem that restores the standard inference for the t-tests on the parameters associated with the useful factors and for the test of correct model specification. In particular, we propose an easy-to-implement sequential procedure that allows us to eliminate the useless factors from the model and show its asymptotic validity. Monte Carlo simulation results suggest that our sequential model selection procedure is effective in retaining useful factors in the model and eliminating factors that are either useless or do not reduce the HJ-distance. As a result, our proposed method is robust to both model misspecification and presence of useless factors in the analysis.[3]

Several remarks regarding our theoretical results are in order. We should stress that, similarly to White (1982) in a maximum likelihood framework, our misspecification-robust approach to inference allows for the model to be correctly specified and is asymptotically valid (albeit possibly slightly conservative) even when the model holds. This is important because a pre-test for correct model specification lacks power in distinguishing between correctly specified and misspecified models when a useless factor is included in the model. This leaves the misspecification-robust approach as the only feasible way to conduct inference, especially if a reduced rank test suggests an identification failure of the model. Another important issue that requires some clarification concerns our definition of a useless factor. While the paper studies the knife-edge case of a factor that is independent of the

[3]While we study explicitly only the GMM estimator based on the HJ-distance, our results continue to hold for the class of optimal GMM estimators. Some simulation results for the optimal GMM case are provided in the online appendix.

returns on the test assets, in practice all factors exhibit some nonzero correlation in finite samples and are probably better characterized as near-useless. The analysis then should be performed using a local-to-zero asymptotic framework that would provide a continuous transition between the useless and useful factor cases as in the literature on weak instruments and near unit root processes. The drawback of this local-to-zero approach is that the limiting distributions depend on a host of nuisance parameters (localizing constants) that are not consistently estimable. The number of these localizing constants depends on the dimensionality of the vector of test asset returns which could be very large (for example, 43 test assets are employed in our empirical analysis). Therefore, it proves to be more convenient to analyze the knife-edge case of a useless factor which well characterizes the behavior of near-useless factors as it is the case in the literature on weak instruments and integrated processes. Simulation results for factors that exhibit a low (but nonzero) correlation with the test asset returns are qualitatively similar to the ones for the useless factor case reported in this paper and are available from the authors upon request.

Empirically, our interest is in robust estimation of several prominent asset pricing models with macroeconomic and financial factors, also studied in Kan, Robotti, and Shanken (2013), using the HJ-distance measure. In addition to the basic CAPM and consumption CAPM (CCAPM), the theory-based models considered in our main empirical analysis are the CCAPM conditioned on the consumption-wealth ratio (CC-CAY) of Lettau and Ludvigson (2001), a time-varying version of the CAPM with human capital (C-LAB) of Jagannathan and Wang (1996), where the state variable driving the time variation in the SDF coefficients is the consumption-wealth ratio, the durable consumption model (D-CCAPM) of Yogo (2006), and the five-factor implementation of the intertemporal CAPM (ICAPM) used by Petkova (2006). We also study the well-known "three-factor model" of Fama and French (FF3, 1993). Although this model was primarily motivated by empirical observation, its size and book-to-market factors are sometimes viewed as proxies for more fundamental economic factors.

Our main empirical analysis uses the one-month T-bill, the monthly gross returns on the 25 Fama-French size and book-to-market portfolios and the monthly gross returns on the 17 Fama-French industry portfolios from February 1959 until December 2012. The industry portfolios are included to provide a greater challenge to the various asset pricing models, as recommended by Lewellen, Nagel, and Shanken (2010). The HJ-distance test rejects the hypothesis of correct specification for all models. In addition, the test for reduced rank indicates that only CAPM and FF3, two models with traded factors only, are properly identified. This clearly points to the need for statistical methods that are robust to model misspecification and weak identification. We show empirically that when misspecification-robust standard errors are employed, several macroeconomic factors – notably, the durable and nondurable consumption factors, the consumption-wealth factor of Lettau and Ludvigson (2001) and its interaction with nondurable consumption, labor income and the market return, the default premium in ICAPM – do not appear to be priced at the 5% significance level. The only factors that survive our sequential procedure, which eliminates useless factors and the factors with zero risk premia, are the market factor in CAPM and FF3, the book-to-market factor in FF3 and the term premium in ICAPM.

It is important to stress that the useless factor problem is not an isolated problem limited to the data and asset pricing models considered in our main empirical analysis. We show that qualitatively similar pricing conclusions can be reached using different test assets and SDF specifications. Overall, our results suggest that the statistical evidence on the pricing ability of many macroeconomic and financial factors is weak and their usefulness in explaining the cross-section of asset returns should be interpreted with caution.

The rest of the paper is organized as follows. Section 1 reviews some of the main results for asymptotically valid inference under potential model misspecification. In Section 2, we introduce a useless factor in the analysis and present limiting results for the parameters of interest and their t-statistics under both correct model specification and model misspecification. In Section 3, we discuss some practical implications of our theoretical analysis and suggest an easy-to-implement and asymptotically valid model selection procedure. Section 4 reports results from a Monte Carlo simulation experiment. In Section 5, we investigate the performance of some popular asset pricing models with traded and nontraded factors. Section 6 concludes.

1. Asymptotic Inference with Useful Factors

This section introduces the notation and reviews some main results that will be used in the subsequent analysis. Let

$$y_t(\gamma_1) = \tilde{f}_t'\gamma_1 \tag{1}$$

be a candidate linear SDF, where $\tilde{f}_t = [1,\ f_t']'$ is a K-vector with f_t being a $(K-1)$-vector of risk factors, and γ_1 is a K-vector of SDF parameters with generic element γ_{1i} for $i = 1, \ldots, K$. The specification in (1) is general enough to allow $\tilde{f}_t$ to include cross-product terms (using lagged state variables as scaling factors); see Cochrane (1996).

Also, let x_t be the random payoffs of N assets at time t and $q \neq 0_N$ be a vector of their original costs. This setup covers the case of gross returns on the test assets. For the case of excess returns ($q = 0_N$), the mean of the SDF cannot be identified and researchers have to choose some normalization of the SDF (see, for example, Kan and Robotti, 2008, and Burnside, 2010). The theoretical and simulation results for the case of excess returns are very similar to those of the gross returns case presented below and are provided in the online appendix. We assume throughout that the second moment matrix of x_t, $U = E[x_t x_t']$, is nonsingular so that none of the test assets is redundant.

Define the model pricing errors as

$$e(\gamma_1) = E[x_t \tilde{f}_t' \gamma_1 - q] = B\gamma_1 - q, \tag{2}$$

where $B = E[x_t \tilde{f}_t']$. If there exists no value of γ_1 for which $e(\gamma_1) = 0_N$, the model is misspecified. This corresponds to the case when q is not in the span of the column space of B. The pseudo-true parameter vector γ_1^* is defined as the solution to the quadratic minimization problem

$$\gamma_1^* = \underset{\gamma_1 \in \Gamma_1}{\arg\min}\ e(\gamma_1)' W e(\gamma_1) \tag{3}$$

for some symmetric and positive-definite weighting matrix W, where Γ_1 denotes the parameter space.

The HJ-distance is obtained when $W = U^{-1}$ and is given by

$$\delta = \sqrt{e(\gamma_1^*)' U^{-1} e(\gamma_1^*)}. \tag{4}$$

Given the computational simplicity and the nice economic and maximum pricing error interpretation of the HJ-distance, this measure of model misspecification is often used in applied work for estimation and evaluation of asset pricing models. For this reason, we consider explicitly only the case of the HJ-distance although results for the optimal GMM estimator are also available from the authors upon request.

The estimator $\tilde{\gamma}_1$ of γ_1^* is obtained by minimizing the sample analog of (3):

$$\tilde{\gamma}_1 = \underset{\gamma_1 \in \Gamma_1}{\arg\min}\ \hat{e}(\gamma_1)'\hat{U}^{-1}\hat{e}(\gamma_1), \tag{5}$$

where $\hat{U} = \frac{1}{T}\sum_{t=1}^{T} x_t x_t'$, $\hat{e}(\gamma_1) = \hat{B}\gamma_1 - q$ and

$$\hat{B} = \frac{1}{T}\sum_{t=1}^{T} x_t \tilde{f}_t'. \tag{6}$$

Then, the solution to the above minimization problem is given by

$$\tilde{\gamma}_1 = (\hat{B}'\hat{U}^{-1}\hat{B})^{-1}\hat{B}'\hat{U}^{-1}q. \tag{7}$$

Let $e_t(\gamma_1^*) = x_t\tilde{f}_t'\gamma_1^* - q$ and $S = E[e_t(\gamma_1^*)e_t(\gamma_1^*)']$. Assuming that $[x_t',\ f_t']'$ are jointly stationary and ergodic processes with finite fourth moments, $e_t(\gamma_1^*) - e(\gamma_1^*)$ forms a martingale difference sequence and B is of full column rank, Kan and Robotti (2009) show that

$$\sqrt{T}(\tilde{\gamma}_1 - \gamma_1^*) \xrightarrow{d} N(0_K, \Sigma_{\tilde{\gamma}_1}), \tag{8}$$

where $\Sigma_{\tilde{\gamma}_1} = E[h_t h_t']$,

$$h_t = (B'U^{-1}B)^{-1}B'U^{-1}e_t(\gamma_1^*) + (B'U^{-1}B)^{-1}(\tilde{f}_t - B'U^{-1}x_t)u_t \tag{9}$$

and

$$u_t = e(\gamma_1^*)'U^{-1}x_t. \tag{10}$$

Note that if the model is correctly specified (i.e., $u_t = 0$), the expression for h_t specializes to

$$h_t^0 = (B'U^{-1}B)^{-1}B'U^{-1}e_t(\gamma_1^*) \tag{11}$$

and the asymptotic covariance matrix of $\sqrt{T}(\tilde{\gamma}_1 - \gamma_1^*)$ is simplified to

$$\Sigma_{\tilde{\gamma}_1}^0 = E[h_t^0 h_t^{0\prime}] = (B'U^{-1}B)^{-1}B'U^{-1}SU^{-1}B(B'U^{-1}B)^{-1}. \tag{12}$$

Suppose now that the interest lies in testing hypotheses on the individual parameters of the form $H_0 : \gamma_{1i} = \gamma_{1i}^*$ (for $i = 1, \ldots, K$) and define a selector vector $\boldsymbol{\iota}_i$ with one for its i-th element and zero otherwise (the length of $\boldsymbol{\iota}_i$ is implied by the matrix that it is multiplied to). Then, the t-statistic for $\tilde{\gamma}_{1i}$ with standard error computed under potential model misspecification is asymptotically distributed as

$$t_m(\tilde{\gamma}_{1i}) = \frac{\tilde{\gamma}_{1i} - \gamma_{1i}^*}{\sqrt{\boldsymbol{\iota}_i' \hat{\Sigma}_{\tilde{\gamma}_1} \boldsymbol{\iota}_i}} \xrightarrow{d} N(0,1), \tag{13}$$

where $\hat{\Sigma}_{\tilde{\gamma}_1}$ is a consistent estimator of $\Sigma_{\tilde{\gamma}_1}$. Note that this result is valid irrespective of whether the model is misspecified or correctly specified.

In applied work, it is a common practice to test parameter restrictions using t-tests based on standard errors computed under the assumption of correct model specification. For this reason, it is instructive to consider the large sample behavior of the t-test

$$t_c(\tilde{\gamma}_{1i}) = \frac{\tilde{\gamma}_{1i} - \gamma_{1i}^*}{\sqrt{\boldsymbol{\iota}_i' \hat{\Sigma}^0_{\tilde{\gamma}_1} \boldsymbol{\iota}_i}}, \tag{14}$$

where $\hat{\Sigma}^0_{\tilde{\gamma}_1}$ is a consistent estimator of $\Sigma^0_{\tilde{\gamma}_1}$. If the model is indeed correctly specified, the t-test $t_c(\tilde{\gamma}_{1i})$ is asymptotically distributed as a standard normal random variable

$$t_c(\tilde{\gamma}_{1i}) \xrightarrow{d} N(0,1). \tag{15}$$

However, using the result in (8)–(9), we have that under misspecified models

$$t_c(\tilde{\gamma}_{1i}) \xrightarrow{d} N\left(0, \frac{\boldsymbol{\iota}_i' \Sigma_{\tilde{\gamma}_1} \boldsymbol{\iota}_i}{\boldsymbol{\iota}_i' \Sigma^0_{\tilde{\gamma}_1} \boldsymbol{\iota}_i}\right). \tag{16}$$

Furthermore, under the assumption that x_t and f_t are multivariate elliptically distributed, it can be shown (Kan and Robotti, 2009) that $(\boldsymbol{\iota}_i' \Sigma_{\tilde{\gamma}_1} \boldsymbol{\iota}_i)/(\boldsymbol{\iota}_i' \Sigma^0_{\tilde{\gamma}_1} \boldsymbol{\iota}_i) > 1$, which implies that standard inference based on critical values from the $N(0,1)$ distribution would tend to over-reject the null hypothesis $H_0 : \gamma_{1i} = \gamma_{1i}^*$. This also suggests that the testing procedure based on t_c would reject too often the null hypothesis that the SDF parameter is equal to zero.

We conclude this section with several observations that emerge from a closer inspection of the function h_t in (9) which is used for computing the covariance matrix $\Sigma_{\hat{\gamma}_1}$ under misspecification. It proves useful to rewrite h_t as

$$h_t = h_t^0 + (B'U^{-1}B)^{-1}(\tilde{f}_t - B'U^{-1}x_t)u_t. \tag{17}$$

The adjustment term $(B'U^{-1}B)^{-1}(\tilde{f}_t - B'U^{-1}x_t)u_t$ contains three components: (i) a misspecification component u_t, (ii) a spanning component $\tilde{f}_t - B'U^{-1}x_t$ that measures the degree to which the factors are mimicked by the returns on the test assets, and (iii) a component $(B'U^{-1}B)^{-1}$ that measures the usefulness of factors. The adjustment term is zero if the model is correctly specified ($u_t = 0$) or if the factors are fully mimicked by the returns ($\tilde{f}_t = B'U^{-1}x_t$). If the factors are nearly uncorrelated with the returns (i.e., B is close to zero), the component $(B'U^{-1}B)^{-1}$ can be large and the adjustment term tends to dominate the behavior of h_t. Intuitively, near-uncorrelatedness between the factors and the returns magnifies the consequences of small model specification errors and imperfectly mimicked factors.

2. Asymptotic Inference in the Presence of a Useless Factor

As argued in the introduction, many popular asset pricing models include macroeconomic risk factors that often exhibit very low correlations with the returns on the test assets. For this reason, we now consider a candidate SDF which is given by

$$y_t = \tilde{f}_t'\gamma_1 + g_t\gamma_2, \tag{18}$$

where g_t is assumed to be a useless factor such that it is independent of x_t and f_t for all time periods. Note that the independence between g_t and x_t implies $d = E[x_t g_t] = 0_N$.

Now let $D = [B,\ d]$, $\gamma = [\gamma_1',\ \gamma_2]'$, $e(\gamma) = D\gamma - q$, $\hat{d} = \frac{1}{T}\sum_{t=1}^{T} x_t g_t$, and $\hat{D} = [\hat{B},\ \hat{d}]$. Note that since $d = 0_N$, the vector of pricing errors

$$e(\gamma) = B\gamma_1 + d\gamma_2 - q = B\gamma_1 - q \tag{19}$$

is independent of the choice of γ_2. For the pseudo-true values of the SDF parameters, we can set γ_1^* as in (3) but the parameter associated with the useless factor (γ_2^*) cannot be identified. In the following, we set $\gamma_2^* = 0$, which is a natural choice because we show that $\hat{\gamma}_2$ is symmetrically distributed around zero (see the online appendix for details). While the pseudo-true value of γ_2^* is not identifiable, the sample estimates of the SDF parameters are always identified and are given by

$$\hat{\gamma} = (\hat{D}'\hat{U}^{-1}\hat{D})^{-1}\hat{D}'\hat{U}^{-1}q. \tag{20}$$

We make the following assumptions.

Assumption 1. *Assume that (i) $N > K + 1$; (ii) $[x_t',\ f_t',\ g_t]'$ are jointly stationary and ergodic processes with finite fourth moments; (iii) $e_t(\gamma_1^*) - e(\gamma_1^*)$ forms a martingale difference sequence; and (iv) the matrices B $(N \times K)$ and D $(N\times(K + 1))$ have a column rank K.*

Assumption 2. *Let* $\epsilon_t = x_t - B(E[\tilde{f}_t\tilde{f}_t'])^{-1}\tilde{f}_t$ *and assume that* $E[\epsilon_t\epsilon_t'|\tilde{f}_t] = \Sigma$ *(conditional homoskedasticity).*

Assumption 1 imposes relatively mild restrictions on the data. The martingale difference sequence assumption allows for time-varying second- and higher-order moments of the pricing errors. This martingale difference sequence structure can be further relaxed to allow for serially correlated $e_t(\gamma_1^*) - e(\gamma_1^*)$ at the cost of more tedious notation. Part (iv) of Assumption 1 accommodates our setup in which g_t is a useless factor. Assumption 2 is made for convenience in order to simplify some limiting results and can also be relaxed.

Our first results are concerned with the limiting behavior of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ under correctly specified and misspecified models. Before we present these results, note that D is not of full column rank due to the presence of a useless factor. Therefore, the identification of the parameter vector γ^* fails and the sufficient conditions for the consistency and the asymptotic normality of the GMM estimator are not satisfied.

Proposition 1. *Assume that the conditions in Assumption 1 are satisfied.*

(a) *If* $\delta = 0$*, i.e., the model is correctly specified, we have* $\hat{\gamma}_1 - \gamma_1^* = O_p(T^{-1/2})$ *and* $\hat{\gamma}_2 = O_p(1)$.

(b) *If* $\delta > 0$*, i.e., the model is misspecified, we have* $\hat{\gamma}_1 - \gamma_1^* = O_p(1)$ *and* $\hat{\gamma}_2 = O_p(T^{1/2})$.

The explicit forms of the limiting distributions of the estimators and the proof of Proposition 1 are available in the online appendix. All of the asymptotic distributions are non-normal and only the rate of convergence for $\hat{\gamma}_1$ under correctly specified models is standard. The estimator $\hat{\gamma}_2$ for the parameter on the useless factor converges to a bounded random variable and, hence, it is inconsistent.[4] The presence of model misspecification further exacerbates the inference problems with the estimator $\hat{\gamma}_1$ becoming inconsistent and the estimator $\hat{\gamma}_2$ diverging at rate $T^{\frac{1}{2}}$.

Despite the highly non-standard limits of the SDF parameter estimates, it is possible that their t-statistics are well behaved. To investigate this, we define two types of t-statistics: (i) $t_c(\hat{\gamma}_{1i})$, for $i = 1, \ldots, K$, and $t_c(\hat{\gamma}_2)$ that use standard errors obtained under the assumption that the model

[4]Convergence to a random variable arises in other contexts such as spurious regressions with nonstationary data and instrumental variable models with weak instruments.

is correctly specified, and (ii) $t_m(\hat{\gamma}_{1i})$, for $i = 1, \ldots, K$, and $t_m(\hat{\gamma}_2)$ that use standard errors under potentially misspecified models. The two types of t-statistics are based on the estimated covariance matrices $\hat{\Sigma}^0_{\hat{\gamma}} = \frac{1}{T}\sum_{t=1}^T \hat{h}^0_t \hat{h}^{0\prime}_t$ and $\hat{\Sigma}_{\hat{\gamma}} = \frac{1}{T}\sum_{t=1}^T \hat{h}_t \hat{h}'_t$, where

$$\hat{h}^0_t = (\hat{D}'\hat{U}^{-1}\hat{D})^{-1}\hat{D}'\hat{U}^{-1}\hat{e}_t, \tag{21}$$

$$\hat{h}_t = \hat{h}^0_t + (\hat{D}'\hat{U}^{-1}\hat{D})^{-1}([\tilde{f}'_t,\ g_t]' - \hat{D}'\hat{U}^{-1}x_t)\hat{e}'\hat{U}^{-1}x_t, \tag{22}$$

$\hat{e}_t = x_t(\tilde{f}'_t\hat{\gamma}_1 + g_t\hat{\gamma}_2) - q$ and $\hat{e} = \frac{1}{T}\sum_{t=1}^T \hat{e}_t$. We explicitly consider the behavior of $t_c(\hat{\gamma}_{1i})$ and $t_c(\hat{\gamma}_2)$ because it is a common practice for researchers to assume correct specification when computing the t-statistics.

In particular, the t-statistics of $H_0 : \gamma_{1i} = \gamma^*_{1i}$ and $H_0 : \gamma_2 = 0$ under the assumption of a correctly specified model have the form

$$t_c(\hat{\gamma}_{1i}) = \frac{\sqrt{T}(\hat{\gamma}_{1i} - \gamma^*_{1i})}{\sqrt{\boldsymbol{\iota}'_i \hat{\Sigma}^0_{\hat{\gamma}} \boldsymbol{\iota}_i}} \tag{23}$$

and

$$t_c(\hat{\gamma}_2) = \frac{\sqrt{T}\hat{\gamma}_2}{\sqrt{\boldsymbol{\iota}'_{K+1} \hat{\Sigma}^0_{\hat{\gamma}} \boldsymbol{\iota}_{K+1}}}. \tag{24}$$

Kan and Zhang (1999b) studied the limiting behavior of $t_c(\hat{\gamma}_2)$, when no useful factor is present in the model, and showed that under $H_0 : \gamma_2 = 0$, $t_c(\hat{\gamma}_2)^2$ is dominated by χ^2_1 for correctly specified models and $t_c(\hat{\gamma}_2)^2 \xrightarrow{d} \chi^2_{N-K}$ for misspecified models. As stated below, these results continue to hold in the presence of useful factors. We further extend the results in Kan and Zhang (1999b) by deriving the explicit form of the limiting distributions of $t_c(\hat{\gamma}_2)$ for correctly specified models and of $t_c(\hat{\gamma}_{1i})$ for correctly specified and misspecified models (see the online appendix) which allows us to compute the limiting rejection probabilities of these tests when $N(0, 1)$ critical values are used for inference.

The t-statistics of $H_0 : \gamma_{1i} = \gamma^*_{1i}$ and $H_0 : \gamma_2 = 0$ under a potentially misspecified model are given by

$$t_m(\hat{\gamma}_{1i}) = \frac{\sqrt{T}(\hat{\gamma}_{1i} - \gamma^*_{1i})}{\sqrt{\boldsymbol{\iota}'_i \hat{\Sigma}_{\hat{\gamma}} \boldsymbol{\iota}_i}} \tag{25}$$

and

$$t_m(\hat{\gamma}_2) = \frac{\sqrt{T}\hat{\gamma}_2}{\sqrt{\boldsymbol{\iota}'_{K+1} \hat{\Sigma}_{\hat{\gamma}} \boldsymbol{\iota}_{K+1}}}. \tag{26}$$

One of the main contributions of this paper is to establish that the use of misspecification robust standard errors in constructing the t-tests restores the standard inference on γ_2 in misspecified models. In the other three cases (inference on γ_2 in correctly models or on γ_1 in correctly specified and misspecified models), the misspecification-robust t-tests are still asymptotically valid but they tend to be conservative. The results for $t_m(\hat{\gamma}_{1i})$ and $t_m(\hat{\gamma}_2)$ are presented in the following proposition. For completeness, we also include the results for $t_c(\hat{\gamma}_{1i})$ and $t_c(\hat{\gamma}_2)$ under correctly specified and misspecified models. For presentational convenience, Proposition 2 states the limiting results for the squared t-tests (Wald tests).

Proposition 2.

(a) *Suppose that the conditions in Assumptions 1 and 2 hold. If $\delta = 0$, i.e., the model is correctly specified, then $t_c(\hat{\gamma}_{1i})^2$, $t_c(\hat{\gamma}_2)^2$, $t_m(\hat{\gamma}_{1i})^2$ and $t_m(\hat{\gamma}_2)^2$ are dominated by χ^2_1.*

(b) *Suppose that the conditions in Assumption 1 hold. If $\delta > 0$, i.e., the model is misspecified, then $t_c(\hat{\gamma}_{1i})^2$ and $t_m(\hat{\gamma}_{1i})^2$ are dominated by χ^2_1, and*

$$t_c(\hat{\gamma}_2)^2 \xrightarrow{d} \chi^2_{N-K}, \tag{27}$$

$$t_m(\hat{\gamma}_2)^2 \xrightarrow{d} \chi^2_1. \tag{28}$$

Explicit expressions for the limiting distributions of the four tests statistics in part (a) and $t_c(\hat{\gamma}_{1i})$ and $t_c(\hat{\gamma}_2)$ in part (b) as well as the proof of Proposition 2 are provided in the online appendix. Proposition 2 illustrates the implications of using standard inference procedures (critical values from $N(0,1)$) for testing the individual statistical significance of the SDF parameters γ in the presence of a useless factor. Apart from $t_m(\hat{\gamma}_2)$ in misspecified models, all the other statistics are not asymptotically distributed as standard normal random variables. For example, in misspecified models, the test statistic $t_c(\hat{\gamma}_2)$ will over-reject the null hypothesis when $N(0,1)$ is used as a reference distribution and this over-rejection increases with the number of test assets N (see Equation (27)). As a result, researchers will conclude erroneously (with high probability) that the factor g_t is important and should be included in the model. In order to visualize the source of the over-rejection problem, Figure 1 plots the probability density function of $t_c(\hat{\gamma}_2)$ for $N - K = 7$ when the model is misspecified (see Theorem 2 in the online appendix).

Figure 1 about here

Given the bimodal shape and a variance of 7 for the limiting distribution of $t_c(\hat{\gamma}_2)$, using the critical values from the standard normal distribution would obviously result in highly misleading inference. Importantly, part (b) of Proposition 2 shows that the t-statistic under potentially misspecified models, $t_m(\hat{\gamma}_2)$, retains its standard normal asymptotic distribution even when the factor is useless and Figure 1 provides a graphical illustration of this result. The reduction in the degrees of freedom from $N-K$ for the asymptotic chi-squared distribution of $t_c(\hat{\gamma}_2)^2$ to 1 for the asymptotic chi-squared distribution of $t_m(\hat{\gamma}_2)^2$ is striking.

Proposition 2 also suggests that the presence of a useless factor renders the inference on all the remaining parameters non-standard. Testing the statistical significance of the parameters on the useful factors, in both correctly specified and misspecified models, against the standard normal critical values would lead to under-rejection of the null hypothesis and conservative inference.

The main conclusion that emerges from these results is that one should use misspecification-robust t-statistics when testing the statistical significance of individual SDF parameters. This will ensure that the statistical decision from this test is robust to possible model misspecification and useless factors. If the model happens to be correctly specified, this will result in conservative inference but the useless factor will be removed with probability greater than $1-\alpha$, where α is the size of the test. If a useless factor is not present in the model, the standard normal asymptotics for the misspecification-robust test is restored as discussed in Section 1.[5]

3. Model Selection Procedure

It is worthwhile stressing an important aspect of the approach adopted in this paper. Economic theory often dictates which risk factors should be included in the model. While in a regression framework, the inclusion of economic factors that are irrelevant does not affect the statistical inference in the model and only results in slightly inflated standard errors, this is not the case in the SDF setup studied in this paper. Our theoretical analysis suggests that the inclusion of an irrelevant factor, even if there is a compelling reason for its presence from an economic point of

[5] In the online appendix, we also show that the presence of a useless factor tends to distort the inference on the HJ-distance test for correct model specification and that this test is inconsistent under the alternative.

view, has detrimental effects on the statistical inference on the remaining SDF parameters, their associated t-statistics and the model specification test. We show that the presence of a useless factor renders the remaining parameter estimates inconsistent and causes their t-statistics under both correct model specification and model misspecification to under-reject the null. Only after the useless factor is identified and removed using the misspecification-robust t-test, the validity of the inference and the consistency of the parameters are restored.

