Journal of Financial Econometrics, 2020, Vol. 18, No. 4, 736–775
doi: 10.1093/jjfinec/nbz009
Advance Access Publication Date: 20 March 2019
Commentary

OXFORD

Comment on: Pseudo-True SDFs in Conditional Asset Pricing Models. Comparing Fixed- versus Vanishing-Bandwidth Estimators of Pseudo-True SDFs*

Patrick Gagliardini[1,2] and Diego Ronchetti[3]

[1]Faculty of Economics, Università della Svizzera Italiana, [2]Swiss Finance Institute and [3]Faculty of Economics and Business, University of Groningen

Address correspondence to D. Ronchetti, Faculty of Economics and Business, University of Groningen, Nettelbosje 2, 9747 AE, Groningen, Netherlands, or e-mail d.ronchetti@rug.nl

accepted February 25, 2019

The paper by Antoine, Proulx, and Renault (2018) (APR) deals with the econometric definition, economic interpretation, and statistical estimation of the pseudo-true stochastic discount factor (SDF) in misspecified conditional asset pricing models. The paper revolves around fundamental issues like the role of conditioning information and omitted risk factors, and has non-trivial interactions with the current debate in the literature on the impact of weak factors (weak identification) for assessing asset pricing models. Building on, and substantially extending, previous contributions in the literature, the approach of the authors to define a pseudo-true SDF relies on the minimizers of econometric criteria based on a *conditional* version of the Hansen–Jagannathan (HJ) distance, that is, an average across states of squared conditional pricing errors. The authors provide an insightful discussion of the economic interpretation of pseudo-true SDFs. APR advocate the use of a fixed bandwidth (i.e., independent of the sample size) when estimating the conditional pricing errors by kernel regression methods to facilitate statistical analysis. This route leads to bandwidth-dependent pseudo-true SDF parameters and estimators thereof.

In our discussion, we investigate the different definitions of pseudo-true SDFs and interpret the fixed-bandwidth proposal as a model calibration which down-weights high-frequency Fourier components of the conditional pricing errors (Section 1). We compare the statistical properties of pseudo-true SDF parameters' estimators relying on vanishing versus fixed bandwidth, and provide a condition under which the former have a smaller

* We are very grateful to the Editors for the invitation to contribute to the discussion of Antoine, Proulx and Renault (2018). We thank E. Renault and O. Scaillet for very useful discussions.

asymptotic variance than the latter (or viceversa). We look at these topics through the lens of misspecified conditional linear SDF models in which priced risk factors are omitted using both simulated and real data (Section 3). We skip regularity conditions and relegate some technical derivations in the Appendix of the paper.

1 Estimation of SDF Pseudo-True Parameter Values

Let $m_{t+1}(\theta)$ denote a parametric family of SDFs between dates t and $t+1$, and $I(t)$ the conditioning information set at date t. The SDF model is misspecified for pricing a set of n test assets with gross returns vector R_{t+1} if there is no vector θ in the parameter space $\Theta \subset \mathbb{R}^p$ such that the conditional pricing error $e[I(t), \theta] = E\left[\Psi_{t+1}(\theta)|I(t)\right]$ vanishes almost surely, where

$$\Psi_{t+1}(\theta) := m_{t+1}(\theta) R_{t+1} - 1_n$$

is the n-dimensional conditional moment vector function. To define the concept of pseudo-true SDF parameter, the authors consider the conditional HJ distance introduced by Gagliardini and Ronchetti (2016) (GR), that is

$$\delta := \min_{\theta \in \Theta} \ \delta(\theta), \tag{1}$$

where

$$\delta(\theta)^2 := E\left[e[I(t), \theta]' \Omega^{-1}[I(t)] e[I(t), \theta]\right]$$

for $\Omega[I(t)] := E\left[R_{t+1} R'_{t+1}|I(t)\right]$. One definition of pseudo-true SDF parameter considered by APR is the argument that minimizes the criterion $\delta(\theta)$, namely[1]

$$\theta^* := \arg\min_{\theta \in \Theta} \ \delta(\theta). \tag{2}$$

APR deploy the first-order condition of the minimization problem in Equation (2) to get economic interpretations for the pseudo-true SDF parameter. For the sake of conciseness, in our discussion we focus on an interpretation in terms of "optimal" instruments. Specifically, the pseudo-true SDF $m_{t+1}(\theta^*)$ yields a set of exactly identified *unconditional* moment restrictions for a set of managed portfolios

$$E\left[z_t(\theta^*) \Psi_{t+1}(\theta^*)\right] = 0_p,$$

for the $(p \times n)$-dimensional instrument matrix

$$z_t(\theta) := E\left[\frac{\partial \Psi_{t+1}(\theta)}{\partial \theta'} |I(t)\right]' \Omega^{-1}[I(t)]. \tag{3}$$

Thus, under misspecification the focus moves from the GMM optimal instrument matrix

$$E\left[\frac{\partial \Psi_{t+1}(\theta_0)}{\partial \theta'} \middle| I(t)\right]' \mathrm{Var}\left[\Psi_{t+1}(\theta_0)|I(t)\right]^{-1}$$

to the "HJ-optimal" instrument matrix $z_t(\theta^*)$.

1 APR also considers the minimizer θ_t of the state-dependent criterion $\delta(\theta)[I(t)] = e\left[I(t), \theta\right]' \Omega^{-1}\left[I(t)\right] e\left[I(t), \theta\right]$.

At this stage, it is convenient for the measurement of misspecification to assume that the conditioning information $I(t)$ can be summarized by the m-dimensional state variables vector x_t, so that $e[I(t),\theta] = e(x_t,\theta)$ for all $\theta \in \Theta$ and $\Omega[I(t)] = \Omega(x_t)$.[2] In this setting, APR considers three variations of the pseudo-true SDF parameter and consistent estimators of them.

1. The local GMM estimator $\hat{\theta}$. This estimator, which is studied by GR, is defined as

$$\hat{\theta} := \arg\min_{\theta \in \Theta} \frac{1}{T}\sum_{t=1}^{T} 1(x_t)\hat{E}\left[\Psi_{t+1}(\theta)|x_t\right]'\hat{\Omega}^{-1}(x_t)\hat{E}\left[\Psi_{t+1}(\theta)|x_t\right],$$

   where $1(x_t)$ is a trimming factor, $\hat{\Omega}(x_t) = \hat{E}\left[R_{t+1}R'_{t+1}|x_t\right]$, and $\hat{E}[\cdot|x_t]$ denotes the Nadaraya–Watson kernel regression estimator. Vector $\hat{\theta}$ is a consistent estimator for the minimizer of criterion

$$Q_\infty(\theta) = E\left[1(x_t)e(x_t,\theta)'\Omega^{-1}(x_t)e(x_t,\theta)\right] \tag{4}$$

   in which the trimming factor is included in the expectation operator. By a slight abuse of notation, we denote the minimizer of criterion Q_∞ also by θ^*.
2. The GMM estimator with HJ-optimal instruments $\tilde{\theta}$. This estimator, which is similar to the Nagel and Singleton (2011) estimator but uses the HJ-optimal instrument matrix for the misspecified setting, solves the first-order condition

$$\frac{1}{T}\sum_{t=1}^{T} 1(x_t)\hat{z}_t(\tilde{\theta})\Psi_{t+1}(\tilde{\theta}) = 0_p,$$

   where

$$\hat{z}_t(\theta) := \hat{E}\left[\frac{\partial \Psi_{t+1}(\theta)}{\partial \theta'}|x_t\right]'\hat{\Omega}^{-1}(x_t)$$

   is an estimator of the instrument matrix $z_t(\theta)$ defined in Equation (3) obtained by Nadaraya–Watson kernel regression functions.
3. The smooth minimum distance (SMD) estimator with fixed kernel bandwidth $\hat{\theta}(h)$. In motivating this estimator, APR build on the insight of Hall and Inoue (2003) who highlighted that for a misspecified unconditional moment restriction setting the estimation of the Jacobian and weighting matrices affects the distributional properties of the GMM estimator even asymptotically. APR write on p. 30 of their paper that "... this message is ominous regarding the impact of misspecification" in conditional moment restriction setting and might complicate the study of the estimators' asymptotic properties. To cope with this potential issue, they advocate the use of an estimator with fixed bandwidth. To define the latter, APR write the criterion underlying the square of the conditional HJ distance in Equation (1) as $\delta^2(\theta) = E[(\Omega(x_t)^{-1/2}e(x_t,\theta))'$

2 More precisely, either we assume that the model is such that the dependence of $e[I(t),\theta]$ and $\Omega[I(t)]$ on x_t alone holds when the information set $I(t)$ is the sigma-field $I(t) = \sigma\{R^t, Y^t, x^t\}$, where Y is the vector of risk factors in the SDF, and $x^t := \{x_t, x_{t-1}, \ldots\}$, or $I(t) = \sigma\{x_t\}$ is a "reduced" conditioning information set with respect to (w.r.t.) which the econometrician measures model misspecification.

