Journal of Financial Econometrics, 2018, 1–59
doi: 10.1093/jjfinec/nby017

OXFORD

Pseudo-True SDFs in Conditional Asset Pricing Models*

Bertille Antoine[1], Kevin Proulx[2] and Eric Renault[2]
[1]Simon Fraser University and [2]Brown University

Address correspondence to Eric Renault, Department of Economics, Brown University, 64 Waterman Street, Providence, RI 02912, USA, or e-mail: eric_renault@brown.edu.

Abstract

This article is motivated by the need to bridge some gap between modern asset pricing theory and recent developments in econometric methodology. While asset pricing theory enhances the use of conditional pricing models, econometric inference of conditional models can be challenging due to misspecification or weak identification. To tackle the case of misspecification, we utilize the conditional Hansen and Jagannathan (1997) (HJ) distance as studied by Gagliardini and Ronchetti (2016), but we set the focus on interpretation and estimation of the pseudo-true value defined as the argument of the minimum of this distance. While efficient Generalized Method of Moments (GMM) has no meaning for estimation of a pseudo-true value, the HJ-distance not only delivers a meaningful loss function, but also features an additional advantage for the interpretation and estimation of managed portfolios whose exact pricing characterizes the pseudo-true pricing kernel (stochastic discount factor (SDF)). For conditionally affine pricing kernels, we can display some managed portfolios which are well-defined independently of the pseudo-true value of the parameters, although their exact pricing is achieved by the pseudo-true SDF. For the general case of nonlinear SDFs, we propose a smooth minimum distance (SMD) estimator (Lavergne and Patilea, 2013) that avoids a focus on specific directions as in the case of managed portfolios. Albeit based on kernel smoothing, the SMD approach avoids instabilities and the resulting need of trimming strategies displayed by classical local GMM estimators when the density function of the conditioning variables may take arbitrarily small values. In addition, the fact that SMD may allow fixed bandwidth asymptotics is helpful regarding the curse of dimensionality. In contrast with the true unknown value for a well-specified model, the estimated pseudo-true value, albeit defined in a time-invariant (unconditional) way, may actually depend on the choice of the state variables that define fundamental factors and their scaling weights. Therefore, we may not want to be overly parsimonious about the set of explanatory variables. Finally, following Antoine and Lavergne (2014), we show how SMD can be further robustified to deal with weaker identification contexts. Since SMD can be seen as a local extension of the method of jackknife GMM (Newey and Windmeijer, 2009), we characterize the Gaussian

* We gratefully acknowledge very helpful comments by Lars Peter Hansen and Sydney Ludvigson on an earlier draft of this article. We are also grateful to participants of the 2017 EC[2] conference in Amsterdam for insightful questions and remarks.

asymptotic distribution of the estimator of the pseudo-true value using classical U-statistic theorems.

Key words: misspecification, Hansen-Jagannathan distance, pseudo-true pricing kernel, local factors

JEL classification: C13, C14, G12

1 Introduction

It has been known since Hansen and Richard (1987) that the empirical content of any asset pricing model can be summarized by a stochastic discount factor (SDF). If one were always able to identify a SDF that could price all financial assets, any asset pricing model would be a one factor model. While such a unique accurate factor is elusive, multifactor models are popular because they suggest a limited set of factors that may hopefully span an accurate SDF.

However, as stressed by Ludvigson (2013), all these factor models are "abstractions and therefore by definition misspecified" so that "methods that permit statistical comparisons of the magnitude of misspecification among multiple competing models" are more practically relevant than "hypothesis tests of the null of correct specification against the alternative of incorrect specification."Hansen and Jagannathan (1997) (HJ hereafter) have precisely developed some ways to answer the question: "How large is the misspecification of the stochastic discount factor proxy?" They have also applied this measure of model performance to the cases where the SDF proxy is provided by a linear factor model when model factors can only span a proxy of an accurate SDF.

The initial motivation of this article is a follow up on the above quotations: it is fine to "permit comparisons of the magnitude of misspecification among multiple competing models," but what next? Once we have concluded that one specific model, albeit misspecified, is our preferred one, what do we do with it? Following the logic that the empirical content of an asset pricing model is encapsulated in some given SDF, one should normally associate with their model choice a specific value of the SDF, picked among the set of possible SDFs (in general indexed by a vector $\theta \in \Theta \subset \mathbb{R}^p$ of some free parameters). However, since all models are misspecified, there is no such thing as a true SDF ("true" in the sense that it is able to price accurately all financial assets), so one can only elicit a pseudo-true SDF. In contrast with the true unknown value of some well-identified parameters, the pseudo-true value is not defined in an intrinsic way. Its definition can only be objective-driven; the elicited pseudo-true value θ^* in the set Θ of possible values of the parameters is the one that does the best job in respect to a specific application like pricing, hedging, forecasting, explaining, etc.

Even if the goal is to minimize pricing errors, as in the logic of the measures of specification errors proposed by HJ, these authors rightly note that "comparing the specification error in two different proxies is not possible without taking a stand on the relative importance of the various assets." Our approach leads us to set the focus on a HJ pseudo-true SDF that is characterized by the fact that it is able to price exactly the mimicking portfolios of the factors. We make this statement in a very general setting including not only linear factor models, but also exponentially affine factor models, and even more generally local factors that can always be defined in the neighborhood of the parameter value of interest. In doing so, we remain true to the argument that Lettau and Ludvigson (2001b) present as the key motivation of their work: irrespective of some well-documented empirical shortcomings, "the reputation of the theoretical paradigm" of some popular factor models "remain well preserved."

Moreover, they stress that "an asset's risk is determined not by its unconditional correlation with the model's underlying factor, but rather by its correlation conditional on the state of the economy." We conclude then that we should elicit a pseudo-true SDF that is characterized by its ability to price accurately the fundamental factors (or the mimicking portfolios, including some actively managed ones) while taking into account the relevant conditioning information set. More precisely, fundamental factors may come with time-varying scaling weights that capture the relevant state variables.

As far as technical results are concerned, the contribution of the article is three-fold. First, it is often believed that it takes well-specified models to be able to apply the law of iterated expectations to zero conditional expectations, and thus to get a parameter true value whose definition does not imply that we take a stand on the conditional information set used by investors. However, we show that our characterization of the pseudo-true SDF as the only one able to price properly the scaled fundamental factors implies that its definition is immune to any shrinking of the information set, insofar as the relevant scaling weights are still included. Moreover, we are also able to characterize the pseudo-true SDF as minimizing the so-called conditional HJ distance that is defined, following Gagliardini and Ronchetti (2016), as the unconditional expectation of the conditional measure that one would have computed for a given information set.

Second, we propose several interpretations of this conditional HJ distance that complements the ones put forward by Gagliardini and Ronchetti (2016). A possible criticism against this conditional HJ distance is that it takes the unconditional expectation of the squared of the pricing errors on the most mispriced portfolios. We argue that, after all, when faced with a time series of conditional values of the HJ distance, the researcher must take a stand on the relevant summary of this time series of conditional assessments. The conditional HJ distance is mean squared error, meaning that it takes into account not only the magnitude of pricing errors for the most mispriced portfolios for each possible value of the conditioning information, but also the variance of these conditional pricing errors. It sounds sensible to elicit conditional asset pricing models whose level of misspecification is controlled not only through its expectation, but also through its volatility. There are obvious concerns of parsimony and out-of-sample performance that justify this mean-variance trade-off on conditional model misspecification. Moreover, we argue that there is no simple alternative to this conditional HJ measure. While managed portfolios have often been put forward to be able to characterize conditional pricing errors while using unconditional models, we document that this approach may be misleading in several respects. First, a set of managed portfolios that would by chance capture the right pseudo-true value would lead to an unconditional HJ measure that would, in contrast with the conditional measure, necessarily give no weight to the pricing errors on portfolios that are zero-beta with respect to the elicited managed portfolios. Moreover, in nonlinear models, these managed portfolios depend on the unknown parameter value and do not provide a clear guidance for estimation of the pseudo-true value. As documented by Hall and Inoue (2003), in case of misspecified moment conditions, there is no reason why iterated Generalized Method of Moments (GMM) would deliver at each step a consistent estimator of the same pseudo-true value.

Third, our characterization of the pseudo-true SDF as minimizer of the conditional HJ distance paves the way for several estimation strategies. First, it would be possible to adapt the approach of Nagel and Singleton (2011) to the context of estimation of a pseudo-true value. While these authors propose a nonparametric estimation of managed portfolios that are optimal for the purpose of efficient GMM, one could apply a similar approach, albeit with different managed portfolios, for the purpose of minimization of the HJ distance. However, as acknowledged by Nagel and Singleton (2011), their kernel-based nonparametric approach suffers from the curse of dimensionality that prevents them from including all the relevant state variables in the set of explanatory variables for nonparametric regression. Gagliardini and Ronchetti (2016) resort to the more promising approach of local GMM, extending to their context the asymptotic theory of Hall and Inoue (2003) for estimation with misspecified moment conditions. However, they need to maintain a martingale difference sequence hypothesis that means that, in a way somewhat opposite to Nagel and Singleton (2011), they must include all the needed state variables in the nonparametric regression, irrespective of the curse of dimensionality. We propose to circumvent this quandary by following Lavergne and Patilea (2013) to resort to a fixed bandwidth approach. Their smooth minimum distance (SMD) approach can be seen as a conditional extension of the jackknife GMM proposed by Newey and Windmeijer (2009). This jackknife GMM is quite convenient in that it allows us to resort to the asymptotic theory of U-statistics to revisit the theory of GMM under misspecification as developed by Hall and Inoue (2003), 15 years after the seminal work of Gallant and White (1988). Beyond the specific goal of this article, there is some value added of general interest to be able to settle the asymptotic theory of GMM under misspecification, including dynamic, conditional and SMD settings in a very concise and transparent way thanks to the extant Central Limit Theorems for U-statistics.

The rest of this article is organized as follows. Throughout the article, the focus of interest is a parametric family of sequences of SDFs $m_{t+1}(\theta)$, $\theta \in \Theta \subset \mathbb{R}^p$. In Section 2, we introduce the concept of "local factors" through the gradient $(\partial m_{t+1}(\theta)/\partial\theta)$. They coincide with the (possibly scaled) fundamental factors in the case of a conditionally affine SDF model and to a discounted value of them in the case of an exponentially conditionally affine SDF model. We note that minimizing a conditional HJ distance for each possible value of the conditional information set (with a possibly time varying parameter $\theta = \theta_t$) would amount to price exactly the local factors in each state of the world. It is then natural to define a pseudo-true value of the SDF parameters θ by the accurate pricing on average of the local factors. In Section 3, we show that this pseudo-true SDF can also be characterized as minimizing the conditional HJ distance defined, following Gagliardini and Ronchetti (2016), as the unconditional expectation (over possible values of conditioning information sets) of the time-varying conditional measures. We argue that this conditional HJ distance provides the right mean-variance trade-off regarding the time series of conditional HJ distances for the realized sequence of information sets at each date. In Section 4, we argue that the common practice of analyzing conditional pricing errors through the unconditional HJ distance computed for some specific actively managed portfolios may be misleading in several respects. First, the unconditional HJ distance underestimates the real level of conditional pricing errors, by overlooking some other portfolios. Second, inference on the pseudo true value may be flawed by the dependence on unknown parameters of the relevant actively managed portfolios. In Section 5, we discuss the various aforementioned approaches to estimating a pseudo-true SDF. In Section 6, we interpret a fixed bandwidth pseudo-true SDF in an i.i.d. setting and through a dynamic state-variable framework. In Section 7, we revisit the theory of GMM under misspecification with a jackknife alternative, which leads to the derivation of the asymptotic distributional theory of the SMD estimator of a fixed bandwidth pseudo-true value. In Section 8, we provide two numerical experiments: a finite sample comparison of local GMM and SMD for estimating a misspecified option pricing model; and a finite sample comparison of local GMM, efficient GMM, and weighted minimum distance (WMD) for estimating a linear factor model with semi-strong factors. We also provide empirical evidence for semi-strong identification in popular linear factor models. Section 9 concludes.

2 Pseudo-True SDF and Local Factors

2.1 General Framework

Let

$$R_t = \left[R_{i,t}\right]_{1 \leq i \leq n}$$

be the vector of gross returns for the primitive assets. We assume that all asset returns belong to the linear space $L^2[I(t)]$ of real random variables with finite second moment and measurable with respect to some σ-field $I(t)$. We dub an "$I-$ admissible SDF" any stochastic process $M_{t,t+1}$ that at any date $t = 1, 2, \ldots,$ satisfies $M_{t,t+1} \in L^2[I(t+1)]$ and the n no-arbitrage restrictions:

$$E\left[M_{t,t+1} R_{t+1} | I(t)\right] = 1_n, \tag{2.1}$$

where 1_n is a vector in $\mathbb{R}^n$ whose components are all equal to 1. Note that we implicitly maintain a stationarity assumption such that the no-arbitrage restriction (2.1) is independent of the date t. We denote by $\overline{\mathbf{M}}(I)$ the set of $I-$ admissible SDFs. As does most of the empirical asset pricing literature, we overlook the positivity constraint on the SDF. This allows us in particular to consider the example of linear factor models.

We consider the set $G_{t+1|t}(I)$ of payoffs that can be obtained from $I-$ actively managed portfolios based on the primitive assets:

$$G_{t+1|t}(I) = \left\{ g_{t+1} \in L^2[I(t+1)]; \exists z_{i,t} \in L^2[I(t)], i = 1, .., n : g_{t+1} = \sum_{i=1}^{n} z_{i,t} R_{i,t+1} \right\}.$$

Note that $I-$ actively managed portfolios are portfolios with weights $z_{i,t}, i = 1, \ldots, n$ that are updated at date t within the information set $I(t)$. For any payoff $g_{t+1} \in G_{t+1|t}(I)$, its market price is given by:

$$g_{t+1} = \sum_{i=1}^{n} z_{i,t} R_{i,t+1} \Rightarrow \pi_t(g_{t+1}) = \sum_{i=1}^{n} z_{i,t}.$$

Let $\{m_{t+1}(\theta); \theta \in \Theta \subset \mathbb{R}^p\}$ stand for a parametric family of candidate SDFs. These candidate SDFs will be assessed through the associated vector of $I-$ pricing errors on primitive assets:

$$e[I(t), \theta] = E[m_{t+1}(\theta) R_{t+1} - 1_n | I(t)].$$

Obviously, this vector of pricing errors is not modified if the SDF $m_{t+1}(\theta)$ is replaced by its $L^2[I(t+1)]$ orthogonal projection on the closed subspace $G_{t+1|t}(I)$. Therefore, we can assume without loss of generality that:

$$m_{t+1}(\theta) \in G_{t+1|t}(I), \forall \theta \in \Theta. \tag{2.2}$$

2.2 Local Factors

For a given parametric model $\{m_{t+1}(\theta); \theta \in \Theta \subset \mathbb{R}^p\}$, assuming that the function $\theta \to m_{t+1}(\theta)$ is differentiable for almost surely all sample paths, we can define p local factors as the components of the vector:

$$F_{t+1}^L(\theta) = \frac{\partial m_{t+1}(\theta)}{\partial \theta}, \theta \in \text{Int}(\Theta), \tag{2.3}$$

where Int(Θ) stands for the interior set of Θ. From (2.2) (and an integrability assumption), we deduce that:

$$\frac{\partial m_{t+1}(\theta)}{\partial \theta} \in G_{t+1|t}(I), \forall \theta \in \text{Int}(\Theta).$$

The terminology "local factor" is justified by the following examples.