These considerations suggest that a sequential procedure based on the misspecification-robust t-statistics is necessary. One possibility is to select a subset of risk factors that survive a sequential testing procedure of individual significance of the SDF parameters γ based on p-values that are obtained from the quantiles of the $N(0,1)$ distribution and compared to a common significance level α. Specifically, in the first stage, the full model is estimated and the factor with the smallest (in absolute value) t-statistic, for which the null of zero risk premium is not rejected at the pre-specified nominal level, is eliminated from the model. The model is then re-estimated with only the factors that survive the first stage. This procedure is repeated until either all factors are eliminated or the SDF parameter estimates on the remaining factors are found to be statistically significant at the desired nominal level when using the misspecification-robust t-test.

However, this model selection procedure does not account for the multitude of tests and could result in a substantially inflated rate of false discoveries (i.e., falsely identifying risk factors as being priced) depending on the number of tested hypotheses, the number of irrelevant factors and the dependence structure of the individual tests. The common solution to this multiple testing problem is to devise adjustments in the testing procedure that control the probability of one or more false discoveries, which is referred to in the statistical literature as familywise error rate. For a review of the large literature on controlling the familywise error rate in multiple testing problems and model selection, see Romano, Shaikh, and Wolf (2008). In this paper, we adopt the Bonferroni method for controlling the number of false discoveries and ensuring the asymptotic validity of our model selection procedure. While several refinements of the Bonferroni procedure have been proposed in the literature (see Benjamini and Hochberg, 1995; Goeman and Solari, 2010; and Romano, Shaikh, and Wolf, 2008, among others), the Bonferroni method has some advantages that are appealing to applied researchers.

First, the Bonferroni modification to the multiple testing problem is rather straightforward.

Let K be the total number of SDF parameters being tested and let the p-value for testing the hypothesis $H_0 : \gamma_i = 0$ be denoted by $p_{T,i}$. Recall that all of the hypotheses $H_0 : \gamma_i = 0$ for $i = 1, ..., K$ are tested by comparing the misspecification-robust t-statistic against the critical value of the $N(0, 1)$ distribution for a significance level α. Then, the Bonferroni method rejects the null hypothesis if $p_{T,i} \leq \alpha/K$ instead of $p_{T,i} \leq \alpha$ as in the standard approach. Second, while some of the proposed methods require independence or some knowledge of the dependence structure of the individual test statistics, the Bonferroni method is robust to any form of dependence by assuming a worst-case dependence structure (Romano, Shaikh, and Wolf, 2008). Naturally, the cost of this robustness is conservative inference. However, given that the number of risk factors in asset pricing models rarely exceeds five, we show in our simulations that the Bonferroni method is only mildly conservative (i.e., it tends to select the relevant factors with high probability) while it fully controls the familywise error rate.

The Bonferroni method is asymptotically valid if (Romano, Shaikh, and Wolf, 2008)

$$\lim\,\sup_{T\to\infty} P\{p_{T,i} \leq u\} \leq u,$$

for any uniformly distributed random variable u on the interval $(0, 1)$. The asymptotic validity of our proposed model selection procedure then follows from combining this asymptotic control of the Bonferroni method with the result in Proposition 2 which states that the limiting distribution of $t_m(\hat{\gamma}_i)^2$ is stochastically dominated by the χ^2_1 distribution.

Instead of eliminating factors with insignificant t-ratios one at a time, one may be tempted to drop all the factors with insignificant t-ratios in each stage. Unlike our proposed method, this alternative procedure can lead to the undesirable outcome of eliminating multiple useful factors when a linear combination of them is useless. In this situation, only one of these useful factors should be dropped to restore the full rank condition for the remaining factors. The effectiveness of our model selection procedure in eliminating useless factors (and factors with zero risk premia) and retaining useful factors in the model is analyzed in the simulation section below.

4. Monte Carlo Simulations

In this section, we undertake a Monte Carlo experiment to assess the small-sample properties of the various test statistics in models with useful and useless factors. In our simulations, we also evaluate the effectiveness of the sequential model selection procedure described above in retaining useful factors and eliminating useless factors and factors with zero risk premia.

4.1 Size of tests of parameter restrictions

For the analysis of the SDF parameter and specification tests, we consider three linear models: (i) a model with a constant term and a useful factor, (ii) a model with a constant term and a useless factor, and (iii) a model with a constant term, a useful factor and a useless factor. For each model, we consider two separate cases: the case in which the model is correctly specified and the case in which the model is misspecified. The returns on the test assets, denoted by R_t, and the useful factor f_t are drawn from a multivariate normal distribution with mean μ and covariance matrix V, where

$$\mu = E\begin{bmatrix} f_t \\ R_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \tag{29}$$

and

$$V = \text{Var}\begin{bmatrix} f_t \\ R_t \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}. \tag{30}$$

Let $\hat{\mu}$ and $\hat{V}$ denote the sample estimates obtained from actual data. In all simulation designs, the covariance matrix of the simulated test asset returns V_{22} is set equal to the estimated covariance matrix $\hat{V}_{22}$ from the 1959:2–2012:12 sample of monthly gross returns on the one-month T-bill, the 25 Fama-French size and book-to-market ranked portfolios and the 17 Fama-French industry portfolios (from Kenneth French's website). Note that in this case, $N = 43$ and $q = 1_N$. For the simulated useful factor, we calibrate its mean μ_1 and variance V_{11} to the sample mean $\hat{\mu}_1$ and sample variance $\hat{V}_{11}$ of the value-weighted market excess return. The covariances between the useful factor and the returns, V_{12}, are chosen based on the covariances estimated from the data, $\hat{V}_{12}$.

The pseudo-true values of the SDF parameters on the constant and the useful factor, $\gamma^* = [\gamma_0^*,\ \gamma_1^*]'$, are chosen as

$$\gamma^* = (X'\hat{V}_{22}^{-1}X)^{-1}X'\hat{V}_{22}^{-1}1_N,$$

where $X = [\hat{\mu}_2,\ \hat{V}_{21} + \hat{\mu}_2\hat{\mu}_1]$. Since the pricing errors of the model are given by

$$e(\gamma^*) = E[R_t(\gamma_0^* + f_t\gamma_1^*)] - 1_N = \mu_2(\gamma_0^* + \mu_1\gamma_1^*) + V_{21}\gamma_1^* - 1_N, \tag{31}$$

we set $\mu_2 = (1_N - V_{21}\gamma_1^*)/(\gamma_0^* + \mu_1\gamma_1^*)$ for the pricing errors to be zero and the model to be correctly specified. For misspecified models, the means of the simulated returns, μ_2, are set equal to the sample means of the actual returns $\hat{\mu}_2$. Finally, the useless factor is generated as a standard normal random variable independent of the returns and the useful factor. Using this approach to generating data allows the models in our simulation experiment to exhibit a similar degree of misspecification as some benchmark asset pricing models such as the CAPM. More specifically, the HJ-distance for models (i) and (iii) is 0.523 while the HJ-distance for model (ii) is 0.535.

The time-series sample sizes that we consider in the simulations are $T = 200$, 600, and 1000. These choices of T cover the range of sample sizes that are typically encountered in empirical work with quarterly ($T = 200$) and monthly ($T = 600$) data. The sample size $T = 1000$ is used to assess the quality of our asymptotic approximations. We also present the limiting rejection probabilities based on our asymptotic results in Theorems 2 and 3 in the online appendix.[6]

In Tables 1 to 3, we report the probabilities of rejection (based on 100,000 simulations) of $H_0 : \gamma_i = \gamma_i^*$ for models (i), (ii), and (iii), respectively, where the γ_i^*'s for the constant and the useful factor are the chosen pseudo SDF parameters, and the γ_i^* for the useless factor is set equal to zero. We present results by comparing two different t-statistics with the standard normal distribution, the one computed under the assumption that the model is correctly specified, $t_c(\hat{\gamma}_i)$, and the one computed under the assumption that the model is potentially misspecified, $t_m(\hat{\gamma}_i)$. For each table, Panel A reports the probabilities of rejection when the model is correctly specified and Panel B reports the probabilities of rejection when the model is misspecified.

Table 1 about here

The results in Table 1.A show that for models that are correctly specified and contain only useful factors, the standard asymptotics provides an accurate approximation of the finite-sample behavior of the t-tests. Since the useful factor, calibrated to the properties of the value-weighted market excess return, is closely replicated by the returns on the test assets, the differences between the t-tests under correctly specified models (t_c) and the t-tests under potentially misspecified models (t_m) exhibit no differences even when the model fails to hold exactly (see Panel B).

Tables 2 and 3 present the empirical size of the t-tests in the presence of a useless factor. The simulation results for the t-tests on the parameters of the useful factor (and the constant term)

[6] The limiting rejection probabilities of t_c in Table 1.B are computed based on (16) assuming that the factor and the returns are multivariate normally distributed.

confirm our theoretical findings that the null hypothesis is under-rejected when $N(0,1)$ is used as a reference distribution. This is the case for correctly specified and misspecified models. For example, the under-rejections are particularly pronounced for the t_m test on the constant term in Tables 2 and 3.

Tables 2 and 3 about here

Similarly, the inference on the useless factor proves to be conservative when the model is correctly specified. However, when the model is misspecified, there are substantial differences between t_c and t_m for the useless factor. Under this scenario, we argued in Section 2 that the t-statistics under correct model specification have a non-normal asymptotic distribution while the misspecification-robust t-statistic for the parameter on the useless factor has a $N(0,1)$ asymptotic distribution. Since the t_c test on the useless factor is asymptotically distributed (up to a sign) as $\sqrt{\chi^2_{N-K}}$, it tends to over-reject severely when the critical values from $N(0,1)$ are used and the degree of over-rejection increases with the sample size. In contrast, the t_m test on the useless factor has good size properties although, for small sample sizes, it slightly under-rejects. As the sample size increases, the empirical rejection rates approach the limiting rejection probabilities (as shown in the rows for $T = \infty$) computed from the corresponding asymptotic distributions in Theorem 2 in the online appendix.

Some comments about the under-rejections reported in Tables 2 and 3 for useful (in correctly specified and misspecified models) and useless (for correctly specified models) factors are in order. Starting with the useless factor, it turns out that the statistical under-rejection of the hull hypothesis of a zero SDF coefficient is actually economically beneficial for our selection procedure since this useless factor will be selected even less often than the nominal size of the test. For the useful factor, the under-rejections may raise more concerns. However, note that the t-tests for the constant and the useful factors in Tables 1 to 3 are testing the null hypothesis $H_0 : \gamma_i = \gamma_i^*$. In practice, the researchers are interested in testing the null hypothesis of whether the factor is priced or not, i.e., $H_0 : \gamma_i = 0$. While the under-rejections in Tables 2 and 3 will affect the local power of the test, it is still quite possible that these tests reject with reasonably high probability the null $H_0 : \gamma_i = 0$ for useful factors with a nonzero risk premium parameter. This is illustrated in the model selection procedure below which selects factors based on whether their parameter estimates are significantly different from zero or not. The results show that the useful factors with nonzero risk premium parameters tend to be selected with sufficiently high probability (97%-99% for $T = 1000$).

4.2 Survival rates of risk factors in the sequential testing procedure

Our findings suggest that the misspecification-robust t-test should be used when it is uncertain whether the factor is useful or useless and the model is correctly specified or misspecified. However, since the procedure based on the t_m test is often conservative, some useful factors might be erroneously excluded from the model. The frequency at which this happens is evaluated in the model selection procedure presented below.

Table 4 reports the survival rates of different factors when using the sequential procedure described in Section 3. In particular, we compare the survival rates from using the t_m test to the survival rates from using the t_c test. The false discovery rate of the multiple testing problem is controlled using the Bonferroni method. In our simulations, we consider a linear model with a constant term, two useful factors with $\gamma_i^* \neq 0$, a useful factor with $\gamma_i^* = 0$, and a useless factor. The mean and variance of the useful factors with $\gamma_i^* \neq 0$ are calibrated to the mean and variance of the excess market return and the high-minus-low factor of Fama and French (1993). The mean and variance of the useful factor with $\gamma_i^* = 0$ are calibrated to the mean and variance of the small-minus-big factor of Fama and French (1993). As in Tables 1–3, the returns and the useful factors are drawn from a multivariate normal distribution.[7] Finally, the useless factor is generated as a standard normal random variable, independent of the test asset returns and the useful factors. The time-series sample size is taken to be $T = 200$, 600, and 1000. The nominal level of the sequential testing procedure is set equal to 5%. The probability that at least one irrelevant (useless or unpriced) factor survives the model selection (MS) procedure is reported in the last two columns of the table with MS_c denoting the survival probability computed from the t_c tests and MS_m denoting the survival probability computed from the t_m tests.

Table 4 about here

Panel A shows that when the model is correctly specified, the procedures based on t_c and t_m

[7] See Gospodinov, Kan, and Robotti (2013) for a description on how to impose zero restrictions on the parameters in correctly specified and misspecified models. The HJ-distance of the misspecified model used in this section is 0.522.

do a similarly good job in retaining the useful factors with nonzero SDF parameters in the model (the survival probabilities are 99%–100% for $T = 600$) and eliminating the useless factor and the factor that does not reduce the HJ-distance. This indicates that using the t_c test in the presence of a useless factor is not problematic when the underlying model holds exactly. However, as shown in Panel B, the situation drastically changes when the model is misspecified. In this case, the procedures based on t_c and t_m still retain the useful factors with similarly high probability (97%–100% for $T = 600$), but they produce very different results when it comes to the useless factor. For example, despite its conservative nature (due to the Bonferroni adjustment), the procedure based on t_c will retain the useless factor 30% of the time for $T = 1000$. In contrast, the procedure based on t_m will retain the useless factor only about 0.9% of the time for $T = 1000$. Similarly, the probability of at least one irrelevant factor being selected in the final specification of the model is 31% (2.1%) for $T = 1000$ when the t_c (t_m) test is used and the model is misspecified. It should be emphasized that the effectiveness of the proposed sequential procedure in retaining the useful factors in the model depends on the correlation between the useful factors and the returns on the test assets and on the magnitude of the SDF coefficient associated with the useful factor. Our procedure will be more effective in retaining a useful factor in the model, the higher this correlation and the larger the SDF coefficient on the useful factor.

Table 5 about here

Table 5 reports the results from a similar exercise but this time the linear asset pricing model consists of a constant term, two useful factors with $\gamma_i^* \neq 0$ and two useless factors. This setup serves to illustrate the usefulness of combining the misspecification-robust t-tests and the Bonferroni method in controlling the false discovery rate which is about 52% (the probability that at least one useless factor is deemed priced) for the t-tests constructed under correct model specification when the true model is misspecified. In contrast, the misspecification-robust model selection procedure with the Bonferroni adjustment retains one or both useless factors only 2% of the time. Another important observation is that the presence of more useless factors does not have a tangible impact on the survival probabilities of the useful factors in misspecified models which remain high, for the t_m tests, at 90%–95% for $T = 600$.

Some aspects of our model selection procedure deserve additional remarks. It should be stressed that the objective of our proposed sequential test is to find the most parsimonious model with the same HJ-distance as the full model. The fact that our method eliminates unpriced useful factors along with useless factors is not of particular concern since these factors do not contribute to the reduction of the HJ-distance. As a result, our model selection procedure purges the model from useless factors (that give rise to inference problems) and unpriced factors (that do not improve the pricing ability of the model) and restores the standard asymptotic theory for the remaining factors in the SDF.

Finally, we consider a scenario in which a linear combination of two useful factors is useless. Although our theoretical setup in Section 2 is not specifically designed to deal with this type of situation, it is still interesting to examine how our sequential model selection procedure fares in this framework. Each factor is created by adding a normally distributed error to the excess market return. The error term in each factor has a mean of zero and a variance of 4% of the variance of the excess market return. The two error terms are independent of each other as well as of the returns on the test assets and the market portfolio. As in Tables 4 and 5, the returns and the factors are drawn from a multivariate normal distribution. We are interested in determining the probability that (i) both factors survive, (ii) only one factor survives, and (iii) no factor survives using the sequential procedure (with the Bonferroni adjustment) based on misspecification-robust t-tests. For comparison, we also report results of the sequential procedure based on t-tests under correct model specification. The nominal level of the sequential testing procedures is set equal to 5%. Ideally, in this framework, only one factor should survive the testing procedures described above.

Table 6 about here

Panel A of Table 6 shows that when the model is correctly specified, the procedures based on t_c and t_m do a similarly good job in retaining only one factor in the model. For example, for $T = 1000$, the probability that only one factor survives is either 89% or 90% depending on whether we use t_c or t_m. For this sample size, the probabilities that both factors survive and no factor survives are very low and similar across procedures. However, when the model is misspecified (see Panel B), the procedures based on t_c and t_m deliver very different results for the "Both factors survive" and "One factor survives" cases. For $T = 1000$, the probability that both factors survive the model selection procedure based on t_c is 37.5% while the probability that both factors survive the model selection procedure based on t_m is 1.7%. This difference in probabilities becomes larger as the sample size is allowed to grow. Importantly, the probabilities that only one factor survives are markedly different across procedures. For example, when $T = 1000$, the probability that only one factor survives is about 89% when using t-tests under misspecified models while it is only 56.6% when using t-tests under correctly specified models.[8] In summary, our selection procedure based on t-tests that are robust to misspecification continues to perform reasonably well even when no single factor is useless but a linear combination of them is.

5. Empirical Analysis

Our theoretical and simulation results point out some serious pitfalls in the empirical analysis of asset pricing models with nontraded factors. In the following, we use monthly data to demonstrate the relevance of our theoretical results.

To show that our findings are not specific to the test assets and SDFs considered in the main empirical application, we also use an alternative set of test assets and SDFs with macroeconomic risk factors that are available only at quarterly frequency.

5.1 Main application

First, we describe the data used in the empirical analysis and outline the different specifications of the asset pricing models considered. Then, we present our results.

5.1.1 Data and asset pricing models

As in the Monte Carlo simulations, the test asset returns are the monthly gross returns on the one-month T-bill, the value-weighted 25 Fama-French size and book-to-market ranked portfolios and the 17 industry portfolios from Kenneth French's website.[9] The data are from February 1959

[8] In an unreported empirical example of the liquidity-augmented CAPM of Liu (2006), the market factor and the liquidity factor of Pastor and Stambaugh (2003) appear to be individually useful but jointly cause a model identification failure. Our proposed model selection procedure proves to be effective in retaining only one useful factor (the market factor in this case) and restoring the full rank condition necessary for identification.

[9] Using only the 25 Fama-French size and book-to-market ranked portfolios as test assets yields qualitatively similar results.

until December 2012. We analyze seven asset pricing models starting with the simple static CAPM. The SDF specification for this model is

$$y_t^{CAPM}(\gamma) = \gamma_0 + \gamma_1 vw_t, \tag{32}$$

where vw is the excess return (in excess of the one-month T-bill rate) on the value-weighted stock market index (NYSE-AMEX-NASDAQ) from Kenneth French's website. The CAPM performed well in the early tests, e.g., Fama and MacBeth (1973), but has fared poorly since.

One extension that has performed better is our second model, the three-factor model (FF3) of Fama and French (1993). This model extends the CAPM by including the two empirically-motivated factors smb and hml, where smb is the return difference between portfolios of stocks with small and large market capitalizations, and hml is the return difference between portfolios of stocks with high and low book-to-market ratios ("value" and "growth" stocks, respectively) from Kenneth French's website. The SDF specification is

$$y_t^{FF3}(\gamma) = \gamma_0 + \gamma_1 vw_t + \gamma_2 smb_t + \gamma_3 hml_t. \tag{33}$$

The third model (ICAPM) is an empirical implementation of Merton's (1973) intertemporal extension of the CAPM based on Campbell (1996), who argues that innovations in state variables that forecast future investment opportunities should serve as factors. The five-factor specification proposed by Petkova (2006) is

$$y_t^{ICAPM}(\gamma) = \gamma_0 + \gamma_1 vw_t + \gamma_2 term_t + \gamma_3 def_t + \gamma_4 div_t + \gamma_5 rf_t, \tag{34}$$

where $term$ is the difference between the yields of ten-year and one-year government bonds (from the Board of Governors of the Federal Reserve System), def is the difference between the yields of long-term corporate Baa bonds and long-term government bonds (from Ibbotson Associates), div is the dividend yield on the Center for Research in Security Prices (CRSP) value-weighted stock market portfolio, and rf is the one-month T-bill yield (from CRSP, Fama Risk Free Rates). The actual factors for $term$, def, div, and rf are their innovations from a VAR(1) system of seven state variables that also includes vw, smb, and hml.

Our fourth model is the scaled CAPM specification including human capital (C-LAB) also considered by Lettau and Ludvigson (2001). This model incorporates measures of the return on human capital as well as the change in financial wealth, and allows the conditional SDF coefficients to vary with the consumption-aggregate wealth ratio (cay) of Lettau and Ludvigson (2001).[10] The SDF specification is

$$y_t^{C-LAB}(\gamma) = \gamma_0 + \gamma_1 cay_{t-1} + \gamma_2 vw_t + \gamma_3 lab_t + \gamma_4 vw_t \cdot cay_{t-1} + \gamma_5 lab_t \cdot cay_{t-1}, \tag{35}$$

where lab is the growth rate in per capita labor income, L, defined as the difference between total personal income and dividend payments, divided by the total population (from the Bureau of Economic Analysis). Following Jagannathan and Wang (1996), we use a two-month moving average to construct the growth rate $lab_t = (L_{t-1} + L_{t-2})/(L_{t-2} + L_{t-3}) - 1$, for the purpose of minimizing the influence of measurement error. The SDF specification in (35) is obtained by scaling the constant term, the vw and the lab factors by a constant and cay. Scaling factors by instruments is one popular way of allowing factor risk premia to vary over time. See Cochrane (1996), among others.

Next, we consider consumption-based models. Our fifth model (CCAPM) is the unconditional consumption model with

$$y_t^{CCAPM}(\gamma) = \gamma_0 + \gamma_1 c_{nd,t}, \tag{36}$$

where c_{nd} is the growth rate in real per capita nondurable consumption (seasonally adjusted at annual rates) from the Bureau of Economic Analysis. This model has generally not performed well empirically (see Lettau and Ludvigson, 2001, for a summary of the poor empirical performance of CCAPM). Therefore, we also examine two other consumption models that have yielded more encouraging results.

One such model (CC-CAY) is a conditional version of the CCAPM due to Lettau and Ludvigson (2001). The relation is

$$y_t^{CC-CAY}(\gamma) = \gamma_0 + \gamma_1 c_{nd,t} + \gamma_2 cay_{t-1} + \gamma_3 c_{nd,t} \cdot cay_{t-1}. \tag{37}$$

The SDF specification in (37) is obtained by scaling the constant term and the c_{nd} factor by a constant and cay.

[10] Following Vissing–Jørgensen and Attanasio (2003), we linearly interpolate the quarterly values of cay to permit analysis at the monthly frequency.

The last model (D-CCAPM), due to Yogo (2006), highlights the cyclical role of *durable* consumption in asset pricing. The specification is

$$y_t^{D-CCAPM}(\gamma) = \gamma_0 + \gamma_1 vw_t + \gamma_2 c_{nd,t} + \gamma_3 c_{d,t}, \tag{38}$$

where c_d) is the growth rate in real per capita durable consumption (seasonally adjusted at annual rates) from the Bureau of Economic Analysis.

5.1.2 Results

Before presenting the estimation results for the SDF parameters, we first investigate whether the various risk factors are correlated with the asset returns and whether the seven models described above are properly identified. As mentioned in the theoretical section of the paper, the presence of a useless factor leads to a violation of the crucial identification condition that the $N \times K$ matrix $B = E[x_t \tilde{f}_t']$ is of full column rank. Therefore, it is of interest to test if B is of (reduced) rank $K - 1$. Since $\hat{B}$ is not invariant to rescaling of the data, we first perform a normalization on $\hat{B}$. Define $\hat{\Theta} = \hat{U}^{-\frac{1}{2}} \hat{B} \hat{S}_{\tilde{f}}^{-\frac{1}{2}}$ and its corresponding population counterpart $\Theta = U^{-\frac{1}{2}} B S_{\tilde{f}}^{-\frac{1}{2}}$, where $\hat{S}_{\tilde{f}} = \frac{1}{T}\sum_{t=1}^{T} \tilde{f}_t \tilde{f}_t'$ and $S_{\tilde{f}} = E[\tilde{f}_t \tilde{f}_t']$. Note that $\hat{\Theta}$ is invariant to rescaling of the data and $\text{rank}(\Theta) = \text{rank}(B)$.

Let $\hat{\Pi}$ be a consistent estimator of the asymptotic covariance matrix of $\sqrt{T}\text{vec}(\hat{\Theta} - \Theta)$, where $\text{vec}(\cdot)$ is the vec operator. Following Ratsimalahelo (2002) and Kleibergen and Paap (2006), we perform a singular value decomposition on $\hat{\Theta}$ such that $\hat{\Theta} = \tilde{U}\tilde{S}\tilde{V}'$, where $\tilde{U}'\tilde{U} = I_N$, $\tilde{V}'\tilde{V} = I_K$, and $\tilde{S}$ is an $N \times K$ matrix with the singular values of $\hat{\Theta}$ in decreasing order on its diagonal. Let $\tilde{U}_2$ be the last $N - K + 1$ columns of $\tilde{U}$, $\tilde{V}_2$ be the last column of $\tilde{V}$, and

$$\tilde{\Pi} = (\tilde{V}_2' \otimes \tilde{U}_2')\hat{\Pi}(\tilde{V}_2 \otimes \tilde{U}_2). \tag{39}$$

Then, a test of $H_0 : \text{rank}(\Theta) = \text{rank}(B) = K - 1$ is given by

$$\mathcal{W}^* = T\tilde{s}_K^2 \tilde{\Pi}^{11} \stackrel{d}{\to} \chi^2_{N-K+1}, \tag{40}$$

where $\tilde{s}_K$ is the smallest singular value of $\hat{\Theta}$, and $\tilde{\Pi}^{11}$ is the $(1, 1)$ element of $\tilde{\Pi}^{-1}$.