$(\Omega(x_t)^{-1/2}\Psi_{t+1}(\theta))]$. Moreover, they introduce a weighting of the argument within the expectation operator equal to the state vector density $f(x_t)$ to deal with boundary effects. Then, to get a sample counterpart of the criterion, the outer expectation is replaced by a sample average, and the quantity $f(x_t)\Omega^{-1/2}(x_t)e(x_t;\theta)$ for date t is replaced by the jackknife kernel estimator for kernel function K and bandwidth h:

$$\frac{1}{(T-1)h^m}\sum_{s=1,s\neq t}^{T}\hat{\Omega}^{-1/2}(x_s)\Psi_{s+1}(\theta)K\left(\frac{x_s-x_t}{h}\right). \tag{5}$$

The SMD estimator is defined as

$$\hat{\theta}(h):=\arg\min_{\theta\in\Theta}\frac{1}{T(T-1)h^m}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\left(\hat{\Omega}^{-1/2}(x_s)\Psi_{s+1}(\theta)\right)'\left(\hat{\Omega}^{-1/2}(x_t)\Psi_{t+1}(\theta)\right)K\left(\frac{x_s-x_t}{h}\right).$$

The motivation of APR to consider the bandwidth as a fixed parameter (i.e., independent of T) is essentially a statistical one, related to establishing the large sample distributional properties of the estimator.

What is the parameter of interest implied by the SMD estimator? Under regularity conditions for $T\to\infty$ and h fixed, the estimator $\hat{\theta}(h)$ converges to a pseudo-true SDF parameter $\theta^*(h)$ which is the minimizer of the large sample limit criterion:

$$\theta^*(h)=\arg\min_{\theta\in\Theta}\ Q_\infty(\theta,h).$$

This pseudo-true parameter is bandwidth-dependent. APR refers to it as a "calibration parameter" (APR, p. 24). It is an interesting question to explore the structural implications of such dependency on the bandwidth parameter. APR investigate the large sample limit criterion of the SMD estimator by leveraging on the analogy with the i.i.d. case, and introducing Assumptions A1, A2, and A3 reported in Section 6.2 including the strict exogeneity and Markov property of process $\{x_t\}$.

Here, we offer a different perspective to characterize the limit criterion Q_∞. We replace $f(x_t)\Omega^{-1/2}(x_t)e(x_t,\theta)$ by the large sample equivalent of the kernel estimator with fixed bandwidth h in Expression (5), namely $C(x_t,\theta,h)$ where

$$C(x,\theta,h)=\frac{1}{h^m}E\left[\Omega^{-1/2}(x_s)\Psi_{s+1}(\theta)K\left(\frac{x_s-x}{h}\right)\right]. \tag{6}$$

Let us assume that the kernel function K is symmetric and in the class of positive definite kernels (Andrews, 1991), and let $\overline{K}(\xi):=\frac{1}{(2\pi)^m}\int_{\mathbb{R}^m}K(x)\exp[-i\xi'x]\mathrm{d}x$ denote its Fourier transform. Then, by the inverse Fourier representation of function K, we get

$$C(x_t,\theta,h)=\frac{1}{h^m}\int_{\mathbb{R}^m}E\left[\Omega^{-1/2}(x_s)\Psi_{s+1}(\theta)\exp\left[\frac{i}{h}\xi'x_s\right]\right]\exp\left[-\frac{i}{h}\xi'x_t\right]\overline{K}(\xi)\mathrm{d}\xi. \tag{7}$$

Then, by plugging Equation (7) in the criterion underlying the squared density-weighted conditional HJ distance, and rearranging terms we get:

$$\begin{aligned}Q_\infty(\theta,h)&=E\left[C(x_t,\theta,h)'\Omega^{-1/2}(x_t)\Psi_{t+1}(\theta)\right]\\&=\frac{1}{h^m}\int_{\mathbb{R}^m}\left\|E\left[\Omega(x_t)^{-1/2}\Psi_{t+1}(\theta)\exp\left[-\frac{i}{h}\xi'x_t\right]\right]\right\|^2\overline{K}(\xi)\mathrm{d}\xi,\end{aligned} \tag{8}$$

where $\|\cdot\|$ is the standard norm for complex vectors. The property $\overline{K}(\xi) > 0$ for all $\xi \in \mathbb{R}^m$ of a positive definite kernel K is key for $Q_\infty(\theta, h)$ to be a suitable criterion function. APR recall that a conditional moment restriction is equivalent to a continuum of unconditional moment restrictions for instruments that are complex exponential transformations of the conditioning variable (Bierens, 1982). In fact, $Q_\infty(\theta, h)$ is the GMM limit criterion corresponding to a continuum of unconditional moment restrictions, with instrument matrix $\Omega(x_t)^{-1/2}$ multiplied by sinus and cosinus functions of $\xi' x_t / h$ and diagonal weighting operator. In other words, criterion $Q_\infty(\theta, h)$ involves the squared unconditional Fourier transforms of the scaled conditional pricing error vector averaged across frequencies of the process $\{x_t\}$.

It is instructive to consider the limits of criterion $Q_\infty(\theta, h)$ and pseudo-true parameter $\theta^*(h)$ when the bandwidth parameter either h vanishes or diverges to infinity. When $h \to 0$, the criterion $Q_\infty(\theta, h)$ converges to criterion $Q_\infty(\theta)$ in Equation (4), with $1(x_t)$ replaced by $f(x_t)$, since $C(x, \theta, h)$ converges to $f(x)\Omega^{-1/2}(x)e(x, \theta)$. Under uniform convergence, this implies that

$$\lim_{h \to 0} \theta^*(h) = \theta^*.$$

When $h \to \infty$, for $n \geq p$ and under a global identification condition, vector $\theta^*(h)$ converges to the "unconditional" pseudo-true SDF parameter

$$\theta_u^* := \arg\min_{\theta \in \Theta} \; E\left[\Omega^{-1/2}(x_t)\Psi_{t+1}(\theta)\right]' E\left[\Omega^{-1/2}(x_t)\Psi_{t+1}(\theta)\right], \tag{9}$$

which minimizes a quadratic form of the unconditional moments of the scaled pricing errors. Note that when the second moments of the gross returns are time invariant, θ_u^* is the minimizer of the unconditional HJ distance. Hence, by considering a fixed bandwidth, we are introducing a one-parameter family of pseudo-true SDF parameters that interpolate between the conditional and unconditional solutions.

The bottom line of this analysis is that, keeping the bandwidth parameter h fixed in the asymptotics, we can interpret an asymptotically biased estimator for the pseudo-true parameter θ^* as a consistent estimator of the pseudo-true parameter $\theta^*(h)$. When h increases, the criterion $Q_\infty(\theta, h)$ gives less weight to the unconditional Fourier transforms of the pricing errors for large frequencies of the process $\{x_t\}$. It is an open question to understand whether the pseudo-true SDF $m_{t+1}(\theta^*(h))$ for some $h > 0$ can be preferable to SDF $m_{t+1}(\theta^*)$ for a specific economic modeling purpose.

2 Large Sample Distributions of the Estimators

APR compare the different estimators of the pseudo-true SDFs in terms of their asymptotic distributions for $T \to \infty$. Here, we complete their study in Sections 5–7 by providing the large sample distributional properties of the local GMM estimator and of the GMM estimator with HJ-optimal instruments.