2.2.1 Example 1: conditionally affine SDF model

Nagel and Singleton (2011) define a conditionally affine model of SDFs as:

$$m_{t+1}(\theta) = \phi_t^{0,F}(\theta) + \phi_t^F(\theta)' F_{t+1}, \tag{2.4}$$

where F_{t+1} is a $K-$ dimensional vector of observed priced risk factors and the $(K+1)$ coefficients $(\phi_t^{0,F}(\theta), \phi_t^F(\theta)') = (\phi_t^{0,F}(\theta), \phi_t^{1,F}(\theta), \ldots, \phi_t^{K,F}(\theta))$ of the conditionally affine function (2.4) are in the information set $I(t)$.

Typically, the K "fundamental factors" F_{t+1} may include not only a market portfolio return, but also consumption growth, labor income growth as advocated by Jagannathan and Wang (1996), and/or two additional portfolio returns of Fama and French (1995, 1996), namely HML (long in value stocks) and SMB (long in small firms). Following Lettau and Ludvigson (2001b), we call these factors "fundamental" in contrast with the additional scaled factors (scaled value of a fundamental factor) that contribute due to stochastic time variation in the affine function (2.4). Some of these factors will be considered in our numerical experiments (Section 8) regarding the performance of our inference procedures in front of possible weak identification (due to factors that may feature some heterogeneous identification strengths).

Generally speaking, it is worth considering three cases, from specific to general.

Case 1

$$\left(\phi_t^0(\theta), \phi_t^F(\theta)'\right) = \left(\theta_1, \theta_2, \ldots, \theta_p\right).$$

Then:

$$F_{t+1}^L(\theta)' = \frac{\partial m_{t+1}(\theta)}{\partial \theta'} = \left(1, F_{t+1}'\right).$$

The local factors are nothing but the observed priced risk factors F_{t+1} augmented with the constant payoff 1. Note, however, that we will not confuse this model with an unconditional factor model. On the contrary, this factor model provides conditional pricing:

$$1_n = \left[\theta_1 + \sum_{j=2}^{p} \theta_j E\left[F_{j,t+1}|I(t)\right]\right] E[R_{t+1}|I(t)] + \sum_{j=2}^{p} \theta_j \text{Cov}\left[R_{t+1}, F_{j,t+1}|I(t)\right]. \tag{2.5}$$

Case 2 For stochastically time varying weights $W_t = (w_{j,t})_{1 \le j \le J}$ that belong to $L^2[I(t)]$ and deterministic time-invariant coefficients $(c_{0,k}, c'_{1,k}), k = 0, 1, \ldots, K$ we define:

$$\phi_t^{k,F}(\theta) = c_{0,k} + \sum_{j=1}^{J} c_{1,k}^{j} w_{j,t} = c_{0,k} + c'_{1,k} W_t, k = 0, 1, \ldots, K$$
$$\theta = \left(\theta'_0, \theta'_1, \ldots, \theta'_J, \theta'_{J+1}\right)$$
$$\theta_0 = (c_{0,k})_{1 \le k \le K}, \theta_j = \left(c_{1,k}^{j}\right)_{1 \le k \le K}, j = 1, \ldots, J, \theta'_{J+1} = \left(c_{0,0}, c'_{1,0}\right).$$

Following the terminology introduced by Lettau and Ludvigson (2001b), this is the case of a scaled multifactor model:

$$m_{t+1}(\theta) = c_{0,0} + c'_{1,0} W_t + \sum_{k=1}^{K} \left[c_{0,k} + c'_{1,k} W_t\right] F_{k,t+1}$$
$$= \theta'_0 F_{t+1} + \sum_{j=1}^{J} \theta'_j w_{j,t} F_{t+1} + \theta'_{J+1} \left[1, W'_t\right]'.$$

As noticed by Lettau and Ludvigson (2001b), we can then rewrite it as a model conformable to case 1 with an extended set $F_{t+1}^{L}(\theta)$ of factors defined by the constant and the components of W_t, of F_{t+1}, and of $W_t \otimes F_{t+1}$. The vector of local factors precisely gathers this extended set of factors. It is worth noting that when rewriting this factor pricing model as in (2.5), we realize that this conditional factor model is somewhat observationally equivalent to a model with conditional beta coefficients that are time-varying as an affine function of underlying state variables W_t (see, e.g., Ferson and Harvey 1999).

Note, however, that only the components of F_{t+1} and of $W_t \otimes F_{t+1}$ are replicated by risky portfolios and as such will have a priced risk. This is the reason why we will set the focus on the following local factors:

$$\frac{\partial m_{t+1}(\theta)}{\partial \theta_0} = F_{t+1}$$
$$\frac{\partial m_{t+1}(\theta)}{\partial \theta_j} = w_{j,t} F_{t+1}, j = 1, \ldots, J.$$

Case 3 Local factors can be interpreted as rescaled factors in the general case (2.4), and are then given by the components of $\frac{\partial \phi_t^0(\theta)}{\partial \theta}$ and $\frac{\partial \phi_t^{k,F}(\theta)}{\partial \theta} \otimes F_{t+1}, k = 1, \ldots, K$. The only difference with case 2 is that the rescaling of the factors is not necessarily defined as an affine function of some state variables.

2.2.2 Example 2: exponentially conditionally affine SDF model

This case can be described as:

$$m_{t+1}(\theta) = \exp\left[\tilde{m}_{t+1}(\theta)\right],$$

where $\tilde{m}_{t+1}(\theta)$ is conformable to the conditionally affine structure described in example 1 above. The exponential affine structure is well suited to ensure positivity of the SDF.

Since this feature is especially important for option pricing, we will perform in Section 8 some numerical experiments with a discrete time exponentially affine option pricing model for which fundamental factors are the return on the underlying asset jointly with a factor driving its stochastic volatility process. It is worth noting that the exponential transformation of a conditionally affine function of the fundamental factors preserves a large part of the financial interpretation of local factors still defined as the components of:

$$\frac{\partial m_{t+1}(\theta)}{\partial \theta} = m_{t+1}(\theta) \frac{\partial \tilde{m}_{t+1}(\theta)}{\partial \theta}.$$

In other words, local factors can now be seen as discounted values of local factors with a functional form conformable to example 1. Therefore, example 2 can be treated in a way similar to example 1, up to a preliminary discounting of the local factors.

2.3 Pseudo-True SDF

To compute a pseudo-true SDF, it is worth revisiting in a conditional setting the Hansen and Jagannathan (1997) assessment of specification errors in the SDF model. We are then led to compute a conditional stochastic (squared) HJ-distance defined by:

$$\delta^2(\theta)[I(t)] = \inf_{M_{t,t+1} \in \overline{\mathbf{M}}} (\mathbf{I}) E\left\{ \left[M_{t,t+1} - m_{t+1}(\theta) \right]^2 | I(t) \right\}. \tag{2.6}$$

Even though they have been derived in an unconditional framework, the formulas of Hansen and Jagannathan (1997) can be applied for each conditioning value, leading to the solution:

$$M^*_{t,t+1}(\theta) = m_{t+1}(\theta) - \{\lambda(\theta)[I(t)]\}' R_{t+1}$$

$$\lambda(\theta)[I(t)] = \Omega^{-1}[I(t)] e[I(t), \theta]$$

$$\Omega[I(t)] = E\left[R_{t+1} R'_{t+1} | I(t)\right]$$

$$e[I(t), \theta] = E[m_{t+1}(\theta) R_{t+1} - 1_n | I(t)]$$

and, by plugging into (2.6), to:

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

A state-dependent pseudo-true SDF $m_{t+1}(\theta_t)$ could then be defined as minimizer of the state-dependent conditional HJ distance; that is as a solution of the first-order conditions:

$$E\left[\frac{\partial m_{t+1}(\theta_t)}{\partial \theta} R'_{t+1} | I(t) \right]]\Omega^{-1}[I(t)] E[m_{t+1}(\theta_t) R_{t+1} - 1_n | I(t)] = 0. \tag{2.7}$$

A similar use of a state dependent SDF parameter θ_t has also been promoted by Gagliardini, Gourieroux, and Renault (2011) (see also Fang, Ren, and Yuan, 2011). However, for several reasons detailed below, we will set more focus on the minimization of an averaged state dependent conditional HJ distance:

$$\delta^2(\theta) = E\left\{ \delta^2(\theta)[I(t)] \right\}$$

whose minimization leads to the definition of a fixed pseudo-true value θ^* of the SDF parameters as solution of:

$$E\left\{E\left[\frac{\partial m_{t+1}(\theta^*)}{\partial\theta}R'_{t+1}|I(t)\right]\Omega^{-1}[I(t)]E[m_{t+1}(\theta^*)R_{t+1}-1_n|I(t)]\right\}=0. \qquad (2.8)$$

In order to interpret these pseudo-true values $\theta_t, t=1,\ldots,T$ and θ^*, it is first worth realizing that the conditional stochastic HJ distance is invariant by actively recombining portfolios. More precisely, let us define a vector of gross returns on n actively managed portfolios:

$$\tilde{R}_{t+1}=A_tR_{t+1},$$

where A_t is a nonsingular matrix of size n, whose stochastically time varying coefficients belong to $L^2[I(t)]$ and are such that:

$$A_t1_n=1_n.$$

Then, if we define:

$$\begin{aligned}\tilde{e}[I(t),\theta]&=E\left[m_{t+1}(\theta)\tilde{R}_{t+1}-1_n|I(t)\right]\\ \tilde{\Omega}[I(t)]&=E\left[\tilde{R}_{t+1}\tilde{R}'_{t+1}|I(t)\right]\\ \tilde{\delta}^2(\theta)[I(t)]&=\tilde{e}[I(t),\theta]'\tilde{\Omega}^{-1}[I(t)]\tilde{e}[I(t),\theta].\end{aligned}$$

we obviously have:

$$\tilde{\delta}^2(\theta)[I(t)]=\delta^2(\theta)[I(t)].$$

This invariance property will help us to interpret the pseudo-true of the SDF parameters θ_t that we can define from the state-dependent conditional HJ distance $\delta^2(\theta)[I(t)]$. More precisely, for any of the factor models defined above, since the local factors $F^L_{t+1}(\theta_t)$ belong to $G_{t+1|t}(I)$, it is always possible to exhibit a $q-$ dimensional subvector $\tilde{F}^L_{t+1}(\theta_t)$ and a nonsingular matrix A_t such that $\tilde{R}_{t+1}=A_tR_{t+1}$ can be written:

$$\tilde{R}_{t+1}=\begin{bmatrix}\tilde{F}^L_{t+1}(\theta_t)\\ X_{t+1}\end{bmatrix} \qquad (2.9)$$

with:

$$E\left[\tilde{F}^L_{t+1}(\theta_t)X'_{t+1}|I(t)\right]=0.$$

Of course, a similar construction is possible with $\tilde{F}^L_{t+1}(\theta^*)$. Note that we have isolated a subvector $\tilde{F}^L_{t+1}(\theta_t)$ of $F^L_{t+1}(\theta_t)$ (resp. $\tilde{F}^L_{t+1}(\theta^*)$ of $F^L_{t+1}(\theta^*)$) in order to discard the factor values that may be known at time t and to set the focus on factors $\tilde{F}^L_{t+1}(\theta_t)$ (resp. $\tilde{F}^L_{t+1}(\theta^*)$) that are mimicked by risky assets at time t. Then, with a notation that overlooks the difference between primitive asset returns R_{t+1} and returns $\tilde{R}_{t+1}$ on actively managed portfolios, we can assume without loss of generality that the vector R_{t+1} of primitive asset returns can be written as (2.9) (or similarly with $\tilde{F}^L_{t+1}(\theta^*)$) so that:

$$\Omega[I(t)]=E\left[R_{t+1}R'_{t+1}|I(t)\right]=\begin{bmatrix}E\left[\tilde{F}^L_{t+1}(\theta_t)\tilde{F}^{L\prime}_{t+1}(\theta_t)|I(t)\right] & 0\\ 0 & E\left[X_{t+1}X'_{t+1}|I(t)\right]\end{bmatrix}. \qquad (2.10)$$

The decomposition (2.10) allows us to simplify the expression of first-order conditions (2.7) and (2.8) by setting the focus on the pricing of factors:

$$E\Big[m_{t+1}(\theta_t)\tilde{F}^L_{t+1}(\theta_t)|I(t)\Big] = 1_q \tag{2.11}$$

$$E\Big[m_{t+1}(\theta^*)\tilde{F}^L_{t+1}(\theta^*)\Big] = 1_q. \tag{2.12}$$

Equations (2.11) and (2.12) can be interpreted as pricing equations for the local factors. Note that our simplification has allowed us to get conditional pricing Equations (2.11) which, as standard Euler equations, remain true when the econometrician underestimates the information set $I(t)$ used by investors. As far as the time-invariant pseudo-true value θ^* is concerned, its characterization (2.12) does not even depend on the specific definition of any information set $I(t)$. To clarify this statement, we will document it below within the framework of case 2 of conditionally affine SDF models. However, all the factor models described in Section 2.2 could be accommodated similarly, up to possibly heavier notations.