Table 7.A reports the rank restriction test ($\mathcal{W}^*$) and its p-value (p-val) of the null that $E[x_t(1,\ f_{it})]$ has a column rank of one. The panel shows that we cannot reject the null of a column rank of

one at the 5% significance level for seven out of 14 macroeconomic and financial factors. This finding suggests that several risk factors in ICAPM, C-LAB, CCAPM, CC-CAY, and D-CCAPM can be reasonably considered as useless and that our asymptotic results on useless factors are of practical importance. Panel B further shows that only CAPM and FF3 pass the test of full rank condition at the 5% nominal level. This is consistent with the fact that *vw*, *smb* and *hml* are highly correlated with the returns on the test assets while most factors in the other models are not. Panel B also shows that no model passes the HJ-distance specification test at conventional levels of significance. Since the HJ-distance test has been shown to substantially over-reject under the null in realistic simulation settings with many test assets, we also consider an alternative test of $H_0 : \lambda = U^{-1}e = 0_N$ (which is equivalent to the test of $H_0 : \delta = 0$). Gospodinov, Kan, and Robotti (2013) show that this alternative Lagrange multiplier (LM) test has excellent size and power properties. The results in Panel B indicate that one would reach the same conclusions using the LM and HJ-distance tests. Therefore, the model rejections documented in Table 7.B do not seem to be driven by the finite-sample properties of the HJ-distance test. Overall, these empirical findings suggest that valid inference should account for the fact that some of the models are potentially misspecified and poorly identified.

Table 7 about here

Although the rank restriction test serves as a useful pre-test for possible identification problems, this test does not allow us to unambiguously identify which factor contributes to the identification failure of the model. In addition, this test does not address the question of which risk factors are important in explaining the cross-sectional differences in asset returns. Our misspecification-robust test of $H_0 : \gamma_i = 0$ proves to be of critical importance in (i) providing the direction of the identification failure and (ii) allowing us to determine whether a given risk factor is priced. Panels C and D of Table 7 present the t-tests under correct model specification and potential model misspecification as well as the model selection procedure described in Section 3. Using t-tests under correct specification, the results in Panel C suggest that the factors *term* in ICAPM, $lab \cdot cay$ in C-LAB, *cay* in CC-CAY, and c_{nd} in CCAPM, CC-CAY, and D-CCAPM survive the sequential testing procedure at the 5% significance level using the Bonferroni correction. However, given the violation of the full rank condition for these models, the standard normal distribution is not the appropriate reference distribution in this case. Since some of these factors are very weakly correlated with the returns on the test assets and effectively behave as useless factors, they tend to be included in the model much more often than they should (see the simulation results for the useless factors in Panel B of Table 4 and Table 5). Therefore, the model selection procedure needs to be implemented using misspecification-robust t-tests.

Panel D shows that, from all of the nontraded factors listed above, only *term* in ICAPM survives the sequential procedure based on misspecification-robust t-tests at the 5% significance level. Finally, for traded factors, we find strong evidence of pricing for the *vw* factor in CAPM and the *vw* and *hml* factors in FF3 in both Panels C and D.[11]

5.2 Additional empirical evidence

The results in Table 7 suggest that the statistical evidence on the pricing ability of several macroeconomic and financial factors is weak and their usefulness in explaining the cross-section of asset returns should be interpreted with caution. In this section, we further emphasize the importance of accounting for model misspecification and weak identification in empirical work. The following application uses an alternative set of test assets and SDFs that include macroeconomic risk factors whose data are available only at quarterly frequency.

The test asset returns are the quarterly gross returns on the one-month T-bill, the value-weighted 6 Fama-French size and book-to-market ranked portfolios, the 17 industry portfolios and the 10 momentum portfolios from Kenneth French's website. The sample period is from 1952:Q2 until 2012:Q4. We consider the following asset pricing specifications: (i) the conditional CCAPM (CC-CAY) version of Lettau and Ludvigson (2001) with c_{nd}, cay and $c_{nd}\cdot cay$ as described in the previous section; (ii) the conditional CAPM (C-ML) of Santos and Veronesi (2006) with *vw* and *vw* scaled by the labor income-consumption ratio (ml) as risk factors; (iii) a version of the conditional consumption CAPM (CC-MY) proposed by Lustig and Van Nieuwerburgh (2005) with the housing collateral ratio (my), c_{nd}, and the interaction term $c_{nd}\cdot my$ as risk factors; and (iv) the sector investment model (SIM) of Li, Vassalou, and Xing (2006) with the log investment growth rates

[11]It is possible that the individual risk factors considered in this section do not capture adequately the risk incorporated in all of the macroeconomic data that is available to market participants. One approach to extract parsimoniously the common variation in macroeconomic variables is the factor analysis advocated by Stock and Watson (2002). In unreported results, we follow this approach and construct three orthogonal factors that summarize the dynamics of 130 U.S. macroeconomic time series for the period March 1960 – December 2011. Using the same set of tests assets, we find that no factor survives the misspecification-robust model selection procedure at the 5% significance level using the Bonferroni adjustment.

for households (i_h), non-financial corporations (i_c), and non-corporate sector (i_{nc}) as risk factors. We include the CC-CAY model again since the original data for the *cay* factor are available at a quarterly frequency. These four models with nontraded factors have yielded encouraging results in cross-sectional asset pricing.[12]

The empirical results for quarterly data are reported in Table 8, with Panel A showing that for all factors except for *vw* and *vw·ml*, we cannot reject the null that $E[x_t(1,\ f_{it})]$ has a column rank of one at the 5% significance level. In addition, the results in Panel B indicate that we cannot reject the null of reduced rank for all models, except for C-ML, and that all models are rejected by the HJ-distance and *LM* specification tests. This clearly points to the need of statistical procedures that are robust to model misspecification and weak identification.

Table 8 about here

Panel D of Table 8 shows that all factors except for $vw \cdot ml$ do not survive the model selection procedure based on the misspecification-robust *t*-test.[13] This stands in sharp contrast to the results in Panel C of Table 8 where the *t*-test under correctly specified models is employed. However, our theoretical and simulation analyses clearly showed that relying on the *t*-test under correct specification is grossly inappropriate when the underlying model is misspecified and the factors are very weakly correlated with the returns on the test assets. As one example, consider CC-MY. In the final stage of the model selection procedure in Panel C, both *cg* and *my* seem to be priced. On the contrary, no factor in CC-MY survives the model selection procedure based on misspecification-robust *t*-tests. Taken together, these results serve as a warning signal to researchers that are interested in estimating and analyzing SDF parameters on nontraded risk factors.

6. Conclusion

It is well known that asset returns are, at best, only weakly correlated with many macroeconomic factors. Nonetheless, researchers in finance have typically relied on inference methods that are not

[12] We also estimated the CAPM, FF3, CCAPM and D-CCAPM models considered in the previous section using quarterly data and the results are very similar to the ones for the monthly application.

[13] Although *vw·ml* is in principle a nontraded factor, it is very highly correlated with the returns on the test assets. This is a situation in which, as discussed at the end of Section 1, the *t*-ratios under correctly specified and misspecified models are likely to deliver a similar answer.

robust to weak identification and model misspecification when evaluating the incremental pricing ability of these factors. Our paper demonstrates that when a model is misspecified, the standard t-test of statistical significance will lead us to erroneously conclude, with high probability, that a useless factor is relevant and should be included in the model. Importantly, we show that the t-test of statistical significance will be valid only if it is computed using misspecification-robust standard errors. Furthermore, we argue that the presence of a useless factor affects the inference on the remaining model parameters and the test of correct specification. In particular, when a useless factor is present in the model, the limiting distributions of the t-statistics for the useful factors are non-standard and the HJ-distance specification test is inconsistent.

In order to overcome these problems, we propose an easy-to implement sequential model selection procedure based on misspecification-robust t-tests that restores the standard inference on the parameters of interest. We show via simulations that the proposed procedure is effective in eliminating useless factors as well as factors that do not improve the pricing ability of the model.

Finally, we employ our methodology to investigate the empirical performance of several prominent asset pricing models with traded and nontraded factors. While the market factor and the book-to-market factor of Fama and French (1993) are often found to be priced, the statistical evidence on the pricing ability of many nontraded factors is rather weak when using the model selection procedure based on misspecification-robust t-tests.

References

Benjamini, Y., and Y. Hochberg, 1995, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society Series B*, 57, 289–300.

Burnside, C. A., 2010, "Identification and Inference in Linear Stochastic Discount Factor Models," NBER working paper 16634.

Burnside, C. A., 2011, "The Cross-Section of Foreign Currency Risk Premia and Consumption Growth Risk: Comment," *American Economic Review*, 101, 3456–3476.

Campbell, J. Y., 1996, "Understanding Risk and Return," *Journal of Political Economy*, 104, 298–345.

Cochrane, J. H., 1996, "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy*, 104, 572–621.

Fama, E. F., and J. D. MacBeth, 1973, "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 71, 607–636.

Fama, E. F., and K. R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

Goeman, J. J., and A. Solari, 2010, "The Sequential Rejection Principle of Familywise Error Control," *Annals of Statistics*, 38, 3782–3810.

Gospodinov, N., R. Kan, and C. Robotti, 2013, "Chi-Squared Tests for Evaluation and Comparison of Asset Pricing Models," *Journal of Econometrics*, 173, 108–125.

Hansen, L. P., and R. Jagannathan, 1997, "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance*, 52, 557–590.

Jagannathan, R., and Z. Wang, 1996, "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance*, 51, 3–53.

Kan, R., and C. Zhang, 1999a, "Two-Pass Tests of Asset Pricing Models with Useless Factors," *Journal of Finance*, 54, 204–235.

Kan, R., and C. Zhang, 1999b, "GMM Tests of Stochastic Discount Factor Models with Useless Factors," *Journal of Financial Economics*, 54, 103–127.

Kan, R., and C. Robotti, 2008, "Specification Tests of Asset Pricing Models Using Excess Returns," *Journal of Empirical Finance*, 15, 816–838.

Kan, R., and C. Robotti, 2009, "Model Comparison Using the Hansen-Jagannathan Distance," *Review of Financial Studies*, 22, 3449–3490.

Kan, R., C. Robotti, and J. Shanken, 2013, "Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology," *Journal of Finance*, 68, 2615–2647.

Khalaf, L. A., and H. Schaller, 2011, "How Fama-MacBeth Can Go Wrong – And an Informative Solution," working paper, Carleton University.

Kleibergen, F., and R. Paap, 2006, "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97–126.

Kleibergen, F., 2009, "Tests of Risk Premia in Linear Factor Models," *Journal of Econometrics*, 149, 149–173.

Kleibergen, F., 2010, "Reality Checks for and of Factor Pricing," working paper, Brown University.

Lettau, M., and S. C. Ludvigson, 2001, "Resurrecting the (C)CAPM: A Cross-Sectional Test when Risk Premia Are Time-Varying," *Journal of Political Economy*, 109, 1238–1287.

Lewellen, J. W., S. Nagel, and J. Shanken, 2010, "A Skeptical Appraisal of Asset-Pricing Tests," *Journal of Financial Economics*, 96, 175–194.

Li, Q., M. Vassalou, and Y. Xing, 2006, "Sector Investment Growth Rates and the Cross-Section of Equity Returns," *Journal of Business*, 79, 1637–1665.

Liu, W., 2006, "A Liquidity-Augmented Capital Asset Pricing Model, " *Journal of Financial Economics*, 82, 631–671.

Ludvigson, S. C., 2013, "Advances in Consumption-Based Asset Pricing: Empirical Tests," Chapter 12 in *Handbook of the Economics of Finance*, G. M. Constantinides, M. Harris and R. M. Stulz (eds.), Vol. 2, North Holland: Amsterdam, 799–906.

Lustig, H., and S. Van Nieuwerburgh, 2005, "Housing Collateral, Consumption Insurance, and Risk Premia: An Empirical Perspective," *Journal of Finance*, 60, 1167–1219.

Merton, R. C., 1973, "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867–887.

Pastor, L., and R. F. Stambaugh, 2003, "Liquidity Risk and Expected Stock Returns, " *Journal of Political Economy*, 111, 642–685.

Petkova, R., 2006, "Do the Fama-French Factors Proxy for Innovations in Predictive Variables?," *Journal of Finance*, 61, 581–612.

Ratsimalahelo, Z., 2002, "Rank Test Based on Matrix Perturbation Theory," working paper, University of Franche-Comté.

Romano, J. P., A. M. Shaikh, and M. Wolf, 2008, "Formalized Data Snooping Based on Generalized Error Rates," *Econometric Theory*, 24, 404–447.

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

Stock, J. H., and M. W. Watson, 2002, "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

Vissing–Jørgensen, A., and O. P. Attanasio, 2003, "Stock-Market Participation, Intertemporal Substitution, and Risk-Aversion," *American Economic Review Papers and Proceedings*, 93, 383–391.

White, H., 1982, "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

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

Table 1
Empirical size of the t-tests in a model with a useful factor

Panel A: Correctly specified model

		$\hat{\gamma}_0$			$\hat{\gamma}_1$		
t-test	T	10%	5%	1%	10%	5%	1%
t_c	200	0.168	0.139	0.108	0.098	0.049	0.010
	600	0.134	0.098	0.063	0.099	0.049	0.009
	1,000	0.123	0.081	0.044	0.101	0.049	0.009
	∞	0.100	0.050	0.010	0.100	0.050	0.010
t_m	200	0.168	0.139	0.108	0.097	0.049	0.010
	600	0.134	0.098	0.063	0.099	0.049	0.009
	1,000	0.123	0.081	0.044	0.101	0.049	0.009
	∞	0.100	0.050	0.010	0.100	0.050	0.010

Panel B: Misspecified model

		$\hat{\gamma}_0$			$\hat{\gamma}_1$		
t-test	T	10%	5%	1%	10%	5%	1%
t_c	200	0.173	0.144	0.112	0.099	0.048	0.009
	600	0.137	0.099	0.063	0.098	0.049	0.010
	1000	0.123	0.081	0.043	0.098	0.048	0.010
	∞	0.100	0.050	0.010	0.100	0.050	0.010
t_m	200	0.173	0.144	0.112	0.099	0.048	0.009
	600	0.137	0.099	0.063	0.098	0.049	0.010
	1000	0.123	0.081	0.043	0.098	0.048	0.010
	∞	0.100	0.050	0.010	0.100	0.050	0.010

The table presents the empirical size of the t-tests of $H_0 : \gamma_i = \gamma_i^*$ $(i = 0, 1)$ in a model with a constant and a useful factor. γ_0 is the coefficient on the constant term and γ_1 is the coefficient on the useful factor. t_c denotes the t-test constructed under the assumption of correct model specification and t_m denotes the misspecification-robust t-test. We report results for different levels of significance (10%, 5% and 1%) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns, the 17 Fama-French industry portfolio returns and the one-month T-bill rate for the period 1959:2–2012:12. The various t-statistics are compared to the critical values from a standard normal distribution.

Table 2
Empirical size of the t-tests in a model with a useless factor

Panel A: Correctly specified model

t-test	T	$\hat{\gamma}_0$ 10%	$\hat{\gamma}_0$ 5%	$\hat{\gamma}_0$ 1%	$\hat{\gamma}_1$ 10%	$\hat{\gamma}_1$ 5%	$\hat{\gamma}_1$ 1%
t_c	200	0.008	0.002	0.000	0.131	0.067	0.012
	600	0.002	0.000	0.000	0.099	0.046	0.006
	1000	0.002	0.000	0.000	0.098	0.045	0.007
	∞	0.001	0.000	0.000	0.088	0.039	0.005
t_m	200	0.001	0.000	0.000	0.036	0.012	0.001
	600	0.000	0.000	0.000	0.022	0.006	0.000
	1000	0.000	0.000	0.000	0.022	0.006	0.000
	∞	0.000	0.000	0.000	0.018	0.004	0.000

Panel B: Misspecified model

t-test	T	$\hat{\gamma}_0$ 10%	$\hat{\gamma}_0$ 5%	$\hat{\gamma}_0$ 1%	$\hat{\gamma}_1$ 10%	$\hat{\gamma}_1$ 5%	$\hat{\gamma}_1$ 1%
t_c	200	0.020	0.006	0.000	0.328	0.234	0.103
	600	0.020	0.005	0.000	0.472	0.383	0.228
	1000	0.024	0.007	0.000	0.557	0.476	0.327
	∞	0.088	0.039	0.005	1.000	1.000	1.000
t_m	200	0.002	0.000	0.000	0.081	0.036	0.005
	600	0.001	0.000	0.000	0.081	0.038	0.006
	1000	0.001	0.000	0.000	0.086	0.041	0.007
	∞	0.001	0.000	0.000	0.100	0.050	0.010

The table presents the empirical size of the t-tests of $H_0 : \gamma_i = \gamma_i^*$ $(i = 0, 1)$ in a model with a constant and a useless factor. γ_0 is the coefficient on the constant term and γ_1 is the coefficient on the useless factor. t_c denotes the t-test constructed under the assumption of correct model specification and t_m denotes the misspecification-robust t-test. We report results for different levels of significance (10%, 5% and 1%) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns, the 17 Fama-French industry portfolio returns and the one-month T-bill rate for the period 1959:2–2012:12. The various t-statistics are compared to the critical values from a standard normal distribution. The rejection rates for the limiting case ($T = \infty$) are based on the asymptotic distributions given in Theorem 2 in the online appendix.

Table 3
Empirical size of the t-tests in a model with a useful and a useless factor

Panel A: Correctly specified model

		$\hat{\gamma}_0$			$\hat{\gamma}_1$			$\hat{\gamma}_2$		
t-test	T	10%	5%	1%	10%	5%	1%	10%	5%	1%
t_c	200	0.054	0.023	0.007	0.094	0.045	0.008	0.128	0.066	0.012
	600	0.059	0.028	0.008	0.096	0.047	0.009	0.100	0.046	0.006
	1000	0.056	0.025	0.007	0.096	0.046	0.009	0.095	0.043	0.005
	∞	0.052	0.020	0.002	0.096	0.047	0.009	0.088	0.039	0.005
t_m	200	0.027	0.011	0.004	0.090	0.042	0.007	0.036	0.012	0.001
	600	0.037	0.016	0.006	0.092	0.045	0.008	0.022	0.006	0.000
	1000	0.037	0.016	0.005	0.092	0.044	0.008	0.019	0.005	0.000
	∞	0.037	0.014	0.002	0.092	0.045	0.008	0.018	0.004	0.000

Panel B: Misspecified model

		$\hat{\gamma}_0$			$\hat{\gamma}_1$			$\hat{\gamma}_2$		
t-test	T	10%	5%	1%	10%	5%	1%	10%	5%	1%
t_c	200	0.056	0.023	0.005	0.095	0.046	0.008	0.319	0.227	0.098
	600	0.057	0.023	0.004	0.095	0.046	0.009	0.464	0.376	0.224
	1000	0.057	0.021	0.003	0.094	0.046	0.009	0.550	0.469	0.319
	∞	0.088	0.039	0.005	0.088	0.039	0.005	1.000	1.000	1.000
t_m	200	0.016	0.006	0.002	0.086	0.040	0.006	0.080	0.036	0.005
	600	0.013	0.005	0.002	0.079	0.037	0.006	0.082	0.038	0.006
	1000	0.009	0.003	0.001	0.071	0.032	0.005	0.087	0.042	0.007
	∞	0.001	0.000	0.000	0.001	0.000	0.000	0.100	0.050	0.010

The table presents the empirical size of the t-tests of $H_0 : \gamma_i = \gamma_i^*$ $(i = 0, 1, 2)$ in a model with a constant, a useful and a useless factor. γ_0 is the coefficient on the constant term, γ_1 is the coefficient on the useful factor, and γ_2 is the coefficient on the useless factor. t_c denotes the t-test constructed under the assumption of correct model specification and t_m denotes the misspecification-robust t-test. We report results for different levels of significance (10%, 5% and 1%) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns, the 17 Fama-French industry portfolio returns and the one-month T-bill rate for the period 1959:2–2012:12. The various t-tests are compared to the critical values from a standard normal distribution. The rejection rates for the limiting case ($T = \infty$) are based on the asymptotic distributions given in Theorem 2 in the online appendix.

Table 4
Survival rates of risk factors: two useful, one unpriced and one useless factors

Panel A: Correctly specified model

	Useful ($\gamma_1^* \neq 0$)		Useful ($\gamma_2^* \neq 0$)		Useful ($\gamma_3^* = 0$)		Useless		Prob.	
T	$t_c(\hat{\gamma}_1)$	$t_m(\hat{\gamma}_1)$	$t_c(\hat{\gamma}_2)$	$t_m(\hat{\gamma}_2)$	$t_c(\hat{\gamma}_3)$	$t_m(\hat{\gamma}_3)$	$t_c(\hat{\gamma}_4)$	$t_m(\hat{\gamma}_4)$	MS_c	MS_m
200	0.519	0.517	0.674	0.661	0.027	0.023	0.023	0.002	0.049	0.025
600	0.986	0.987	0.999	0.999	0.014	0.013	0.009	0.001	0.023	0.014
1000	1.000	1.000	1.000	1.000	0.013	0.013	0.008	0.000	0.022	0.013

Panel B: Misspecified model

	Useful ($\gamma_1^* \neq 0$)		Useful ($\gamma_2^* \neq 0$)		Useful ($\gamma_3^* = 0$)		Useless		Prob.	
T	$t_c(\hat{\gamma}_1)$	$t_m(\hat{\gamma}_1)$	$t_c(\hat{\gamma}_2)$	$t_m(\hat{\gamma}_2)$	$t_c(\hat{\gamma}_3)$	$t_m(\hat{\gamma}_3)$	$t_c(\hat{\gamma}_4)$	$t_m(\hat{\gamma}_4)$	MS_c	MS_m
200	0.498	0.502	0.648	0.633	0.033	0.025	0.108	0.008	0.138	0.033
600	0.971	0.981	0.993	0.996	0.018	0.014	0.206	0.008	0.220	0.022
1000	0.996	0.998	0.999	0.999	0.015	0.012	0.295	0.009	0.306	0.021

The table presents the survival rates of the useful and useless factors in a model with a constant, two useful factors (with $\gamma_1^* \neq 0$ and $\gamma_2^* \neq 0$), a useful factor that does not contribute to pricing (with $\gamma_3^* = 0$) and a useless factor (with γ_4^* unidentified). The sequential procedure is implemented by using the misspecification-robust t-tests ($t_m(\hat{\gamma}_i)$ column) as well as the t-tests under correctly specified models ($t_c(\hat{\gamma}_i)$ column). The false discovery rate of the multiple testing procedure is controlled using the Bonferroni method. The last two columns of the table report the probability that at least one useless or unpriced useful factor survives using the t-tests under correctly specified models (MS_c) and misspecification-robust t-tests (MS_m). The nominal level of the sequential testing procedure is set equal to 5%. Panels A and B are for correctly specified and misspecified models, respectively. We report results for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns, the 17 Fama-French industry portfolio returns and the one-month T-bill rate for the period 1959:2–2012:12.

Table 5
Survival rates of risk factors: two useful and two useless factors

Panel A: Correctly specified model

	Useful ($\gamma_1^* \neq 0$)		Useful ($\gamma_2^* \neq 0$)		Useless		Useless		Prob.	
T	$t_c(\hat{\gamma}_1)$	$t_m(\hat{\gamma}_1)$	$t_c(\hat{\gamma}_2)$	$t_m(\hat{\gamma}_2)$	$t_c(\hat{\gamma}_3)$	$t_m(\hat{\gamma}_3)$	$t_c(\hat{\gamma}_4)$	$t_m(\hat{\gamma}_4)$	MS_c	MS_m
200	0.520	0.521	0.666	0.658	0.023	0.002	0.024	0.002	0.046	0.004
600	0.988	0.989	0.999	0.999	0.009	0.001	0.010	0.001	0.019	0.001
1000	1.000	1.000	1.000	1.000	0.008	0.000	0.008	0.000	0.016	0.001

Panel B: Misspecified model

	Useful ($\gamma_1^* \neq 0$)		Useful ($\gamma_2^* \neq 0$)		Useless		Useless		Prob.	
T	$t_c(\hat{\gamma}_1)$	$t_m(\hat{\gamma}_1)$	$t_c(\hat{\gamma}_2)$	$t_m(\hat{\gamma}_2)$	$t_c(\hat{\gamma}_3)$	$t_m(\hat{\gamma}_3)$	$t_c(\hat{\gamma}_4)$	$t_m(\hat{\gamma}_4)$	MS_c	MS_m
200	0.300	0.308	0.398	0.389	0.105	0.007	0.106	0.007	0.205	0.015
600	0.853	0.900	0.921	0.951	0.202	0.009	0.204	0.009	0.384	0.018
1000	0.959	0.983	0.981	0.992	0.279	0.010	0.282	0.010	0.517	0.020

The table presents the survival rates of the useful and useless factors in a model with a constant, two useful factors (with $\gamma_1^* \neq 0$ and $\gamma_2^* \neq 0$), and two useless factors (with γ_3^* and γ_4^* unidentified). The sequential procedure is implemented by using the misspecification-robust t-tests ($t_m(\hat{\gamma}_i)$ column) as well as the t-tests under correctly specified models ($t_c(\hat{\gamma}_i)$ column). The false discovery rate of the multiple testing procedure is controlled using the Bonferroni method. The last two columns of the table report the probability that at least one useless factor survives using the t-tests under correctly specified models (MS_c) and misspecification-robust t-tests (MS_m). The nominal level of the sequential testing procedure is set equal to 5%. Panels A and B are for correctly specified and misspecified models, respectively. We report results for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns, the 17 Fama-French industry portfolio returns and the one-month T-bill rate for the period 1959:2–2012:12.

Table 6
Survival rates when a linear combination of the factors is useless

Panel A: Correctly specified model

	Both factors survive		One factor survives		No factor survives	
T	t_c	t_m	t_c	t_m	t_c	t_m
200	0.025	0.002	0.250	0.251	0.726	0.746
600	0.015	0.001	0.680	0.688	0.305	0.311
1000	0.014	0.001	0.889	0.900	0.098	0.100

Panel B: Misspecified model

	Both factors survive		One factor survives		No factor survives	
T	t_c	t_m	t_c	t_m	t_c	t_m
200	0.138	0.013	0.229	0.255	0.633	0.732
600	0.277	0.015	0.505	0.685	0.217	0.300
1000	0.375	0.017	0.566	0.888	0.059	0.095

The table presents the probability that both factors survive, only one factor survives, and no factor survives in a model in which a linear combination of two useful factors is useless. The sequential procedure is implemented by using the misspecification-robust t-test (t_m column) as well as the t-test under correctly specified models (t_c column). The false discovery rate of the multiple testing procedure is controlled using the Bonferroni method. The nominal level of the sequential testing procedure is set equal to 5%. Panels A and B are for correctly specified and misspecified models, respectively. We report results for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns, the 17 Fama-French industry portfolio returns and the one-month T-bill rate for the period 1959:2–2012:12.