1. The asymptotic distribution of the local GMM estimator is provided by GR, Lemma C.2. Under regularity conditions—including restrictions on the convergence rate of the bandwidth h to 0—we have the asymptotic expansion

$$\sqrt{T}(\hat{\theta} - \theta^*) = -(H^*)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varphi_{t+1}(\theta^*) + o_p(1), \tag{10}$$

where

$$H^* := \frac{1}{2}\frac{\partial^2 Q_\infty(\theta^*)}{\partial\theta\partial\theta'} \quad (11)$$

for the criterion Q_∞ defined in Equation (4) and the function

$$\begin{aligned}\varphi_{t+1}(\theta^*) := {} & 1(x_t)J(x_t,\theta^*)'\Omega^{-1}(x_t)\Psi_{t+1}(\theta^*) + 1(x_t)\left(\frac{\partial\Psi_{t+1}(\theta^*)}{\partial\theta'}\right)'\Omega^{-1}(x_t)e(x_t,\theta^*) \\ & - 1(x_t)J(x_t,\theta^*)'\Omega^{-1}(x_t)R_{t+1}R'_{t+1}\Omega^{-1}(x_t)e(x_t,\theta^*),\end{aligned} \quad (12)$$

with

$$J(x_t,\theta) := E\left[\frac{\partial\Psi_{t+1}(\theta)}{\partial\theta'}|x_t\right]. \quad (13)$$

Then, under a CLT for dependent data, $\sqrt{T}(\hat\theta - \theta^*)$ is asymptotically Gaussian with zero mean and variance $\text{AsVar}[\hat\theta] = (H^*)^{-1}\Sigma^*(H^*)^{-1}$, where

$$\Sigma^* \equiv \text{LRVar}[\varphi_t(\theta^*)] := \sum_{j=-\infty}^{\infty} \text{Cov}\left[\varphi_t(\theta^*), \varphi_{t+j}(\theta^*)\right]$$

is the long-run variance of the process $\{\varphi_t(\theta^*)\}$. No additional martingale difference assumption is necessary to derive this asymptotic distribution (see APR page 23). The nonparametric estimation of the Jacobian and weighting matrices does have an effect asymptotically—by means of the second and third terms in the right-hand side (r.h.s.) of Equation (12)—but this does not prevent a root-T asymptotically Gaussian distribution for the local GMM estimator. The reason is that the kernel factors in the estimation criterion are averaged across the sample, thus recovering a parametric convergence rate.[3]

2. Under regularity conditions, the GMM estimator $\tilde\theta$ implementing the optimal HJ instruments is asymptotically equivalent to the local GMM, that is, $\tilde\theta = \hat\theta + o_p(1/\sqrt{T})$ (see Appendix A.1 for a sketch of the proof).
3. APR in Section 7.5 provide the asymptotic distribution of the SMD estimator when the weighting matrix $\Omega(x_t)$ is assumed known and has not to be estimated. The asymptotic expansion is

$$\sqrt{T}\big(\hat\theta(h) - \theta^*(h)\big) = -H^*(h)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varphi_{t+1}\big(\theta^*(h), h\big) + o_p(1), \quad (14)$$

where with another slight abuse of notation, we have

$$H^*(h) := \frac{1}{2}\frac{\partial^2 Q_\infty\big(\theta^*(h), h\big)}{\partial\theta\partial\theta'}$$

3 See in particular Lemma 6 in GR, which uses similar arguments as in Kitamura, Tripathi, and Ahn (2004, p. 1696–1698) and Gospodinov and Otsu (2012, p. 487). The zero conditional mean of the moment function is restored by recentering $\Psi_{t+1}(\theta^*) - e(x_t, \theta^*)$, and similarly for the gradient term and for the term arising from the estimation of the weighting matrix.

and

$$\varphi_{t+1}(\theta^*(h),h) := A(x_t,\theta^*(h),h)'\Omega^{-1/2}(x_t)\Psi_{t+1}(\theta^*(h)) + \left(\frac{\partial\Psi_{t+1}(\theta^*(h))}{\partial\theta'}\right)'\Omega^{-1/2}(x_t)C(x_t,\theta^*(h),h), \tag{15}$$

for

$$A(x,\theta,h) := \frac{1}{h^m}E\left[\Omega^{-1/2}(x_s)\frac{\partial\Psi_{s+1}(\theta)}{\partial\theta'}K\left(\frac{x_s-x}{h}\right)\right],$$

and vector $C(x,\theta,h)$ is defined in Equation (6). Then, the asymptotic variance of estimator $\hat{\theta}(h)$ is $\mathrm{AsVar}[\hat{\theta}(h)] = H^*(h)^{-1}\Sigma^*(h)H^*(h)^{-1}$, where $\Sigma^*(h)$ is the long-run variance of process $\{\varphi_t(\theta^*(h),h)\}$. APR obtain the asymptotic expansion in Equation (14) by an elegant argument based on the theory of U-statistics. In Appendix A.2, we show that we can derive Equation (14) also by deploying the interpretation of $\hat{\theta}(h)$ as GMM estimator for a continuum of unconditional moments minimizing the sample analog of the criterion in Equation (8).

Let us now compare the asymptotic variances of the local GMM estimator (or equivalently, of the GMM estimator with optimal HJ instruments) and that of the SMD estimator for a small bandwidth parameter. This is interesting since, as found in APR and confirmed in our numerical experiments in Section 3, there are empirically relevant frameworks in which some of the components of vector $\theta^*(h)$ vary slowly over relatively small values of the parameter h. We focus on the case in which the weighting matrix is not estimated and density weighting is used, since APR provide the asymptotic distribution of the SMD estimator in this case. Then, the third term in the r.h.s. of Equation (12) has to be replaced by its conditional expectation, and the score of the local GMM estimator becomes

$$\varphi_{t+1}(\theta^*) = 1(x_t)J(x_t,\theta^*)'\Omega^{-1}(x_t)\Psi_{t+1}(\theta^*) + 1(x_t)\left(\frac{\partial\Psi_{t+1}(\theta^*)}{\partial\theta'}\right)'\Omega^{-1}(x_t)e(x_t,\theta^*) - 1(x_t)J(x_t,\theta^*)'\Omega^{-1}(x_t)e(x_t,\theta^*), \tag{16}$$

with $1(x_t) = f(x_t)$. We do not find that $\varphi_{t+1}(\theta^*)$ is the limit of $\varphi_{t+1}(\theta^*(h),h)$ as $h\to 0$ because of the third term in the r.h.s. of Equation (16). Therefore, in general the asymptotic variances of the local GMM estimator and SMD estimator differ, even when we select a small bandwidth parameter for the latter estimator.

To investigate further the asymptotic variances of the estimators, we deploy the next assumption.

Assumption S: *The $n\times(p+1)$-dimensional vector $Z_{t+1}(\theta) := \mathrm{vec}[\Psi_{t+1}(\theta) \;:\; \frac{\partial\Psi_{t+1}(\theta)}{\partial\theta'}]$ is such that*

$$E[Z_{t+1}(\theta)|I(t),s^{t+j}] = E[Z_{t+1}(\theta)|s_t], \tag{17}$$

for any parameter value $\theta\in\Theta$ and any integer $j\geq 0$, where the information $I(t) = \sigma\{R^t,Y^t,s^t\}$ includes the history up to time t of the gross returns R_t, the risk factors Y_t in the SDF, and a (possibly latent) state variables vector s_t, and $s^t := \{s_t,s_{t-1},\ldots\}$. The observed variable x_t is such that $x_t = \pi(s_t)$ for a continuous function π.

APR consider a similar assumption (see their Assumptions A1, A2, and A3 in Section 6.2) and their arguments show that Equation (17) holds if process $\{s_t\}$ is strictly exogenous, Markov, and such that (R_{t+1}, Y_{t+1}) is independent of s_{t+1} conditional on s^t. Assumption S is more general than Assumptions A1, A2, A3 in APR because it accommodates for the fact that the conditioning variable x_t used by the econometrician can be a sub-component of the state vector s_t. Therefore, Assumption S allows us to study the effect of reducing the conditioning information when measuring the amount of misspecification.