Example 1: Case 2 (continued)

In the context of the conditionally affine model with scaled factors described above, we obviously have:

$$\frac{\partial m_{t+1}(\theta)}{\partial \theta_0} = F_{t+1}$$
$$\frac{\partial m_{t+1}(\theta)}{\partial \theta_j} = \omega_{j,t}F_{t+1}, \qquad j = 1, \ldots, J$$

so that:

$$\tilde{F}^L_{t+1}(\theta) = \begin{bmatrix} 1 \\ W_t \end{bmatrix} \otimes F_{t+1}.$$

In other words, the pricing Equations (2.12) can be written as:

$$E\left\{ m_{t+1}(\theta^*)\begin{bmatrix} 1 \\ W_t \end{bmatrix} \otimes F_{t+1} \right\} = 1_q, \qquad q = K(J+1). \tag{2.13}$$

This formula generalizes to a conditional multifactor model a result that Peñaranda and Sentana (2015) (see their Lemma 1 and subsequent comments, end of page 417) have recently put forward for an unconditional factor model: the exact pricing of the factors (both fundamental factors F_{t+1} and scaled factors $w_{j,t}F_{t+1}, j = 1, \ldots, J$) leads to "minimize the sample counterpart of the HJ distance, irrespective of the distribution of returns and the validity of the asset pricing model," since this minimization is by definition achieved by the pseudo-true value θ^*.

Moreover:

$$m_{t+1}(\theta) = \{[1\,W_t'] \otimes F_{t+1}'\}\tilde{\theta} + [1 \quad W_t']\theta_{J+1}$$
$$\tilde{\theta}' = \Big[\theta_0', \theta_1', \ldots, \theta_J'\Big], \theta' = \Big[\tilde{\theta}', \theta_{J+1}'\Big].$$

Therefore (2.13) can be seen as a linear system of equations that determines $\tilde{\theta}^*$ for a given value of θ^*_{J+1}:

$$E\left[x_{t+1}x'_{t+1}\right]\tilde{\theta}^* = 1_q - E\left\{\left[1 \quad W'_t\right]\theta^*_{J+1}x_{t+1}\right\}$$
$$x_{t+1} = \begin{bmatrix} 1 \\ W_t \end{bmatrix} \otimes F_{t+1}. \tag{2.14}$$

Note that the matrix $E\left[x_{t+1}x'_{t+1}\right]$ of this system can safely be assumed to be nonsingular since with obvious notations (and $\omega_{0,t} = 1$)

$$\alpha' x_{t+1} = \sum_{j=0}^{J} \alpha'_j \omega_{j,t} F_{t+1} \equiv 0 \Rightarrow \alpha = \left(\alpha_j\right)_{0 \le j \le J} = 0.$$

Otherwise, some factors among the scaled factors $\omega_{j,t}F_{t+1}, j = 0, 1, \ldots, J$ would be redundant.

One important by-product of the Cramer system of Equations (2.14) is to prove that HJ pseudo-true values θ^*_j , $j = 0, 1, \ldots, J$, albeit defined through the conditional HJ distance (that depends of the information set $I(t)$) do not even depend on the specification I of the information set for the characterization of I-admissible SDFs. Note that we discuss the pseudo-true values of $\theta^*_j, j = 0, 1, \ldots, J$ only for a given value of θ^*_{J+1}. While the former is our focus of interest for conditional pricing models on the period $[t, t+1]$, the latter pertains to the determination of the term structure of interest rates (pricing of assets whose future payoffs is known a time t) which is beyond the scope of this article. Our way to overlook the determination of θ^*_{J+1} can be compared with the discussion in Hansen and Jagannathan (1997) about the parameter γ_0 that is "the price assigned to a unit payoff." They ask "why should we constrain the family of discount factors to price the unit payoff in precisely the same way the proxy does." We are doing something similar in a conditional context. The complete determination of the pseudo-true value θ^* of $\left(\theta_j\right)_{0 \le j \le J+1}$ will be described in the next section.

3 Pseudo-True SDF and HJ-Distance

Following Gagliardini and Ronchetti's (2016) conditional extension of the HJ-distance, we define the conditional distance between our parametric model and the set of admissible SDFs as:

$$\delta^2 = \inf_{\theta \in \Theta} \delta^2(\theta)$$
$$\delta^2(\theta) = \inf_{M_{t,t+1} \in \overline{\mathbf{M}}(I)} E\left[M_{t,t+1} - m_{t+1}(\theta)\right]^2. \tag{3.1}$$

Note that this HJ-distance is conditional, albeit defined by an unconditional expectation, due to the Definition (2.1) of the set $\overline{\mathbf{M}}(I)$ of admissible SDFs by conditional moment restrictions. More precisely, by virtue of the law of iterated expectations, a sufficient condition for $M_{t,t+1}$ to reach the minimum in (3.1) is to solve for (almost surely) any given value of the conditioning information $I(t)$:

$$\delta^2(\theta)[I(t)] = \inf_{M_{t,t+1} \in \overline{\mathbf{M}}(\mathbf{I})} E\left\{\left[M_{t,t+1} - m_{t+1}(\theta)\right]^2 | I(t)\right\}$$
$$\delta^2(\theta) = E\left\{\delta^2(\theta)[I(t)]\right\}. \tag{3.2}$$

Therefore, from the formulas of the previous section:

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

and thus:

$$\delta^2(\theta) = E\{e[I(t),\theta]'\Omega^{-1}[I(t)]e[I(t),\theta]\}. \tag{3.3}$$

This interpretation allows us to revisit in a conditional setting an important result of Hansen and Jagannathan (1997) (their Proposition 2.1): the distance $\delta(\theta)[I(t)]$ is the largest possible (conditional) pricing error (among all payoffs $g_{t+1} \in G_{t+1|t}(I)$ of unit norm) when using the candidate SDF $m_{t+1}(\theta)$:

$$\delta(\theta)[I(t)] = \max_{g_{t+1} \in G_{t+1|t}(I)} \{|E[m_{t+1}(\theta)g_{t+1}|I(t)] - \pi_t(g_{t+1})|; E[g_{t+1}^2|I(t)] = 1\}.$$

More precisely, Hansen and Jagannathan (1997) prove that this maximization gives a sharp bound, in the sense that there exists a payoff $g^*_{t+1}(\theta) \in G_{t+1|t}(I)$ such that:

$$\delta(\theta)[I(t)] = |E[m_{t+1}(\theta)g^*_{t+1}(\theta)|I(t)] - \pi_t(g^*_{t+1}(\theta))| \equiv PE_t(g^*_{t+1}(\theta))$$

that measures the absolute pricing error at time t on the payoff $g^*_{t+1}(\theta)$. Therefore, our definition (3.1) of the conditional HJ-distance can be interpreted as follows:

$$\delta^2(\theta) = E[\delta^2(\theta)[I(t)]] = [E\{PE_t(g^*_{t+1}(\theta))\}]^2 + \mathrm{Var}\{PE_t(g^*_{t+1}(\theta))\}.$$

In other words, our minimization (3.1) not only amounts to a minimization of the averaged maximum pricing error, but also it penalizes excess volatility in this time-varying pricing error. While it is often said that an advantage of HJ-distance by comparison with efficient GMM is that "the HJ-distance does not reward SDF volatility" (Ludvigson, 2013), the conditional setting leads to put even more emphasis on the concern for volatility of pricing errors. This approach is similar in spirit to popular calibration exercises in applied economics when a model-based (possibly simulated) sample path $y_t(\theta), t = 1, \ldots, T$, of a variable of interest can be compared to its real-world observation $y_t, t = 1, \ldots, T$, where here $y_t(\theta) = PE_t(g^*_{t+1}(\theta))$ and $y_t = 0$. It is then common to minimize with respect to the unknown parameters θ the mean squared error:

$$\frac{1}{T}\sum_{t=1}^{T}[y_t(\theta) - y_t]^2.$$

We can see this approach as minimizing a mean squared error of prediction of asset prices. Thus, even though the conditional HJ-distance is a population average quantity, it is true that it sets the focus on conditional pricing errors. Following Lettau and Ludvigson (2001b), the SDF is "expected to price assets [...] conditionally," leading in particular to "conditional rather than fixed linear factor models." It is precisely the HJ pseudo-true value θ^* of θ that does the best job in this respect. Beyond the examples considered in Section 2, it is the solution (assumed to be unique) of the minimization (3.1), characterized by the first-order conditions:

$$E\left\{\frac{\partial e[I(t),\theta^*]'}{\partial\theta}\Omega^{-1}[I(t)]e[I(t),\theta^*]\right\} = 0. \tag{3.4}$$

Irrespective of the interpretation of the conditional HJ-distance $\delta^2(\theta)$, it is worth keeping in mind that, by virtue of the results of Section 2 above, the pseudo-true SDF $m_{t+1}(\theta^*)$ is mainly determined by the fact that it prices factors and scaled factors accurately on average. More precisely, by the law of iterated expectations argument already mentioned above (see (3.2)), we know from Section 2 that for the case of conditionally affine SDF model the pseudo-true value θ^* is determined as solution of the unconditional equations:

First:

$$E\left[m_{t+1}(\theta^*)\omega_{j,t}F_{t+1}\right] = 1_K, \qquad j = 0, 1, \ldots, J. \tag{3.5}$$

Second:

$$E\left\{\begin{bmatrix} 1 \\ W_t \end{bmatrix} E\left[R'_{t+1}|I(t)\right]\Omega^{-1}[I(t)]e[I(t), \theta]\right\} = 0. \tag{3.6}$$

Therefore, up to the dependence on θ^*_{J+1} whose characterization involves Equation (3.6), the pseudo-true value θ^* is essentially determined by the factor pricing Equation (3.5) that do not depend on the specification of the conditioning information I, but only on its summary by state variables W_t. In this respect, we are conformable with the philosophy forcefully promoted by Lettau and Ludvigson (2001b):

On the one hand, irrespective of some well-documented empirical shortcomings, "the reputation of the theoretical paradigm" of some popular factor models "remains well preserved." There are some "fundamental factors" F_{t+1} that should be able to characterize the relevant priced risk in the economy.

On the other hand, "an asset's risk is determined not by its unconditional correlation with the model's underlying factor, but rather by its correlation conditional on the state of the economy," as captured by time-varying scaling weights W_t.

Then, it seems sensible to dub as a "pseudo-true SDF" a SDF that is able to price accurately, on average, both the fundamental factors and their scaled values. It is a fair way to characterize the pricing message of our conditional factor model, that both the fundamental factors and their scaled counterparts with relevant weights (e.g., *cay* in Lettau and Ludvigson, 2001a,b)) exhibit explanatory power. We have proved that it is a convenient property of the conditional HJ-distance, that has been overlooked so far, to deliver such a pseudo-true SDF as a result of its minimization. This result can be seen as a conditional extension of a result of Hansen and Jagannathan (1997) (see their minimization (39)): the pseudo-true value of the parameters is set "to price correctly the factor mimicking payoffs."

4 Pseudo-True SDF and Managed Portfolios

While it has been proved in Section 2 that the HJ pseudo-true value is characterized by the exact pricing of the (rescaled) factors of a conditionally affine SDF model when the factors are traded, the concept of managed portfolios allows us to extend this interpretation for general SDF models.

4.1 Mimicking Portfolios for Local Factors

A $(q \times n)$ matrix $Z_t = \left[Z_{i,t}'\right]_{1 \leq i \leq q}$ whose coefficients are all in $L^2[I(t)]$ allows us to define a vector of returns on q "managed" portfolios with time $t+1$ payoffs $Z'_{i,t}R_{t+1} \in G_{t+1|t}(I)$

and time t prices $Z'_{i,t}1_n \in L^2[I(t)]$, $i = 1, \ldots, q$. We then define the set $\overline{\mathbf{M}}_Z$ of $Z-$ admissible SDFs as the set of variables $M_{t,t+1}$ in $L^2[I(t+1)]$ such that:

$$E\left[M_{t,t+1} Z_t R_{t+1} - Z_t 1_n\right] = 0. \tag{4.1}$$

Obviously, the no-arbitrage condition (4.1) is implied by (2.1). Conversely, (2.1) means that (4.1) is fulfilled by any possible set of managed portfolios Z_t.

This enhances the interpretation of the conditional pseudo-true value θ^* that we have put forward in the former subsection. According to the first-order conditions (3.4), the pseudo-true SDF $m[I(t+1), \theta^*]$ is the only $Z-$ admissible SDF in the parametric family $\{m[I(t+1), \theta]; \theta \in \Theta \subset \mathbb{R}^p\}$ when we want to price properly the p managed portfolios $Z_t = \left[Z_{i,t}{}'\right]_{1 \le i \le p}$ with respective payoffs:

$$g_{i,t+1} = Z_{i,t}(\theta^*)' R_{t+1} \tag{4.2}$$

$$Z_{i,t}(\theta)' = \frac{\partial e[I(t), \theta]'}{\partial \theta_i} \Omega^{-1}[I(t)], i = 1, \ldots, p \tag{4.3}$$

$$Z_t(\theta) = \left[Z_{i,t}(\theta)'\right]_{1 \le i \le p}. \tag{4.4}$$

These managed portfolios $Z_t(\theta) = \left[Z_{i,t}(\theta)'\right]_{1 \le i \le p}$ are "HJ-optimal" in the sense that their pricing characterizes the HJ pseudo-true value since we can rewrite (3.4) as:

$$E\{Z_t(\theta^*)[m_{t+1}(\theta^*) R_{t+1} - 1_n]\} = 0. \tag{4.5}$$

HJ-optimal portfolios are in general different from the managed portfolios $\tilde{Z}_t(\theta) = \left[\tilde{Z}_{i,t}(\theta)'\right]_{1 \le i \le p}$ that characterize optimal instruments for efficient GMM (see, e.g., Nagel and Singleton, 2011). The latter would be defined as:

$$\tilde{Z}_{i,t}(\theta)' = \frac{\partial e[I(t), \theta]'}{\partial \theta_i} \Sigma^{-1}[I(t), \theta], i = 1, \ldots, p$$
$$\Sigma[I(t), \theta] = \mathrm{Var}[m_{t+1}(\theta) R_{t+1} | I(t)].$$

The HJ-optimal managed portfolios are actually tightly related to the local factors defined in (2.3):

$$\begin{aligned} Z_t(\theta) &= \frac{\partial e[I(t), \theta]'}{\partial \theta} \Omega^{-1}[I(t)] \\ &= E\left[F^L_{t+1}(\theta) R'_{t+1} | I(t)\right] \left\{E\left[R_{t+1} R'_{t+1} | I(t)\right]\right\}^{-1} R_{t+1}. \end{aligned} \tag{4.6}$$

In other words, the HJ-optimal managed portfolios are the mimicking portfolios (i.e., the conditional linear regressions on the set of primitive returns) of the local factors. Note these linear regressions are indeed affine regressions when, as it is generally the case, one component of the vector $F^L_{t+1}(\theta)$ of local factors is just the constant. This extends a result already noted in the case of a conditionally affine model by Peñaranda, Rodríguez-Poo, and Sperlich (2017) (see their Appendix B1) to a general SDF model.