Table 7
Monthly analysis of some popular linear asset pricing models

Panel A: Rank test for individual factors

Test	vw	smb	hml	$term$	def	div	rf	cay	lab	$vw{\cdot}cay$	$lab{\cdot}cay$	c_{nd}	$c_{nd}{\cdot}cay$	c_d
$\mathcal{W}^*$	205.7	199.9	189.7	51.5	88.0	186.2	42.0	54.0	54.8	83.1	49.7	54.6	43.0	58.6
p-val	0.000	0.000	0.000	0.149	0.000	0.000	0.470	0.101	0.088	0.000	0.193	0.092	0.430	0.046

Panel B: HJ-distance, Lagrange multiplier, and rank tests

Model	$\hat{\delta}$	p-val	LM	p-val	$\mathcal{W}^*$	p-val
CAPM	0.523	0.000	139.766	0.000	205.7	0.000
FF3	0.487	0.000	121.265	0.000	190.7	0.000
ICAPM	0.446	0.005	73.171	0.000	50.5	0.084
C-LAB	0.481	0.000	78.892	0.000	31.1	0.780
CCAPM	0.513	0.000	112.888	0.000	54.6	0.092
CC-CAY	0.484	0.000	81.523	0.000	40.9	0.430
D-CCAPM	0.510	0.000	114.966	0.000	46.4	0.225

Table 7 (continued)
Monthly analysis of some popular linear asset pricing models

Panel C: Model selection procedure using standard errors under correct model specification

Model	*vw*	*smb*	*hml*	*term*	*def*	*div*	*rf*	*cay*	*lab*	*vw·cay*	*lab·cay*	c_{nd}	c_{nd}·*cay*	c_d
CAPM	**-2.64**													
FF3	-3.43	-0.57	-4.59											
	-3.69		**-4.62**											
ICAPM	1.27			-3.35	0.62	1.32	0.56							
	1.24			-4.00	0.43	1.25								
	1.17			-4.02		1.19								
				-4.11		0.49								
				-4.32										
C-LAB	-0.72							0.15	0.59	2.46	-2.22			
	-0.71								0.60	2.45	-3.36			
	-0.80									2.39	-3.43			
										2.38	-3.63			
											-3.79			
CCAPM												**-3.23**		
CC-CAY								-3.13				-3.05	0.97	
								-3.24				**-3.03**		
D-CCAPM	-1.12											-2.41		-0.67
	-1.14											-2.58		
												-3.23		

Panel D: Model selection procedure using model misspecification-robust standard errors

Model	*vw*	*smb*	*hml*	*term*	*def*	*div*	*rf*	*cay*	*lab*	*vw·cay*	*lab·cay*	c_{nd}	c_{nd}·*cay*	c_d
CAPM	**-2.64**													
FF3	-3.35	-0.53	-4.40											
	-3.68		**-4.45**											
ICAPM	0.97			-2.43	0.49	1.03	0.42							
	0.95			-3.05	0.34	0.99								
	0.93			-3.13		0.98								
				-3.21		0.48								
				-3.54										
C-LAB	-0.64							0.08	0.36	1.88	-1.09			
	-0.61								0.37	1.88	-1.81			
	-0.71									1.87	-1.80			
										1.89	-2.07			
											-1.98			
CCAPM												-1.75		
CC-CAY								-1.85				-1.88	0.47	
								-1.89				-1.84		
								-1.89						
D-CCAPM	-0.90											-1.22		-0.41
	-0.90											-1.27		
												-1.75		

Table 7 (continued)
Monthly analysis of some popular linear asset pricing models

The table presents the estimation and testing results of the seven asset pricing models described in Section 5.1.1. The models are estimated using monthly gross returns on the 25 size and book-to-market Fama-French portfolios, the 17 Fama-French industry portfolios and the one-month T-bill. The data are from 1959:2 until 2012:12. Panel A reports the rank restriction test ($\mathcal{W}^*$) and its p-value (p-val) of the null that $E[x_t(1,\ f_{it})]$ has a column rank of one. In Panel B, we report the sample HJ-distance ($\hat{\delta}$), the Lagrange multiplier (LM) test, and the rank restriction test ($\mathcal{W}^*$) with the corresponding p-values (p-val) for each model. The t-tests of the model selection procedures based on the standard errors under correct model specification and model misspecification are in Panels C and D, respectively. We use boldface to highlight those cases in which the factors survive the model selection procedure at the 5% significance level using the Bonferroni adjustment.

Table 8
Quarterly Analysis of Some Popular Linear Asset Pricing Models

Panel A: Rank Test for Individual Factors

Test	vw	$vw{\cdot}ml$	c_{nd}	cay	$c_{nd}{\cdot}cay$	my	$c_{nd}{\cdot}my$	i_h	i_c	i_{nc}
W^*	92.2	92.5	46.0	43.9	33.6	39.0	39.6	42.4	33.5	28.1
p-val	0.000	0.000	0.066	0.098	0.438	0.218	0.199	0.128	0.444	0.708

Panel B: HJ-Distance, Lagrange Multiplier, and Rank Tests

Model	$\hat{\delta}$	p-val	LM	p-val	W^*	p-val
C-ML	0.790	0.000	95.378	0.000	54.2	0.008
CC-CAY	0.761	0.000	73.471	0.000	29.8	0.527
CC-MY	0.751	0.002	86.353	0.000	40.8	0.111
SIM	0.776	0.000	87.339	0.000	24.3	0.800

Panel C: t-tests Using Standard Errors Under Correct Model Specification

Model	vw	$vw{\cdot}ml$	c_{nd}	cay	$c_{nd}{\cdot}cay$	my	$c_{nd}{\cdot}my$	i_h	i_c	i_{nc}
C-ML	1.91	-1.97								
		-2.99								
CC-CAY			-2.77	-2.17	0.51					
			-2.95	-2.33						
			-3.36							
CC-MY			-2.74			2.51	-0.99			
			-2.91			**2.41**				
SIM								-2.89	-0.59	0.26
								-3.04	-0.60	
								-3.42		

Panel D: t-tests Using Model Misspecification-Robust Standard Errors

Model	vw	$vw{\cdot}ml$	c_{nd}	cay	$c_{nd}{\cdot}cay$	my	$c_{nd}{\cdot}my$	i_h	i_c	i_{nc}
C-ML	1.34	-1.38								
		-2.97								
CC-CAY			-1.80	-1.57	0.31					
			-1.92	-1.64						
			-1.99							
CC-MY			-1.88			1.64	-0.64			
			-1.97			1.51				
			-1.99							
SIM								-1.92	-0.38	0.15
								-2.04	-0.38	
								-2.27		

Table 8 (continued)
Quarterly analysis of some popular linear asset pricing models
The table presents the estimation and testing results of the four asset pricing models described in Section 5.2. The models are estimated using quarterly gross returns on the 6 size and book-to-market Fama-French portfolios, the 17 Fama-French industry portfolios, the 10 momentum portfolios and the one-month T-bill. The data are from 1952:Q2 until 2012:Q4. Panel A reports the rank restriction test ($\mathcal{W}^*$) and its p-value (p-val) of the null that $E[x_t(1,\ f_{it})]$ has a column rank of one. In Panel B, we report the sample HJ-distance ($\hat{\delta}$), the Lagrange multiplier (LM) test, and the rank restriction test ($\mathcal{W}^*$) with the corresponding p-values (p-val) for each model. The t-tests of the model selection procedures based on the standard errors under correct model specification and model misspecification are in Panels C and D, respectively. We use boldface to highlight those cases in which the factors survive the model selection procedure at the 5% significance level using the Bonferroni adjustment.

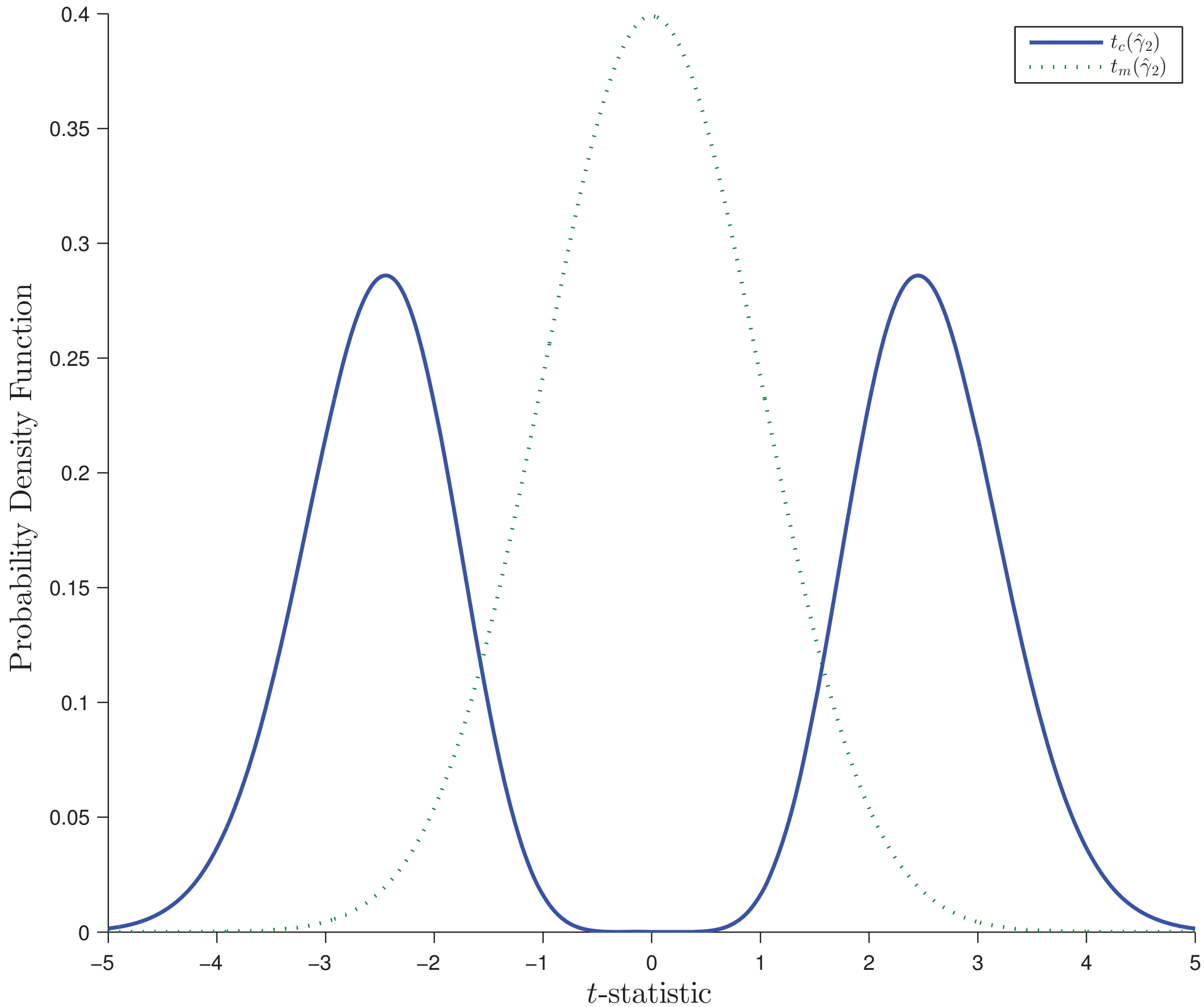

Figure 1
Asymptotic distributions of $t_c(\hat{\gamma}_2)$ and $t_m(\hat{\gamma}_2)$ under misspecified models. The figure presents the probability density functions of the limiting distributions of $t_c(\hat{\gamma}_2)$ and $t_m(\hat{\gamma}_2)$, the t-statistics for the useless factor that use standard errors constructed under correctly specified and potentially misspecified models, respectively, for $N-K=7$ (see part (b) of Theorem 2 in the online appendix).

Online Appendix for

"Misspecification-Robust Inference in Linear Asset Pricing Models with Irrelevant Risk Factors"

NIKOLAY GOSPODINOV, RAYMOND KAN, and CESARE ROBOTTI

September 2013

In this online appendix, we derive the limiting distributions of the parameter estimates and their corresponding t-statistics as well as the HJ-distance test for correct model specification when a useless factor is present in the model. We follow the same notation as in the paper.

Theoretical Results for Gross Returns

First, we provide theoretical results for the gross returns case. Consider a candidate SDF which is given by

$$y_t = \tilde{f}_t'\gamma_1 + g_t\gamma_2, \tag{1}$$

where $\tilde{f}_t = [1,\ f_t']'$, f_t is a $(K-1)$-vector of useful risk factors and g_t denotes a useless factor which is independent of x_t and f_t for all time periods. For ease of exposition, we assume that $E[g_t] = 0$ and $\text{Var}[g_t] = 1$.[1] Let $B = E[x_t\tilde{f}_t']$ and note that the independence between g_t and x_t implies

$$d = E[x_t g_t] = 0_N \tag{2}$$

and

$$E[x_t x_t' g_t^2] = E[E[x_t x_t'|g_t]g_t^2] = UE[g_t^2] = U. \tag{3}$$

Now let $D = [B,\ d]$, $\gamma = [\gamma_1',\ \gamma_2]'$, $e(\gamma) = D\gamma - q$, $\hat{d} = \frac{1}{T}\sum_{t=1}^{T} x_t g_t$, $\hat{B} = \frac{1}{T}\sum_{t=1}^{T} x_t\tilde{f}_t'$ and $\hat{D} = [\hat{B},\ \hat{d}]$. Note that since $d = 0_N$, the vector of pricing errors

$$e(\gamma) = B\gamma_1 + d\gamma_2 - q = B\gamma_1 - q \tag{4}$$

is independent of the choice of γ_2. The pseudo-true value of the SDF parameter associated with the useless factor (γ_2^*) cannot be identified. In the following, we set $\gamma_2^* = 0$, which is a natural choice because in Theorem 1 we will show that $\hat{\gamma}_2$ is symmetrically distributed around zero. While the pseudo-true value γ_2^* is not identified, the sample estimates of the SDF parameters are always identified and they are given by

$$\hat{\gamma} = (\hat{D}'\hat{U}^{-1}\hat{D})^{-1}\hat{D}'\hat{U}^{-1}q. \tag{5}$$

[1] This assumption does not affect our asymptotic results on statistical inference for the slope parameters of the linear SDF. It does, however, affect the limiting distribution of the estimated SDF's intercept and the statistical inference on it. The limiting results derived under a generic mean and variance of the useless factor are available from the authors upon request.

Note that the estimator in (5) can be obtained equivalently by running an ordinary least squares (OLS) regression of $\hat{U}^{-\frac{1}{2}}q$ on $\hat{U}^{-\frac{1}{2}}\hat{B}$ and $\hat{U}^{-\frac{1}{2}}\hat{d}$. In order to construct $\hat{\gamma}_2$, we can project $\hat{U}^{-\frac{1}{2}}q$ and $\hat{U}^{-\frac{1}{2}}\hat{d}$ on $\hat{U}^{-\frac{1}{2}}\hat{B}$, and then regress the residuals from the first projection on the residuals from the second projection. It follows that

$$\hat{\gamma}_2 = \frac{\hat{d}'\hat{U}^{-\frac{1}{2}}[I_N - \hat{U}^{-\frac{1}{2}}\hat{B}(\hat{B}'\hat{U}^{-1}\hat{B})^{-1}\hat{B}'\hat{U}^{-\frac{1}{2}}]\hat{U}^{-\frac{1}{2}}q}{\hat{d}'\hat{U}^{-\frac{1}{2}}[I_N - \hat{U}^{-\frac{1}{2}}\hat{B}(\hat{B}'\hat{U}^{-1}\hat{B})^{-1}\hat{B}'\hat{U}^{-\frac{1}{2}}]\hat{U}^{-\frac{1}{2}}\hat{d}}. \tag{6}$$

Similarly, the parameter vector $\hat{\gamma}_1$ is obtained by projecting $\hat{U}^{-\frac{1}{2}}q$ and $\hat{U}^{-\frac{1}{2}}\hat{B}$ on $\hat{U}^{-\frac{1}{2}}\hat{d}$ and then regressing the residuals from the first projection on the residuals from the second projection, which yields

$$\begin{aligned}\hat{\gamma}_1 &= (\hat{B}'\hat{U}^{-\frac{1}{2}}[I_N - \hat{U}^{-\frac{1}{2}}\hat{d}(\hat{d}'\hat{U}^{-1}\hat{d})^{-1}\hat{d}'\hat{U}^{-\frac{1}{2}}]\hat{U}^{-\frac{1}{2}}\hat{B})^{-1} \\ &\quad \times \hat{B}'\hat{U}^{-\frac{1}{2}}[I_N - \hat{U}^{-\frac{1}{2}}\hat{d}(\hat{d}'\hat{U}^{-1}\hat{d})^{-1}\hat{d}'\hat{U}^{-\frac{1}{2}}]\hat{U}^{-\frac{1}{2}}q. \end{aligned} \tag{7}$$

We make the following assumptions.

Assumption 1. *Assume that (i) $N > K+1$; (ii) $[x_t',\ f_t',\ g_t]'$ are jointly stationary and ergodic processes with finite fourth moments; (iii) $e_t(\gamma_1^*) - e(\gamma_1^*)$ forms a martingale difference sequence; and (iv) the matrices B $(N \times K)$ and D $(N\times(K+1))$ have a column rank K.*

Assumption 2. *Let $\epsilon_t = x_t - B(E[\tilde{f}_t\tilde{f}_t'])^{-1}\tilde{f}_t$ and assume that $E[\epsilon_t\epsilon_t'|\tilde{f}_t] = \Sigma$ (conditional homoskedasticity).*

Our first results are concerned with the limiting behavior of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ under correctly specified and misspecified models. We adopt the following notation. Let $\tilde{B} = U^{-\frac{1}{2}}B$, $\tilde{q} = U^{-\frac{1}{2}}q$, and P be an $N \times (N-K)$ orthonormal matrix whose columns are orthogonal to $\tilde{B}$ so that $PP' = I_N - \tilde{B}(\tilde{B}'\tilde{B})^{-1}\tilde{B}'$. Also, let $z \sim N(0_N, I_N)$ and $y \sim N(0_N, U^{-\frac{1}{2}}SU^{-\frac{1}{2}})$, and they are independent of each other. Finally, we define $w = P'z \sim N(0_{N-K}, I_{N-K})$, $s = (\tilde{q}'Pw)/(\tilde{q}'PP'\tilde{q})^{\frac{1}{2}} \sim N(0,1)$, $u = P'y \sim N(0_{N-K}, V_u)$ with $V_u = P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P$, and $r = (\tilde{B}'\tilde{B})^{-\frac{1}{2}}\tilde{B}'y \sim N(0_K, V_r)$ with $V_r = (\tilde{B}'\tilde{B})^{-\frac{1}{2}}\tilde{B}'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}\tilde{B}(\tilde{B}'\tilde{B})^{-\frac{1}{2}}$.

Theorem 1. *Assume that the conditions in Assumption 1 are satisfied.*

(a) *If $\delta = 0$, i.e., the model is correctly specified, we have*

$$\sqrt{T}(\hat{\gamma}_1 - \gamma_1^*) \xrightarrow{d} (\tilde{B}'\tilde{B})^{-\frac{1}{2}}\left[r - \frac{w'u}{w'w}(\tilde{B}'\tilde{B})^{-\frac{1}{2}}\tilde{B}'z\right], \tag{8}$$

and

$$\hat{\gamma}_2 \xrightarrow{d} \frac{w'u}{w'w}. \tag{9}$$

(b) *If $\delta > 0$, i.e., the model is misspecified, we have*

$$\hat{\gamma}_1 - \gamma_1^* \xrightarrow{d} -\frac{\delta s}{w'w}(\tilde{B}'\tilde{B})^{-1}\tilde{B}'z, \tag{10}$$

and

$$\frac{1}{\sqrt{T}}\hat{\gamma}_2 \xrightarrow{d} \frac{\delta s}{w'w}. \tag{11}$$

Proof. See the Appendix.

The results in Theorem 1 subsume the results in Proposition 1 in the paper and can be summarized as follows. First, for correctly specified models, Theorem 1 shows that $\hat{\gamma}_2$ converges to a bounded random variable rather than the constant zero.[2] While the parameter estimates for the useful factors are consistently estimable, they are asymptotically non-normally distributed. Second, the presence of a useless factor further exacerbates the inference problems when the model is misspecified. In this case, the estimator $\hat{\gamma}_1$ is inconsistent while the estimator $\hat{\gamma}_2$ diverges at rate $T^{\frac{1}{2}}$.

We next derive the limiting distributions of two types of t-statistics (as defined in the paper): (i) $t_c(\hat{\gamma}_{1i})$ of $H_0 : \gamma_{1i} = \gamma_{1i}^*$ for $i = 1, \ldots, K$, and $t_c(\hat{\gamma}_2)$ of $H_0 : \gamma_2 = 0$ that use standard errors obtained under the assumption that the model is correctly specified, and (ii) $t_m(\hat{\gamma}_{1i})$ of $H_0 : \gamma_{1i} = \gamma_{1i}^*$ for $i = 1, \ldots, K$, and $t_m(\hat{\gamma}_2)$ of $H_0 : \gamma_2 = 0$ that use standard errors under potentially misspecified models. The two types of t-statistics are based on the estimated covariance matrices $\hat{\Sigma}_{\hat{\gamma}}^0 = \frac{1}{T}\sum_{t=1}^T \hat{h}_t^0 \hat{h}_t^{0\prime}$ and $\hat{\Sigma}_{\hat{\gamma}} = \frac{1}{T}\sum_{t=1}^T \hat{h}_t \hat{h}_t'$, where

$$\hat{h}_t^0 = (\hat{D}'\hat{U}^{-1}\hat{D})^{-1}\hat{D}'\hat{U}^{-1}\hat{e}_t, \tag{12}$$

$$\hat{h}_t = \hat{h}_t^0 + (\hat{D}'\hat{U}^{-1}\hat{D})^{-1}([\tilde{f}_t', \; g_t]' - \hat{D}'\hat{U}^{-1}x_t)\hat{e}'\hat{U}^{-1}x_t, \tag{13}$$

$\hat{e}_t = x_t(\tilde{f}_t'\hat{\gamma}_1 + g_t\hat{\gamma}_2) - q$ and $\hat{e} = \frac{1}{T}\sum_{t=1}^T \hat{e}_t$.

The results presented below are driven, to a large extent, by the limiting behavior of the matrix $\hat{S} = \frac{1}{T}\sum_{t=1}^T \hat{e}_t\hat{e}_t'$. In the presence of a useless factor, the results in Theorem 1 imply that for

[2]The limiting random variable has mean zero and variance $\mathrm{tr}(V_u)/[(N-K)(N-K-2)]$, where $\mathrm{tr}(\cdot)$ is the trace operator.

misspecified models

$$\hat{e}_t = (T^{-\frac{1}{2}}\hat{\gamma}_2)(T^{\frac{1}{2}}x_t g_t) + O_p(1) \tag{14}$$

and

$$\frac{\hat{S}}{T} = (T^{-\frac{1}{2}}\hat{\gamma}_2)^2 U + o_p(1), \tag{15}$$

so $\hat{S}$ diverges at rate T. In contrast, for correctly specified models, we have

$$\hat{S} = S + \hat{\gamma}_2^2 U + o_p(1), \tag{16}$$

so that $\hat{S}$ converges to a random matrix.

In addition to the random variables and matrices defined before Theorem 1, we introduce the following notation. Let $\tilde{u} \sim N(0,1)$, $\tilde{r}_i \sim N(0,1)$, $\tilde{z}_i \sim N(0,1)$, $v \sim \chi^2_{N-K-1}$, and they are independent of each other and w. Theorem 2 and Corollary 1 (Proposition 2 in the paper) below provide the limiting distributions of the t-statistics under correctly specified and misspecified models.

Theorem 2.

(a) *Suppose that the conditions in Assumptions 1 and 2 hold.*[3] *If $\delta = 0$, i.e., the model is correctly specified, we have*

$$t_c(\hat{\gamma}_{1i}) \xrightarrow{d} \frac{\tilde{u}\tilde{z}_i + \sqrt{\lambda_i}\sqrt{w'w}\tilde{r}_i}{\left[\lambda_i w'w + \tilde{z}_i^2 + \tilde{u}^2\left(1 + \frac{\tilde{z}_i^2}{w'w}\right)\right]^{\frac{1}{2}}}, \tag{17}$$

$$t_m(\hat{\gamma}_{1i}) \xrightarrow{d} \frac{\tilde{u}\tilde{z}_i + \sqrt{\lambda_i}\sqrt{w'w}\tilde{r}_i}{\left[\lambda_i w'w + \tilde{z}_i^2 + \tilde{u}^2\left(1 + \frac{\tilde{z}_i^2}{w'w}\right) + \frac{\tilde{z}_i^2 v}{w'w}\right]^{\frac{1}{2}}}, \tag{18}$$

$$t_c(\hat{\gamma}_2) \xrightarrow{d} \frac{\tilde{u}}{\left(1 + \frac{\tilde{u}^2}{w'w}\right)^{\frac{1}{2}}}, \tag{19}$$

$$t_m(\hat{\gamma}_2) \xrightarrow{d} \frac{\tilde{u}}{\left(1 + \frac{\tilde{u}^2 + v}{w'w}\right)^{\frac{1}{2}}}, \tag{20}$$

where λ_i is a positive constant and its explicit expression is given in the Appendix.

[3] The limiting distribution of $t_c(\hat{\gamma}_2)$ does not depend on the conditional homoskedasticity assumption. The expressions for the limiting distributions of the other t-statistics under conditional heteroskedasticity are more involved, and the results are available upon request.

(b) *Suppose that the conditions in Assumption 1 hold and denote the sign operator by* $\text{sgn}(\cdot)$. *If* $\delta > 0$, *i.e., the model is misspecified, we have*

$$t_c(\hat{\gamma}_{1i}) \xrightarrow{d} \frac{\tilde{z}_i}{\left(1 + \frac{\tilde{z}_i^2}{w'w}\right)^{\frac{1}{2}}}, \tag{21}$$

$$t_m(\hat{\gamma}_{1i}) \xrightarrow{d} N\left(0, \frac{1}{4}\right), \tag{22}$$

$$t_c(\hat{\gamma}_2) \xrightarrow{d} \text{sgn}(s)\sqrt{w'w}, \tag{23}$$

$$t_m(\hat{\gamma}_2) \xrightarrow{d} N(0, 1). \tag{24}$$

Proof. See the Appendix.

Corollary 1.

(a) *Suppose that the conditions in Assumptions 1 and 2 hold. Then, for correctly specified models, the limiting distributions of* $t_c^2(\hat{\gamma}_{1i})$, $t_m^2(\hat{\gamma}_{1i})$, $t_c^2(\hat{\gamma}_2)$, *and* $t_m^2(\hat{\gamma}_2)$ *are stochastically dominated by* χ_1^2.

(b) *Suppose that the conditions in Assumption 1 hold. Then, for misspecified models, the limiting distributions of* $t_c^2(\hat{\gamma}_{1i})$ *and* $t_m^2(\hat{\gamma}_{1i})$ *are stochastically dominated by* χ_1^2.

Proof. See the Appendix.

Finally, it is instructive to investigate whether the presence of a useless factor affects the limiting behavior of the specification test based on the sample squared HJ-distance

$$\hat{\delta}^2 = \hat{e}'\hat{U}^{-1}\hat{e}. \tag{25}$$

In the absence of a useless factor, it is well known that under a correctly specified model (Jagannathan and Wang, 1996)

$$T\hat{\delta}^2 \xrightarrow{d} \sum_{i=1}^{N-K} \xi_i X_i, \tag{26}$$

where the X_i's are independent chi-squared random variables with one degree of freedom and the ξ_i's are the $N - K$ nonzero eigenvalues of

$$S^{\frac{1}{2}}U^{-1}S^{\frac{1}{2}} - S^{\frac{1}{2}}U^{-1}B(B'U^{-1}B)^{-1}B'U^{-1}S^{\frac{1}{2}}. \tag{27}$$

In practice, the specification test based on the HJ-distance is performed by comparing $T\hat{\delta}^2$ with the critical values of $\sum_{i=1}^{N-K} \hat{\xi}_i X_i$, where the $\hat{\xi}_i$'s are the nonzero eigenvalues of

$$\hat{S}^{\frac{1}{2}}\hat{U}^{-1}\hat{S}^{\frac{1}{2}} - \hat{S}^{\frac{1}{2}}\hat{U}^{-1}\hat{B}(\hat{B}'\hat{U}^{-1}\hat{B})^{-1}\hat{B}'\hat{U}^{-1}\hat{S}^{\frac{1}{2}}. \tag{28}$$

When the model is misspecified, Hansen, Heaton, and Luttmer (1995) show that the sample squared HJ-distance has a limiting normal distribution. However, in the presence of a useless factor, the above results do not hold. In the next theorem, we add to the existing literature (Kan and Zhang, 1999) by characterizing the limiting behavior of the sample squared HJ-distance in the presence of a useless factor.