Under Assumption S, for the local GMM estimator we can write the score in Equation (16) as

$$\varphi_{t+1}(\theta^*) = \xi_{t+1} + \delta(s_t) + g(x_t), \tag{18}$$

where

$$\begin{aligned}
\xi_{t+1} &:= \varphi_{t+1}(\theta^*) - E\left[\varphi_{t+1}(\theta^*)|I(t)\right] = 1(x_t)J(x_t,\theta^*)'\boldsymbol{\Omega}^{-1}(x_t)\left[\boldsymbol{\Psi}_{t+1}(\theta^*) - \mu(s_t,\theta^*)\right] \\
&\quad + 1(x_t)\left[\frac{\partial \boldsymbol{\Psi}_{t+1}(\theta^*)}{\partial\theta'} - G(s_t,\theta^*)\right]'\boldsymbol{\Omega}^{-1}(x_t)e(x_t,\theta^*), \\
\delta(s_t) &:= E\left[\varphi_{t+1}(\theta^*)|I(t)\right] - E\left[\varphi_{t+1}(\theta^*)|x_t\right] = 1(x_t)J(x_t,\theta^*)'\boldsymbol{\Omega}^{-1}(x_t)[\mu(s_t,\theta^*) - e(x_t,\theta^*)] \\
&\quad + 1(x_t)\left[G(s_t,\theta^*) - J(x_t,\theta^*)\right]'\boldsymbol{\Omega}^{-1}(x_t)e(x_t,\theta^*), \\
g(x_t) &:= E\left[\varphi_{t+1}(\theta^*)|x_t\right] = 1(x_t)J(x_t,\theta^*)'\boldsymbol{\Omega}^{-1}(x_t)e(x_t,\theta^*),
\end{aligned}$$

with $\mu(s_t,\theta) := E[\boldsymbol{\Psi}_{t+1}(\theta)|s_t]$ and $G(s_t,\theta) := E\left[\frac{\partial\boldsymbol{\Psi}_{t+1}(\theta)}{\partial\theta'}\Big|s_t\right]$, for any $\theta\in\boldsymbol{\Theta}$. Equation (18) yields a decomposition of the score in three mutually orthogonal components: the innovation ξ_{t+1} w.r.t. the information $I(t)$, the difference $\delta(s_t)$ between the score's conditional expectation given $I(t)$ and its projection on x_t and the score's conditional expectation given x_t denoted $g(x_t)$.

In a similar vein, for the score of the SMD estimator in Equation (15) we have

$$\varphi_{t+1}\left(\theta^*(h),h\right) = \xi_{t+1}(h) + \delta(s_t,h) + 2g(x_t,h), \tag{19}$$

where

$$\begin{aligned}
\xi_{t+1}(h) &:= A\left(x_t,\theta^*(h),h\right)'\boldsymbol{\Omega}^{-1/2}(x_t)\left[\boldsymbol{\Psi}_{t+1}(\theta^*(h)) - \mu\left(s_t,\theta^*(h)\right)\right] \\
&\quad + \left[\frac{\partial\boldsymbol{\Psi}_{t+1}(\theta^*(h))}{\partial\theta'} - G\left(s_t,\theta^*(h)\right)\right]'\boldsymbol{\Omega}^{-1/2}(x_t)C\left(x_t,\theta^*(h),h\right), \\
\delta(s_t,h) &:= A\left(x_t,\theta^*(h),h\right)'\boldsymbol{\Omega}^{-1/2}(x_t)\left[\mu\left(s_t,\theta^*(h)\right) - e\left(x_t,\theta^*(h)\right)\right] \\
&\quad + \left[G\left(s_t,\theta^*(h)\right) - J\left(x_t,\theta^*(h)\right)\right]'\boldsymbol{\Omega}^{-1/2}(x_t)C\left(x_t,\theta^*(h),h\right), \\
g(x_t,h) &:= \frac{1}{2}\left[A\left(x_t,\theta^*(h),h\right)'\boldsymbol{\Omega}^{-1/2}(x_t)e\left(x,\theta^*(h)\right) + J\left(x_t,\theta^*(h)\right)'\boldsymbol{\Omega}^{-1/2}(x_t)C\left(x_t,\theta^*(h),h\right)\right].
\end{aligned}$$

Under regularity conditions, as $h\to 0$, the processes $\{g(x_t,h)\}$, $\{\delta(s_t,h)\}$, and $\{\xi_{t+1}(h)\}$ tend to the processes $\{g(x_t)\}$, $\{\delta(s_t)\}$, and $\{\xi_{t+1}\}$. Hence, in that limit the difference in the asymptotic expansions of the local GMM and SMD estimators is the factor 2 in front of term $g(x_t,h)$ appearing in the score $\varphi_{t+1}\left(\theta^*(h),h\right)$ of the SMD estimator.

Under Assumption S, process $\{\xi_{t+1}\}$ is not autocorrelated, and it is uncorrelated with processes $\{\delta(s_t)\}$ and $\{g(x_t)\}$ at all leads and lags. Similar properties hold for the components of the score of the SMD estimator. Thus, from Equations (18) and (19), the asymptotic variances of the scores for the local GMM and SMD estimators are

$$\Sigma^* = \text{Var}[\xi_{t+1}] + \text{LRVar}[\delta(s_t) + g(x_t)], \tag{20}$$

$$\Sigma^*(h) = \text{Var}[\xi_{t+1}(h)] + \text{LRVar}[\delta(s_t, h) + 2g(x_t, h)], \tag{21}$$

respectively. In the limit $h \to 0$, we have $\text{AsVar}[\hat{\theta}] \leq \text{AsVar}[\hat{\theta}(h)]$ if, and only if, $\Sigma^* \leq \Sigma^*(h)$, where $\leq$ is the standard ordering for symmetric matrices. This condition holds if, and only if,

$$\sum_{j=-\infty, j\neq 0}^{\infty} \left(\text{Cov}[\delta(s_t), g(x_{t+j})] + \text{Cov}[g(x_t), \delta(s_{t+j})] \right) + 3\text{LRVar}[g(x_t)] \geq 0_{p\times p}. \tag{22}$$

Loosely speaking, the latter inequality is satisfied if the prediction error $\delta(s_t)$ is not too large, or is not too negatively correlated with leads and lags of the prediction $g(x_t)$ itself. In particular, this holds true when $x_t = s_t$. A more detailed investigation of the inequality (22) would be worthwhile.

3 A Numerical and Empirical Study with Conditionally Linear SDFs

We investigate the patterns of the pseudo-true SDF parameters θ^* and $\theta^*(h)$, and the asymptotic variances of estimators $\hat{\theta}$ and $\hat{\theta}(h)$, in a conditionally linear factor model with omitted factors. We rely on numerical experiments and Monte-Carlo simulations calibrated on a real dataset which we also use for estimation. This data set consists of the monthly cum-dividend returns of the $n = 6$ size- and value-based Fama–French (FF-) research portfolios of U.S. publicly traded equities deflated by the Consumer Price Index for All Urban Consumers rate from July 1963 to October 2012, which yields $T = 592$ monthly observations. The five factors mimick the market, size, value, investment, and profitability factors, that we denote as $F_t = \left(F_{\text{mkt},t}, F_{\text{smb},t}, F_{\text{hml},t}, F_{\text{rmw},t}, F_{\text{cma},t}\right)'$. We consider $m = 2$ variables in the state vector $x_t = (cay_t, YC_t)'$, and we identify them with demeaned consumption to wealth ratio cay_t (Lettau and Ludvigson, 2001) and demeaned labor income to consumption ratio YC_t (Santos and Veronesi, 2006).

3.1 Numerical Experiments and Monte-Carlo Simulations

In our numerical experiments and Monte-Carlo simulations, the assumed Data Generating Process (DGP) for the assets returns and the SDF is

$$\begin{aligned} R_{t+1} &= R_{f,t} 1_n + \tilde{\mu} + B(F_{t+1} - Cx_t) + \varepsilon_{t+1}, \\ F_{t+1} &= Cx_t + u_{t+1}, \\ x_t &= Rx_{t-1} + v_t, \\ m_{t+1} &= \nu_0 + \delta' x_t + \nu' F_{t+1} + x_t' D F_{t+1}, \end{aligned} \tag{23}$$

for scalar $\nu_0 > 0$. The risk-free gross return $R_{f,t}$ is such that $R_{f,t}^{-1} = E[m_{t+1}|x_t] = \nu_0 + \delta' x_t + \nu' Cx_t + x_t' DCx_t$. The processes $\{\varepsilon_t\}$, $\{u_t\}$, and $\{v_t\}$ are mutually independent Gaussian white noise processes with covariance matrices Σ_ε, Σ_u, and Σ_v. The conditional mean of the n_F-dimensional factor F_t is driven by an exogenous Markov state vector x_t with Gaussian autoregressive dynamic. The state is exogenous and strictly stationary if the eigenvalues of matrix R are inside the unit circle. The SDF m_{t+1} is linear in the factors conditionally on the state vector, for example as in the SDF specifications considered by Nagel and Singleton (2011). Assumption S is met in this model with $s_t \equiv x_t$.