It is an additional advantage of the conditional HJ-distance approach with respect to the efficient GMM à la Nagel and Singleton (2011). The regression formulas in (4.6) clearly characterize the HJ-optimal managed portfolios. This can be seen as a generalization of the interpretation of the pseudo-true value of the SDF put forward in Section 2 in the case of conditionally affine factor models. Even when there is no such thing as fundamental factors (and scaled values of them), we still characterize the HJ pseudo-true value of the SDF in terms of pricing of local factors (or their mimicking portfolios). Of course, in nonaffine cases, this characterization is complicated by the fact that local factors and their mimicking portfolios may depend upon unknown parameters (and conditioning information for the definition of mimicking portfolios). The practical consequences of this dependence are discussed in the next subsection 4.2. However, in any case, and in contrast with efficient GMM, we resort to mimicking portfolios of local factors that are defined in a model-invariant way.

4.2 Sensitivity of Mimicking Portfolios to Unknown Parameters and Conditioning Information

4.2.1 Dependence on unknown parameters

We know from (4.5) that the conditional HJ pseudo-true value θ^* is characterized by the exact pricing of the HJ-optimal managed portfolios, meaning that it is the only solution θ of the equation:

$$e_Z[\theta^*, \theta] = 0,$$

where:

$$e_Z\left[\theta^1, \theta^2\right] = E\left\{Z_t\left(\theta^1\right)\left[m\left[I(t+1), \theta^2\right]R_{t+1} - 1_n\right]\right\}. \tag{4.7}$$

Moreover, it is obvious that:

$$e_Z\left[\theta^1, \theta^2\right] = 0 \not\Rightarrow \theta^2 = \theta^1.$$

As discussed in Section 5 below, this difficulty is the conditional analog of the issue of two-step GMM estimators with misspecified moment models in Hall and Inoue (2003): a two-step estimator does not converge toward the same pseudo true value as the first step estimator used to compute it. In our case, a GMM estimator based on instruments estimated with a first step consistent estimator of a pseudo true value θ^1 does not converge in general toward θ^1. However, we do know from the first-order conditions (3.4) that the conditional pseudo-true value of interest can be characterized as a fixed point.

This characterization of the pseudo-true value θ^* as a fixed point may suggest a GMM-based estimation of θ^* through an iterated sequence of minimization problems based on unconditional HJ distance. More precisely, under standard regularity conditions, we can prove the following implication of the implicit function theorem:

Theorem 1 There exists a neighborhood $\aleph(\theta^*)$ of θ^* such that an implicit function $\overline{\theta}(.)$ can be defined as solution of:

$$\begin{gathered} e_Z\left[\theta, \overline{\theta}(\theta)\right] = 0, \forall \theta \in \aleph(\theta^*) \\ \overline{\theta}(\theta^*) = \theta^* \end{gathered} \tag{4.8}$$

In particular:

$$\frac{\partial e_Z(\theta^*, \theta^*)}{\partial \theta^1} + \Omega_Z(\theta^*)\frac{\partial \overline{\theta}(\theta^*)}{\partial \theta} = 0 \tag{4.9}$$

with:

$$\Omega_Z(\theta) = E\left[Z_t(\theta)R_{t+1}R'_{t+1}Z'_t(\theta)\right] = E\left[E\left[\frac{\partial \Psi'_{t+1}(\theta)}{\partial \theta}|I(t)\right]\Omega^{-1}[I(t)]E\left[\frac{\partial \Psi_{t+1}(\theta)}{\partial \theta'}|I(t)\right]\right].$$

Theorem 1 suggests that it may be sensible to assume that the function $\overline{\theta}(.)$ is a contraction mapping in neighborhood of θ^*. While each of the two terms of (4.9) is zero in the case of a conditional affine structure of the SDF, the common practice of linearizing nonlinear factor models (see, e.g., linearized version of the model of Lustig and Van Nieuwerburgh (2005) mentioned in Nagel and Singleton (2011)) leads us to reckon that the first term is often small enough to get the contraction mapping inequality:

$$\left\|\frac{\partial \overline{\theta}(\theta^*)}{\partial \theta}\right\| = \left\|\Omega_Z^{-1}(\theta^*)\frac{\partial e_Z(\theta^*,\theta^*)}{\partial \theta^1}\right\| < 1. \tag{4.10}$$

Note in particular that:

$$\frac{\partial e_Z(\theta^*,\theta^*)}{\partial \theta^1} = E\left\{E\left[\frac{\partial F'_{t+1}(\theta^*)}{\partial \theta}R'_{t+1}|I(t)\right]\left\{E\left[R_{t+1}R'_{t+1}|I(t)\right]\right\}^{-1}[m[I(t+1),\theta^*]R_{t+1}-1_n]\right\}. \tag{4.11}$$

Assuming a small norm for the matrix (4.11), and as a consequence the contraction mapping property (4.10), amounts to assume that local factors, or at least their mimicking portfolios, do not vary much with parameters in the neighborhood of the pseudo-true value θ^*.

When this condition is fulfilled, it should ensure the convergence of an iterated sequence of GMM estimators based on managed portfolios, corresponding to the sample counterpart of the iterations defined by:

$$e_Z\left[\theta^k,\theta^{k+1}\right] = 0.$$

4.2.2 Dependence on conditioning information

For sake of illustration, let us follow Nagel and Singleton (2011) and consider an extended consumption-based SDF in which c_t denotes the logarithm of consumption and:

$$\begin{gathered} m[I(t+1),\theta] = (\beta_1+\gamma_1 n_t) + (\beta_2+\gamma_2 n_t)\Delta c_{t+1} \\ \theta = (\beta_1,\gamma_1,\beta_2,\gamma_2)'. \end{gathered}$$

The model in Lettau and Ludvigson (2001b), for example, is the special case with n_t equal to *cay*. For this setup, with notations:

$$\begin{aligned} \Psi_{t+1}(\theta) &= m[I(t+1),\theta]R_{t+1} - 1_n \\ e[I(t),\theta] &= E[\Psi_{t+1}(\theta)|I(t)]. \end{aligned}$$

we have:

$$\frac{\partial \Psi_{t+1}(\theta)'}{\partial \theta} = \begin{bmatrix} R'_{t+1} \\ n_t R'_{t+1} \\ \Delta c_{t+1} R'_{t+1} \\ n_t \Delta c_{t+1} R'_{t+1} \end{bmatrix}.$$

so that:

$$\frac{\partial e[I(t),\theta^*]'}{\partial\theta} = E\left[\frac{\partial\Psi_{t+1}(\theta^*)'}{\partial\theta}|I(t)\right] = \begin{bmatrix} E\left[R'_{t+1}|I(t)\right] \\ n_t E\left[R'_{t+1}|I(t)\right] \\ E\left[\Delta c_{t+1}R'_{t+1}|I(t)\right] \\ n_t E\left[\Delta c_{t+1}R'_{t+1}|I(t)\right] \end{bmatrix}. \tag{4.12}$$

As already noted by Nagel and Singleton (2011), the managed portfolios of interest are constructed with weights that are computed from the components of the conditional first and second moments of the joint vector of returns and factors (Δc_{t+1} in this case). An additional factor f_{t+1} in the SDF specification would lead to add weights based on $E\left[f_{t+1}R'_{t+1}|I(t)\right]$. When the factors themselves are returns, we are computing conditional first and second moment of returns.

In the case of a misspecified model, since the pseudo true value may depend on the conditioning information, we should rather say that an additional factor f_{t+1} in the SDF specification would lead to add weights based on $E\left[f_{t+1}R'_{t+1}|\bar{I}(t)\right]$, where $\bar{I}(t)$ is an augmented information set defined by:

$$\bar{I}(t) = I(t) \vee \{f_\tau, \tau \leq t\}.$$

For the same reason, all the conditional moments in (4.12) should now be computed given $\bar{I}(t)$. This remark is somewhat at odds with the empirical strategy as described by Nagel and Singleton (2011) (see page 889). They reveal that they use nonparametric local polynomial regression estimators of the conditional moments in (4.12) (as well as some sieve methods). They stress that "computational considerations" (and we shall add the curse of dimensionality in terms of rates of convergence) "typically dictate that nonparametric estimation must focus on a small number of conditioning variables." We consider that this parsimony in terms of the choice of conditioning variables is far from innocuous in the misspecified case, and this is the reason why we will propose in the next section an alternative nonparametric strategy with fixed bandwidth that is less sensitive to the curse of dimensionality.

4.3 Pricing of Zero-Beta Portfolios

The characterization (3.4) of the conditional pseudo-true value θ^* through well-chosen managed portfolios may lead to the spurious conclusion that an unconditional HJ-distance, computed with well-suited managed portfolios, conveys the same information as the conditional HJ-distance. We will show in this subsection that it is not the case, and that similarly to conditional moment specification tests, an unconditional distance based on a finite number of arbitrary managed portfolios is only "directional," in the sense that it is unable to assess some pricing errors in other directions, even though they may deliver the optimal pseudo-true SDF.

To see that, let us consider the unconditional HJ-distance associated with a given set Z_t of managed portfolios (with Z_t being a $(q \times n)$ matrix with coefficients in $L^2[I(t)]$):

$$\begin{aligned} \delta_Z^2 &= \inf_{\theta\in\Theta} \delta_Z^2(\theta) \\ \delta_Z^2(\theta) &= \inf_{M_{t,t+1}\in\overline{\mathbf{M}}_Z} E\left[M_{t,t+1} - m[I(t+1),\theta]\right]^2. \end{aligned} \tag{4.13}$$

Again, this quadratic optimization problem can be solved as follows:

$$\begin{aligned} M^{*Z}_{t,t+1}(\theta) &= m[I(t+1),\theta] - \gamma(\theta)' Z_t R_{t+1} \\ \gamma(\theta) &= \Omega_Z^{-1} e_Z[\theta] \\ \Omega_Z &= E\left[Z_t R_{t+1} R'_{t+1} Z'_t\right] = E\left[Z_t \Omega[I(t)] Z'_t\right] \\ e_Z[\theta] &= E\{Z_t[m[I(t+1),\theta]R_{t+1} - 1_n]\} = E\{Z_t e[I(t),\theta]\} \end{aligned}$$

which, by plugging into (4.13), leads to:

$$\delta_Z^2(\theta) = e_Z[\theta]' \Omega_Z^{-1} e_Z[\theta].$$

For any choice of instruments Z_t (with possible overidentification), that would allow us to compute not only the unconditional pseudo true value:

$$\theta_Z = \arg \min_{\theta \in \Theta} \delta_Z^2(\theta)$$

but also the conditional one (because $\theta^* = \theta_Z$). However, the coincidence between θ^* and θ_Z does not mean that the unconditional HJ-distance $\delta_Z^2 = \delta_Z^2(\theta_Z)$ is an accurate assessment of the conditional HJ-distance $\delta^2 = \delta^2(\theta^*)$. To see that, it is worth following Gagliardini and Ronchetti (2016) by reinterpreting the HJ-distances in terms of orthogonal projections. For this purpose, we introduce a Hilbert space $H[I(t)]$ of n-dimensional functions $\xi[I(t)]$ whose components are all elements of $L^2[I(t)]$, and we endow $H[I(t)]$ with the scalar product:

$$\langle \xi, \zeta \rangle = E\left[\xi[I(t)]' \Omega^{-1}[I(t)] \varsigma[I(t)]\right].$$

From Equation (3.3), the conditional HJ-distance can be written as the minimized $H[I(t)]$-norm of the conditional pricing error vector w.r.t. the parameter θ:

$$\begin{aligned} &\delta^2(\theta) = ||e[.,\theta]||^2 \\ &\delta^2 = \min{}_\theta ||e[.,\theta]||^2 = ||e[.,\theta^*]||^2. \end{aligned}$$

Then Proposition 1 below is a corollary of Proposition 1 in Gagliardini and Ronchetti (2016). However, we provide the direct proof in Appendix for the purpose of self-containedness:

Proposition 1 *If a $(q \times n)$ matrix Z_t that defines a vector of payoffs $Z_t R_{t+1}$ on q "managed" portfolios is such that the implied pseudo true value θ_Z coincides with the conditional pseudo true value θ^*, then:*

$$\delta_Z^2 = ||P_C e[.,\theta^*]||^2,$$

where P_C stands for the orthogonal projection on the columns of the $(n \times q)$ matrix:

$$C[I(t)] = \Omega[I(t)] Z'_t.$$

The message of Proposition 1 is not surprising. Even if, by chance, a well-chosen set of q managed portfolios ($q \geq p$) allows us to characterize the conditional pseudo true value θ^*, the (squared) HJ-distance induced by these portfolios underestimates the (squared) conditional HJ-distance by an amount equal to the orthogonal projection of the vector of pricing errors on the space orthogonal to the one spanned by these portfolios. By virtue of Pythagoras' theorem:

$$\delta^2 - \delta_Z^2 = ||P_{C^\perp} e[., \theta^*]||^2.$$

The unconditional HJ-distance based on q managed portfolios simply overlooks the pricing errors for portfolios whose payoffs are (conditionally) uncorrelated with the payoffs of these q particular portfolios. For all $i = 1, 2, \dots, q$ and $j = 1, 2, \dots, n - q$:

$$0 = \text{Cov}\left[\left(Z_t^{i\prime} R_{t+1}\right), \left(Z_t^{j\prime} R_{t+1}\right) | I(t)\right] = \langle C^i(.), C^{\perp j}(.)\rangle$$

if:

$$C[I(t)] = \Omega[I(t)]\left[Z_t^1, Z_t^2, \dots, Z_t^q\right]$$
$$C^\perp[I(t)] = \Omega[I(t)]\left[Y_t^1, Y_t^2, \dots, Y_t^{n-q}\right]$$

and we assume without loss of generality that these managed portfolios deliver payoffs with zero conditional mean.

The unconditional HJ-distance δ_Z^2 based on a set of q reference portfolios Z_t is not reliable since it does not take into account the (conditional) pricing errors on assets that are (conditionally) zero-beta with respect to these q portfolios. Note that this negative result is nothing but the flipside of the result put forward in Section 2. Since the HJ pseudo-true value θ^* is characterized by the exact pricing of fundamental factors and their scaled values (or more generally of local factors), it is not informative about the pricing of assets that are zero-beta with respect to the mimicking portfolios of these factors.