Theorem 3. *Let* $Q_1 \sim \text{Beta}\left(\frac{N-K}{2}, \frac{1}{2}\right)$ *with density* $f_{Q_1}(\cdot)$, $Q_2 \sim \text{Beta}\left(\frac{N-K-1}{2}, \frac{1}{2}\right)$ *with density* $f_{Q_2}(\cdot)$ *and* c_α *be the* $100(1-\alpha)$-th *percentile of* χ^2_{N-K-1}.

(a) *Suppose that the assumptions in part (a) of Theorem 2 hold. If* $\delta = 0$, *we have*

$$T\hat{\delta}^2 \stackrel{d}{\to} E[(\tilde{f}_t'\gamma_1^*)^2]\chi^2_{N-K-1} \tag{29}$$

and the limiting probability of rejecting $H_0 : \delta^2 = 0$ *by the HJ-distance test of size* α *is*

$$\int_0^1 P\left[\chi^2_{N-K-1} > \frac{c_\alpha}{q}\right] f_{Q_1}(q)\mathrm{d}q < \alpha. \tag{30}$$

(b) *Suppose that the assumptions in Theorem 1 hold. If* $\delta > 0$, *we have*

$$\hat{\delta}^2 \stackrel{d}{\to} \delta^2 Q_2 \tag{31}$$

and the limiting probability of rejecting $H_0 : \delta^2 = 0$ *by the HJ-distance test of size* α *is*

$$\int_0^1 P\left[\chi^2_{N-K} > \frac{c_\alpha q}{1-q}\right] f_{Q_2}(q)\mathrm{d}q < 1. \tag{32}$$

Proof. See the Appendix.

An immediate consequence of the result in Theorem 3 is that the presence of a useless factor tends to distort the inference on the specification test as well. More specifically, part (b) of Theorem 3 reveals that the HJ-distance test of correct model specification is inconsistent under the alternative.

Note that the limiting probabilities of rejection in (30) and (32) are only functions of the significance level α and the degree of over-identification $N-K$. Figure 1 plots these probabilities for different significance levels ($\alpha = 0.01,\ 0.05$, and 0.1) and $N-K$ ranging from 2 to 20.

Figure 1 about here

The top panel of Figure 1 reveals that under a correctly specified model, the limiting probability of rejection of the HJ-distance test is below its nominal level when a useless factor is present. When the model is misspecified, the bottom panel of Figure 1 shows that the probability of rejection of the HJ-distance test will not approach one even in large samples. In fact, there is a nonzero probability that the HJ-distance test will favor the null of correct specification, and this probability is particularly high when $N-K$ is small. As a result, the presence of a useless factors makes it more difficult for the HJ-distance test to detect a misspecified model.

Theoretical Results for Excess Returns

In the following analysis, we provide theoretical results for the excess returns case. The proofs are similar to the gross returns case and are omitted but are available from the authors upon request.

Let x_t be the excess returns on N test assets at time t with mean μ and covariance matrix V. It is well known that when only excess returns are used as test assets, it is not possible to identify the mean of the candidate SDF and some normalization of the SDF becomes necessary. As a result, we follow Kan and Robotti (2008) and define the candidate SDF as

$$y_t = 1 - (f_t - \mu_f)'\gamma_1 - (g_t - \mu_g)\gamma_2, \tag{33}$$

where f_t is a vector of K systematic factors with mean μ_f and covariance matrix S_f, and g_t is a useless factor with mean μ_g and variance σ_g^2, such that it is independent of f_t and x_t for all time periods.[4]

The pseudo-true value of γ_1 under the modified HJ-distance measure is given by

$$\gamma_1^* = (B'V^{-1}B)^{-1}B'V^{-1}\mu, \tag{34}$$

[4]Note that here the number of useful factors is set equal to K. This differs from the analysis in the previous section where the number of useful factors is set equal to $K-1$.

where $B = \text{Cov}[x_t, f_t']$. We set the pseudo-true value of γ_2, γ_2^*, equal to 0 even though it is not identified (see Section 2 of the paper for a discussion of this issue). Let $d = \text{Cov}[x_t, g_t] = 0_N$, $\hat{\mu} = \frac{1}{T}\sum_{t=1}^T x_t$, $\hat{V} = \frac{1}{T}\sum_{t=1}^T (x_t - \hat{\mu})(x_t - \hat{\mu})'$, and

$$\hat{D} = \left[\frac{1}{T}\sum_{t=1}^T x_t(f_t - \hat{\mu}_f)',\ \frac{1}{T}\sum_{t=1}^T x_t(g_t - \hat{\mu}_g)\right] \equiv [\hat{B},\ \hat{d}]. \tag{35}$$

The sample estimator of $\gamma = [\gamma_1',\ \gamma_2]'$ is given by

$$\hat{\gamma} = \begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix} = (\hat{D}'\hat{V}^{-1}\hat{D})^{-1}\hat{D}'\hat{V}^{-1}\hat{\mu}. \tag{36}$$

It is straightforward to show that

$$\begin{aligned} \hat{\gamma}_1 &= (\hat{B}'\hat{V}^{-\frac{1}{2}}[I_N - \hat{V}^{-\frac{1}{2}}\hat{d}(\hat{d}'\hat{V}^{-1}\hat{d})^{-1}\hat{d}'\hat{V}^{-\frac{1}{2}}]\hat{V}^{-\frac{1}{2}}\hat{B})^{-1} \\ &\quad \times \hat{B}'\hat{V}^{-\frac{1}{2}}[I_N - \hat{V}^{-\frac{1}{2}}\hat{d}(\hat{d}'\hat{V}^{-1}\hat{d})^{-1}\hat{d}'\hat{V}^{-\frac{1}{2}}]\hat{V}^{-\frac{1}{2}}\hat{\mu} \end{aligned} \tag{37}$$

and

$$\hat{\gamma}_2 = \frac{\hat{d}'\hat{V}^{-\frac{1}{2}}[I_N - \hat{V}^{-\frac{1}{2}}\hat{B}(\hat{B}'\hat{V}^{-1}\hat{B})^{-1}\hat{B}'\hat{V}^{-\frac{1}{2}}]\hat{V}^{-\frac{1}{2}}\hat{\mu}}{\hat{d}'\hat{V}^{-\frac{1}{2}}[I_N - \hat{V}^{-\frac{1}{2}}\hat{B}(\hat{B}'\hat{V}^{-1}\hat{B})^{-1}\hat{B}'\hat{V}^{-\frac{1}{2}}]\hat{V}^{-\frac{1}{2}}\hat{d}}. \tag{38}$$

Finally, Kan and Robotti (2008) suggest that a modification of the traditional HJ-distance is needed when using the de-meaned factors. Their proposed measure, the modified HJ-distance, employs the inverse of the covariance matrix (instead of the second moment matrix) of the excess returns as the weighting matrix and is given by

$$\delta_m = \sqrt{e(\gamma_1^*)'V^{-1}e(\gamma_1^*)}, \tag{39}$$

where $e(\gamma_1^*) = \mu - B\gamma_1^*$. The sample version of the model misspecification measure in (39) is given by

$$\hat{\delta}_m = \sqrt{\hat{e}'\hat{V}^{-1}\hat{e}}, \tag{40}$$

where $\hat{e} = \hat{\mu} - \hat{D}\hat{\gamma}$.

In deriving the limiting behavior of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ under correctly specified and misspecified models, we adopt the following notation. Let $\tilde{B} = V^{-\frac{1}{2}}B$, $\tilde{\mu} = V^{-\frac{1}{2}}\mu$, $e_t(\gamma_1^*) = x_t y_t^*$, $y_t^* = 1 - (f_t - \mu_f)'\gamma_1^*$, $S = E[e_t(\gamma_1^*)e_t(\gamma_1^*)']$, and P be an $N \times (N-K)$ orthonormal matrix whose columns are orthogonal to $\tilde{B}$ so that $PP' = I_N - \tilde{B}(\tilde{B}'\tilde{B})^{-1}\tilde{B}'$. Also, let $z \sim N(0_N, I_N)$ and $y \sim N(0_N, V^{-\frac{1}{2}}SV^{-\frac{1}{2}})$, and they are independent of each other. Finally, we define $w = P'z \sim N(0_{N-K}, I_{N-K})$, $s =$

$(\tilde{\mu}'Pw)/(\tilde{\mu}'PP'\tilde{\mu})^{\frac{1}{2}} \sim N(0,1)$, $u = P'y \sim N(0_{N-K}, V_u)$ with $V_u = P'V^{-\frac{1}{2}}SV^{-\frac{1}{2}}P$, and $r = (\tilde{B}'\tilde{B})^{-\frac{1}{2}}\tilde{B}'y \sim N(0_K, V_r)$ with $V_r = (\tilde{B}'\tilde{B})^{-\frac{1}{2}}\tilde{B}'V^{-\frac{1}{2}}SV^{-\frac{1}{2}}\tilde{B}(\tilde{B}'\tilde{B})^{-\frac{1}{2}}$.

Theorem 4. *Assume that the conditions in Assumption 1 are satisfied.*

(a) *If $\delta_m = 0$, i.e., the model is correctly specified, we have*

$$\sqrt{T}(\hat{\gamma}_1 - \gamma_1^*) \xrightarrow{d} (\tilde{B}'\tilde{B})^{-\frac{1}{2}} \left[r - \frac{w'u}{w'w}(\tilde{B}'\tilde{B})^{-\frac{1}{2}}\tilde{B}'z \right], \tag{41}$$

and

$$\hat{\gamma}_2 \xrightarrow{d} \frac{w'u}{\sigma_g w'w}. \tag{42}$$

(b) *If $\delta_m > 0$, i.e., the model is misspecified, we have*

$$\hat{\gamma}_1 - \gamma_1^* \xrightarrow{d} -\frac{\delta_m s}{w'w}(\tilde{B}'\tilde{B})^{-1}\tilde{B}'z, \tag{43}$$

and

$$\frac{1}{\sqrt{T}}\hat{\gamma}_2 \xrightarrow{d} \frac{\delta_m s}{\sigma_g w'w}. \tag{44}$$

As in the case of gross returns, we define two types of t-statistics: (i) $t_c(\hat{\gamma}_{1i})$, for $i = 1, \ldots, K$, and $t_c(\hat{\gamma}_2)$ that use standard errors obtained under the assumption that the model is correctly specified, and (ii) $t_m(\hat{\gamma}_{1i})$, for $i = 1, \ldots, K$, and $t_m(\hat{\gamma}_2)$ that use standard errors under potentially misspecified models. The two types of t-statistics are based on the estimated covariance matrices $\hat{\Sigma}^0_{\hat{\gamma}} = \frac{1}{T}\sum_{t=1}^T \hat{h}^0_t \hat{h}^{0\prime}_t$ and $\hat{\Sigma}_{\hat{\gamma}} = \frac{1}{T}\sum_{t=1}^T \hat{h}_t \hat{h}'_t$, where

$$\begin{aligned} \hat{h}^0_t &= (\hat{D}'\hat{V}^{-1}\hat{D})^{-1}\hat{D}'\hat{V}^{-1}\tilde{e}_t, & (45) \\ \hat{h}_t &= \hat{h}^0_t + (\hat{D}'\hat{V}^{-1}\hat{D})^{-1}\Big([(f_t - \hat{\mu}_f)',\ g_t - \hat{\mu}_g]' - \hat{D}'\hat{V}^{-1}(x_t - \hat{\mu})\Big)\hat{u}_t, & (46) \end{aligned}$$

$\tilde{e}_t = (x_t - \hat{\mu})\hat{y}_t + \hat{\mu}$, $\hat{y}_t = 1 - (f_t - \hat{\mu}_f)'\hat{\gamma}_1 - (g_t - \hat{\mu}_g)\hat{\gamma}_2$, and $\hat{u}_t = \hat{e}'\hat{V}^{-1}(x_t - \hat{\mu})$.

In addition to the random variables and matrices defined before Theorem 4, we introduce the following notation. Let $\tilde{u} \sim N(0,1)$, $\tilde{r}_i \sim N(0,1)$, $\tilde{z}_i \sim N(0,1)$, $v \sim \chi^2_{N-K-1}$, and they are independent of each other and w. Let c_i and $\hat{c}_i$ be the i-th diagonal elements of C and $\hat{C}$, respectively, where

$$\begin{aligned} C &= S_f^{-1}\mathrm{Cov}[(f_t - \mu_f)(f_t - \mu_f)', y_t^{*2}]S_f^{-1} + \gamma_1^* E[(f_t - \mu_f)y_t^{*2}]'S_f^{-1} \\ &\quad + S_f^{-1}E[(f_t - \mu_f)y_t^{*2}]\gamma_1^{*\prime} + E[y_t^{*2}]\gamma_1^*\gamma_1^{*\prime} & (47) \end{aligned}$$

and

$$\hat{C} = S_f^{-1}\text{Cov}[(f_t - \mu_f)(f_t - \mu_f)', y_t^{*2}]S_f^{-1} - \gamma_1^*\gamma_1^{*\prime}. \tag{48}$$

Define

$$\lambda_i = 1 + \frac{c_i}{E[y_t^{*2}]b_i}, \tag{49}$$

$$\hat{\lambda}_i = 1 + \frac{\hat{c}_i}{E[y_t^{*2}]b_i}, \tag{50}$$

where b_i is the i-th diagonal element of $(\tilde{B}'\tilde{B})^{-1}$. Theorem 5 below provides the limiting distributions of the t-statistics under correctly specified and misspecified models. Let the following assumption replace Assumption 2.

Assumption 2'. *Let* $\epsilon_t = (x_t - \mu) - BS_f^{-1}(f_t - \mu_f)$ *and assume that* $E[\epsilon_t|f_t] = 0_N$ *and* $\text{Cov}[\epsilon_t\epsilon_t', y_t^{*2}] = 0_{N\times N}$.

Theorem 5.

(a) *Suppose that the conditions in Assumptions 1 and 2' hold. If* $\delta_m = 0$*, i.e., the model is correctly specified, we have*

$$t_c(\hat{\gamma}_{1i}) \xrightarrow{d} \frac{\tilde{u}\tilde{z}_i + \sqrt{\lambda_i}\sqrt{w'w}\tilde{r}_i}{\left[\hat{\lambda}_i w'w + \tilde{z}_i^2 + \tilde{u}^2\left(1 + \frac{\tilde{z}_i^2}{w'w}\right)\right]^{\frac{1}{2}}}, \tag{51}$$

$$t_m(\hat{\gamma}_{1i}) \xrightarrow{d} \frac{\tilde{u}\tilde{z}_i + \sqrt{\lambda_i}\sqrt{w'w}\tilde{r}_i}{\left[\hat{\lambda}_i w'w + \tilde{z}_i^2 + \tilde{u}^2\left(1 + \frac{\tilde{z}_i^2}{w'w}\right) + \frac{\tilde{z}_i^2 v}{w'w}\right]^{\frac{1}{2}}}, \tag{52}$$

$$t_c(\hat{\gamma}_2) \xrightarrow{d} \frac{\tilde{u}}{\left(1 + \frac{\tilde{u}^2}{w'w}\right)^{\frac{1}{2}}}, \tag{53}$$

$$t_m(\hat{\gamma}_2) \xrightarrow{d} \frac{\tilde{u}}{\left(1 + \frac{\tilde{u}^2 + v}{w'w}\right)^{\frac{1}{2}}}. \tag{54}$$

(b) *Suppose that the conditions in Assumption 1 hold and denote the sign operator by* $\text{sgn}(\cdot)$*. If*

$\delta_m > 0$, *i.e., the model is misspecified, we have*

$$t_c(\hat{\gamma}_{1i}) \quad \xrightarrow{d} \quad \frac{\tilde{z}_i}{\left(1+\frac{\tilde{z}_i^2}{w'w}\right)^{\frac{1}{2}}}, \tag{55}$$

$$t_m(\hat{\gamma}_{1i}) \quad \xrightarrow{d} \quad N\left(0, \frac{1}{4}\right), \tag{56}$$

$$t_c(\hat{\gamma}_2) \quad \xrightarrow{d} \quad \operatorname{sgn}(s)\sqrt{w'w}, \tag{57}$$

$$t_m(\hat{\gamma}_2) \quad \xrightarrow{d} \quad N(0,1). \tag{58}$$

In the next theorem, we characterize the limiting behavior of the sample squared modified HJ-distance in the presence of a useless factor for the excess returns case.

Theorem 6. *Let* $Q_1 \sim \text{Beta}\left(\frac{N-K}{2}, \frac{1}{2}\right)$ *with density* $f_{Q_1}(\cdot)$, $Q_2 \sim \text{Beta}\left(\frac{N-K-1}{2}, \frac{1}{2}\right)$ *with density* $f_{Q_2}(\cdot)$ *and* c_α *be the* $100(1-\alpha)$-th *percentile of* χ^2_{N-K-1}.

(a) *Suppose that the assumptions in part (a) of Theorem 5 hold. If* $\delta_m = 0$, *we have*

$$T\hat{\delta}_m^2 \xrightarrow{d} E[y_t^{*2}]\chi^2_{N-K-1} \tag{59}$$

and the limiting probability of rejecting $H_0 : \delta_m^2 = 0$ *by the modified HJ-distance test of size* α *is*

$$\int_0^1 P\left[\chi^2_{N-K-1} > \frac{c_\alpha}{q}\right] f_{Q_1}(q)\mathrm{d}q < \alpha. \tag{60}$$

(b) *Suppose that the assumptions in Theorem 4 hold. If* $\delta_m > 0$, *we have*

$$\hat{\delta}_m^2 \xrightarrow{d} \delta_m^2 Q_2 \tag{61}$$

and the limiting probability of rejecting $H_0 : \delta_m^2 = 0$ *by the modified HJ-distance test of size* α *is*

$$\int_0^1 P\left[\chi^2_{N-K} > \frac{c_\alpha q}{1-q}\right] f_{Q_2}(q)\mathrm{d}q < 1. \tag{62}$$

Overall, the results for excess returns are very similar to the results for gross returns in the paper. The only noticeable differences are for the t-tests on $\hat{\gamma}_{1i}$ in part (a) of Theorem 5. This implies that the nature of the problem (and the solution) is essentially the same regardless of whether one uses gross returns or excess returns in the analysis.

Simulation Results

In this section, we undertake Monte Carlo experiments to assess the small-sample properties of the test statistics based on the modified HJ-distance in models with useful and useless factors. In addition, we analyze the finite-sample properties of some optimal GMM estimators. The simulation designs, data, and models are the same as the ones considered in Tables 1–6 of the paper.

Modified HJ-distance with excess returns

The results in Panel A of Table 1 show that for models that are correctly specified and contain only useful factors, the standard asymptotics provides an accurate approximation of the finite-sample behavior of the t-tests.

Table 1 about here

Since the useful factor, calibrated to the properties of the value-weighted market excess return, is closely replicated by the returns on the test assets, the differences between the t-tests under correctly specified models (t_c) and the t-tests under potentially misspecified models (t_m) are negligibly small even when the model fails to hold exactly.

Panel B of Table 1 and Table 2 present the empirical size of the t-tests in the presence of a useless factor.

Table 2 about here

The simulation results for the t-tests on the parameters of the useful factor confirm our theoretical findings that the null hypothesis is under-rejected when $N(0,1)$ is used as a reference distribution. This is the case for correctly specified and misspecified models.

Similarly, the inference on the useless factor proves to be conservative when the model is correctly specified. However, when the model is misspecified, there are substantial differences between t_c and t_m for the useless factor. Since the t_c test for significance of the useless factor is asymptotically distributed (up to a sign) as $\sqrt{\chi^2_{N-K}}$, it tends to over-reject severely when the critical values from $N(0,1)$ are used and the degree of over-rejection increases with the sample size. In contrast, the t_m test on the useless factor has good size properties although, for small sample sizes, it slightly under-rejects. As the sample size increases, the empirical rejection rates approach the limiting rejection probabilities (as shown in the rows for $T = \infty$) computed from the corresponding asymptotic distributions in Theorem 5.

Tables 3 and 4 report the survival rates of different factors when using the sequential procedure described in Section 3 of the paper.

Table 3 about here

Panel A of Table 3 shows that when the model is correctly specified, the procedures based on t_c and t_m do a similarly good job in retaining the useful factors with nonzero SDF parameters in the model and eliminating the useless factor and the factor that does not reduce the HJ-distance. However, as shown in Panel B, the situation drastically changes when the model is misspecified. In this case, the procedures based on t_c and t_m still retain the useful factors with similarly high probability, but they produce very different results when it comes to the useless factor. For example, despite its conservative nature (due to the Bonferroni adjustment), the procedure based on t_c will retain the useless factor 30% of the time for $T = 1000$. In contrast, the procedure based on t_m will retain the useless factor only about 0.8% of the time for $T = 1000$. Similarly, the probability of at least one irrelevant factor being selected in the final specification of the model is 30% (1.5%) for $T = 1000$ when the t_c (t_m) test is used and the model is misspecified.

Table 4 about here

Table 4 reports the results from a similar exercise but this time the linear asset pricing model consists of a constant term, two useful factors with $\gamma_i^* \neq 0$ and two useless factors. This setup serves to illustrate the usefulness of combining the misspecification-robust t-tests and the Bonferroni method in controlling the false discovery rate which is about 48% (the probability that at least one useless factor is deemed priced) for the t-tests constructed under correct model specification when the true model is misspecified. In contrast, the misspecification-robust model selection procedure with the Bonferroni adjustment retains one or both useless factors only 1% of the time.

Finally, we consider a scenario in which a linear combination of two useful factors is useless.

Table 5 about here

Panel A of Table 5 shows that when the model is correctly specified, the procedures based on t_c and t_m are both effective in retaining only one factor in the model. However, when the model is misspecified (see Panel B), the procedures based on t_c and t_m deliver very different results. For $T = 1000$, the probability that both factors survive the model selection procedure based on t_c is about 38% while the probability that both factors survive the model selection procedure based on t_m is about 2%. Importantly, the probabilities that only one factor survives are very different across procedures. For example, when $T = 1000$, the probability than only one factor survives is about 89% when using t-tests under misspecified models while it is only about 56% when using t-tests under correctly specified models.

Optimal GMM with gross returns

In this subsection, we use the same notation as in the paper and set the number of useful factors equal to $K-1$. The optimal s-step ($s \geq 2$) GMM estimator of the SDF parameters is defined as

$$\hat{\gamma}^{(s)} = \left(\hat{D}'\hat{S}_{(s-1)}^{-1}\hat{D}\right)^{-1}\hat{D}'\hat{S}_{(s-1)}^{-1}q, \tag{63}$$

where

$$\hat{D} = \left[\frac{1}{T}\sum_{t=1}^{T} x_t\tilde{f}_t',\ \frac{1}{T}\sum_{t=1}^{T} x_t g_t\right] \tag{64}$$

and

$$\hat{S}_{(s-1)} = \frac{1}{T}\sum_{t=1}^{T}\left[e_t\left(\hat{\gamma}^{(s-1)}\right) - e\left(\hat{\gamma}^{(s-1)}\right)\right]\left[e_t\left(\hat{\gamma}^{(s-1)}\right) - e\left(\hat{\gamma}^{(s-1)}\right)\right]' \tag{65}$$

with $e_t\left(\hat{\gamma}^{(s-1)}\right) = x_t\left[\tilde{f}_t'\hat{\gamma}_1^{(s-1)} + g_t\hat{\gamma}_2^{(s-1)}\right] - q = x_t y_t\left(\hat{\gamma}^{(s-1)}\right) - q$, $e\left(\hat{\gamma}^{(s-1)}\right) = T^{-1}\sum_{t=1}^{T} e_t\left(\hat{\gamma}^{(s-1)}\right) = \hat{D}\hat{\gamma}^{(s-1)} - q$.

Let $\hat{u}_t = e\left(\hat{\gamma}^{(s)}\right)'\hat{S}_{(s-1)}^{-1}x_t$ and $\hat{z}_t = e\left(\hat{\gamma}^{(s)}\right)'\hat{S}_{(s-1)}^{-1}\left(e_t\left(\hat{\gamma}^{(s-1)}\right) - e\left(\hat{\gamma}^{(s-1)}\right)\right)$. A consistent estimator of the asymptotic variance of the SDF parameters under misspecified models is given by (a proof of this result is available upon request) $\hat{\Sigma}_{\hat{\gamma}^{(s)}} = \frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{h}_t'$, where

$$\hat{h}_t = (\hat{D}'\hat{S}_{(s-1)}^{-1}\hat{D})^{-1}\left[\hat{D}'\hat{S}_{(s-1)}^{-1}\left(x_t y_t\left(\hat{\gamma}^{(s)}\right) - e_t\left(\hat{\gamma}^{(s-1)}\right)\hat{z}_t\right) + [\tilde{f}_t',\ g_t]'\hat{u}_t\right] - \hat{\gamma}^{(s)}. \tag{66}$$

When the model is correctly specified, the $\hat{h}_t$ expression simplifies to

$$\hat{h}_t^0 = \left(\hat{D}'\hat{S}_{(s-1)}^{-1}\hat{D}\right)^{-1}\hat{D}'\hat{S}_{(s-1)}^{-1}e_t\left(\hat{\gamma}^{(s)}\right). \tag{67}$$

In addition, the GMM test of correct model specification is given by

$$Te\left(\hat{\gamma}^{(s)}\right)'\hat{S}_{(s-1)}^{-1}e\left(\hat{\gamma}^{(s)}\right). \tag{68}$$

In the absence of a useless factor, it is well known that under a correctly specified model this test is asymptotically chi-squared distributed with $N-K$ degrees of freedom.

Tables 6 to 11 about here

In our simulations, we use the identity matrix to compute the first-step GMM estimator and analyze the finite-sample properties of the optimal 3-step GMM estimator and specification test in models with useful and useless factors. Our Monte Carlo simulations (see Tables 6–11) show that the results for optimal GMM are broadly consistent with the ones for the estimators and test statistics based on the HJ-distance. In addition, the rejection rates for the limiting case ($T=\infty$) are equivalent to those based on the asymptotic distributions given in Theorem 2 in the first section of this online appendix. This implies that our robust model selection procedure is also applicable to the class of optimal GMM estimators.