Let $I(t) = \sigma\{R^t, F^t, x^t\}$ be the conditioning information. In Appendix A.3.1, we show that the conditional restriction $E[m_{t+1}R_{t+1} - 1_n|I(t)] = 0_n$ is met if, and only if, the vector of excess returns $\tilde{\mu}$ in Equations (23) is such that

$$\tilde{\mu} = -\frac{1}{\nu_0}B\Sigma_u\nu, \tag{24}$$

and the SDF parameters are such that either $C'\nu = 0$ and $D = \frac{1}{\nu_0}\delta\nu'$, or

$$\delta = -C'\nu \qquad \text{and} \qquad \mathrm{D} = 0_{\mathrm{m}\times\mathrm{n_F}}. \tag{25}$$

We focus on the case described by Equations (24) and (25), as it simplifies the numerical computation of the pseudo-true parameter values, which are given in closed form (see below). In this case, the risk-free gross return is constant $R_f = \nu_0^{-1}$, and the test assets' expected gross returns vector $\mu = \nu_0^{-1}1_n + \tilde{\mu}$ and their second moments matrix $\Omega = \mu\mu' + B\Sigma_u B' + \Sigma_\varepsilon$ are time-invariant as well. We calibrate the DGP parameters ν_0, ν, B, C, R, Σ_ε, Σ_u, and Σ_ν to mimic the monthly dataset described at the beginning of the section.

To define the misspecification, let us suppose that the econometrician overlooks the importance of the profitability factor mimicked by $F_{\mathrm{cma},t}$. Moreover, she considers the possibility that the risk premium of the market factor is driven by the state variables. Specifically, she considers the following three misspecified parametric SDF models:

$$\mathcal{M}_1 : m_{t+1}(\theta) = \theta_0 + \overline{\theta}' F_{1,t+1}, \tag{26}$$

$$\mathcal{M}_2 : m_{t+1}(\theta) = \theta_0 + \overline{\theta}' F_{1,t+1} + \theta_5 cay_t F_{3,t+1}, \tag{27}$$

$$\mathcal{M}_3 : m_{t+1}(\theta) = \theta_0 + \overline{\theta}' F_{1,t+1} + (\theta_5 cay_t + \theta_6 YC_t) F_{3,t+1}, \tag{28}$$

where $F_{1,t} = \left(F_{\mathrm{mkt},t}, F_{\mathrm{smb},t}, F_{\mathrm{hml},t}, F_{\mathrm{rmw},t}\right)'$ and $F_{3,t} = F_{\mathrm{mkt},t}$, the omitted factor is $F_{2,t} = F_{\mathrm{cma},t}$, and $\overline{\theta} = (\theta_1, \ldots, \theta_4)'$.

In Appendix A.3.2, we derive the pseudo-true parameter vector θ^* computed with uniform weight $1(x_t) = 1$ in the three models. The one for model $\mathcal{M}_3$ is

$$\theta^* = \eta_1 + \begin{bmatrix} \tilde{\Gamma}_1'\Omega^{-1}\tilde{\Gamma}_1 + \kappa\tilde{C}_1\Sigma_x\tilde{C}'_1 & \left(\tilde{\Gamma}_1'\Omega^{-1}\mu\right)C_3\Sigma_x + \tilde{C}_1\Sigma_x\left(\mu'\Omega^{-1}\Gamma_3\right) \\ \Sigma_x C_3'\left(\mu'\Omega^{-1}\tilde{\Gamma}_1\right) + \left(\Gamma_3'\Omega^{-1}\mu\right)\Sigma_x\tilde{C}_1' & \Sigma_x\left(\Gamma_3'\Omega^{-1}\Gamma_3\right) + \kappa\left(2\Sigma_x C_3' C_3\Sigma_x + C_3\Sigma_x C_3'\Sigma_x\right) \end{bmatrix}^{-1}$$
$$\times \begin{bmatrix} \tilde{\Gamma}_1'\Omega^{-1}\Gamma_2 + \kappa\tilde{C}_1\Sigma_x C_2' & \kappa\tilde{C}_1\Sigma_x \\ (\Gamma_3\Omega^{-1}\mu)\Sigma_x C_2' + \Sigma_x C_3'\left(\mu'\Omega^{-1}\Gamma_2\right) & \left(\Gamma_3'\Omega^{-1}\mu\right)\Sigma_x \end{bmatrix} \eta_2, \tag{29}$$

where $\eta_1 = \left(\tilde{\nu}_1', 0_m'\right)'$ is the true parameter vector augmented by zeros, $\eta_2 = \left(\nu_2', \delta'\right)'$, with $\tilde{\nu}_1 = \left(\nu_0, \nu_1'\right)'$, ν_1 and ν_2 are the subvectors in ν corresponding to sub-components $F_{1,t}$ and $F_{2,t}$ of F_t, $\tilde{C}_1 = \left[0_m \ : \ C_1'\right]'$, $\tilde{\Gamma}_1 = [\mu \ : \ \Gamma_1]$, and $\kappa = \mu'\Omega^{-1}\mu$. The discrepancy between the true SDF parameters in η_1 and the pseudo-true SDF parameters in θ^* involves the matrices C_i and Γ_i, for i=1, 2, 3, which are the blocks of matrices C and $\Gamma = B\Sigma_u$ corresponding to factors $F_{i,t}$, and the matrix Σ_x, which is the stationary variance of the state variables vector.

If density weighting is applied, we should consider $1(x_t) = f(x_t)$, and in this Gaussian framework the same formulas hold but with Σ_x replaced by $\Sigma_x/2$.

To compute the pseudo-true parameter vector with fixed bandwidth, we choose the m-dimensional standard Gaussian pdf as kernel function K and $h\Sigma_x^{1/2}$ as bandwidth matrix. The kernel factor is then $K((h\Sigma_x^{1/2})^{-1}(x_t - x_s))$ with scalar $h > 0$. Then we have (see Appendix A.3.4 for the proof):

$$\theta^*(h) = \eta_1 + \begin{bmatrix} \tilde{\Gamma}_1'\Omega^{-1}\tilde{\Gamma}_1 + \alpha(h)\kappa\tilde{C}_1\Sigma_x\tilde{C}_1' & (1-\alpha(h))\left(\tilde{\Gamma}_1'\Omega^{-1}\mu\right)C_3\Sigma_x + \alpha(h)(\mu'\Omega^{-1}\Gamma_3)\tilde{C}_1\Sigma_x \\ \left\{\begin{matrix} (1-\alpha(h))\Sigma_x C_3'\left(\mu'\Omega^{-1}\tilde{\Gamma}_1'\right) \\ +\alpha(h)\left(\Gamma_3'\Omega^{-1}\mu\right)\Sigma_x\tilde{C}_1' \end{matrix}\right\} & \left\{\begin{matrix} \alpha(h)\left(\Gamma_3'\Omega^{-1}\Gamma_3\right)\Sigma_x + \kappa\left(\left(1-2\alpha(h)+2\alpha(h)^2\right)\right. \\ \left. \times\Sigma_x C_3' C_3\Sigma_x + \alpha(h)^2(C_3\Sigma_x C_3')\Sigma_x\right) \end{matrix}\right\} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \tilde{\Gamma}_1'\Omega^{-1}\Gamma_2 + \alpha(h)(\mu'\Omega^{-1}\mu)\tilde{C}_1\Sigma_x C_2' & \alpha(h)(\mu'\Omega^{-1}\mu)\tilde{C}_1\Sigma_x \\ \alpha(h)\left(\Gamma_3'\Omega^{-1}\mu\right)\Sigma_x C_2' + (1-\alpha(h))\Sigma_x C_3'(\mu'\Omega^{-1}\Gamma_2) & \alpha(h)\left(\Gamma_3'\Omega^{-1}\mu\right)\Sigma_x \end{bmatrix}\eta_2,$$

where $\alpha(h) = \frac{1}{2+h^2}$. The magnitude of h affects the terms which involve the variance Σ_x, reflecting the changing relevance of the conditioning information as h varies. Some of these terms disappear when $h \to \infty$ and $\theta^*(h)$ converges to the unconditional pseudo-true parameter vector θ_u^*. For $h \to 0$, vector $\theta^*(h)$ converges to θ^* in Equation (29), with matrix Σ_x replaced by $\Sigma_x/2$ because of density weighting.