5 GMM Estimation of Pseudo-True SDF

We sketch in this section three kernel smoothing-based GMM strategies for inference about the HJ pseudo-true SDF. We start from the traditional approach of optimal managed portfolios, even though in contrast with Nagel and Singleton (2011), we set the focus on HJ-optimality, and not on optimal instruments for efficient GMM. We subsequently argue that a one-step local GMM may have a better finite sample performance. Finally, we discuss how our preferred approach, the SMD estimator promoted by Lavergne and Patilea (2013), can be seen as a jackknife version of local GMM. It may allow for a fixed bandwidth point of view that is well-suited to address the curse of dimensionality for kernel smoothing.

5.1 Estimation based on Nonparametric Estimation of Managed Portfolios

The characterization of the pseudo-true value θ^* as a fixed point may suggest a GMM-based estimation of θ^* through an iterated sequence of minimization problems based on the unconditional HJ-distance. When the contraction mapping condition (4.10) is fulfilled, it should ensure the convergence of an iterated sequence of GMM estimators based on managed portfolios, corresponding to the sample counterpart of the iterations defined by:

$$e_Z\left[\theta^k, \theta^{k+1}\right] = 0.$$

In particular, with a number of iterations $k(T)$ going to infinity with the sample size T, we should end up with a consistent estimator of the pseudo-true value θ^*. The asymptotic theory of such an estimator could be easily derived by applying the general asymptotic theory in Pastorello, Patilea, and Renault (2003) and its semi-parametric extension in Frazier (2018).

Note that the latter reference is especially relevant since the sample counterpart of the above iterations leads us to solve for each iteration $k = 1, 2, \ldots, k(T)$ the equations (with unknown θ):

$$\sum_{t=1}^{T} \hat{Z}_{t,T}\left(\theta^{(k)}\right)[m[I(t+1),\theta]R_{t+1} - 1_n] = 0, \tag{5.1}$$

where $\hat{Z}_{t,T}(\theta)$ is a nonparametric estimator of the managed portfolio $Z_t(\theta)$ defined in (4.6). As in Frazier (2018), the iterative approach allows us to keep solving equations for an unknown θ which does not enter nonparametric estimators like $\hat{Z}_{t,T}(\theta)$, but only well-behaved functions of θ given by the SDF $m[I(t+1),\theta]$.

More precisely, if we use kernel estimators, we will need to assume that there exists a vector X_t of m state variables such that:

$$\begin{aligned} \Omega[I(t)] = E\left[R_{t+1}R'_{t+1}|I(t)\right] &\approx E\left[R_{t+1}R'_{t+1}|X_t\right] \\ E\left[\frac{\partial \Psi'_{t+1}\left(\theta^{(k)}\right)}{\partial \theta}|I(t)\right] &\approx E\left[\frac{\partial \Psi'_{t+1}\left(\theta^{(k)}\right)}{\partial \theta}|X_t\right] \end{aligned} \tag{5.2}$$

such that we can compute Nadaraya–Watson kernel regression estimators for each given value $\theta^{(k)}$ of θ along the iteration sequence:

$$\begin{aligned} \hat{\Omega}_{t,T} &= \sum_{s=1}^{T} \omega[X_t, X_s, h_T]R_{s+1}R'_{s+1} \\ \hat{E}_{t,T}\left[\frac{\partial \Psi'_{t+1}\left(\theta^{(k)}\right)}{\partial \theta}|I(t)\right] &= \sum_{s=1}^{T} \omega[X_t, X_s, h_T]\frac{\partial \Psi'_{t+1}\left(\theta^{(k)}\right)}{\partial \theta} \\ \hat{Z}_{t,T}\left(\theta^{(k)}\right) &= \hat{E}_{t,T}\left[\frac{\partial \Psi'_{t+1}\left(\theta^{(k)}\right)}{\partial \theta}|I(t)\right]\hat{\Omega}_{t,T}^{-1} \end{aligned}$$

with kernel weights:

$$\omega[x_0, x, h] = \frac{K\left[\frac{x-x_0}{h}\right]}{\sum_{s=1}^{T} K\left[\frac{x_s-x_0}{h}\right]}.$$

Note that we have used the approximation notation $\approx$ in (5.2) to stress that, on the one hand, the curse of dimensionality for nonparametric estimation may prevent us from eliciting a vector X of state variables large enough to capture all the relevant conditioning information $I(t)$, and that, on the other hand, we are not keen on restricting the information set $I(t)$ since this would also modify the definition of the pseudo-true value. In other words, we will have to live with the fact that our nonparametric estimator may not estimate exactly the conditional expectation of interest. Moreover, it is worth keeping in mind that, in case one would be afraid that the above iterative procedure would not properly converge (because the contraction mapping property may not be warranted), it is also possible to perform a one-step GMM in the spirit of Nagel and Singleton (2011), even though we maintain the difference of using the arguably more robust HJ-distance. Our GMM estimator $\hat{\theta}_T$ would then be the solution of the equations:

$$\sum_{t=1}^{T} \hat{Z}_{t,T}\left(\hat{\theta}_T\right)\left[m\left[I(t+1),\hat{\theta}_T\right]R_{t+1}-1_n\right]=0. \tag{5.3}$$

In the one-step GMM form (5.3), one can clearly figure out the challenge of the estimation of a pseudo-true value that entails nonparametric estimation of managed portfolios. As Nagel and Singleton (2011) have rightly noted, if z_t stands for "the conditioning variable [...] that appears in the pricing kernel," then "the dependence of the SDF weights on z_t means that, if these models are correctly specified, conditional moments of returns and consumption are likely to vary with z_t." This remark is even more relevant in the case of a misspecified asset pricing model, since the mere interest of the pseudo-true value is to capture that asset prices are satisfactorily explained by a SDF $m[I(t+1),\theta^*]$. If the conditioning variables X_t used for kernel smoothing do not properly capture the relevant dynamics of $m[I(t+1),\theta]$ through z_{t+1}, one may be afraid that the pseudo-true value will not play its expected role. To see that, note that the GMM first-order conditions (5.3) stipulate that the "local pricing errors" (local alphas) should be orthogonal to the (estimated) "local betas." When the latter coefficients are estimated with an overly parsimonious vector of state variables (the extreme case being Nagel and Singleton (2011) who claim to restrict themselves to just one conditioning variable), one may be afraid that the elicitation of the pseudo true value may be more based on restoring the too poor dynamics of conditional betas than on minimizing pricing errors. The orthogonality relationship (5.3) may have a distorted interpretation if, for all possible value of θ, $\hat{Z}_{t,T}(\theta)$ cannot efficiently track the dynamics of $m[I(t+1),\theta]R_{t+1}$ because it depends on a much smaller vector X_t of state variables.

5.2 Local GMM

Lewbel (2007) has arguably been the first to propose an estimation strategy dubbed local GMM. Fang, Ren, and Yuan (2011) have applied this method for a nonparametric estimation of the pricing kernel. The idea is to see the SDF parameters as functions of the state variables. Gagliardini, Gourieroux, and Renault (2011) have extended this idea by allowing the identification of some fixed parameters jointly with some functional ones.

We rather use here the terminology "local GMM" in the sense of Gospodinov and Otsu (2012) who revisit in a time series context an estimator initially introduced by Antoine, Bonnal, and Renault (2007) and Smith (2007) as a byproduct of an Euclidean empirical likelihood approach, following the empirical likelihood based inference in conditional moment restrictions models of Kitamura, Tripathi, and Ahn (2004). Ai and Chen (2003, 2007) have introduced a very similar estimator that they have also revisited with misspecified conditional moment restrictions, albeit only in an i.i.d. setting. In all of these cases, the local GMM estimator is defined as minimizer of:

$$\hat{\delta}_T^2(\theta)=\frac{1}{T}\sum_{t=1}^{T} 1_{t,T}\hat{e}[I(t),\theta]'\hat{\Omega}^{-1}[I(t)]\hat{e}[I(t),\theta] \tag{5.4}$$

when we assume that:

$$e[I(t),\theta]=E[\Psi_{t+1}(\theta)|I(t)]\approx E[\Psi_{t+1}(\theta)|X_t].$$

Gagliardini and Ronchetti (2016) note that (5.4) is nothing but the sample counterpart of the conditional HJ-distance (3.3), where $\hat{e}[I(t),\theta]$ stands for a nonparametric estimator of $E[\Psi_{t+1}(\theta)|X_t]$, and $1_{t,T}$ is a trimming term to discard observations for which the nonparametric estimator of the marginal density of X_t is overly small (below some well chosen threshold c_T). In this article, we also follow Gagliardini and Ronchetti (2016) by only considering kernel smoothing:

$$\hat{e}[X_t,\theta]=\sum_{s=1}^{T}\omega[X_t,X_s,h_T]\Psi_{s+1}(\theta).$$

We note that Ai and Chen (2003, 2007) rather promote sieve estimation, and therefore call the minimization of (5.4) a sieve minimum distance estimation.

It is then worth comparing the first-order conditions that characterize the estimator $\hat{\theta}_T^*$ of the pseudo-true value obtained by minimization of (5.4) with the first-order conditions (5.3) that had characterized the GMM estimator $\hat{\theta}_T$ associated with the optimal managed portfolios $\hat{Z}_{t,T}\left(\hat{\theta}_T\right)$. While (5.3) can be rewritten:

$$\sum_{t=1}^{T}\frac{\partial\hat{e}'\left[X_t,\hat{\theta}_T\right]}{\partial\theta}\hat{\Omega}_{t,T}^{-1}\Psi_{t+1}\left(\hat{\theta}_T\right)=0 \tag{5.5}$$

we now characterize $\hat{\theta}_T^*$ as solution of the first-order conditions associated with the minimization of (5.4):

$$\sum_{t=1}^{T}1_{t,T}\frac{\partial\hat{e}'\left[X_t,\hat{\theta}_T^*\right]}{\partial\theta}\hat{\Omega}_{t,T}^{-1}\hat{e}'\left[X_t,\hat{\theta}_T^*\right]=0. \tag{5.6}$$

Up to trimming, the only difference between (5.5) and (5.6) is the replacement of the sample counterpart of:

$$\Psi_{t+1}(\theta)=m[I(t+1),\theta]R_{t+1}-1_n$$

by a kernel estimator of:

$$e[X_t,\theta]=E[\Psi_{t+1}(\theta)|X_t].$$

In a context where we cannot afford to run nonparametric regressions on a set X_t of regressors as large as the relevant conditional information set $I(t)$, which should include all pricing factors and all asset prices as well, one may intuitively consider that the latter approach (5.6) is safer than the former one (5.5). As already explained, one may be afraid that a pseudo-true value computed from the population analog of (5.5) would be more focused on tracking the rich dynamics of the vector $m[I(t+1),\theta]R_{t+1}$ through a reduced set X_t of state variables present in the managed portfolios (or equivalently in $\sum_{t=1}^{T}\frac{\partial\hat{e}'\left[X_t,\hat{\theta}_T^*\right]}{\partial\theta}\hat{\Omega}_{t,T}^{-1}$) than on minimizing the pricing errors. At the least, (5.6) has the advantage to treat symmetrically the information sets included on both sides of the orthogonality relationship.

This intuition may lead to prefer, especially to deal with misspecified conditional asset pricing models, local GMM as Gagliardini and Ronchetti (2016) do, instead of GMM with nonparametric estimation of instruments as in Nagel and Singleton (2011). In particular, Gagliardini and Ronchetti (2016) extend the asymptotic distributional theory of GMM with misspecification that had been developed in the unconditional case by Hall and Inoue (2003). However, the price to pay for this asymptotic theory is the need to resort to martingale difference sequences that are not provided, in case of misspecification, by the conditional moment restrictions themselves. They need to basically assume (see their proof of Proposition 5) that differences:

$$\Psi_{t+1}(\theta) - e[X_t, \theta]$$

provide a martingale difference sequence w.r.t the process $\{X_t\}$ for any θ, meaning that $m[I(t+1), \theta]R_{t+1}$ is $\sigma[X_s, s \leq t+1]$ measurable for all θ. This is likely to imply that all the pricing factors in $m[I(t+1), \theta]$ and all the returns components of R_{t+1} are themselves $\sigma[X_s, s \leq t+1]$ measurable. As explained above, it is an assumption that we do not want to maintain due to the curse of dimensionality in nonparametric regression. This is one of the main motivations for the introduction below of the alternative SMD approach.

5.3 A Jackknife GMM Alternative to Local GMM

While local GMM as defined by (5.4) amounts to minimizing a direct sample counterpart of the conditional HJ-distance:

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

where:

$$e[I(t), \theta] = E[\Psi_{t+1}(\theta)|I(t)].$$

It is worth noting that we can resort less to kernel smoothing by, before computing sample counterparts, using as follows a square root matrix of $\Omega^{-1}[I(t)]$ and the law of iterated expectations:

$$\begin{aligned}\delta^2(\theta) &= E\{E[B(X_t)\Psi_{t+1}(\theta)|I(t)]'E[B(X_t)\Psi_{t+1}(\theta)|I(t)]\} \\ &= E\{\Psi'_{t+1}(\theta)B'(X_t)B(X_t)E[\Psi_{t+1}(\theta)|I(t)]\} \\ &= E\{\Phi_{t+1}(\theta)'E[\Phi_{t+1}(\theta)|I(t)]\},\end{aligned}$$

where we maintain the Assumption (5.2):

$$\Omega[I(t)] \approx E[R_{t+1}R'_{t+1}|X_t]$$

and define accordingly a square root matrix $B(X_t)$ and associated sphericized moment functions $\Phi_{t+1}(\theta)$ by:

$$\begin{aligned}\Omega^{-1}[I(t)] &= B(X_t)'B(X_t) \\ \Phi_{t+1}(\theta) &= B(X_t)\Psi_{t+1}(\theta).\end{aligned} \tag{5.7}$$

Using this new version of the conditional HJ-distance, we can revisit the local GMM estimator as minimizing over θ:

$$\frac{1}{T}\sum_{t=1}^{T} 1_{t,T}\Phi_{t+1}(\theta)'\hat{E}_{t,T}[\Phi_{t+1}(\theta)|X_t], \tag{5.8}$$

where the notation $\hat{E}_{t,T}[\Phi_{t+1}(\theta)|X_t]$ stands for a kernel counterpart of the conditional expectation $E[\Phi_{t+1}(\theta)|I(t)]$. Note that in practice the sphericization (5.7) may also take a kernel counterpart of the matrix $\Omega[I(t)]$ before computing a square root $B(X_t)$ of its inverse. Recall that in (5.4) and its modification (5.8) $1_{t,T}$ is a trimming term to discard observations for which the nonparametric estimator $\hat{f}_T(X_t)$ of the marginal density of X_t is overly small (below some well-chosen threshold c_T). We refer to Zheng (1996) to note that "as in Powell, Stock, and Stoker (1989) the inclusion of the density function avoids the problem of trimming the small values of the density function." In other words, a sensible smoother alternative to the trimming terms $1_{t,T}$ might be to weight the various terms of the sum by the estimated density itself:

$$SM_T^*(\theta) = \frac{1}{T}\sum_{t=1}^{T}\hat{f}_T(X_t)\Phi_{t+1}(\theta)'\hat{E}_{t,T}[\Phi_{t+1}(\theta)|X_t]$$

$$\hat{f}_T(X_t) = \frac{1}{(T-1)h_T^m}\sum_{1\le s\le T, s\ne t} K\left[\frac{X_s - x_0}{h_T}\right].$$