Appendix: Preliminary Lemma and Proofs of Main Results

A.1 Preliminary Lemma

Lemma A.1. *Let*

$$x_t = BS_{\tilde{f}}^{-1}\tilde{f}_t + \epsilon_t, \tag{A.1}$$

where $B = E[x_t\tilde{f}_t']$, $S_{\tilde{f}} = E[\tilde{f}_t\tilde{f}_t']$ *and* $E[\epsilon_t|\tilde{f}_t] = 0_N$. *Suppose* $\mathrm{Cov}[\epsilon_t\epsilon_t', (\tilde{f}_t'\gamma_1^*)^2] = 0_{N\times N}$ *(a sufficient condition for this to hold is* $E[\epsilon_t\epsilon_t'|\tilde{f}_t] = \Sigma$*, i.e., conditional homoskedasticity). When the model is correctly specified, we have*

$$S = E[(x_t\tilde{f}_t'\gamma_1^* - q)(x_t\tilde{f}_t'\gamma_1^* - q)'] = E[(\tilde{f}_t'\gamma_1^*)^2]U + BCB', \tag{A.2}$$

where $U = E[x_tx_t']$ *and* C *is a symmetric* $K \times K$ *matrix.*

Proof of Lemma A.1. Under a correctly specified model, we have $q = B\gamma_1^*$. It follows that

$$S = E[x_tx_t'(\tilde{f}_t'\gamma_1^*)^2] - qq' = E[x_tx_t'(\tilde{f}_t'\gamma_1^*)^2] - B\gamma_1^*\gamma_1^*B'. \tag{A.3}$$

For the first term, we have

$$\begin{aligned} E[x_tx_t'(\tilde{f}_t'\gamma_1^*)^2] &= E[x_tx_t']E[(\tilde{f}_t'\gamma_1^*)^2] + \mathrm{Cov}[x_tx_t', (\tilde{f}_t'\gamma_1^*)^2] \\ &= E[(\tilde{f}_t'\gamma_1^*)^2]U + \mathrm{Cov}[BS_{\tilde{f}}^{-1}\tilde{f}_t\tilde{f}_t'S_{\tilde{f}}^{-1}B' + \epsilon_t\epsilon_t', (\tilde{f}_t'\gamma_1^*)^2] \\ &= E[(\tilde{f}_t'\gamma_1^*)^2]U + BS_{\tilde{f}}^{-1}\mathrm{Cov}[\tilde{f}_t\tilde{f}_t', (\tilde{f}_t'\gamma_1^*)^2]S_{\tilde{f}}^{-1}B', \end{aligned} \tag{A.4}$$

where the last equality follows from the assumption that $\mathrm{Cov}[\epsilon_t\epsilon_t', (\tilde{f}_t'\gamma_1^*)^2] = 0_{N\times N}$. Therefore, we have

$$S = E[(\tilde{f}_t'\gamma_1^*)^2]U + BCB', \tag{A.5}$$

where

$$C = S_{\tilde{f}}^{-1}\mathrm{Cov}[\tilde{f}_t\tilde{f}_t', (\tilde{f}_t'\gamma_1^*)^2]S_{\tilde{f}}^{-1} - \gamma_1^*\gamma_1^{*\prime}. \tag{A.6}$$

This completes the proof.

A.2 Proofs of Theorems and Corollary 1

Proof of Theorem 1.

part (a): We start with the limiting distribution of $\sqrt{T}(\hat{\gamma}_1 - \gamma_1^*)$. Under the assumptions in Theorem 1, we have

$$\sqrt{T}\hat{U}^{-\frac{1}{2}}\hat{d} \stackrel{d}{\to} z \sim N(0_N, I_N) \tag{A.7}$$

and

$$-\sqrt{T}\hat{U}^{-\frac{1}{2}}(\hat{B}\gamma_1^* - q) \stackrel{d}{\to} y \sim N(0_N, V_y), \tag{A.8}$$

where $V_y = E[m_t m_t']$ is the covariance matrix of y, and

$$m_t = U^{-\frac{1}{2}}(x_t \tilde{f}_t' \gamma_1^* - q) = U^{-\frac{1}{2}} e_t(\gamma_1^*). \tag{A.9}$$

Therefore, we have $V_y = U^{-\frac{1}{2}} S U^{-\frac{1}{2}}$ for correctly specified models. In addition, y and z are independent of each other. Using y and z, we can write (7) as

$$\begin{aligned}
\sqrt{T}(\hat{\gamma}_1 - \gamma_1^*) &= (\hat{B}'\hat{U}^{-\frac{1}{2}}[I_N - \hat{U}^{-\frac{1}{2}}\hat{d}(\hat{d}'\hat{U}^{-1}\hat{d})^{-1}\hat{d}'\hat{U}^{-\frac{1}{2}}]\hat{U}^{-\frac{1}{2}}\hat{B})^{-1} \\
&\quad \times \hat{B}'\hat{U}^{-\frac{1}{2}}[I_N - \hat{U}^{-\frac{1}{2}}\hat{d}(\hat{d}'\hat{U}^{-1}\hat{d})^{-1}\hat{d}'\hat{U}^{-\frac{1}{2}}]\sqrt{T}\hat{U}^{-\frac{1}{2}}(q - \hat{B}\gamma_1^*) \\
&\stackrel{d}{\to} (\tilde{B}'[I_N - z(z'z)^{-1}z']\tilde{B})^{-1}\tilde{B}'[I_N - z(z'z)^{-1}z']y \\
&= (\tilde{B}'[I_N - z(z'z)^{-1}z']\tilde{B})^{-1}\tilde{B}'[I_N - z(z'z)^{-1}z'][PP' + \tilde{B}(\tilde{B}'\tilde{B})^{-1}\tilde{B}']y \\
&= -(\tilde{B}'[I_N - z(z'z)^{-1}z']\tilde{B})^{-1}\frac{\tilde{B}'zz'PP'y}{z'z} + (\tilde{B}'\tilde{B})^{-1}\tilde{B}'y.
\end{aligned} \tag{A.10}$$

Let $w = P'z \sim N(0_{N-K}, I_{N-K})$, $u = P'y \sim N(0_{N-K}, V_u)$ with $V_u = P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P$, $r = (\tilde{B}'\tilde{B})^{-\frac{1}{2}}\tilde{B}'y \sim N(0_K, V_r)$ with $V_r = (\tilde{B}'\tilde{B})^{-\frac{1}{2}}\tilde{B}'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}\tilde{B}(\tilde{B}'\tilde{B})^{-\frac{1}{2}}$. Making use of the identity

$$(\tilde{B}'[I_N - z(z'z)^{-1}z']\tilde{B})^{-1} = (\tilde{B}'\tilde{B})^{-1} + \frac{(\tilde{B}'\tilde{B})^{-1}\tilde{B}'zz'\tilde{B}(\tilde{B}'\tilde{B})^{-1}}{w'w} \tag{A.11}$$

and $z'z = z'\tilde{B}(\tilde{B}'\tilde{B})^{-1}\tilde{B}'z + w'w$, we obtain

$$\sqrt{T}(\hat{\gamma}_1 - \gamma_1^*) \stackrel{d}{\to} (\tilde{B}'\tilde{B})^{-\frac{1}{2}}\left[-\frac{w'u}{w'w}(\tilde{B}'\tilde{B})^{-\frac{1}{2}}\tilde{B}'z + r\right]. \tag{A.12}$$

For the derivation of the limiting distribution of $\hat{\gamma}_2$, we define $M = I_N - U^{-\frac{1}{2}}B(B'U^{-1}B)^{-1}B'U^{-\frac{1}{2}}$ and $\hat{M} = I_N - \hat{U}^{-\frac{1}{2}}\hat{B}(\hat{B}'\hat{U}^{-1}\hat{B})^{-1}\hat{B}'\hat{U}^{-\frac{1}{2}}$. Using that $\hat{M}\hat{U}^{-\frac{1}{2}}\hat{B} = 0_{N\times K}$, we obtain

$$\sqrt{T}\hat{M}\hat{U}^{-\frac{1}{2}}q = \sqrt{T}\hat{M}\hat{U}^{-\frac{1}{2}}(q - \hat{B}\gamma_1^*) \stackrel{d}{\to} My, \tag{A.13}$$

and we can rewrite $\hat{\gamma}_2$ as

$$\hat{\gamma}_2 = \frac{(\sqrt{T}\hat{U}^{-\frac{1}{2}}\hat{d})'(\sqrt{T}\hat{M}\hat{U}^{-\frac{1}{2}}(B - \hat{B})\gamma_1^*)}{(\sqrt{T}\hat{U}^{-\frac{1}{2}}\hat{d})'\hat{M}(\sqrt{T}\hat{U}^{-\frac{1}{2}}\hat{d})}. \tag{A.14}$$

Then, from (A.7), (A.8) and $\hat{M} \xrightarrow{p} M = PP'$, we get

$$\hat{\gamma}_2 \xrightarrow{d} \frac{z'My}{z'Mz} = \frac{(P'z)'(P'y)}{(P'z)'(P'z)} = \frac{w'u}{w'w}. \tag{A.15}$$

This completes the proof of part (a) of Theorem 1.

part (b): Using the fact that $\hat{U}^{-\frac{1}{2}}\hat{B} \xrightarrow{\text{a.s.}} \tilde{B}$ and $\sqrt{T}\hat{U}^{-\frac{1}{2}}\hat{d} \xrightarrow{d} z$, we can obtain the limiting distribution of $\hat{\gamma}_1$ in (7) as

$$\hat{\gamma}_1 \xrightarrow{d} (\tilde{B}'[I_N - z(z'^{-1}z']\tilde{B})^{-1}\tilde{B}'[I_N - z(z'^{-1}z']\tilde{q}. \tag{A.16}$$

Using (A.11) and the fact that $\gamma_1^* = (\tilde{B}'\tilde{B})^{-1}\tilde{B}'\tilde{q}$, we obtain

$$\begin{aligned}
\hat{\gamma}_1 - \gamma_1^* &\xrightarrow{d} \left[(\tilde{B}'\tilde{B})^{-1} + \frac{(\tilde{B}'\tilde{B})^{-1}\tilde{B}'zz'\tilde{B}(\tilde{B}'\tilde{B})^{-1}}{w'w}\right]\left(\tilde{B}'\tilde{q} - \frac{\tilde{B}'zz'\tilde{q}}{z'z}\right) - (\tilde{B}'\tilde{B})^{-1}\tilde{B}'\tilde{q} \\
&= -(\tilde{B}'\tilde{B})^{-1}\tilde{B}'z\frac{z'\tilde{q}}{z'z} + (\tilde{B}'\tilde{B})^{-1}\tilde{B}'z\frac{z'\tilde{B}(\tilde{B}'\tilde{B})^{-1}\tilde{B}'\tilde{q}}{w'w} - (\tilde{B}'\tilde{B})^{-1}\tilde{B}'z\frac{z'\tilde{q}}{z'z}\frac{z'\tilde{B}(\tilde{B}'\tilde{B})^{-1}\tilde{B}'z}{w'w} \\
&= -(\tilde{B}'\tilde{B})^{-1}\tilde{B}'z\frac{z'\tilde{q}}{w'w} + (\tilde{B}'\tilde{B})^{-1}\tilde{B}'z\frac{z'\tilde{B}(\tilde{B}'\tilde{B})^{-1}\tilde{B}'\tilde{q}}{w'w} \\
&= -\frac{z'M\tilde{q}}{w'w}(\tilde{B}'\tilde{B})^{-1}\tilde{B}'z \\
&= -\frac{\delta s}{w'w}(\tilde{B}'\tilde{B})^{-1}\tilde{B}'z,
\end{aligned} \tag{A.17}$$

and the last equality follows because $\delta^2 = \tilde{q}'PP'\tilde{q}$ and $s = \tilde{q}'PP'z/(\tilde{q}'PP'\tilde{q})^{\frac{1}{2}}$.

For the limiting distribution of $\hat{\gamma}_2$, we have

$$T^{-\frac{1}{2}}\hat{\gamma}_2 = \frac{(\sqrt{T}\hat{d}'\hat{U}^{-\frac{1}{2}})\hat{M}\hat{U}^{-\frac{1}{2}}q}{(\sqrt{T}\hat{d}'\hat{U}^{-\frac{1}{2}})\hat{M}(\sqrt{T}\hat{U}^{-\frac{1}{2}}\hat{d})} \xrightarrow{d} \frac{z'M\tilde{q}}{z'Mz} = \frac{\delta s}{w'w}. \tag{A.18}$$

This completes the proof of part (b) of Theorem 1.

Proof of Theorem 2.

part (a): Using Lemma A.1, we have

$$S = E[(\tilde{f}_t'\gamma_1^*)^2]U + BCB' \tag{A.19}$$

under the conditional homoskedasticity assumption. It follows that

$$\begin{aligned}
V_u &= P'^{-\frac{1}{2}}SU^{-\frac{1}{2}}P = E[(\tilde{f}_t'\gamma_1^*)^2]I_{N-K}, && \text{(A.20)} \\
V_r &= (\tilde{B}'\tilde{B})^{-\frac{1}{2}}\tilde{B}'^{-\frac{1}{2}}SU^{-\frac{1}{2}}\tilde{B}(\tilde{B}'\tilde{B})^{-\frac{1}{2}} = E[(\tilde{f}_t'\gamma_1^*)^2]I_K + (\tilde{B}'\tilde{B})^{\frac{1}{2}}C(\tilde{B}'\tilde{B})^{\frac{1}{2}}, && \text{(A.21)} \\
\text{Cov}[u, r'] &= P'^{-\frac{1}{2}}SU^{-\frac{1}{2}}\tilde{B}(\tilde{B}'\tilde{B})^{-\frac{1}{2}} = 0_{(N-K)\times K}. && \text{(A.22)}
\end{aligned}$$

Let $\tilde{u} = w'u/(w'V_u w)^{\frac{1}{2}} = E[(\tilde{f}_t'\gamma_1^*)^2]^{-\frac{1}{2}} w'u/(w'w)^{\frac{1}{2}}$. It is easy to show that $\tilde{u} \sim N(0,1)$ and it is independent of w, z and r. Using $\tilde{u}$, we can simplify the limiting distribution of $\sqrt{T}(\hat{\gamma}_1 - \gamma_1^*)$ in (A.12) to

$$\sqrt{T}(\hat{\gamma}_1 - \gamma_1^*) \stackrel{d}{\to} -E[(\tilde{f}_t'\gamma_1^*)^2]^{\frac{1}{2}} \frac{\tilde{u}}{(w'w)^{\frac{1}{2}}} (\tilde{B}'\tilde{B})^{-1}\tilde{B}'z + (\tilde{B}'\tilde{B})^{-\frac{1}{2}} r. \tag{A.23}$$

The estimated covariance matrix of $\hat{\gamma}$ for a potentially misspecified model is given by

$$\hat{V}_m(\hat{\gamma}) = \frac{1}{T^2} \sum_{t=1}^{T} \hat{h}_t \hat{h}_t', \tag{A.24}$$

where

$$\hat{h}_t = (\hat{D}'\hat{U}^{-1}\hat{D})^{-1}\hat{D}'\hat{U}^{-1}\hat{e}_t + (\hat{D}'\hat{U}^{-1}\hat{D})^{-1}([\tilde{f}_t',\ g_t]' - \hat{D}'\hat{U}^{-1}x_t)\hat{u}_t, \tag{A.25}$$

and $\hat{u}_t = \hat{e}'\hat{U}^{-1}x_t$. In order to derive the limiting distribution of $\hat{h}_t$, we need to obtain the limiting representations of $(\hat{D}'\hat{U}^{-1}\hat{D})^{-1}$, $(\hat{D}'\hat{U}^{-1}\hat{D})^{-1}\hat{D}'\hat{U}^{-1}$, and $\hat{u}_t$.

It is straightforward to show that

$$\hat{D}'\hat{U}^{-1} = \begin{bmatrix} \tilde{B}'U^{-\frac{1}{2}} + O_p(T^{-\frac{1}{2}}) \\ \frac{1}{\sqrt{T}} z'U^{-\frac{1}{2}} + O_p(T^{-1}) \end{bmatrix}, \tag{A.26}$$

$$\hat{D}'\hat{U}^{-1}\hat{D} = \begin{bmatrix} \tilde{B}'\tilde{B} + O_p(T^{-\frac{1}{2}}) & \frac{1}{\sqrt{T}}\tilde{B}'z + O_p(T^{-1}) \\ \frac{1}{\sqrt{T}} z'\tilde{B} + O_p(T^{-1}) & \frac{z'z}{T} + O_p(T^{-\frac{3}{2}}) \end{bmatrix}. \tag{A.27}$$

Then, using the partitioned matrix inverse formula, we have

$$(\hat{D}'\hat{U}^{-1}\hat{D})^{-1} = \begin{bmatrix} H + O_p(T^{-\frac{1}{2}}) & -\sqrt{T}\frac{(\tilde{B}'\tilde{B})^{-1}\tilde{B}'z}{w'w} + O_p(1) \\ -\sqrt{T}\frac{z'\tilde{B}(\tilde{B}'\tilde{B})^{-1}}{w'w} + O_p(1) & \frac{T}{w'w} + O_p(T^{\frac{1}{2}}) \end{bmatrix}, \tag{A.28}$$

where

$$H = (\tilde{B}'[I_N - z(z'z)^{-1}z']\tilde{B})^{-1} = (\tilde{B}'\tilde{B})^{-1} + \frac{(\tilde{B}'\tilde{B})^{-1}\tilde{B}'zz'\tilde{B}(\tilde{B}'\tilde{B})^{-1}}{w'w}. \tag{A.29}$$

After simplification, we obtain

$$(\hat{D}'\hat{U}^{-1}\hat{D})^{-1}\hat{D}'\hat{U}^{-1} = \begin{bmatrix} (\tilde{B}'\tilde{B})^{-1}\tilde{B}'^{-\frac{1}{2}} - \frac{(\tilde{B}'\tilde{B})^{-1}\tilde{B}'zw'P'^{-\frac{1}{2}}}{w'w} + O_p(T^{-\frac{1}{2}}) \\ \frac{\sqrt{T}w'P'^{-\frac{1}{2}}}{w'w} + O_p(1) \end{bmatrix}. \tag{A.30}$$

With the above expressions, we now derive the limiting distribution of $\hat{u}_t$. Note that the vector of sample pricing errors is given by

$$\hat{e} = \hat{D}\hat{\gamma} - q = \hat{D}(\hat{D}'\hat{U}^{-1}\hat{D})^{-1}\hat{D}'\hat{U}^{-1}q - q. \tag{A.31}$$

Using (A.13), (A.15), and the identity

$$I_N - \hat{U}^{-\frac{1}{2}}\hat{D}(\hat{D}'\hat{U}^{-1}\hat{D})^{-1}\hat{D}'\hat{U}^{-\frac{1}{2}} = \hat{M} - \hat{M}\hat{U}^{-\frac{1}{2}}\hat{d}(\hat{d}'\hat{U}^{-\frac{1}{2}}\hat{M}\hat{U}^{-\frac{1}{2}}\hat{d})^{-1}\hat{d}'\hat{U}^{-\frac{1}{2}}\hat{M}, \tag{A.32}$$

we can obtain the limiting distribution of $-\sqrt{T}\hat{U}^{-\frac{1}{2}}\hat{e}$ as

$$-\sqrt{T}\hat{U}^{-\frac{1}{2}}\hat{e} = \sqrt{T}\hat{M}\hat{U}^{-\frac{1}{2}}q - \sqrt{T}\hat{M}\hat{U}^{-\frac{1}{2}}\hat{d}\hat{\gamma}_2 \xrightarrow{d} My - Mz\frac{w'u}{w'w} = P\left(I_{N-K} - \frac{ww'}{w'w}\right)u, \tag{A.33}$$

and we have

$$\sqrt{T}\hat{u}_t \xrightarrow{d} -u'\left(I_{N-K} - \frac{ww'}{w'w}\right)P'^{-\frac{1}{2}}x_t. \tag{A.34}$$

Using (A.28), (A.30), (A.34), and the fact that

$$\hat{e}_t = x_t(\tilde{f}_t'\hat{\gamma}_1 + \hat{\gamma}_2 g_t) - q = x_t\tilde{f}_t'\gamma_1^* - q + \frac{w'u}{w'w}x_t g_t + O_p(T^{-\frac{1}{2}}) \tag{A.35}$$

under a correctly specified model, we can write the limiting distribution of $\hat{h}_t = [\hat{h}_{1t}',\ \hat{h}_{2t}]'$, where $\hat{h}_{1t}$ denotes the first K elements of $\hat{h}_t$, as

$$\begin{aligned}\hat{h}_{1t} \quad &\xrightarrow{d} \quad \left[(\tilde{B}'\tilde{B})^{-1}\tilde{B}'U^{-\frac{1}{2}} - \frac{(\tilde{B}'\tilde{B})^{-1}\tilde{B}'zw'P'U^{-\frac{1}{2}}}{w'w}\right]\left(x_t\tilde{f}_t'\gamma_1^* - q + x_t g_t\frac{w'u}{w'w}\right) \\ &\quad + \frac{(\tilde{B}'\tilde{B})^{-1}\tilde{B}'z}{w'w}u'\left(I_{N-K} - \frac{ww'}{w'w}\right)P'^{-\frac{1}{2}}x_t g_t, \end{aligned} \tag{A.36}$$

$$\frac{\hat{h}_{2t}}{\sqrt{T}} \xrightarrow{d} \frac{1}{w'w}w'P'U^{-\frac{1}{2}}\left(x_t\tilde{f}_t'\gamma_1^* - q + x_t g_t\frac{w'u}{w'w}\right) - \frac{1}{w'w}u'\left(I_{N-K} - \frac{ww'}{w'w}\right)P'U^{-\frac{1}{2}}x_t g_t. \tag{A.37}$$

Under the conditional homoskedasticity assumption, we have

$$\frac{1}{T}\sum_{t=1}^{T}(x_t\tilde{f}_t'\gamma_1^* - q)(x_t\tilde{f}_t'\gamma_1^* - q)' \xrightarrow{\text{a.s.}} S = E[(\tilde{f}_t'\gamma_1^*)^2]U + BCB'. \tag{A.38}$$

Together with the fact that

$$\frac{1}{T}\sum_{t=1}^{T}x_t x_t' g_t^2 \xrightarrow{\text{a.s.}} E[x_t x_t' g_t^2] = E[x_t x_t']E[g_t^2] = U, \tag{A.39}$$

we can show that the estimated misspecification-robust covariance matrix of $\hat{\gamma}_1$ has a limiting distribution of

$$\begin{aligned} T\hat{V}_m(\hat{\gamma}_1) \quad &= \quad \frac{1}{T}\sum_{t=1}^{T}\hat{h}_{1t}\hat{h}_{1t}' \\ &\xrightarrow{d} \quad E[(\tilde{f}_t'\gamma_1^*)^2]\left(1 + \frac{\tilde{u}^2}{w'w}\right)\left[(\tilde{B}'\tilde{B})^{-1} + \frac{(\tilde{B}'\tilde{B})^{-1}\tilde{B}'zz'\tilde{B}(\tilde{B}'\tilde{B})^{-1}}{w'w}\right] + C \\ &\quad + u'\left(I_{N-K} - \frac{ww'}{w'w}\right)u\frac{(\tilde{B}'\tilde{B})^{-1}\tilde{B}'zz'\tilde{B}(\tilde{B}'\tilde{B})^{-1}}{(w'w)^2}. \end{aligned} \tag{A.40}$$

Let b_i be the i-th diagonal element of $(\tilde{B}'\tilde{B})^{-1}$. Then, we can readily show that

$$\tilde{z}_i = -\frac{\boldsymbol{\iota}_i'(\tilde{B}'\tilde{B})^{-1}\tilde{B}'z}{\sqrt{b_i}} \sim N(0,1), \tag{A.41}$$

$$v = \frac{u'[I_{N-K} - w(w'w)^{-1}w']u}{E[(\tilde{f}_t'\gamma_1^*)^2]} \sim \chi^2_{N-K-1}, \tag{A.42}$$

and v is independent of $\tilde{u}$, z and w. Using $\tilde{z}_i$ and v, we can express the limiting distribution of $s_m^2(\hat{\gamma}_{1i})$ as

$$Ts_m^2(\hat{\gamma}_{1i}) = T\boldsymbol{\iota}_i'\hat{V}_m(\hat{\gamma}_1)\boldsymbol{\iota}_i \xrightarrow{d} E[(\tilde{f}_t'\gamma_1^*)^2]b_i\left[\left(1+\frac{\tilde{u}^2}{w'w}\right)\left(1+\frac{\tilde{z}_i^2}{w'w}\right)+\frac{\tilde{z}_i^2 v}{(w'w)^2}\right]+c_i, \tag{A.43}$$

where c_i is the i-th diagonal element of C. In addition, by letting

$$\tilde{r}_i = (E[(\tilde{f}_t'\gamma_1^*)^2]b_i + c_i)^{-\frac{1}{2}}\boldsymbol{\iota}_i'(\tilde{B}'\tilde{B})^{-\frac{1}{2}}r \sim N(0,1), \tag{A.44}$$

we can write the i-th element in (A.23) as

$$\sqrt{T}(\hat{\gamma}_{1i} - \gamma_{1i}^*) \xrightarrow{d} (E[(\tilde{f}_t'\gamma_1^*)^2]b_i)^{\frac{1}{2}}\frac{\tilde{u}\tilde{z}_i}{(w'w)^{\frac{1}{2}}} + (E[(\tilde{f}_t'\gamma_1^*)^2]b_i + c_i)^{\frac{1}{2}}\tilde{r}_i. \tag{A.45}$$

Finally, by letting[5]

$$\lambda_i = 1 + \frac{c_i}{E[(\tilde{f}_t'\gamma_1^*)^2]b_i} > 0, \tag{A.46}$$

we can write the limiting distribution of $t_m(\hat{\gamma}_{1i})$ as

$$t_m(\hat{\gamma}_{1i}) = \frac{\hat{\gamma}_{1i} - \gamma_{1i}^*}{s_m(\hat{\gamma}_{1i})} \xrightarrow{d} \frac{\tilde{u}\tilde{z}_i + \sqrt{\lambda_i}\sqrt{w'w}\tilde{r}_i}{\left[\lambda_i(w'w) + \tilde{z}_i^2 + \tilde{u}^2\left(1+\frac{\tilde{z}_i^2}{w'w}\right) + \frac{\tilde{z}_i^2 v}{w'w}\right]^{\frac{1}{2}}}. \tag{A.47}$$

The estimated covariance matrix of $\hat{\gamma}_1$ that assumes a correctly specified model is obtained by dropping the second term in (A.40). Then, it can be shown that

$$Ts_c^2(\hat{\gamma}_{1i}) \xrightarrow{d} E[(\tilde{f}_t'\gamma_1^*)^2]b_i\left[\left(1+\frac{\tilde{u}^2}{w'w}\right)\left(1+\frac{\tilde{z}_i^2}{w'w}\right)\right]+c_i \tag{A.48}$$

and hence

$$t_c(\hat{\gamma}_{1i}) = \frac{\hat{\gamma}_{1i} - \gamma_{1i}^*}{s_c(\hat{\gamma}_{1i})} \xrightarrow{d} \frac{\tilde{u}\tilde{z}_i + \sqrt{\lambda_i}\sqrt{w'w}\tilde{r}_i}{\left[\lambda_i(w'w) + \tilde{z}_i^2 + \tilde{u}^2\left(1+\frac{\tilde{z}_i^2}{w'w}\right)\right]^{\frac{1}{2}}}. \tag{A.49}$$

We now turn our attention to the limiting distributions of $t_c(\hat{\gamma}_2)$ and $t_m(\hat{\gamma}_2)$. From part (a) of Theorem 1, we have

$$\hat{\gamma}_2 \xrightarrow{d} \frac{w'u}{w'w} = \frac{(w'V_u w)^{\frac{1}{2}}}{(w'w)}\tilde{u}, \tag{A.50}$$