We plot in Figures 1 and 2 the patterns of the pseudo-true SDF parameter values for misspecified model $\mathcal{M}_1$ as functions of h, for h in the intervals $[0, 1]$ and $[0, 50]$, respectively. As remarked by APR in their empirical and numerical illustrations, also in our experiments the pseudo-true parameter vector $\theta^*(h)$ is not very sensitive w.r.t. bandwidth h and is close to the conditional pseudo-true parameter vector θ^*, for small values of h.[4] When analyzed over a broader range of values of h, the components $\theta_k^*(h)$ are decreasing functions of h for $k = 1, ..., 4$, that is, the SDF coefficients associated to the four factors become more negative and larger in absolute value as h increases. For large values of h, the pseudo-true value $\theta_0^*(h)$ is above 1, which is not compatible with its interpretation as expected non-defaultable bond price, and implies negative time preferences. For $h \to \infty$ vector $\theta^*(h)$ converges to the unconditional pseudo-true parameter vector θ_u^*. The difference between the conditional and unconditional pseudo-true parameter values is numerically large for the investment factor ($k = 4$). The patterns of the components of $\theta^*(h)$ are similar in the other two misspecified models $\mathcal{M}_2$ and $\mathcal{M}_3$, see Figures 3 and 4. The pseudo-true parameter value $\theta_5^*(h)$, which corresponds to the interaction of *cay* with the market factor and is absent in the DGP, is increasing w.r.t. h. For model $\mathcal{M}_3$, the unconditional pseudo-true parameter θ_u^* is not well-defined, because the unconditional moment restrictions are underidentified ($n = 6 < p = 7$). Hence, the pseudo-true parameter is identifiable from the conditional moment restrictions but not from the unconditional ones.

Let us compute the asymptotic variances of the local GMM estimator $\hat{\theta}$, and of the APR estimator with very small bandwidth [i.e., for $h \to 0$ in Equation (21)] denoted $\hat{\theta}(0)$, for model $\mathcal{M}_1$.[5] From the results in Section 2, the asymptotic variances are

4 Note that the Scott's rule would imply $h = T^{-1/6}$. For $T = 592$ as in our analysis this yields $h \simeq 0.345$.

5 The formulas for models $\mathcal{M}_2$ and $\mathcal{M}_3$ can be derived similarly, but are a bit more cumbersome.

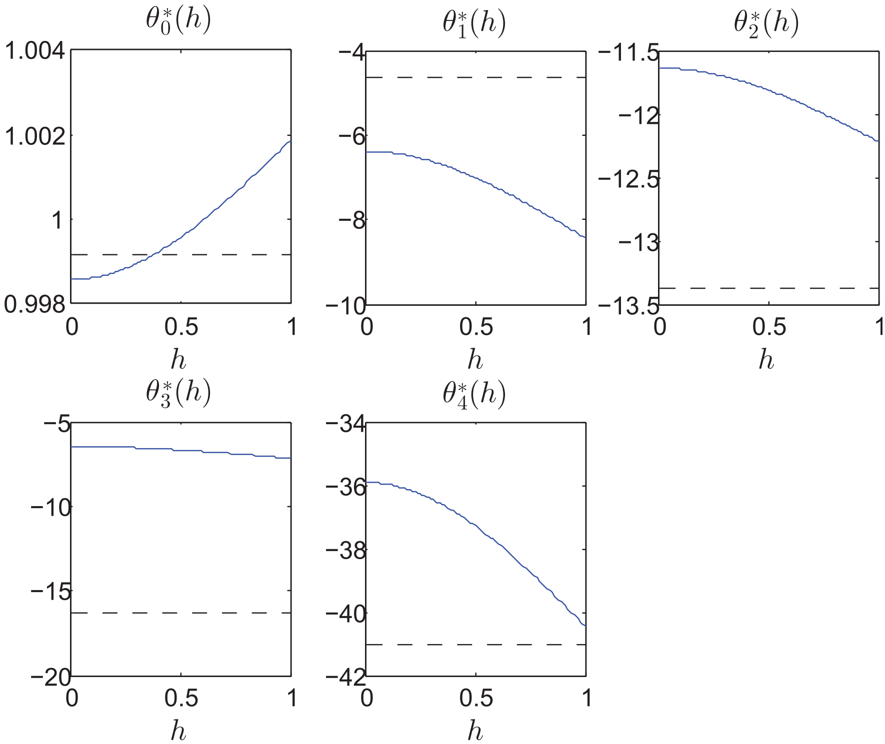

Figure 1 True and pseudo-true SDF parameter values as functions of h for misspecified model $\mathcal{M}_1$. The SDF is $m_{t+1}(\theta) = \theta_0 + \overline{\theta}' F_{1,t+1}$, where $F_{1,t} = (F_{\text{mkt},t}, F_{\text{smb},t}, F_{\text{hml},t}, F_{\text{rmw},t})'$ and $\overline{\theta} = (\theta_1, \ldots, \theta_4)'$. The DGP parameters are calibrated to mimic the inflation- and dividends-adjusted gross returns of the $n = 6$ size- and value-based FF research portfolios of U.S. publicly traded equities, as well as FF factors and the demeaned consumption to wealth ratio and the demeaned labor income to consumption ratio from July 1963 to October 2012. In each panel, the solid curve is the pseudo-true parameter value $\theta_k^*(h)$ as a function of h in the interval $[0, 1]$ for $k = 0, 1, \ldots, 4$. The dashed line represents the corresponding true SDF parameter value.

$\text{AsVar}[\hat{\theta}] = (H^*)^{-1}\{\text{Var}[\xi_{t+1}] + \text{LRVar}[g(x_t)]\}(H^*)^{-1}$ and $\text{AsVar}[\hat{\theta}(0)] = (H^*)^{-1}\{\text{Var}[\xi_{t+1}] + 4\text{LRVar}[g(x_t)]\}(H^*)^{-1}$, since $\delta(s_t) = 0$. It turns out that $\text{Var}[\xi_{t+1}|x_t]$ and $g(x_t)$ are quadratic functions of the Gaussian state vector x_t, which leads to closed-form expressions. We have (see Appendix A.4 for the derivation):

$$\begin{aligned} H^* &= \tilde{\Gamma}_1'\Omega^{-1}\tilde{\Gamma}_1 + \kappa\tilde{C}_1\Sigma_x\tilde{C}_1', \\ \text{LRVar}[g(x_t)] &= A\tilde{\Sigma}_x A' + A\tilde{C}_1\left(R(I_m - R)^{-1}\Sigma_x + \Sigma_x(I_m - R')^{-1}R'\right)\tilde{C}_1'A' \\ &\quad + \kappa^2\left[\tilde{\Sigma}_x\theta^*\theta^{*\prime}\tilde{\Sigma}_x + \tilde{\Sigma}_x\left(\theta^{*\prime}\tilde{\Sigma}_x\theta^*\right)\right] + \kappa^2\tilde{C}_1\left[S_1\Sigma_x + \Sigma_x S_1' + S_2\Sigma_x + \Sigma_x S_2'\right]\tilde{C}_1', \end{aligned}$$

where $A := \mu'\Omega^{-1}(\tilde{\Gamma}_1\theta^* - 1_n)I_p + \tilde{\Gamma}_1'\Omega^{-1}\mu\theta^{*\prime}$ and $\tilde{\Sigma}_x := \tilde{C}_1\Sigma_x\tilde{C}'_1$, the $(m \times m)$-dimensional matrix S_1 is such that $\text{vec}[S_1] = [I_{m^2} - R \otimes R]^{-1}\text{vec}[R\Sigma_x\alpha\alpha' R]$, $S_2 := \sum_{i=1}^m \alpha_i\beta_i\lambda_i$

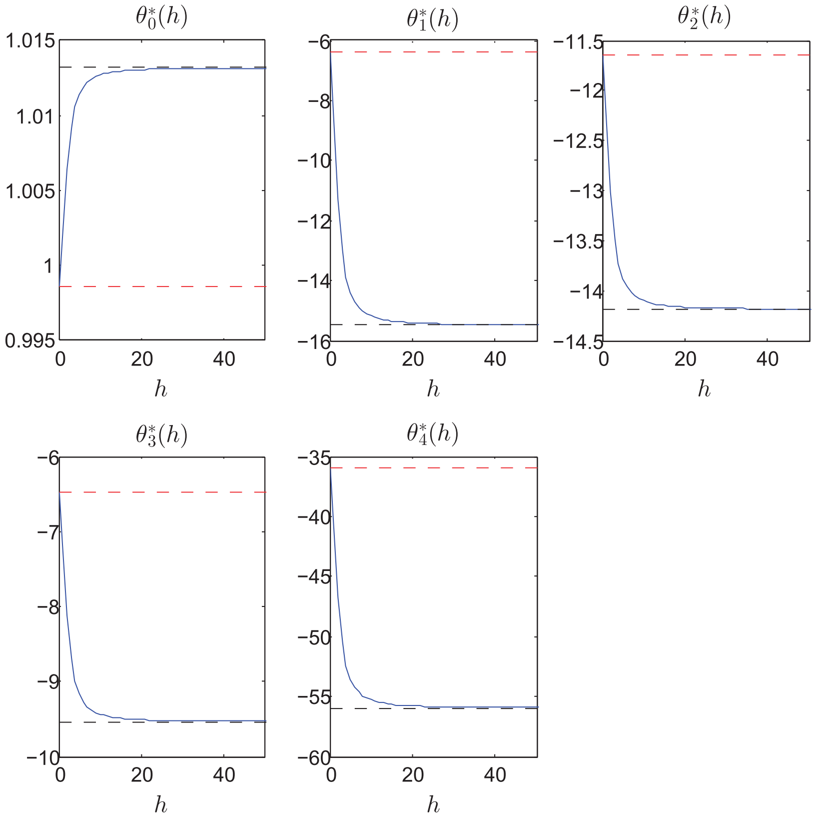

Figure 2 Pseudo-true SDF parameter values and their limits for h tending to zero or to infinity for misspecified model $\mathcal{M}_1$. The SDF specification and the model calibration are as described in the caption of Figure 1. In each panel, the solid curve is the pseudo-true parameter value $\theta_k^*(h)$ as a function of h for $k = 0, 1, \ldots, 4$. The dashed lines represent the conditional and unconditional pseudo-true parameter values θ_k^* and $\theta_{u,k}^*$. In each panel except the first, the highest line indicates the conditional pseudo-true value.