Given our "externalized" (Jones, Davies, and Park, 1994) version of the Nadaraya–Watson estimator of the conditional expectation, we have:

$$\hat{f}_T(X_t)\hat{E}_{t,T}[\Phi_{t+1}(\theta)|X_t] = \frac{1}{(T-1)h_T^m}\sum_{1\le s\le T, s\ne t} K\left[\frac{X_s - X_t}{h_T}\right]\Phi_{s+1}(\theta)$$

so that:

$$SM_T^*(\theta) = \frac{1}{T(T-1)}\frac{1}{h_T^m}\sum_{t=1}^{T}\sum_{s=1, s\ne t}^{T}\Phi_{t+1}(\theta)'\Phi_{s+1}(\theta)K\left[\frac{X_s - X_t}{h_T}\right].$$

This approach can actually be seen as a generalization of the test statistic proposed independently by Fan and Li (1996) and Zheng (1996) for testing the specification of regression functions. Moreover, Fan and Li (2000) complete this approach by addressing the issue of "fixed bandwidth vs. vanishing bandwidth." More generally, the focus of our interest in the next sections will be a possibly bandwidth dependent pseudo-true value $\theta^*(h)$ whose estimator, the SMD estimator promoted by Lavergne and Patilea (2013), is defined by:

$$\hat{\theta}_T(h) = \arg\min_{\theta\in\Theta} \frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{s=1, s\ne t}^{T}[B(X_s)\Psi_{s+1}(\theta)]'[B(X_t)\Psi_{t+1}(\theta)]K\left[\frac{X_s - X_t}{h}\right]. \quad (5.9)$$

Note that the sphericization matrix $B(X_t)$ itself can also be seen as resulting of a choice of a fixed bandwidth in kernel smoothing of $\Omega[I(t)]$. This additional fixed bandwidth parameter is not made explicit for notational simplicity. While our focus of interest is estimation of a pseudo true value, Peñaranda, Rodríguez-Poo, and Sperlich (2017) have recently led a similar discussion for the purpose of nonparametric specification testing of conditional asset pricing models.They first put forward an integral of squared sphericized pricing errors over the possible values of the conditioning variable. Interestingly enough, their suggestion of weighting this squared pricing errors by the value of the probability density of the conditioning variable exactly corresponds to our conditional HJ-distance. In other words, the sample counterpart of this integral coincides with our local GMM (5.4) distance. Moreover, they also note that it is worth downplaying the role of pricing errors "when the data are sparse," such that they re-weight the kernel estimation of the pricing error by the kernel estimator of probability density of the conditioning variable, leading to our discrepancy measure (5.9). They put forward "an automatic procedure for maximizing the power" of the test, leading to see the bandwidth parameter h as a "calibration parameter." This is conformable to our strategy of looking at a pseudo true value $p\lim_{T=\infty} \hat{\theta}_T(h)$ for different possible values of this "calibration parameter" h. In the particular case of linear factor pricing models, our approach can also be related to the work of Roussanov (2014) and Wang (2003).

We will explain in the next two subsections below why this bandwidth dependent pseudo-true value is worth considering. We already note that the new discrepancy function (5.9) can be related to local GMM in a way similar to the extension of GMM dubbed Jackknife GMM by Newey and Windmeijer (2009). To see that, let us consider some unconditional moment restrictions:

$$E[g(Y,\theta)] = 0 \tag{5.10}$$

For a given positive definite matrix W, a GMM estimator is defined as:

$$\hat{\theta}_T = \arg\min_{\theta\in\Theta} Q_T(\theta),$$

where:

$$Q_T(\theta) = \overline{g}_T(\theta)' W \overline{g}_T(\theta)$$
$$\overline{g}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} g(Y_t,\theta).$$

The logic of jackknife GMM can be described from the following decomposition of the GMM criterion function:

$$Q_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left\{g(Y_t,\theta)' W \overline{g}_T(\theta)\right\}.$$

Consistent with the leave-one-out logic, it is natural to consider that in the t^{th} term of this sum the expectation $E[g(Y,\theta)]$ should rather be estimated by:

$$\overline{g}_T^{(t)}(\theta) = \frac{1}{T-1}\sum_{s=1,s\neq t}^{T} g(Y_s,\theta)$$

leading to a modification of the objective function as follows:

$$\tilde{Q}_T(\theta) = \frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\left\{g(Y_t,\theta)' W g(Y_s,\theta)\right\}. \tag{5.11}$$

Following Newey and Windmeijer (2009), an estimator of θ computed as minimizer of $\tilde{Q}_T(\theta)$ will be dubbed a jackknife GMM estimator. As explained by Newey and Windmeijer (2009), this estimator is a generalization of the second jackknife instrumental variables estimator (JIVE2) of Angrist, Imbens, and Krueger (1999) in order to allow for a general weighting matrix W. The analogy between the discrepancy measure (5.11) and (5.9) is obvious, up to the fact that in the context of conditional moment restrictions, naive empirical weights $(1/T)$ must be replaced by kernel weights (see, e.g., Antoine, Bonnal, and Renault, 2007).

This jackknife principle is known to be useful in the context of weak identification. We will show in Section 7 that it is also helpful to deal with the asymptotic distributional theory of a GMM estimator based on misspecified moment conditions. We will develop an asymptotic distributional theory that revisits the theory of Hall and Inoue (2003) in the case (5.10) of unconditional moment restrictions and replaces the theory of local GMM in the context of SMD estimation (5.9) based on misspecified conditional moment restrictions.

6 Bandwidth Dependent Pseudo-True SDF

We provide in the next two subsections an interpretation of the fixed bandwidth pseudo-true value in the i.i.d. case and in the context of a state variable framework. Following Antoine and Lavergne (2014), we also discuss in the last subsection how the SMD estimator can be modified to deal with weak identification contexts in the i.i.d. case.

6.1 Interpretation in the i.i.d. Case

Even though the i.i.d. setting is not our focus of interest in this article, it is worth considering it now to get more intuition about the interpretation of the fixed bandwidth pseudo-true value defined in the former subsection. To do so, let us for the moment replace the conditional moment restrictions of interest:

$$E[\Phi_{t+1}(\theta)|X_t] = 0$$

by the following:

$$E[g(Y_i,\theta)|X_i] = 0 \tag{6.1}$$

when we observe some i.i.d. sample $(Y_l, X_l), l = 1, 2...$ and $g(y,\theta) = \left(g^{(k)}(y,\theta)\right)_{1\leq k\leq n}$ is a vector of n moment functions. Assume that the kernel function $K(.)$ is such that:

$$K(u) = \int_{\mathbb{R}^m} \exp\,(i\zeta' u)d\mu(\zeta)$$

for some measure $\mu(.)$ that is strictly positive (except possibly for a set of isolated points). This condition is not very restrictive for kernel functions based on products of univariate probability density functions (see, e.g., Lavergne and Patilea (2013) for a discussion). Then, if we compute the analog of the quantity (5.9), assuming $h=1$ without loss of generality:

$$\begin{aligned}
&E\{g'(Y_l,\theta)g(Y_j,\theta)K(X_l-X_j)\}\\
&= E\left\{g'(Y_l,\theta)g(Y_j,\theta)\int_{\mathbb{R}^m}\exp\,[i\zeta'(X_l-X_j)]\,d\mu(\zeta)\right\}\\
&= \sum_{k=1}^{n}\int_{\mathbb{R}^m} E\{g^{(k)}(Y_l,\theta)g^{(k)}(Y_j,\theta)\exp\,[i\zeta'(X_l-X_j)]\}d\mu(\zeta)\\
&= \sum_{k=1}^{n}\int_{\mathbb{R}^m} E\{g^{(k)}(Y_l,\theta)\exp\,[i\zeta'X_l]\}E\{g^{(k)}(Y_j,\theta)\exp\,[-i\zeta'X_j]\}d\mu(\zeta)\\
&= \sum_{k=1}^{n}\int_{\mathbb{R}^m} |E\left\{g^{(k)}(Y_l,\theta)\exp\,[i\zeta'X_l]\right\}|^2 d\mu(\zeta).
\end{aligned} \tag{6.2}$$

Therefore, we see that (6.2) can be zero if and only if:

$$E\{g(Y_l,\theta)\exp\,[i\zeta' X_l]\} = 0, \forall \zeta \in \mathbb{R}^m. \tag{6.3}$$

However, it has been known at least since Bierens (1982) that the continuum of unconditional restrictions (6.3) is equivalent to the set (6.1) of conditional moment restrictions. The logical equivalence between (6.2) and (6.3) has first been pointed out by Fan and Li (2000) in the context of specification testing. As far as estimation is concerned, it says that when (6.1) is well-specified and identifies a unique true value θ^0, the nullity of (6.2) identifies the same true value. It is then natural to define a pseudo-true value by minimization of (6.2) or, in our context, minimization of (5.9). This argument does not extend directly to general time series models. However, we can also produce an additional interpretation that we will eventually be able to extend to time series context with well suited state variables.

The key idea is to use the law of iterated expectations to rewrite (6.2) as follows:

$$E\{g'(Y_l,\theta)g(Y_j,\theta)K(X_l - X_j)\} = E\{E[g'(Y_l,\theta)g(Y_j,\theta)|X_l,X_j]K(X_l - X_j)\}.$$

However, the fact that (Y_l, X_l) and (Y_j, X_j) are independent implies that:

$$E[g'(Y_l,\theta)g(Y_j,\theta)|X_l,X_j] = E[g'(Y_l,\theta)|X_l]E[g(Y_j,\theta)|X_j].$$

To see this, it is convenient to assume that (Y_l, X_l) admits a density with respect to a product probability measure. Elementary probability computations then allow us to show that:

$$f(y_l, y_j|x_l, x_j) = f(y_l|x_l)f(y_j|x_j)$$

which implies the announced factorization of conditional expectations.With a nonspecified bandwidth parameter h, we have for the same reason:

$$\begin{aligned} &E\left\{g'(Y_l,\theta)g(Y_j,\theta)K\left(\frac{X_l - X_j}{h}\right)\right\} \\ &= E\left\{E[g'(Y_l,\theta)|X_l]E[g(Y_j,\theta)|X_j]K\left(\frac{X_l - X_j}{h}\right)\right\}, \end{aligned} \tag{6.4}$$

Therefore, from (6.4) we understand that, at least for a sufficiently small bandwidth parameter h, minimizing this quantity amounts to searching for a value $\theta^*(h)$ of θ that comes as close as possible to fulfilling the conditional moment restrictions (6.1).

6.2 Interpretation through State Variables

The goal of this subsection is to show that a natural state variables setting allows us to handle our minimization of interest, namely (5.9), in a way quite similar to the well-documented i.i.d. case. We start by applying the law of iterated expectations to the quantity in (5.9), rewritten as follows:

$$\begin{aligned} &E\left[(B(X_s)\Psi_{s+1}(\theta))'(B(X_t)\Psi_{t+1}(\theta))K\left(\frac{X_t - X_s}{h}\right)\right] \\ &= E\left\{(B(X_s)\Psi(Y_{s+1},\theta))'(B(X_t)\Psi(Y_{t+1},\theta))K\left(\frac{X_t - X_s}{h}\right)\right\}. \end{aligned}$$

For $s < t$:

$$\Delta_{s,t}(h,\theta) = E\left\{(B(X_s)\Psi(Y_{s+1},\theta))'(B(X_t)\Psi(Y_{t+1},\theta))K\left(\frac{X_t - X_s}{h}\right)\right\}$$
$$= E\left\{(B(X_s)\Psi(Y_{s+1},\theta))'B(X_t)E\left[\Psi(Y_{t+1},\theta)|Y_1^t, X_1^{t+1}\right]K\left(\frac{X_t - X_s}{h}\right)\right\}.$$

A common assumption in asset pricing models (see, e.g., Garcia, Luger, and Renault (2003) and references therein) is that exogenous state variables summarize the dynamics of asset returns and pricing kernels; given the path of state variables (for instance, the path of stochastic volatility), the consecutive asset returns and SDFs are serially conditionally independent:

Assumption A1

$$E\left[\Psi(Y_{t+1},\theta)|Y_1^t, X_1^{t+1}\right] = E\left[\Psi(Y_{t+1},\theta)|X_1^{t+1}\right]$$

Then:

$$\Delta_{s,t}(h,\theta) = E\left\{(B(X_s)\Psi(Y_{s+1},\theta))'B(X_t)E\left[\Psi(Y_{t+1},\theta)|X_1^{t+1}\right]K\left(\frac{X_t - X_s}{h}\right)\right\}$$
$$= E\left\{E\left[\Psi'(Y_{s+1},\theta)|X_1^{t+1}\right]B'(X_s)B(X_t)E\left[\Psi(Y_{t+1},\theta)|X_1^{t+1}\right]K\left(\frac{X_t - X_s}{h}\right)\right\}.$$

A second common assumption (see again the discussion in Garcia, Luger, and Renault (2003)) is that state variables are strictly exogenous in the sense that they are not caused by the time series of asset returns and SDFs. Writing this noncausality assumption in the Sims way allows us to write:

Assumption A2

$$s < t \Rightarrow E\left[\Psi'(Y_{s+1},\theta)|X_1^{t+1}\right] = E\left[\Psi'(Y_{s+1},\theta)|X_1^{s+1}\right]$$

Thus, under assumptions A1 and A2:

$$\Delta_{s,t}(h,\theta) = E\left\{E\left[\Psi'(Y_{s+1},\theta)|X_1^{s+1}\right]B'(X_s)B(X_t)E\left[\Psi(Y_{t+1},\theta)|X_1^{t+1}\right]K\left(\frac{X_t - X_s}{h}\right)\right\}.$$

An additional assumption, that is more questionable, states that there is no instantaneous causality between asset returns/SDF and state variables. When state variables include stochastic volatility, this would amount to assuming that there is no leverage effect. Garcia, Luger, and Renault (2003) have extensively documented the consequences of this questionable assumption for the purpose of asset pricing. This assumption can be written:

Assumption A3

$$E\left[\Psi(Y_{t+1},\theta)|X_1^{t+1}\right] = E\left[\Psi(Y_{t+1},\theta)|X_1^{t}\right]$$

If in addition we assume that the state variables process fulfills the Markov property (a natural and hardly restrictive assumption) we conclude that under assumptions A1, A2, and A3:

$$\Delta_{s,t}(h,\theta) = E\left\{ E\left[(B(X_s)\Psi(Y_{s+1},\theta))'|X_s\right] E[B(X_t)\Psi(Y_{t+1},\theta)|X_t] K\left(\frac{X_t - X_s}{h}\right)\right\}.$$

As in the i.i.d. case, we can then conclude that, at least for a sufficiently small bandwidth parameter, the minimization of (5.9) amounts to searching for a value $\theta^*(h)$ of θ that comes as close as possible to fulfilling the conditional moment restrictions of the asset pricing model, when the moment restrictions are properly rescaled by the conditional HJ weighting matrix (whose square root is the matrix $B(X_t)$).