[5] From (A.44), we can see that $E[(\tilde{f}_t'\gamma_1^*)^2]b_i + c_i$ is the variance of $\boldsymbol{\iota}_i'(\tilde{B}'\tilde{B})^{-\frac{1}{2}}r$. Therefore, we have $\lambda_i > 0$.

where $\tilde{u} = w'u/(w'V_uw)^{\frac{1}{2}} \sim N(0,1)$, and it is independent of w. Using (A.37), we obtain

$$\begin{aligned} s_m^2(\hat{\gamma}_2) &= \frac{1}{T^2}\sum_{t=1}^T \hat{h}_{2t}^2 \\ &\stackrel{d}{\to} \frac{1}{(w'w)^2}\left[w'V_uw + \frac{(w'u)^2}{w'w}\right] + \frac{u'[I_{N-K} - w(w'w)^{-1}w']u}{(w'w)^2} \\ &= \frac{w'V_uw + u'u}{(w'w)^2}. \end{aligned} \tag{A.51}$$

Therefore, the t-statistic of $\hat{\gamma}_2$ under the misspecification-robust standard error is given by

$$t_m(\hat{\gamma}_2) = \frac{\hat{\gamma}_2}{s_m(\hat{\gamma}_2)} \stackrel{d}{\to} \frac{\tilde{u}}{\left(1 + \frac{u'u}{w'V_uw}\right)^{\frac{1}{2}}}. \tag{A.52}$$

For $s_c^2(\hat{\gamma}_2)$ which assumes a correctly specified model, we drop the second term in $\hat{h}_{2t}$, and we obtain

$$s_c^2(\hat{\gamma}_2) \stackrel{d}{\to} \frac{1}{(w'w)^2}\left[w'V_uw + \frac{(w'u)^2}{w'w}\right] = \frac{w'V_uw}{(w'w)^2}\left(1 + \frac{\tilde{u}^2}{w'w}\right). \tag{A.53}$$

It follows that

$$t_c(\hat{\gamma}_2) = \frac{\hat{\gamma}_2}{s_c(\hat{\gamma}_2)} \stackrel{d}{\to} \frac{\tilde{u}}{\left(1 + \frac{\tilde{u}^2}{w'w}\right)^{\frac{1}{2}}}. \tag{A.54}$$

Under the conditional homoskedasticity assumption, $V_u = E[(\tilde{f}_t'\gamma_1^*)^2]I_{N-K}$, so we can write

$$t_m(\hat{\gamma}_2) \stackrel{d}{\to} \frac{\tilde{u}}{\left(1 + \frac{\tilde{u}^2 + v}{w'w}\right)^{\frac{1}{2}}}, \tag{A.55}$$

where v is defined in (A.42). This completes the proof of part (a) of Theorem 2.

part (b): We first derive the limiting distribution of $\hat{h}_t$ in (A.25). When a model is misspecified, we can see from part (b) of Theorem 1 that $\hat{\gamma}_2 = O_p(T^{\frac{1}{2}})$ and $\hat{\gamma}_1 = O_p(1)$, so $\hat{\gamma}_2$ is the dominant term. Therefore, using (11), we have

$$\hat{e}_t = x_t(\tilde{f}_t'\hat{\gamma}_1 + g_t\hat{\gamma}_2) - q = x_tg_t\hat{\gamma}_2 + O_p(1) = \frac{\sqrt{T}\delta s}{w'w}x_tg_t + O_p(1). \tag{A.56}$$

In addition, using (A.31), (A.32) and (A.18), we have

$$-\hat{U}^{-\frac{1}{2}}\hat{e} = \hat{M}\hat{U}^{-\frac{1}{2}}q - \hat{M}\hat{U}^{-\frac{1}{2}}\hat{d}\hat{\gamma}_2 \stackrel{d}{\to} M\tilde{q} - \frac{Mzz'M\tilde{q}}{z'Mz} = P[I_{N-K} - w(w'w)^{-1}w']P'\tilde{q}. \tag{A.57}$$

It follows that under a misspecified model,

$$\hat{u}_t = \hat{e}'\hat{U}^{-1}x_t \stackrel{d}{\to} -\tilde{q}'P[I_{N-K} - w(w'w)^{-1}w']P'^{-\frac{1}{2}}x_t. \tag{A.58}$$

Then, using (A.28) and (A.30), we can express the limiting distribution of $\hat{h}_t = [\hat{h}_{1t}',\ \hat{h}_{2t}]'$ as

$$\begin{aligned} \frac{\hat{h}_{1t}}{\sqrt{T}} &\stackrel{d}{\to} \frac{\tilde{q}'Pw}{w'w}(\tilde{B}'\tilde{B})^{-1}\tilde{B}'\left(I_N - \frac{zw'}{w'w}P'\right)U^{-\frac{1}{2}}x_t g_t \\ &\quad + \frac{(\tilde{B}'\tilde{B})^{-1}(\tilde{B}'z)}{w'w}\tilde{q}'P[I_{N-K} - w(w'w)^{-1}w']P'^{-\frac{1}{2}}x_t g_t, \qquad \text{(A.59)} \\ \frac{\hat{h}_{2t}}{T} &\stackrel{d}{\to} \frac{\tilde{q}'Pw}{(w'w)^2}w'P'^{-\frac{1}{2}}x_t g_t - \frac{1}{w'w}\tilde{q}'P[I_{N-K} - w(w'w)^{-1}w']P'^{-\frac{1}{2}}x_t g_t. \qquad \text{(A.60)} \end{aligned}$$

Using the fact that $P'\tilde{B} = 0_{(N-K)\times K}$ and $[I_{N-K} - w(w'w)^{-1}w']w = 0_{N-K}$, we have

$$\tilde{B}'\left(I_N - \frac{zw'}{w'w}P'\right)P[I_{N-K} - w(w'w)^{-1}w']P'\tilde{q} = 0_K, \qquad \text{(A.61)}$$

and we can show that the two terms in the limiting distribution of $\hat{h}_{1t}/\sqrt{T}$ are asymptotically uncorrelated. It follows that

$$\begin{aligned} \hat{V}_m(\hat{\gamma}_1) &= \frac{1}{T^2}\sum_{t=1}^{T}\hat{h}_{1t}\hat{h}_{1t}' \\ &= \frac{(\tilde{q}'Pw)^2}{(w'w)^2}\left[(\tilde{B}'\tilde{B})^{-1} + \frac{(\tilde{B}'\tilde{B})^{-1}\tilde{B}'zz'\tilde{B}(\tilde{B}'\tilde{B})^{-1}}{w'w}\right] \\ &\quad + \frac{1}{(w'w)^2}\left[\tilde{q}'PP'\tilde{q} - \frac{(\tilde{q}'Pw)^2}{w'w}\right](\tilde{B}'\tilde{B})^{-1}\tilde{B}'zz'\tilde{B}(\tilde{B}'\tilde{B})^{-1} \\ &= \frac{\delta^2}{(w'w)^2}\left[s^2(\tilde{B}'\tilde{B})^{-1} + (\tilde{B}'\tilde{B})^{-1}\tilde{B}'zz'\tilde{B}(\tilde{B}'\tilde{B})^{-1}\right]. \qquad \text{(A.62)} \end{aligned}$$

Using $\tilde{z}_i$ as defined in (A.41), we can express the limiting distribution of $s_m^2(\hat{\gamma}_{1i})$ as

$$s_m^2(\hat{\gamma}_{1i}) = \boldsymbol{\iota}_i'\hat{V}_m(\hat{\gamma}_1)\boldsymbol{\iota}_i \stackrel{d}{\to} \frac{\delta^2 b_i}{(w'w)^2}(s^2 + \tilde{z}_i^2). \qquad \text{(A.63)}$$

In addition, we can also use $\tilde{z}_i$ to express the i-th element in (10) as

$$\hat{\gamma}_{1i} - \gamma_{1i}^* \stackrel{d}{\to} \frac{\delta s\sqrt{b_i}\tilde{z}_i}{w'w}. \qquad \text{(A.64)}$$

It follows that when the model is misspecified, $t_m(\hat{\gamma}_{1i})$ has the following limiting distribution:

$$t_m(\hat{\gamma}_{1i}) = \frac{\hat{\gamma}_{1i} - \gamma_{1i}^*}{s_m(\hat{\gamma}_{1i})} \stackrel{d}{\to} \frac{s\tilde{z}_i}{\sqrt{s^2 + \tilde{z}_i^2}}. \qquad \text{(A.65)}$$

To show that $t_m(\hat{\gamma}_{1i}) \stackrel{d}{\to} N(0, 1/4)$, consider the polar transformation $s = \omega\cos(\theta)$ and $\tilde{z}_i = \omega\sin(\theta)$, where $\omega = \sqrt{s^2 + \tilde{z}_i^2}$. The joint density of (ω, θ) is given by

$$f(\omega, \theta) = \frac{\omega e^{-\frac{\omega^2}{2}}}{2\pi}I_{\{\omega>0\}}I_{\{0<\theta<2\pi\}}. \qquad \text{(A.66)}$$

Therefore, ω and θ are independent. Using the polar transformation, we obtain

$$\frac{s\tilde{z}_i}{\sqrt{s^2+\tilde{z}_i^2}} = \omega\cos(\theta)\sin(\theta) = \frac{\omega\sin(2\theta)}{2}. \tag{A.67}$$

Since θ is uniformly distributed over $(0, 2\pi)$, $\sin(\theta)$ and $\sin(2\theta)$ have the same distribution. It follows that $\omega\sin(2\theta) \stackrel{d}{=} \omega\sin(\theta) \sim N(0,1)$. Therefore,

$$t_m(\hat{\gamma}_{1i}) \stackrel{d}{\to} N\left(0, \frac{1}{4}\right). \tag{A.68}$$

The estimated covariance matrix of $\hat{\gamma}_1$ that assumes a correctly specified model is obtained by dropping the second term in the line before (A.62). We can then show that

$$s_c^2(\hat{\gamma}_{1i}) \stackrel{d}{\to} \frac{\delta^2 s^2 b_i}{(w'w)^2}\left(1 + \frac{\tilde{z}_i^2}{w'w}\right). \tag{A.69}$$

Using (A.64), we can then obtain the limiting distribution of $t_c(\hat{\gamma}_{1i})$ as

$$t_c(\hat{\gamma}_{1i}) = \frac{\hat{\gamma}_{1i} - \gamma_{1i}^*}{s_c(\hat{\gamma}_{1i})} \stackrel{d}{\to} \frac{\tilde{z}_i}{\left(1 + \frac{\tilde{z}_i^2}{w'w}\right)^{\frac{1}{2}}}. \tag{A.70}$$

Turning our attention to the limiting distributions of $t_c(\hat{\gamma}_2)$ and $t_m(\hat{\gamma}_2)$, we use (A.60) and the fact that $\delta^2 = \tilde{q}'PP'\tilde{q}$ to obtain

$$\begin{aligned}
\frac{s_m^2(\hat{\gamma}_2)}{T} &= \frac{1}{T^3}\sum_{t=1}^{T}\hat{h}_{2t}^2 \\
&\stackrel{d}{\to} \frac{(\tilde{q}'^2}{(w'w)^4}w'w + \frac{1}{(w'w)^2}\tilde{q}'P\left(I_{N-K} - \frac{ww'}{w'w}\right)P'\tilde{q} \\
&= \frac{\delta^2}{(w'w)^2}.
\end{aligned} \tag{A.71}$$

Therefore, using (11), the t-statistic of $\hat{\gamma}_2$ under the misspecification-robust standard error is given by

$$t_m(\hat{\gamma}_2) = \frac{\hat{\gamma}_2}{s_m(\hat{\gamma}_2)} \stackrel{d}{\to} s \sim N(0,1). \tag{A.72}$$

For $s_c^2(\hat{\gamma}_2)$ which assumes a correctly specified model, we drop the second term of $\hat{h}_{2t}$ in (A.60), and we obtain

$$\frac{s_c^2(\hat{\gamma}_2)}{T} \stackrel{d}{\to} \frac{(\tilde{q}'Pw)^2}{(w'w)^3} = \frac{\delta^2 s^2}{(w'w)^3}. \tag{A.73}$$

It follows that

$$t_c(\hat{\gamma}_2) = \frac{\hat{\gamma}_2}{s_c(\hat{\gamma}_2)} \xrightarrow{d} \text{sgn}(s)\sqrt{w'w}. \tag{A.74}$$

Note that since $s \sim N(0,1)$, sgn(s) has probabilities of 1/2 of taking the values of -1 or 1, and it is independent of s^2. As a result, sgn(s) is also independent of $w'w \sim \chi^2_{N-K}$.[6] This completes the proof of part (b) of Theorem 2.

Proof of Corollary 1 (Proposition 2 in the paper).

We only provide the proof of part (a) since the proof of part (b) is similar for $t_c^2(\hat{\gamma}_{1i})$ and obvious for $t_m^2(\hat{\gamma}_{1i})$. First, comparing the limiting distribution of $t_c^2(\hat{\gamma}_{1i})$ with the limiting distribution of $t_m^2(\hat{\gamma}_{1i})$ in part (a) of Theorem 2, we see that there is an extra positive term $\tilde{z}_i^2 v/(w'w)$ in the denominator. Therefore, the limiting distribution of $t_m^2(\hat{\gamma}_{1i})$ is stochastically dominated by the limiting distribution of $t_c^2(\hat{\gamma}_{1i})$. It remains to be shown that the latter is stochastically dominated by χ^2_1. From part (a) of Theorem 2, we have

$$t_c^2(\hat{\gamma}_{1i}) \xrightarrow{d} \frac{(\tilde{u}\tilde{z}_i + \sqrt{\lambda_i}\sqrt{w'w}\tilde{r}_i)^2}{\lambda_i(w'w) + \tilde{z}_i^2 + \tilde{u}^2\left(1 + \frac{\tilde{z}_i^2}{w'w}\right)}. \tag{A.76}$$

Let $\tilde{t} = \tilde{z}_i/\sqrt{w'w}$. It is easy to see that the limit of $t_c^2(\hat{\gamma}_{1i})$ is stochastically dominated by $(\tilde{t}\tilde{u} + \sqrt{\lambda_i}\tilde{r}_i)^2/(\lambda_i + \tilde{t}^2) \sim \chi^2_1$.

Next, since $1 + \tilde{u}^2/(w'w) > 1$ and $1 + (\tilde{u}^2 + v)/(w'w) > 1$ almost surely, both the limiting distributions of $t_c^2(\hat{\gamma}_2)$ and $t_m^2(\hat{\gamma}_2)$ are stochastically dominated by $\tilde{u}^2 \sim \chi^2_1$. This completes the proof of Corollary 1.

Proof of Theorem 3.

part (a): Using (A.33) in the proof of Theorem 2, we can easily obtain

$$T\hat{\delta}^2 = T\hat{e}'\hat{U}^{-1}\hat{e} \xrightarrow{d} u'[I_{N-K} - w(w'w)^{-1}w']u = u'P_w P_w' u, \tag{A.77}$$

[6] It is straightforward to show that the limiting probability density function of $t_c(\hat{\gamma}_2)$ is

$$f(t) = \frac{|t|^{N-K-1}e^{-\frac{t^2}{2}}}{2^{\frac{N-K}{2}}\Gamma\left(\frac{N-K}{2}\right)}. \tag{A.75}$$

where P_w is an $(N-K)\times(N-K-1)$ orthonormal matrix such that $P_wP_w' = I_{N-K} - w(w'w)^{-1}w'$. Let $\tilde{v} = (P_w'V_uP_w)^{-\frac{1}{2}}P_w'u \sim N(0_{N-K-1}, I_{N-K-1})$, which is independent of w. Then, we have

$$T\hat{\delta}^2 \xrightarrow{d} \tilde{v}'(P_w'V_uP_w)\tilde{v}. \tag{A.78}$$

For testing $H_0 : \delta = 0$, $T\hat{\delta}^2$ is compared with $\sum_{i=1}^{N-K-1}\hat{\xi}_iX_i$, where the X_i's are independent chi-squared random variables with one degree of freedom and the $\hat{\xi}_i$'s are the $N-K-1$ nonzero eigenvalues of

$$\hat{S}^{\frac{1}{2}}\hat{U}^{-1}\hat{S}^{\frac{1}{2}} - \hat{S}^{\frac{1}{2}}\hat{U}^{-1}\hat{D}(\hat{D}'\hat{U}^{-1}\hat{D})^{-1}\hat{D}'\hat{U}^{-1}\hat{S}^{\frac{1}{2}}. \tag{A.79}$$

Using (A.32), we can write the above matrix as

$$\begin{aligned} & \hat{S}^{\frac{1}{2}}\hat{U}^{-\frac{1}{2}}[I_N - \hat{U}^{-\frac{1}{2}}\hat{D}(\hat{D}'\hat{U}^{-1}\hat{D})^{-1}\hat{D}'\hat{U}^{-\frac{1}{2}}]\hat{U}^{-\frac{1}{2}}\hat{S}^{\frac{1}{2}} \\ = \; & \hat{S}^{\frac{1}{2}}\hat{U}^{-\frac{1}{2}}\hat{M}\hat{U}^{-\frac{1}{2}}\hat{S}^{\frac{1}{2}} - \hat{S}^{\frac{1}{2}}\hat{U}^{-\frac{1}{2}}\hat{M}\hat{U}^{-\frac{1}{2}}\hat{d}(\hat{d}'\hat{U}^{-\frac{1}{2}}\hat{M}\hat{U}^{-\frac{1}{2}}\hat{d})^{-1}\hat{d}'\hat{U}^{-\frac{1}{2}}\hat{M}\hat{U}^{-\frac{1}{2}}\hat{S}^{\frac{1}{2}}. \end{aligned} \tag{A.80}$$

Let $\hat{P}$ be an $N\times(N-K)$ orthonormal matrix such that $\hat{P}\hat{P}' = \hat{M}$ and $\hat{P}_w$ be an $(N-K)\times(N-K-1)$ orthonormal matrix such that $\hat{P}_w\hat{P}_w' = I_{N-K} - \hat{P}'^{-\frac{1}{2}}\hat{d}(\hat{d}'\hat{U}^{-\frac{1}{2}}\hat{M}\hat{U}^{-\frac{1}{2}}\hat{d})^{-1}\hat{d}'\hat{U}^{-\frac{1}{2}}\hat{P}$. We can easily show that $\hat{\xi}_i$'s are the nonzero eigenvalues of

$$\hat{S}^{\frac{1}{2}}\hat{U}^{-\frac{1}{2}}\hat{P}\hat{P}_w\hat{P}_w'\hat{P}'\hat{U}^{-\frac{1}{2}}\hat{S}^{\frac{1}{2}}, \tag{A.81}$$

or equivalently the eigenvalues of

$$\hat{P}_w'\hat{P}'\hat{U}^{-\frac{1}{2}}\hat{S}\hat{U}^{-\frac{1}{2}}\hat{P}\hat{P}_w. \tag{A.82}$$

Using (A.35), we can show that

$$\hat{P}'\hat{U}^{-\frac{1}{2}}\hat{e}_t \xrightarrow{d} P'^{-\frac{1}{2}}e_t(\gamma_1^*) + \frac{w'u}{w'w}P'^{-\frac{1}{2}}x_tg_t. \tag{A.83}$$

It follows that

$$\hat{P}'\hat{U}^{-\frac{1}{2}}\hat{S}\hat{U}^{-\frac{1}{2}}\hat{P} \xrightarrow{d} P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P + \frac{(w'u)^2}{(w'w)^2}I_{N-K} = V_u + \frac{(w'V_uw)\tilde{u}^2}{(w'w)^2}I_{N-K}, \tag{A.84}$$

where $\tilde{u} = w'u/(w'V_uw)^{\frac{1}{2}} \sim N(0,1)$ and it is independent of w.

Under the conditional homoskedasticity assumption, we have $V_u = E[(\tilde{f}_t'\gamma_1^*)^2]I_{N-K}$ and hence

$$\begin{aligned} T\hat{\delta}^2 & \xrightarrow{d} E[(\tilde{f}_t'\gamma_1^*)^2]\tilde{v}'\tilde{v} \sim E[(\tilde{f}_t'\gamma_1^*)^2]\chi^2_{N-K-1}, & \text{(A.85)} \\ \hat{P}_w'\hat{P}'\hat{U}^{-\frac{1}{2}}\hat{S}\hat{U}^{-\frac{1}{2}}\hat{P}\hat{P}_w & \xrightarrow{d} E[(\tilde{f}_t'\gamma_1^*)^2]\left(1 + \frac{\tilde{u}^2}{w'w}\right)I_{N-K-1}. & \text{(A.86)} \end{aligned}$$

It follows that

$$\hat{\xi}_i \xrightarrow{d} E[(\tilde{f}_t'\gamma_1^*)^2]\left(1+\frac{\tilde{u}^2}{w'w}\right) = \frac{E[(\tilde{f}_t'\gamma_1^*)^2]}{Q_1}, \tag{A.87}$$

where $Q_1 = w'w/(\tilde{u}^2 + w'w) \sim \text{Beta}\left(\frac{N-K}{2}, \frac{1}{2}\right)$ and it is independent of $\tilde{v}'\tilde{v}$. Therefore, the limiting probability of rejection of the HJ-distance test of size α is

$$\int_0^1 P\left[\chi^2_{N-K-1} > \frac{c_\alpha}{q}\right] f_{Q_1}(q)\mathrm{d}q, \tag{A.88}$$

where c_α is the $100(1-\alpha)$ percentile of χ^2_{N-K-1}. Since $0 < Q_1 < 1$, the limiting probability of rejection is less than α. This completes the proof of part (a) of Theorem 3.

part (b): Using (A.57), the limiting distribution of the squared sample HJ-distance $\hat{\delta}^2 = \hat{e}'\hat{U}^{-1}\hat{e}$ can be obtained as

$$\begin{aligned} \hat{\delta}^2 &\xrightarrow{d} \tilde{q}'P[I_{N-K} - w(w'w)^{-1}w']P'\tilde{q} \\ &= (\tilde{q}'PP'\tilde{q})\frac{w'[I_{N-K} - P'\tilde{q}(\tilde{q}'PP'\tilde{q})^{-1}\tilde{q}'P]w}{w'w} = \delta^2 Q_2, \end{aligned} \tag{A.89}$$

where

$$Q_2 = \frac{w'[I_{N-K} - P'\tilde{q}(\tilde{q}'PP'\tilde{q})^{-1}\tilde{q}'P]w}{w'w} \sim \text{Beta}\left(\frac{N-K-1}{2}, \frac{1}{2}\right) \tag{A.90}$$

and it is independent of w.

From the proof of part (a), we know that the $\hat{\xi}_i$'s are the eigenvalues of

$$\hat{P}_w'\hat{P}'\hat{U}^{-\frac{1}{2}}\hat{S}\hat{U}^{-\frac{1}{2}}\hat{P}\hat{P}_w. \tag{A.91}$$

From (15) and (11), we have

$$\frac{\hat{S}}{T} \xrightarrow{d} \frac{\delta^2 s^2}{(w'w)^2}U, \tag{A.92}$$

which implies

$$\frac{\hat{P}_w'\hat{P}'\hat{U}^{-\frac{1}{2}}\hat{S}\hat{U}^{-\frac{1}{2}}\hat{P}\hat{P}_w}{T} \xrightarrow{d} \frac{\delta^2 s^2}{(w'w)^2} I_{N-K-1} \tag{A.93}$$

and

$$\frac{\hat{\xi}_i}{T} \xrightarrow{d} \frac{\delta^2 s^2}{(w'w)^2} = \frac{\delta^2(1-Q_2)}{w'w}. \tag{A.94}$$

When we compare $T\hat{\delta}^2$ with the distribution of $\sum_{i=1}^{N-K-1}\hat{\xi}_i X_i$, we are effectively comparing Q_2 with $(1-Q_2)/(w'^2_{N-K-1}$, and we will reject $H_0 : \delta = 0$ when

$$w'w > \frac{c_\alpha Q_2}{1-Q_2}. \tag{A.95}$$

Note that $w'w \sim \chi^2_{N-K}$ and it is independent of Q_2, so the limiting probability of rejection for a test with size α is

$$\int_0^1 P\left[\chi^2_{N-K} > \frac{c_\alpha q}{1-q}\right] f_{Q_2}(q)\mathrm{d}q. \tag{A.96}$$

This completes the proof of part (b) of Theorem 3.

References

Hansen, L. P., J. Heaton, and E. G. J. Luttmer, 1995, "Econometric Evaluation of Asset Pricing Models," *Review of Financial Studies*, 8, 237–274.

Jagannathan, R., and Z. Wang, 1996, "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance*, 51, 3–53.

Kan, R., and C. Robotti, 2008, "Specification Tests of Asset Pricing Models Using Excess Returns," *Journal of Empirical Finance*, 15, 816–838.

Kan, R., and C. Zhang, 1999, "GMM Tests of Stochastic Discount Factor Models with Useless Factors," *Journal of Financial Economics*, 54, 103–127.

Table 1
Empirical size of the *t*-tests (modified HJ-distance case)

Panel A: Model with a useful factor

t-test	T	Correctly specified model			Misspecified model		
		10%	5%	1%	10%	5%	1%
t_c	200	0.098	0.049	0.009	0.098	0.049	0.009
	600	0.100	0.050	0.009	0.099	0.048	0.009
	1000	0.097	0.048	0.010	0.099	0.049	0.009
	∞	0.100	0.050	0.010	0.100	0.050	0.010
t_m	200	0.098	0.049	0.009	0.098	0.048	0.009
	600	0.100	0.050	0.009	0.098	0.048	0.009
	1000	0.097	0.048	0.010	0.099	0.049	0.009
	∞	0.100	0.050	0.010	0.100	0.050	0.010

Panel B: Model with a useless factor

t-test	T	Correctly specified model			Misspecified model		
		10%	5%	1%	10%	5%	1%
t_c	200	0.129	0.067	0.013	0.327	0.235	0.101
	600	0.101	0.046	0.007	0.472	0.384	0.231
	1000	0.095	0.044	0.006	0.556	0.477	0.328
	∞	0.088	0.039	0.005	1.000	1.000	1.000
t_m	200	0.037	0.012	0.001	0.080	0.036	0.005
	600	0.022	0.006	0.000	0.082	0.038	0.006
	1000	0.021	0.006	0.000	0.088	0.041	0.007
	∞	0.018	0.004	0.000	0.100	0.050	0.010

The table presents the empirical size of the t-tests of $H_0 : \gamma_1 = \gamma_1^*$ in a model with a useful factor (Panel A) and in a model with a useless factor (Panel B). Each panel considers the case in which the model is correctly specified and the case in which the model is misspecified. t_c denotes the t-test constructed under the assumption of correct model specification and t_m denotes the misspecification-robust t-test. We report results for different levels of significance (10%, 5% and 1%) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the excess returns on the 25 Fama-French size and book-to-market portfolios and the 17 Fama-French industry portfolios for the period 1959:2–2012:12. The various t-statistics are compared to the critical values from a standard normal distribution. In Panel B, the rejection rates for the limiting case ($T = \infty$) are based on the asymptotic distributions given in Theorem 5.