$R(I_m - \lambda_i R)^{-1}$ with $\alpha := U'\tilde{C}_1'\theta^*$ and $\beta := U^{-1}\Sigma_x\tilde{C}_1'\theta^*$, matrix U diagonalizes R, that is, $\Lambda = U^{-1}RU$ is the diagonal matrix of the eigenvalues λ_i, and:

$$\begin{aligned}
\mathrm{Var}[\xi_{t+1}] = & a_0'\left(I_p \otimes \Omega^{-1}\right)V_0\left(I_p \otimes \Omega^{-1}\right)a_0 + (\theta^{*\prime}\tilde{\Sigma}_x\theta^*)\tilde{\Gamma}_1'\Omega^{-1}V_R\Omega^{-1}\tilde{\Gamma}_1 \\
& + \tilde{\Sigma}_x(\tilde{\Gamma}_1\theta^* - 1_n)'\Omega^{-1}V_R\Omega^{-1}(\tilde{\Gamma}_1\theta^* - 1_n) + 8\kappa(1-\kappa)\tilde{\Sigma}_x\theta^*\theta^{*\prime}\tilde{\Sigma}_x \\
& + 4\kappa(1-\kappa)\tilde{\Sigma}_x(\theta^{*\prime}\tilde{\Sigma}_x\theta^*) \\
& + \Big\{ (\theta^{*\prime}\tilde{\Sigma}_x) \otimes \tilde{\Gamma}_1'\Omega^{-1}V_R\Omega^{-1}(\tilde{\Gamma}_1\theta^* - 1_n) \\
& + 2\left(\theta^{*\prime} \otimes \tilde{\Gamma}_1'\Omega^{-1}\right)W\Omega^{-1}\mu(\theta^{*\prime}\tilde{\Sigma}_x) + 2\left(I_p \otimes (\tilde{\Gamma}_1\theta^* - 1_n)'\Omega^{-1}\right)W\Omega^{-1}\mu(\theta^{*\prime}\tilde{\Sigma}_x) \\
& + (\theta^{*\prime}\tilde{\Sigma}_x) \otimes \left[\tilde{\Gamma}_1'\Omega^{-1}W'(\theta^* \otimes \Omega^{-1}\mu)\right] + \tilde{\Sigma}_x\left[(\tilde{\Gamma}_1\theta^* - 1_n)'\Omega^{-1}W'(\theta^* \otimes \Omega^{-1}\mu)\right] \\
& + (\theta^{*\prime}\tilde{\Sigma}_x\theta^*)\left[\tilde{\Gamma}_1'\Omega^{-1}W'\left(I_p \otimes \Omega^{-1}\mu\right)\right] + (\tilde{\Sigma}_x\theta^*) \otimes \left[(\tilde{\Gamma}_1\theta^* - 1_n)'\Omega^{-1}W'\left(I_p \otimes \Omega^{-1}\mu\right)\right]\Big\}_s,
\end{aligned}$$

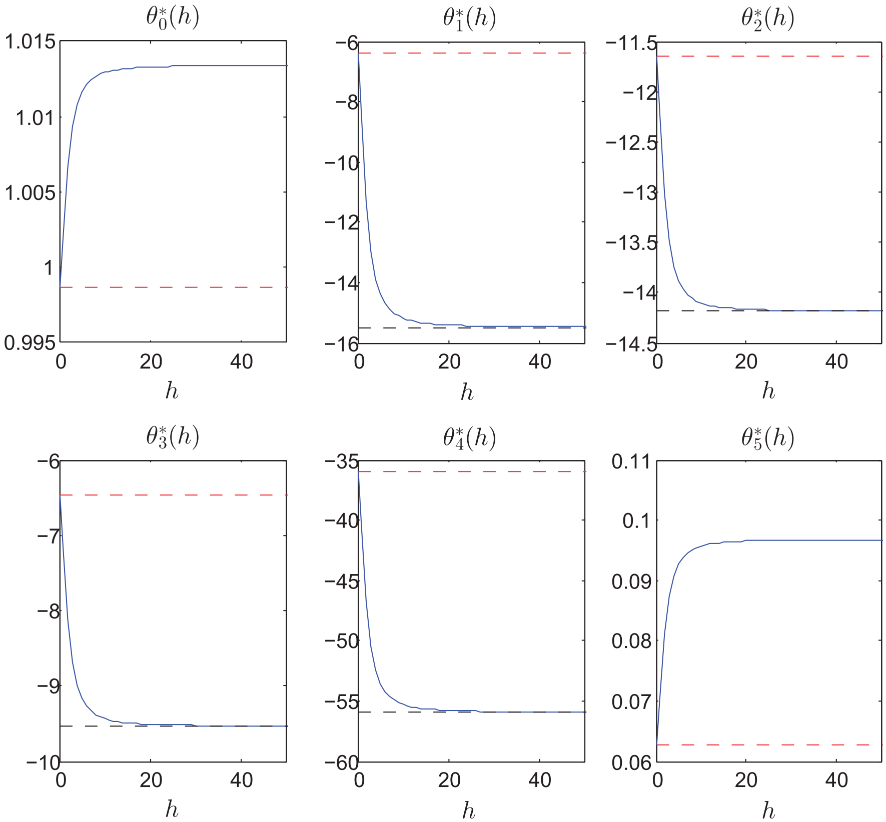

Figure 3 Pseudo-true SDF parameter values and their limits for h tending to zero or to infinity for misspecified model $\mathcal{M}_2$. The SDF is $m_{t+1}(\theta) = \theta_0 + \overline{\theta}' F_{1,t+1} + \theta_5 x_{1,t} F_{3,t+1}$, where $F_{1,t} = (F_{mkt,t}, F_{smb,t}, F_{hml,t}, F_{rmw,t})'$, $F_{3,t} = F_{mkt,t}$, $x_{1,t} = cay_t$, $\overline{\theta} = (\theta_1, \ldots, \theta_4)'$. The DGP parameters are calibrated to mimic the inflation- and dividends-adjusted gross returns of the $n=6$ size- and value-based FF research portfolios of U.S. publicly traded equities, as well as FF factors and the demeaned consumption to wealth ratio and the demeaned labor income to consumption ratio from July 1963 to October 2012. In each panel, the solid curve is the pseudo-true parameter value $\theta_k^*(h)$ as a function of h for $k = 0, 1, \ldots, 5$. The dashed lines represent the conditional and unconditional pseudo-true parameter values θ_k^* and $\theta_{u,k}^*$. In each panel except the first and the last, the highest line indicates the conditional pseudo-true value. In the first and last panel, the line corresponding to the unconditional pseudo-true parameter value is not displayed, because the near singularity of a matrix to be inverted implies numerical instabilities, which affect the computation of $\theta_{u,0}^*$ and $\theta_{u,5}^*$.

where $a_0 := \theta^* \otimes \tilde{\Gamma}_1 + I_p \otimes (\tilde{\Gamma}_1 \theta^* - 1_n)$, $V_0 = \begin{bmatrix} V_R & \Gamma_1 \otimes \mu' \\ \Gamma_1' \otimes \mu & \Sigma_{u,11} \otimes \Omega + K_{p-1,n}(\Gamma_1 \otimes \Gamma_1') \end{bmatrix}$, $V_R := \Omega - \mu\mu'$, $K_{p-1,n}$ is the commutator matrix such that $K_{p-1,n}(\Gamma_1 \otimes \Gamma_1') = (\Gamma_1' \otimes \Gamma_1) K_{n,p-1}$ (see e.g., Magnus and Neudecker, 2007), $\Sigma_{u,11}$ denotes the block of matrix Σ_u corresponding to the shocks on $F_{1,t}$, $W = \begin{bmatrix} V_R \\ (I_{p-1} \otimes \mu)\Gamma_1' \end{bmatrix}$, and $\{A\}_s := A + A'$.