6.3 Weak Identification

In the i.i.d. case, Antoine and Lavergne (2014) show that the bandwidth dependent SMD estimator $\hat{\theta}_T(h)$ defined in (5.9) is consistent under semi-strong identification for any chosen fixed bandwidth h. However, as the gradient of the objective function flattens under semi-strong identification, the solution of the first-order conditions can be numerically quite dispersed and unstable in practice. To avoid such a behavior, Antoine and Lavergne (2014) propose instead the following WMD estimator:

$$\hat{\theta}^*_T(h) = \arg\min_{\theta\in\Theta} \,[WM_T(h,\theta)]$$

$$WM_T(h,\theta) = \frac{\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\left[(B(X_s)\Psi_{s+1}(\theta))'(B(X_t)\Psi_{t+1}(\theta))K\left(\frac{X_t - X_s}{h}\right)\right]}{\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\left[(B(X_s)\Psi_{s+1}(\theta))'(B(X_t)\Psi_{t+1}(\theta))\right]}.$$

The associated first-order conditions combine the gradient of the SMD estimator with the one of a least-squares criterion. This second gradient does not flatten, even under semi-strong identification, and thus yields more stability in estimation for small and moderate samples.

In Section 8.2, we present evidence that the factors commonly used in asset pricing models exhibit semi-strong identification strength. We then illustrate the finite sample properties of our inference procedures in a linear asset pricing model with semi-strong factors. In a (univariate) linear model, say $y = Y'\theta + e$, the WMD estimator actually resembles a k-class estimator. Recall that a general k-class estimator is of the form

$$[Y'(A-\lambda I)Y]^{-1}Y'(A-\lambda I)y,$$

where A is a matrix depending on the exogenous variables. Estimators differ in the choice of A and λ: for instance, 2SLS uses $\lambda = 0$ and $A = P$, the projection matrix on the space spanned by the chosen instruments, while the jackknife IV estimator (JIVE) (see Angrist, Imbens, and Krueger, 1999) uses $\lambda = 0$ and $A = P$, the same projection matrix P, but with its diagonal elements are set to zero; limited information maximum likelihood (LIML) corresponds to $A = P$ and λ equal to the smallest eigenvalue of $(Y'Y)^{-1}(Y'PY)$, while the jackknife LIML, labeled as HLIM by Hausman et al. (2012), corresponds to $A = P$ and λ equal to the smallest eigenvalue of $(Y'Y)^{-1}(Y'PY)$.

Under strong identification, 2SLS (respectively, JIVE) and LIML (respectively, HLIM) are asymptotically equivalent, while they are generally not under semi-strong identification: 2SLS can be inconsistent (see Chao and Swanson, 2005) or asymptotically inefficient relative to LIML (see Hausman et al., 2012) because the correction λ can yield a lower variance. But, λ also depends on the number of instruments, because it involves a projection matrix whose rank is the number of instruments. This yields an interplay between the number of instruments and their strength. In contrast, the WMD estimator does not use projection on instruments, so there is no user-chosen parameter that affects its consistency or asymptotic variance. Nonetheless, it retains the computational simplicity of k-class estimators. An extension to weak identification frameworks as considered in Antoine and Lavergne (2018) are beyond the scope of this article.

7 Asymptotic Distributional Theory for the Bandwidth Dependent Pseudo-True SDF

The asymptotic theory for GMM estimators based either on nonparametric estimators of managed portfolios or on local GMM is well-documented in the case of well-specified moment conditions. The extension of this theory to misspecified moments is problematic. To see that, it is worth referring to the seminal work of Hall and Inoue (2003) in the case of misspecified unconditional moment restrictions.

The main message of Hall and Inoue (2003) is that in the case of misspecification, the estimation errors on the Jacobian matrix of the moment conditions and on the weighting matrix have an impact on the asymptotic distribution of the GMM estimator of the pseudo-true value. Moreover, when this weighting matrix is estimated at a nonparametric rate, this nonparametric rate will contaminate the estimation of the pseudo-true value.

This message is ominous regarding the impact of misspecification when either managed portfolios are estimated at a nonparametric rate or when local GMM leads one to estimate the moment conditions themselves at a nonparametric rate. As already mentioned, the assumption of martingale difference sequence that Gagliardini and Ronchetti (2016) need to maintain is rather unrealistic. To escape this quandary, we develop in this section the asymptotic theory of the SMD estimator of this pseudo-true value computed with a fixed bandwidth. We settle this theory through a jackknife alternative of Hall and Inoue's (2003) asymptotic theory of GMM with misspecification.

7.1 Why a Jackknife Alternative?

In this subsection, we provide some intuition for why a jackknife version of GMM is especially well-suited in the case of misspecification. We develop this intuition in the simplest framework of some unconditional moment restrictions:

$$E[g(Y,\theta)] = 0$$

based on a sequence of i.i.d. variables $Y_t, t = 1, \ldots, T$ for inference about a vector $\theta \in \Theta \subset \mathbb{R}^p$ of unknown parameters.

For a given positive definite matrix W, a GMM estimator of θ is defined as:

$$\hat{\theta}_T = \arg\min_{\theta\in\Theta} Q_T(\theta),$$

where:

$$Q_T(\theta) = \bar{g}_T(\theta)' W \bar{g}_T(\theta).$$

We maintain the assumption of laws of large numbers that are uniform over the parameter space Θ:

$$P \lim_{T=\infty} Q_T(\theta) = Q_\infty(\theta) = E[g(Y,\theta)]' W E[g(Y,\theta)].$$

We also assume that there is a unique pseudo-true value θ^* defined as solution of the population minimization problem:

$$Q_\infty(\theta^*) < Q_\infty(\theta), \forall \theta \in \Theta - \{\theta^*\}.$$

Under standard regularity conditions (compactness of Θ and continuity of the function $g(Y,.)$), the GMM estimator $\hat{\theta}_T$ is a consistent estimator of θ^*. Moreover, θ^* is assumed to be an interior point of Θ, and a solution of the first-order conditions:

$$\frac{\partial Q_\infty(\theta^*)}{\partial \theta} = 2E\left[\frac{\partial g(Y,\theta^*)'}{\partial \theta}\right] W E[g(Y,\theta^*)] = 0. \tag{7.1}$$

The jackknife GMM alternative to the GMM estimator defined above is motivated by the following decomposition of the quadratic form:

$$\begin{aligned} Q_T(\theta) &= Q_T^*(\theta) + \tilde{Q}_T(\theta) \\ Q_T^*(\theta) &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T} g(Y_t,\theta)' W g(Y_s,\theta) \\ \tilde{Q}_T(\theta) &= \frac{1}{T^2}\sum_{t=1}^{T} g(Y_t,\theta)' W g(Y_t,\theta). \end{aligned}$$

Obviously, we expect that uniformly over Θ:

$$P \lim_{T=\infty} Q_T^*(\theta) = E\big[g(Y_t,\theta)' W g(Y_s,\theta)\big] = E[g(Y,\theta)]' W E[g(Y,\theta)] = Q_\infty(\theta) \tag{7.2}$$

while:

$$\tilde{Q}_T(\theta) \sim \frac{1}{T} E\big[g(Y_t,\theta)' W g(Y_t,\theta)\big] = O_P\left(\frac{1}{T}\right). \tag{7.3}$$

Therefore, we have an alternative consistent estimator θ_T^* of θ^* by defining θ_T^* as solution of the minimization program:

$$\theta_T^* = \arg \min_{\theta\in\Theta} Q_T^*(\theta).$$

Following Newey and Windmeijer (2009), θ_T^* will be dubbed a jackknife GMM estimator. We naturally assume that the estimator θ_T^* is characterized (for T sufficiently large) as solution of the first-order conditions:

$$\frac{\partial Q_T^*(\theta_T^*)}{\partial \theta} = 0.$$

With a standard Taylor expansion, we deduce:

$$\sqrt{T}\frac{\partial Q_T^*(\theta^*)}{\partial\theta}+\frac{\partial^2 Q_T^*(\theta^*)}{\partial\theta\partial\theta'}\sqrt{T}\left[\theta_T^*-\theta^*\right]=o_P(1). \tag{7.4}$$

As usual, the estimator θ_T^* will be asymptotically normal as a linear function of the score vector:

$$\begin{aligned}\sqrt{T}\frac{\partial Q_T^*(\theta^*)}{\partial\theta}&=\frac{\sqrt{T}}{T^2}\sum_{1\leq t\neq s\leq T}\left[\frac{\partial g(Y_t,\theta^*)'}{\partial\theta}Wg(Y_s,\theta^*)+\frac{\partial g(Y_s,\theta^*)'}{\partial\theta}Wg(Y_t,\theta^*)\right]\\&=\frac{\sqrt{T}}{T^2}\sum_{1\leq t\neq s\leq T}U(Y_s,Y_t),\end{aligned}$$

where $U(Y_s, Y_t)$ has a zero expectation by virtue of (7.1).

This expression of the score vector displays the advantage of the jackknife approach: the asymptotic normal distribution of the GMM estimator will involve, in a symmetric way, the asymptotic normality of sample means of the moment conditions, as well as the asymptotic normality of sample means of the Jacobian matrix of these moments. This is the key difference with respect to the standard asymptotic theory of GMM in the case of well-specified moments: only the asymptotically normal distribution of sample means of the moment conditions matters, while the sample means of the Jacobian matrix can be replaced by their population values. Hall and Inoue (2003), in revisiting the standard way to derive the asymptotic distribution of GMM, point out that the need to demean misspecified moment conditions by their population expectation (computed at the pseudo-true value θ^*) implies the nonstandard impact of the asymptotic normal distribution of sample means of the Jacobian matrix of moments on the asymptotic distribution of GMM. We will see in the next subsection that the argument is much more transparent when based on sample means of the symmetric function $U(Y_s, Y_t)$. It appears as a direct corollary of classical results for the asymptotic normality of U-statistics of order 2.

7.2 Jackknife GMM and U-Statistics

As shown above:

$$\frac{\partial Q_T^*(\theta^*)}{\partial\theta}=\frac{T-1}{T}\overline{U}_T,$$

where $\overline{U}_T$ is the U-statistic of order 2:

$$\overline{U}_T=\frac{1}{T(T-1)}\sum_{1\leq t\neq s\leq T}U(Y_t,Y_s) \tag{7.5}$$

with the symmetric function $U(Y_t, Y_s)$:

$$U(Y_t,Y_s)=\frac{\partial g(Y_t,\theta^*)'}{\partial\theta}Wg(Y_s,\theta^*)+\frac{\partial g(Y_s,\theta^*)'}{\partial\theta}Wg(Y_t,\theta^*).$$

Assuming that $U(Y_t, Y_s)$ is square-integrable, $\overline{U}_T$ is a U-statistic with zero mean that is nondegenerate insofar as we have a nonzero Hájek projection by computing:

$$\varphi(Y_t) = E[U(Y_t, Y_s)|Y_t] = \frac{\partial g(Y_t, \theta^*)'}{\partial \theta} W\mu^* + G(\theta^*)' W g(Y_t, \theta^*), \tag{7.6}$$

where the above expression has been obtained by using the independence of Y_t and Y_s and the notations:

$$\mu^* = E[g(Y_t, \theta^*)], G(\theta) = E\left[\frac{\partial g(Y_t, \theta)}{\partial \theta'}\right].$$

The Hájek projection (7.6) is obviously nonzero under standard GMM assumptions; namely that $G(\theta^*)$ is of rank p and $S^* = \text{Var}[g(Y_t, \theta^*)]$ is nonsingular. Then, by virtue of the Central Limit Theorem for nondegenerate U-statistics, we know that $\sqrt{T}\overline{U}_T$ is asymptotically equivalent to the asymptotically normal vector:

$$\frac{2}{\sqrt{T}} \sum_{t=1}^{T} \varphi(Y_t).$$

Therefore, $\sqrt{T}\overline{U}_T$ is asymptotically normal with mean zero and variance $4\Sigma^*$, where:

$$\begin{aligned}\Sigma^* = \text{Var}[\varphi(Y_t)] &= \text{Var}\left[G(\theta^*)' W g(Y_t, \theta^*)\right] + \text{Var}\left[\frac{\partial g(Y_t, \theta^*)'}{\partial \theta} W\mu^*\right] \\ &+ \text{Cov}\left[\frac{\partial g(Y_t, \theta^*)'}{\partial \theta} W\mu^*, G(\theta^*)' W g(Y_t, \theta^*)\right] + \text{Cov}\left[G(\theta^*)' W g(Y_t, \theta^*), \frac{\partial g(Y_t, \theta^*)'}{\partial \theta} W\mu\right].\end{aligned}$$

Note that, since all the terms have a zero expectation by virtue of (7.1), all the above variances and covariances are actually expectations of products. Following Hall and Inoue (2003), it is worth introducing the following notations:

$$\begin{aligned}\text{Var}\left[\frac{\partial g(Y_t, \theta^*)'}{\partial \theta} W\mu\right] &= \Omega_{22}(\mu) \\ \text{Cov}\left[\frac{\partial g(Y_t, \theta^*)'}{\partial \theta} W\mu, g(Y_t, \theta^*)\right] &= \Omega_{21}(\mu) \\ \text{Cov}\left[g(Y_t, \theta^*), \frac{\partial g(Y_t, \theta^*)'}{\partial \theta} W\mu\right] &= \Omega_{12}(\mu) = \Omega_{21}'(\mu),\end{aligned}$$

where, in particular:

$$\Omega_{22}(0) = 0, \Omega_{12}(0) = 0 = \Omega_{21}'(0).$$

Then:

$$Var[\varphi(Y_t)] = G(\theta^*)' W S^* W G(\theta^*) + \Omega_{22}(\mu^*) + \Omega_{21}(\mu^*) W G(\theta^*) + G(\theta^*)' W \Omega_{12}(\mu^*).$$