Table 2
Empirical size of the *t*-tests (modified HJ-distance case)

Panel A: Correctly specified model

		$\hat{\gamma}_1$			$\hat{\gamma}_2$		
t-test	T	10%	5%	1%	10%	5%	1%
t_c	200	0.094	0.045	0.008	0.130	0.066	0.012
	600	0.095	0.047	0.009	0.100	0.047	0.007
	1000	0.097	0.048	0.009	0.095	0.043	0.006
	∞	0.092	0.045	0.008	0.088	0.039	0.005
t_m	200	0.090	0.042	0.008	0.036	0.012	0.001
	600	0.091	0.044	0.008	0.023	0.006	0.000
	1000	0.093	0.046	0.008	0.020	0.005	0.000
	∞	0.088	0.042	0.008	0.018	0.004	0.000

Panel B: Misspecified model

		$\hat{\gamma}_1$			$\hat{\gamma}_2$		
t-test	T	10%	5%	1%	10%	5%	1%
t_c	200	0.094	0.046	0.008	0.321	0.230	0.098
	600	0.095	0.047	0.008	0.464	0.374	0.223
	1000	0.094	0.046	0.008	0.553	0.471	0.321
	∞	0.088	0.039	0.005	1.000	1.000	1.000
t_m	200	0.086	0.041	0.007	0.080	0.036	0.005
	600	0.079	0.036	0.006	0.081	0.038	0.006
	1000	0.072	0.032	0.005	0.088	0.041	0.007
	∞	0.001	0.000	0.000	0.100	0.050	0.010

The table presents the empirical size of the t-tests of $H_0 : \gamma_i = \gamma_i^*$ $(i = 1, 2)$ in a model with a useful and a useless factor. γ_1 is the coefficient on the useful factor and γ_2 is the coefficient on the useless factor. t_c denotes the t-test constructed under the assumption of correct model specification and t_m denotes the misspecification-robust t-test. We report results for different levels of significance (10%, 5% and 1%) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the excess returns on the 25 Fama-French size and book-to-market portfolios and the 17 Fama-French industry portfolios for the period 1959:2–2012:12. The various t-tests are compared to the critical values from a standard normal distribution. The rejection rates for the limiting case ($T = \infty$) are based on the asymptotic distributions given in Theorem 5.

Table 3
Survival rates of risk factors: two useful, one unpriced and one useless factors (modified HJ-distance case)

Panel A: Correctly specified model

	Useful ($\gamma_1^* \neq 0$)		Useful ($\gamma_2^* \neq 0$)		Useful ($\gamma_3^* = 0$)		Useless		Prob.	
T	$t_c(\hat{\gamma}_1)$	$t_m(\hat{\gamma}_1)$	$t_c(\hat{\gamma}_2)$	$t_m(\hat{\gamma}_2)$	$t_c(\hat{\gamma}_3)$	$t_m(\hat{\gamma}_3)$	$t_c(\hat{\gamma}_4)$	$t_m(\hat{\gamma}_4)$	MS_c	MS_m
200	0.253	0.239	0.380	0.355	0.010	0.008	0.013	0.001	0.023	0.008
600	0.862	0.852	0.962	0.958	0.010	0.009	0.008	0.000	0.018	0.009
1000	0.986	0.984	0.999	0.999	0.010	0.009	0.006	0.000	0.016	0.009

Panel B: Misspecified model

	Useful ($\gamma_1^* \neq 0$)		Useful ($\gamma_2^* \neq 0$)		Useful ($\gamma_3^* = 0$)		Useless		Prob.	
T	$t_c(\hat{\gamma}_1)$	$t_m(\hat{\gamma}_1)$	$t_c(\hat{\gamma}_2)$	$t_m(\hat{\gamma}_2)$	$t_c(\hat{\gamma}_3)$	$t_m(\hat{\gamma}_3)$	$t_c(\hat{\gamma}_4)$	$t_m(\hat{\gamma}_4)$	MS_c	MS_m
200	0.242	0.213	0.368	0.320	0.013	0.007	0.084	0.005	0.096	0.012
600	0.818	0.776	0.930	0.908	0.013	0.007	0.201	0.006	0.211	0.013
1000	0.958	0.934	0.989	0.983	0.013	0.007	0.295	0.008	0.304	0.015

The table presents the survival rates of the useful and useless factors in a model with a constant, two useful factors (with $\gamma_1^* \neq 0$ and $\gamma_2^* \neq 0$), a useful factor that does not contribute to pricing (with $\gamma_3^* = 0$) and a useless factor (with γ_4^* unidentified). The sequential procedure is implemented by using the misspecification-robust t-tests ($t_m(\hat{\gamma}_i)$ column) as well as the t-tests under correctly specified models ($t_c(\hat{\gamma}_i)$ column). The false discovery rate of the multiple testing procedure is controlled using the Bonferroni method. The last two columns of the table report the probability that at least one useless or unpriced useful factor survives using the t-tests under correctly specified models (MS_c) and misspecification-robust t-tests (MS_m). The nominal level of the sequential testing procedure is set equal to 5%. Panels A and B are for correctly specified and misspecified models, respectively. We report results for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the excess returns on the 25 Fama-French size and book-to-market portfolios and the 17 Fama-French industry portfolios for the period 1959:2–2012:12.

Table 4
Survival rates of risk factors: two useful and two useless factors (modified HJ-distance case)

Panel A: Correctly specified model

	Useful ($\gamma_1^* \neq 0$)		Useful ($\gamma_2^* \neq 0$)		Useless		Useless		Prob.	
T	$t_c(\hat{\gamma}_1)$	$t_m(\hat{\gamma}_1)$	$t_c(\hat{\gamma}_2)$	$t_m(\hat{\gamma}_2)$	$t_c(\hat{\gamma}_3)$	$t_m(\hat{\gamma}_3)$	$t_c(\hat{\gamma}_4)$	$t_m(\hat{\gamma}_4)$	MS_c	MS_m
200	0.272	0.254	0.375	0.344	0.012	0.001	0.012	0.001	0.024	0.001
600	0.891	0.877	0.959	0.951	0.007	0.000	0.007	0.000	0.014	0.001
1000	0.991	0.989	0.999	0.998	0.006	0.000	0.006	0.000	0.012	0.000

Panel B: Misspecified model

	Useful ($\gamma_1^* \neq 0$)		Useful ($\gamma_2^* \neq 0$)		Useless		Useless		Prob.	
T	$t_c(\hat{\gamma}_1)$	$t_m(\hat{\gamma}_1)$	$t_c(\hat{\gamma}_2)$	$t_m(\hat{\gamma}_2)$	$t_c(\hat{\gamma}_3)$	$t_m(\hat{\gamma}_3)$	$t_c(\hat{\gamma}_4)$	$t_m(\hat{\gamma}_4)$	MS_c	MS_m
200	0.252	0.218	0.352	0.294	0.075	0.004	0.075	0.004	0.147	0.008
600	0.812	0.751	0.900	0.857	0.178	0.005	0.179	0.005	0.340	0.010
1000	0.947	0.908	0.976	0.957	0.263	0.006	0.261	0.006	0.482	0.013

The table presents the survival rates of the useful and useless factors in a model with a constant, two useful factors (with $\gamma_1^* \neq 0$ and $\gamma_2^* \neq 0$), and two useless factors (with γ_3^* and γ_4^* unidentified). The sequential procedure is implemented by using the misspecification-robust t-tests ($t_m(\hat{\gamma}_i)$ column) as well as the t-tests under correctly specified models ($t_c(\hat{\gamma}_i)$ column). The false discovery rate of the multiple testing procedure is controlled using the Bonferroni method. The last two columns of the table report the probability that at least one useless factor survives using the t-tests under correctly specified models (MS_c) and misspecification-robust t-tests (MS_m). The nominal level of the sequential testing procedure is set equal to 5%. Panels A and B are for correctly specified and misspecified models, respectively. We report results for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the excess returns on the 25 Fama-French size and book-to-market portfolios and the 17 Fama-French industry portfolios for the period 1959:2–2012:12.

Table 5
Survival rates when a linear combination of the factors is useless (modified HJ-distance case)

Panel A: Correctly specified model

	Both factors survive		One factor survives		No factor survives	
T	t_c	t_m	t_c	t_m	t_c	t_m
200	0.026	0.003	0.247	0.250	0.727	0.747
600	0.015	0.001	0.677	0.685	0.308	0.313
1000	0.013	0.001	0.889	0.900	0.097	0.099

Panel B: Misspecified model

	Both factors survive		One factor survives		No factor survives	
T	t_c	t_m	t_c	t_m	t_c	t_m
200	0.140	0.013	0.228	0.255	0.631	0.733
600	0.275	0.015	0.505	0.684	0.219	0.301
1000	0.377	0.016	0.563	0.890	0.060	0.094

The table presents the probability that both factors survive, only one factor survives, and no factor survives in a model in which a linear combination of two useful factors is useless. The sequential procedure is implemented by using the misspecification-robust t-test (t_m column) as well as the t-test under correctly specified models (t_c column). The false discovery rate of the multiple testing procedure is controlled using the Bonferroni method. The nominal level of the sequential testing procedure is set equal to 5%. Panels A and B are for correctly specified and misspecified models, respectively. We report results for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the excess returns on the 25 Fama-French size and book-to-market portfolios and the 17 Fama-French industry portfolios for the period 1959:2–2012:12.

Table 6
Empirical size of the *t*-tests in a model with a useful factor (optimal GMM case)

Panel A: Correctly specified model

		$\hat{\gamma}_0$			$\hat{\gamma}_1$		
t-test	T	10%	5%	1%	10%	5%	1%
t_c	200	0.176	0.142	0.108	0.114	0.061	0.015
	600	0.140	0.100	0.063	0.103	0.052	0.011
	1000	0.125	0.082	0.043	0.102	0.051	0.010
	∞	0.100	0.050	0.010	0.100	0.050	0.010
t_m	200	0.173	0.141	0.108	0.107	0.055	0.012
	600	0.139	0.100	0.063	0.102	0.051	0.010
	1000	0.125	0.081	0.043	0.101	0.050	0.010
	∞	0.100	0.050	0.010	0.100	0.050	0.010

Panel B: Misspecified model

		$\hat{\gamma}_0$			$\hat{\gamma}_1$		
t-test	T	10%	5%	1%	10%	5%	1%
t_c	200	0.182	0.147	0.110	0.122	0.067	0.018
	600	0.143	0.103	0.065	0.110	0.057	0.013
	1000	0.128	0.085	0.044	0.107	0.055	0.012
	∞	0.100	0.050	0.010	0.100	0.050	0.010
t_m	200	0.175	0.144	0.110	0.109	0.056	0.012
	600	0.140	0.101	0.064	0.103	0.052	0.011
	1000	0.125	0.083	0.044	0.101	0.051	0.010
	∞	0.100	0.050	0.010	0.100	0.050	0.010

The table presents the empirical size of the t-tests of $H_0 : \gamma_i = \gamma_i^*$ $(i = 0, 1)$ in a model with a constant and a useful factor estimated by optimal (3-step) GMM. γ_0 is the coefficient on the constant term and γ_1 is the coefficient on the useful factor. t_c denotes the t-test constructed under the assumption of correct model specification and t_m denotes the misspecification-robust t-test. We report results for different levels of significance (10%, 5% and 1%) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns, the 17 Fama-French industry portfolio returns and the one-month T-bill rate for the period 1959:2–2012:12. The various t-statistics are compared to the critical values from a standard normal distribution.

Table 7
Empirical size of the *t*-tests in a model with a useless factor (optimal GMM case)

Panel A: Correctly specified model

		$\hat{\gamma}_0$			$\hat{\gamma}_1$		
t-test	T	10%	5%	1%	10%	5%	1%
t_c	200	0.012	0.004	0.000	0.150	0.088	0.026
	600	0.003	0.000	0.000	0.107	0.053	0.009
	1000	0.002	0.000	0.000	0.100	0.047	0.007
	∞	0.001	0.000	0.000	0.088	0.039	0.005
t_m	200	0.002	0.000	0.000	0.038	0.015	0.002
	600	0.000	0.000	0.000	0.024	0.007	0.000
	1000	0.000	0.000	0.000	0.018	0.004	0.000
	∞	0.000	0.000	0.000	0.016	0.004	0.000

Panel B: Misspecified model

		$\hat{\gamma}_0$			$\hat{\gamma}_1$		
t-test	T	10%	5%	1%	10%	5%	1%
t_c	200	0.043	0.020	0.004	0.350	0.267	0.146
	600	0.035	0.013	0.002	0.475	0.391	0.248
	1000	0.040	0.015	0.002	0.559	0.481	0.336
	∞	0.088	0.039	0.005	1.000	1.000	1.000
t_m	200	0.007	0.002	0.000	0.079	0.039	0.009
	600	0.003	0.001	0.000	0.083	0.040	0.007
	1000	0.003	0.000	0.000	0.088	0.043	0.008
	∞	0.001	0.000	0.000	0.100	0.050	0.010

The table presents the empirical size of the t-tests of $H_0 : \gamma_i = \gamma_i^*$ $(i = 0, 1)$ in a model with a constant and a useless factor estimated by optimal (3-step) GMM. γ_0 is the coefficient on the constant term and γ_1 is the coefficient on the useless factor. t_c denotes the t-test constructed under the assumption of correct model specification and t_m denotes the misspecification-robust t-test. We report results for different levels of significance (10%, 5% and 1%) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns, the 17 Fama-French industry portfolio returns and the one-month T-bill rate for the period 1959:2–2012:12. The various t-statistics are compared to the critical values from a standard normal distribution. The rejection rates for the limiting case ($T = \infty$) are equivalent to those based on the asymptotic distributions given in Theorem 2.

Table 8
Empirical size of the *t*-tests in a model with a useful and a useless factor (optimal GMM case)

Panel A: Correctly specified model

		$\hat{\gamma}_0$			$\hat{\gamma}_1$			$\hat{\gamma}_2$		
t-test	T	10%	5%	1%	10%	5%	1%	10%	5%	1%
t_c	200	0.064	0.029	0.008	0.118	0.064	0.015	0.153	0.091	0.028
	600	0.061	0.029	0.008	0.101	0.051	0.010	0.108	0.054	0.009
	1000	0.058	0.025	0.006	0.097	0.049	0.009	0.099	0.048	0.007
	∞	0.052	0.020	0.002	0.096	0.047	0.009	0.088	0.039	0.005
t_m	200	0.031	0.013	0.004	0.103	0.052	0.011	0.040	0.016	0.002
	600	0.038	0.017	0.006	0.095	0.047	0.009	0.024	0.006	0.000
	1000	0.037	0.016	0.004	0.092	0.045	0.008	0.021	0.006	0.000
	∞	0.037	0.014	0.002	0.092	0.045	0.008	0.018	0.004	0.000

Panel B: Misspecified model

		$\hat{\gamma}_0$			$\hat{\gamma}_1$			$\hat{\gamma}_2$		
t-test	T	10%	5%	1%	10%	5%	1%	10%	5%	1%
t_c	200	0.086	0.041	0.010	0.144	0.084	0.026	0.350	0.266	0.144
	600	0.077	0.034	0.007	0.124	0.067	0.016	0.471	0.385	0.241
	1000	0.076	0.032	0.006	0.120	0.065	0.015	0.552	0.473	0.330
	∞	0.088	0.039	0.005	0.088	0.039	0.005	1.000	1.000	1.000
t_m	200	0.026	0.010	0.003	0.106	0.056	0.012	0.081	0.040	0.008
	600	0.018	0.006	0.002	0.089	0.042	0.008	0.082	0.040	0.008
	1000	0.013	0.005	0.001	0.080	0.037	0.006	0.089	0.042	0.008
	∞	0.001	0.000	0.000	0.001	0.000	0.000	0.100	0.050	0.010

The table presents the empirical size of the t-tests of $H_0 : \gamma_i = \gamma_i^*$ $(i = 0, 1, 2)$ in a model with a constant, a useful and a useless factor estimated by optimal (3-step) GMM. γ_0 is the coefficient on the constant term, γ_1 is the coefficient on the useful factor, and γ_2 is the coefficient on the useless factor. t_c denotes the t-test constructed under the assumption of correct model specification and t_m denotes the misspecification-robust t-test. We report results for different levels of significance (10%, 5% and 1%) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns, the 17 Fama-French industry portfolio returns and the one-month T-bill rate for the period 1959:2–2012:12. The various t-tests are compared to the critical values from a standard normal distribution. The rejection rates for the limiting case ($T = \infty$) are equivalent to those based on the asymptotic distributions given in Theorem 2.

Table 9
Survival rates of risk factors: two useful, one unpriced and one useless factors (optimal GMM case)

Panel A: Correctly specified model

	Useful ($\gamma_1^* \neq 0$)		Useful ($\gamma_2^* \neq 0$)		Useful ($\gamma_3^* = 0$)		Useless		Prob.	
T	$t_c(\hat{\gamma}_1)$	$t_m(\hat{\gamma}_1)$	$t_c(\hat{\gamma}_2)$	$t_m(\hat{\gamma}_2)$	$t_c(\hat{\gamma}_3)$	$t_m(\hat{\gamma}_3)$	$t_c(\hat{\gamma}_4)$	$t_m(\hat{\gamma}_4)$	MS_c	MS_m
200	0.744	0.708	0.812	0.770	0.042	0.030	0.048	0.004	0.087	0.034
600	0.999	0.999	1.000	1.000	0.016	0.014	0.014	0.001	0.029	0.014
1000	1.000	1.000	1.000	1.000	0.015	0.014	0.011	0.000	0.026	0.014

Panel B: Misspecified model

	Useful ($\gamma_1^* \neq 0$)		Useful ($\gamma_2^* \neq 0$)		Useful ($\gamma_3^* = 0$)		Useless		Prob.	
T	$t_c(\hat{\gamma}_1)$	$t_m(\hat{\gamma}_1)$	$t_c(\hat{\gamma}_2)$	$t_m(\hat{\gamma}_2)$	$t_c(\hat{\gamma}_3)$	$t_m(\hat{\gamma}_3)$	$t_c(\hat{\gamma}_4)$	$t_m(\hat{\gamma}_4)$	MS_c	MS_m
200	0.713	0.639	0.785	0.697	0.062	0.033	0.157	0.013	0.207	0.045
600	0.994	0.995	0.997	0.998	0.026	0.014	0.219	0.009	0.237	0.023
1000	0.999	0.999	1.000	1.000	0.023	0.013	0.299	0.009	0.314	0.022

The table presents the survival rates of the useful and useless factors in a model with a constant, two useful factors (with $\gamma_1^* \neq 0$ and $\gamma_2^* \neq 0$), a useful factor that does not contribute to pricing (with $\gamma_3^* = 0$) and a useless factor (with γ_4^* unidentified) estimated by optimal (3-step) GMM. The sequential procedure is implemented by using the misspecification-robust t-tests ($t_m(\hat{\gamma}_i)$ column) as well as the t-tests under correctly specified models ($t_c(\hat{\gamma}_i)$ column). The false discovery rate of the multiple testing procedure is controlled using the Bonferroni method. The last two columns of the table report the probability that at least one useless or unpriced useful factor survives using the t-tests under correctly specified models (MS_c) and misspecification-robust t-tests (MS_m). The nominal level of the sequential testing procedure is set equal to 5%. Panels A and B are for correctly specified and misspecified models, respectively. We report results for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns, the 17 Fama-French industry portfolio returns and the one-month T-bill rate for the period 1959:2–2012:12.

Table 10
Survival rates of risk factors: two useful and two useless factors (optimal GMM case)

Panel A: Correctly specified model

	Useful ($\gamma_1^* \neq 0$)		Useful ($\gamma_2^* \neq 0$)		Useless		Useless		Prob.	
T	$t_c(\hat{\gamma}_1)$	$t_m(\hat{\gamma}_1)$	$t_c(\hat{\gamma}_2)$	$t_m(\hat{\gamma}_2)$	$t_c(\hat{\gamma}_3)$	$t_m(\hat{\gamma}_3)$	$t_c(\hat{\gamma}_4)$	$t_m(\hat{\gamma}_4)$	MS_c	MS_m
200	0.749	0.715	0.807	0.768	0.048	0.005	0.047	0.005	0.092	0.009
600	0.999	0.999	1.000	1.000	0.014	0.001	0.014	0.001	0.028	0.001
1000	1.000	1.000	1.000	1.000	0.011	0.000	0.010	0.000	0.021	0.001

Panel B: Misspecified model

	Useful ($\gamma_1^* \neq 0$)		Useful ($\gamma_2^* \neq 0$)		Useless		Useless		Prob.	
T	$t_c(\hat{\gamma}_1)$	$t_m(\hat{\gamma}_1)$	$t_c(\hat{\gamma}_2)$	$t_m(\hat{\gamma}_2)$	$t_c(\hat{\gamma}_3)$	$t_m(\hat{\gamma}_3)$	$t_c(\hat{\gamma}_4)$	$t_m(\hat{\gamma}_4)$	MS_c	MS_m
200	0.356	0.283	0.453	0.359	0.153	0.012	0.152	0.011	0.284	0.023
600	0.843	0.873	0.915	0.935	0.224	0.011	0.222	0.011	0.415	0.022
1000	0.951	0.973	0.977	0.987	0.295	0.012	0.292	0.011	0.533	0.023

The table presents the survival rates of the useful and useless factors in a model with a constant, two useful factors (with $\gamma_1^* \neq 0$ and $\gamma_2^* \neq 0$), and two useless factors (with γ_3^* and γ_4^* unidentified) estimated by optimal (3-step) GMM. The sequential procedure is implemented by using the misspecification-robust t-tests ($t_m(\hat{\gamma}_i)$ column) as well as the t-tests under correctly specified models ($t_c(\hat{\gamma}_i)$ column). The false discovery rate of the multiple testing procedure is controlled using the Bonferroni method. The last two columns of the table report the probability that at least one useless factor survives using the t-tests under correctly specified models (MS_c) and misspecification-robust t-tests (MS_m). The nominal level of the sequential testing procedure is set equal to 5%. Panels A and B are for correctly specified and misspecified models, respectively. We report results for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns, the 17 Fama-French industry portfolio returns and the one-month T-bill rate for the period 1959:2–2012:12.

Table 11
Survival rates when a linear combination of the factors is useless (optimal GMM case)

Panel A: Correctly specified model

	Both factors survive		One factor survives		No factor survives	
T	t_c	t_m	t_c	t_m	t_c	t_m
200	0.046	0.006	0.259	0.253	0.695	0.741
600	0.020	0.002	0.673	0.683	0.306	0.315
1000	0.016	0.001	0.887	0.899	0.097	0.099

Panel B: Misspecified model

	Both factors survive		One factor survives		No factor survives	
T	t_c	t_m	t_c	t_m	t_c	t_m
200	0.186	0.017	0.228	0.240	0.586	0.743
600	0.295	0.017	0.489	0.670	0.216	0.313
1000	0.389	0.019	0.552	0.882	0.059	0.099

The table presents the probability that both factors survive, only one factor survives, and no factor survives in a model estimated by optimal (3-step) GMM in which a linear combination of two useful factors is useless. The sequential procedure is implemented by using the misspecification-robust t-test (t_m column) as well as the t-test under correctly specified models (t_c column). The false discovery rate of the multiple testing procedure is controlled using the Bonferroni method. The nominal level of the sequential testing procedure is set equal to 5%. Panels A and B are for correctly specified and misspecified models, respectively. We report results for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns, the 17 Fama-French industry portfolio returns and the one-month T-bill rate for the period 1959:2–2012:12.

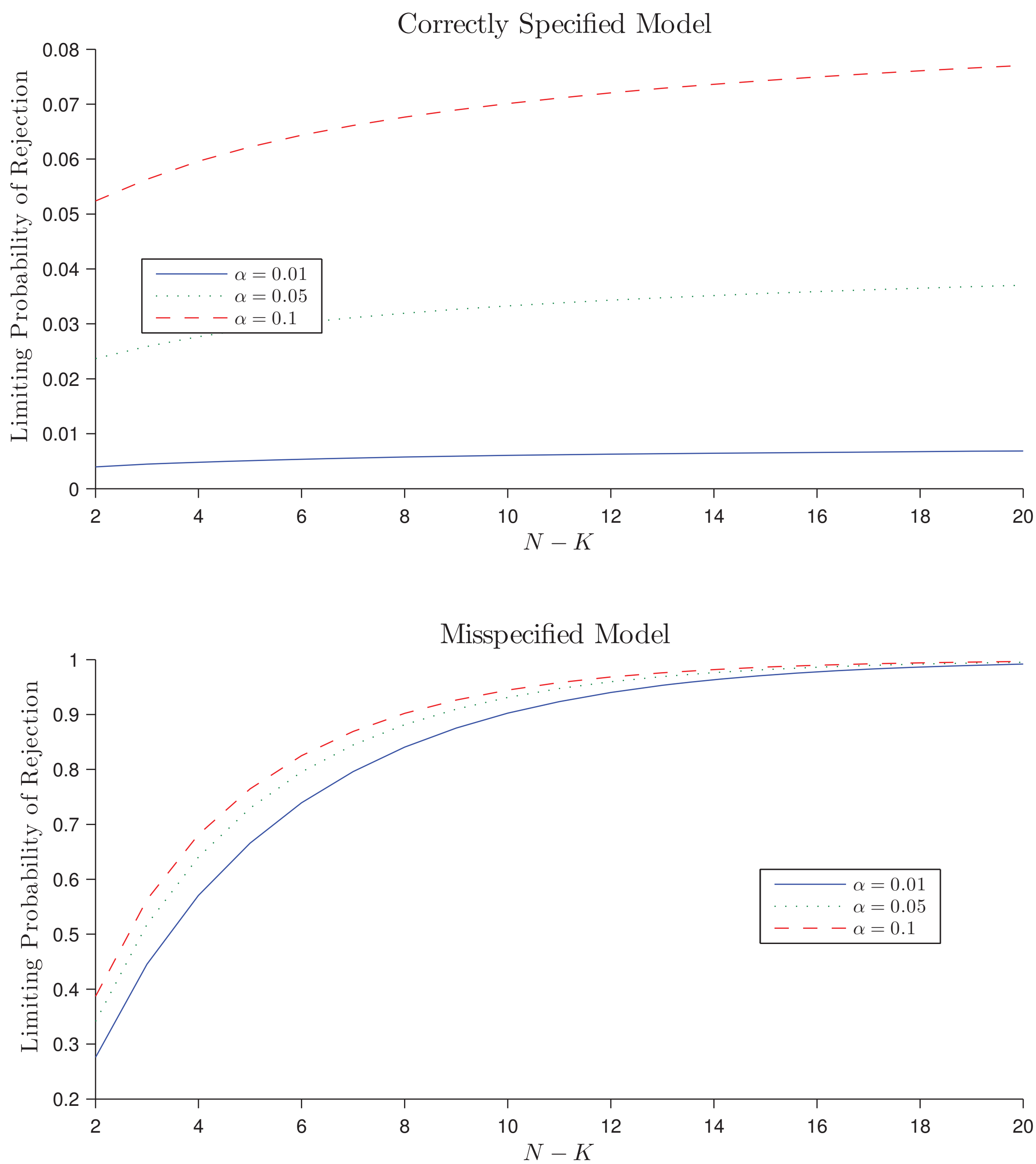

Figure 1
Limiting probabilities of rejection of the HJ-distance test. The figure presents the limiting probabilities of rejection of the HJ-distance test under correctly specified and misspecified models when one of the factors is useless.