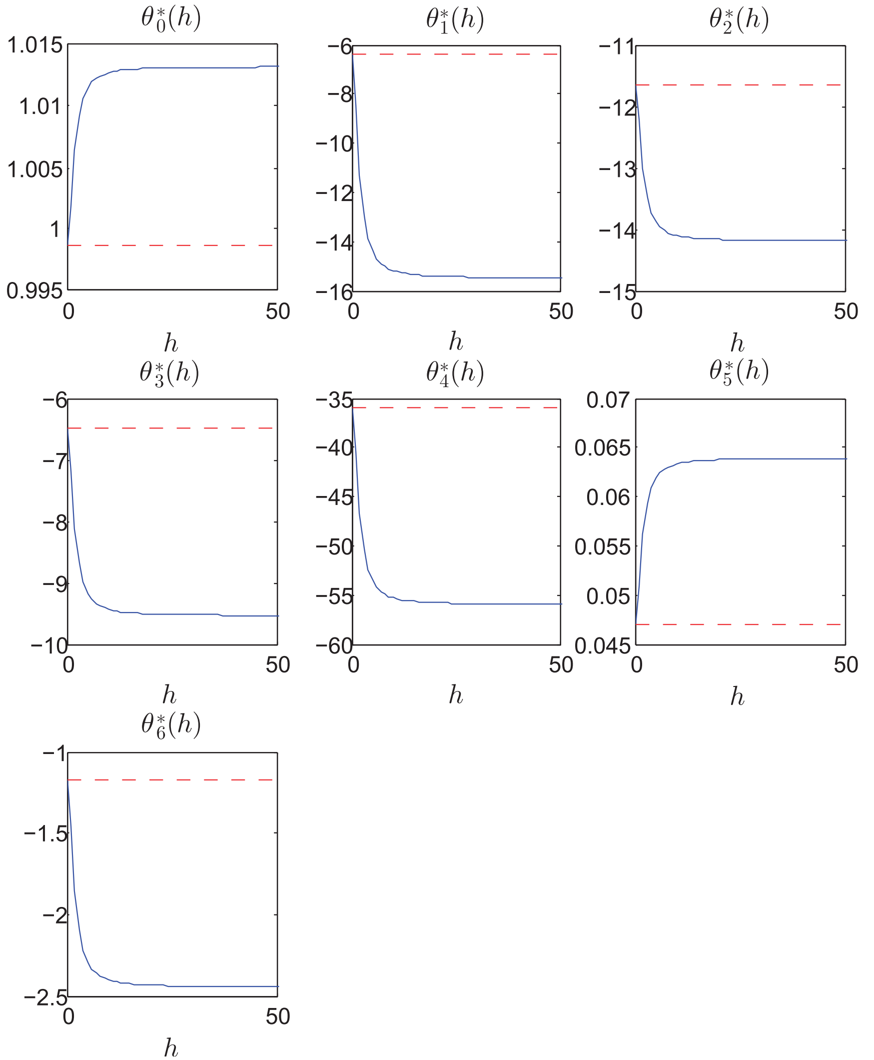

Figure 4 Pseudo-true SDF parameter values and their limits for h tending to zero or to infinity for misspecified model $\mathcal{M}_3$. The SDF is $m_{t+1}(\theta) = \theta_0 + \overline{\theta}' F_{1,t+1} + \theta_5 x_{1,t} F_{3,t+1} + \theta_6 x_{2,t} F_{3,t+1}$, where $F_{1,t} = (F_{\mathrm{mkt},t}, F_{\mathrm{smb},t}, F_{\mathrm{hml},t}, F_{\mathrm{rmw},t})'$, $F_{3,t} = F_{\mathrm{mkt},t}$, $x_{1,t} = cay_t$, $x_{2,t} = YC_t$, $\overline{\theta} = (\theta_1, \ldots, \theta_4)'$. The DGP parameters are calibrated to mimic the inflation- and dividends-adjusted gross returns of the $n = 6$ size- and value-based FF research portfolios of U.S. publicly traded equities, as well as FF factors and the demeaned consumption to wealth ratio and the demeaned labor income to consumption ratio from July 1963 to October 2012. In each panel, the solid curve is the pseudo-true parameter value $\theta_k^*(h)$ as a function of h for $k = 0, 1, \ldots, 6$. The dashed lines represent the conditional pseudo-true parameter values θ_k^*. The unconditional pseudo-true parameter vector θ_u^* is not well-defined for model $\mathcal{M}_3$ because the model parameters are unidentifiable by the unconditional moment restrictions.

Under the calibrated DGP, we get:

$$\frac{1}{T}(H^*)^{-1}\text{LRVar}[g(x_t)](H^*)^{-1} = \begin{bmatrix} 0.001 & -0.04 & -0.01 & -0.01 & -0.09 \\ & 26.04 & 7.22 & 8.72 & 57.05 \\ & & 2.00 & 2.42 & 15.83 \\ & & & 2.92 & 19.11 \\ & & & & 124.98 \end{bmatrix} \tag{30}$$

and

$$\frac{1}{T}(H^*)^{-1}\text{Var}[\xi_{t+1}](H^*)^{-1} = \begin{bmatrix} 0.001 & -0.16 & -0.06 & -0.06 & -0.37 \\ & 106.39 & 25.63 & 36.93 & 213.95 \\ & & 12.83 & 8.96 & 63.75 \\ & & & 16.31 & 73.33 \\ & & & & 470.78 \end{bmatrix}, \tag{31}$$

with $T = 592$. We get 90% asymptotic confidence intervals for the parameters θ_k^* for $k = 0, ..., 4$ estimated with local GMM:

$$[0.92 : 1.08], [-25.26 : 12.49], [-17.96 : -5.32], [-13.66 : 0.72], [-75.94 : 4.12]. \tag{32}$$

The 90% asymptotic confidence intervals for the APR estimator with very small bandwidth are

$$[0.87 : 1.13], [-30.18 : 17.41], [-19.13 : -4.15], [-15.15 : 2.21], [-87.01 : 15.18]. \tag{33}$$

The asymptotic confidence intervals for sample size $T = 592$ are rather large, especially for $k = 4$ (investment factor). The confidence intervals do not contain zero for $k = 0$ and $k = 2$ (the size factor) for both estimators. The diagonal elements of the asymptotic variance part originated by the long-run variance of process $\{g(x_t)\}$ in Equation (30) are smaller than the diagonal elements for the component associated with the variance of $\{\xi_{t+1}\}$ in Equation (31). This explains why the confidence intervals for the two estimators are similar, with the ones for the local GMM estimator being narrower. Note that the limited accuracy of these GMM estimators has to be understood in a setting that differs from the standard one of weak instruments, because of the conditional nature of the moment restrictions and the misspecification.

Let us now consider the finite-sample distributions of the estimators. We report in Table 1 the median, 5-percentile, and 95-percentile of the estimates $\hat{\theta}$, $\hat{\theta}(h)$ for $h = 0.1, 0.4, 0.7, 1, 10, 50$, and $\hat{\theta}_u$ obtained in 10^5 Monte-Carlo repetitions with sample size $T = 592$ for the three model misspecifications. For components $k = 1, ..., 4$ the median values of $\hat{\theta}_k$ and $\hat{\theta}_k(h)$ are generally close to the pseudo-true parameters computed in population, across the three misspecified models (Figures 2–4). The difference is more pronounced for the coefficient of the investment factor ($k = 4$) when h is large. The finite-sample 5–95% interquantile ranges for estimators $\hat{\theta}_k$ and $\hat{\theta}_k(0.1)$ in model $\mathcal{M}_1$ (first and second lines in the upper panel of Table 1) are comparable with the 90% asymptotic confidence intervals in (32) and (33), respectively. The 5–95% interquantile ranges get wider when h increases. This range is large for $k = 4$ when large values of h are considered for the APR estimator, and for the unconditional estimator $\hat{\theta}_{u,4}$. The patterns of the medians of the estimators for parameters θ_5 and θ_6 associated with the interactions of the market factor with