From (7.4), we get that $\sqrt{T}\left[\theta_T^* - \theta^*\right]$ is asymptotically normal with asymptotic variance:

$$\left[\frac{\partial^2 Q_\infty^*(\theta^*)}{\partial \theta \partial \theta'}\right]^{-1} \Sigma^* \left[\frac{\partial^2 Q_\infty^*(\theta^*)}{\partial \theta \partial \theta'}\right]^{-1}$$

with:

$$Q^*_\infty(\theta) = p\lim_{T=\infty} Q^*_T(\theta) = p\lim_{T=\infty} Q_T(\theta) = Q_\infty(\theta),$$

where the second equality comes from (7.3). We then deduce that under standard regularity conditions:

$$\frac{\partial^2 Q_\infty(\theta)}{\partial\theta\partial\theta'} = \frac{\partial}{\partial\theta}\left\{2E\left[g'(Y,\theta)WE\left[\frac{\partial g'(Y_t,\theta)}{\partial\theta}\right]\right]\right\}$$

leading to:

$$\begin{aligned} H^* &= \frac{1}{2}\frac{\partial^2 Q_\infty(\theta^*)}{\partial\theta\partial\theta'} \\ &= G(\theta^*)'WG(\theta^*) + \left(\mu^{*\prime}W \otimes Id_p\right)(\Delta G)(\theta^*) \\ (\Delta G)(\theta) &= E\left\{\frac{\partial}{\partial\theta'}vec\left[\frac{\partial g(Y_t,\theta)}{\partial\theta'}\right]\right\}. \end{aligned}$$

Hall and Inoue (2003) note that in contrast with the well-specified case ($\mu^* = 0$), the classical local identification condition of a rank p Jacobian matrix $G(\theta^*)$ does not guarantee that the matrix H^* is nonsingular. However, we follow Hall (2005) (see his Assumption 4.6) to note that if the second-order condition is satisfied ($Q_\infty(\theta)$ strictly convex in the neighborhood of θ^*), H^* should be positive definite. Therefore, as Hall and Inoue (2003) and Hall (2005), we maintain the assumption that H^* is nonsingular. Hence, we have proved:

Theorem 2 *Assume that $Y_t, t = 1, \ldots, T$ is an i.i.d. sequence and the jackknife GMM estimator θ^*_T is defined as:*

$$\arg\min_{\theta\in\Theta} \sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T} g(Y_t,\theta)'Wg(Y_s,\theta)$$

for some positive definite matrix W. Then, under standard regularity conditions, the asymptotic distribution of $\sqrt{T}\left[\theta^*_T - \theta^*\right]$ is normal with zero mean and variance:

$$(H^*)^{-1}\Sigma^*(H^*)^{-1}$$

with:

$$\begin{aligned} \Sigma^* &= G(\theta^*)'WS^*WG(\theta^*) + \Omega_{22}(\mu^*) + \Omega_{21}(\mu^*)WG(\theta^*) + G(\theta^*)'W\Omega_{12}(\mu^*) \\ H^* &= G(\theta^*)'WG(\theta^*) + \left(\mu^{*\prime}W \otimes Id_p\right)(\Delta G)(\theta^*). \end{aligned}$$

Several remarks are in order. First, when the moments are well-specified ($\mu^* = 0$), we find the standard formula for asymptotic variance of a GMM estimator associated with a weighting matrix W:

$$(H^*)^{-1}\Sigma^*(H^*)^{-1} = \left[G(\theta^*)'WG(\theta^*)\right]^{-1}\left[G(\theta^*)'WS^*WG(\theta^*)\right]\left[G(\theta^*)'WG(\theta^*)\right]^{-1}.$$

Second, in the general case, our asymptotic distribution is identical to the asymptotic distribution provided for the GMM estimator $\sqrt{T}\left[\hat{\theta}_T - \theta^*\right]$ by Hall and Inoue (2003). However, it is worth noting that the use of CLT for U-statistics in the context of jackknife GMM has made the proof much more transparent. The key for this transparency is the symmetric role between moment conditions and their Jacobian matrix provided by the Hájek projection (7.6). As noted by Hall and Inoue (2003), this result can also be seen as a particular case of Gallant and White's (1998) general asymptotic theory of misspecified models (see their Theorem 5.7).

Third, our result is also nested in the variance formula given by Ai and Chen (2007) (see their Theorem 4.1). However, their analysis is more general since they accommodate moment conditions with infinite dimensional unknown parameters. Moreover, they also do not use the simplifying framework of U-statistics.

7.3 Jackknife GMM with an Estimated Weighting Matrix

If the weighting matrix W has to be replaced by a consistent estimator W_T, a jackknife GMM estimator θ_T^* of θ is obtained by minimizing:

$$Q_T^*(\theta) = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T} g(Y_t,\theta)' W_T g(Y_s,\theta)$$

so that:

$$\frac{\partial Q_T^*(\theta^*)}{\partial\theta} = \frac{1}{T^2}\sum_{1\leq t\neq s\leq n} U_T(Y_t, Y_s),$$

where $U_T(Y_t, Y_s)$ is now the sample-size dependent symmetric function:

$$U_T(Y_t, Y_s) = \frac{\partial g(Y_t,\theta^*)'}{\partial\theta} W_T g(Y_s,\theta^*) + \frac{\partial g(Y_s,\theta^*)'}{\partial\theta} W_T g(Y_t,\theta^*).$$

Powell, Stock, and Stoker (1989) have proved a generalization of the CLT for U-statistics for the case of a symmetric function $U_T(Y_t, Y_s)$ that varies with T (see their Lemma 3.1). However, we must keep in mind that the estimator W_T of the weighting matrix W is a function of the random sample $Y_t, t = 1, .., T$. For this reason, the result of Powell, Stock, and Stoker (1989) does not seem very helpful. It involves the Hájek projection computed with the function $U_T(Y_t, Y_s)$, for which we do not know how to take into account the randomness within W_T to compute the needed conditional expectations. As already stressed by Gallant and White (1988) (see their discussion page 12 and 13), "additional complications arise in allowing W_T to be stochastic," precisely because the effect of this randomness remains relevant, even asymptotically, when the model is misspecified. The impact of the asymptotic distribution of the estimator W_T on the asymptotic distribution of θ_T^* has been fully characterized by Hall and Inoue (2003). The purpose of this subsection is to show that the framework of U-statistics still provides a much more transparent proof of the main result of Hall and Inoue (2003). However, this framework will be based on the time-invariant symmetric function $U(Y_t, Y_s)$ of the former section, rather than $U_T(Y_t, Y_s)$.

We start from:

$$\begin{aligned}
\overline{U}_T &= \frac{1}{T(T-1)} \sum_{1 \le t \ne s \le T} \left[\frac{\partial g(Y_t, \theta^*)'}{\partial \theta} W_T g(Y_s, \theta^*) + \frac{\partial g(Y_s, \theta^*)'}{\partial \theta} W_T g(Y_t, \theta^*) \right] \\
&= \frac{1}{T(T-1)} \sum_{1 \le t \ne s \le T} \left[\frac{\partial g(Y_t, \theta^*)'}{\partial \theta} W g(Y_s, \theta^*) + \frac{\partial g(Y_s, \theta^*)'}{\partial \theta} W g(Y_t, \theta^*) \right] \\
&\quad + \frac{1}{T(T-1)} \sum_{1 \le t \ne s \le T} \left[\frac{\partial g(Y_t, \theta^*)'}{\partial \theta} (W_T - W) g(Y_s, \theta^*) + \frac{\partial g(Y_s, \theta^*)'}{\partial \theta} (W_T - W) g(Y_t, \theta^*) \right].
\end{aligned}$$

From the former subsection, we know that:

$$\begin{aligned}
&\frac{1}{T(T-1)} \sum_{1 \le t \ne s \le T} \left[\frac{\partial g(Y_t, \theta^*)'}{\partial \theta} W g(Y_s, \theta^*) + \frac{\partial g(Y_s, \theta^*)'}{\partial \theta} W g(Y_t, \theta^*) \right] \\
&\quad = \frac{2}{T} \sum_{t=1}^{T} \varphi(Y_t) + o_P\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

Moreover, just by assuming that W_T is a consistent estimator of W, we have:

$$\begin{aligned}
&\frac{1}{T(T-1)} \sum_{1 \le t \ne s \le T} \left[\frac{\partial g(Y_t, \theta^*)'}{\partial \theta} (W_T - W) g(Y_s, \theta^*) + \frac{\partial g(Y_s, \theta^*)'}{\partial \theta} (W_T - W) g(Y_t, \theta^*) \right] \\
&= \frac{2}{T(T-1)} \left(\sum_{t=1}^{T} \frac{\partial g(Y_t, \theta^*)'}{\partial \theta} \right) (W_T - W) \left(\sum_{t=1}^{T} g(Y_t, \theta^*) \right) \\
&\quad - \frac{2}{T(T-1)} \sum_{t=1}^{T} \left[\frac{\partial g(Y_t, \theta^*)'}{\partial \theta} (W_T - W) g(Y_t, \theta^*) \right] \\
&= \frac{2}{T(T-1)} \left(\sum_{t=1}^{T} \frac{\partial g(Y_t, \theta^*)'}{\partial \theta} \right) (W_T - W) \left(\sum_{t=1}^{T} g(Y_t, \theta^*) \right) + o_P\left(\frac{1}{T}\right) \\
&= 2G(\theta^*)'(W_T - W)\mu^* + o_P\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

To get the last equality, we have used a standard CLT for sample means of the moment conditions and their Jacobians, such that:

$$\begin{aligned}
\frac{1}{T} \sum_{t=1}^{T} \frac{\partial g(Y_t, \theta^*)}{\partial \theta'} - G(\theta^*) &= O_P\left(\frac{1}{\sqrt{T}}\right) \\
\frac{1}{T} \sum_{t=1}^{T} g(Y_t, \theta^*) - \mu^* &= O_P\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

We end up with:

$$\overline{U}_T = \frac{2}{T} \sum_{t=1}^{T} \varphi(Y_t) + 2G(\theta^*)'(W_T - W)\mu^* + o_P\left(\frac{1}{\sqrt{T}}\right). \tag{7.7}$$

The decomposition (7.7) clearly displays the deep reason for the findings of Hall and Inoue (2003). On the one hand, there is no hope to get a root-T asymptotically normal GMM estimator if we do not maintain the assumption that $\sqrt{T}(W_T - W)\mu^*$ is asymptotically normal. While this assumption is rather innocuous in the i.i.d. unconditional case (under standard regularity conditions), it might be more problematic in a time series and/or a conditional context. This will be discussed in the next two subsections.

On the other hand, when it is asymptotically normal, the asymptotic distribution of $\sqrt{T}\left[\theta_T^* - \theta^*\right]$ is determined by the asymptotic joint normal distribution of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varphi(Y_t)$ and $\sqrt{T}(W_T - W)\mu^*$. Obviously, this asymptotic distribution can be explained, as in Hall and Inoue (2003), by the asymptotic joint normal distribution of sample means of moments and their Jabobians, as well as $\sqrt{T}(W_T - W)\mu^*$.

Following Hall and Inoue (2003), under the maintained assumption that $\sqrt{T}(W_T - W)\mu^*$ is asymptotically normal, it is worth introducing the following notations:

$$\begin{aligned}
\lim_{T=\infty} \operatorname{Var}\left[\sqrt{T}(W_T - W)\mu\right] &= \Omega_{33}(\mu) \\
\lim_{T=\infty} \operatorname{Cov}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g(Y_t,\theta^*), \sqrt{T}(W_T - W)\mu\right] &= \Omega_{13}(\mu), \Omega_{31}(\mu) = \Omega_{13}(\mu)' \qquad (7.8)\\
\lim_{T=\infty} \operatorname{Cov}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \frac{\partial g(Y_t,\theta^*)'}{\partial\theta} W\mu, \sqrt{T}(W_T - W)\mu\right] &= \Omega_{23}(\mu), \Omega_{32}(\mu) = \Omega_{23}(\mu)'.
\end{aligned}$$

Note that, to make an assumption of asymptotic normality of $\sqrt{T}(W_T - W)\mu^*$ useful for feasible inference, we will need to assume that all the above limits exist, at least for $\mu = \mu^*$.

Then, we can deduce easily from (7.7) that $\sqrt{T}U_T$ is asymptotically normal with mean zero and variance $4\tilde{\Sigma}^*$ with:

$$\begin{aligned}
\tilde{\Sigma}^* = {} & \operatorname{Var}[\varphi(Y_t)] + \Omega_{23}(\mu^*)G(\theta^*) + G(\theta^*)'\Omega_{32}(\mu^*) + G(\theta^*)'\Omega_{33}(\mu)G(\theta^*) \\
& + G(\theta^*)'W\Omega_{13}(\mu)G(\theta^*) + G(\theta^*)'\Omega_{31}(\mu)WG(\theta^*).
\end{aligned}$$

We immediately get the following generalization of Theorem 2:

Theorem 3 *Assume that $Y_t, t = 1, \ldots, T$ is an i.i.d. sequence and the jackknife GMM estimator θ_T^* is defined as:*

$$\arg\min_{\theta\in\Theta} \sum_{t=1}^{T} \sum_{s=1, s\neq t}^{T} g(Y_t,\theta)' W_T g(Y_s,\theta)$$

for a sequence of positive definite matrices W_T converging in probability toward a positive definite matrix W, such that $\sqrt{T}(W_T - W)\mu^*$ is asymptotically normal with the existence of limits (7.8) for $\mu = \mu^*$. Then, under standard regularity conditions, the asymptotic distribution of $\sqrt{T}\left[\theta_T^* - \theta^*\right]$ is normal with zero mean and variance:

$$(H^*)^{-1}\tilde{\Sigma}^*(H^*)^{-1}$$

with:

$$\begin{aligned}
\tilde{\Sigma}^* &= \Sigma^* + \Sigma^{**} \\
\Sigma^* &= G(\theta^*)'WS^*WG(\theta^*) + \Omega_{22}(\mu^*) + \Omega_{21}(\mu^*)WG(\theta^*) + G(\theta^*)'W\Omega_{12}(\mu^*) \\
\Sigma^{**} &= G(\theta^*)'W\Omega_{13}(\mu^*)G(\theta^*) + G(\theta^*)'\Omega_{31}(\mu)WG(\theta^*) + \Omega_{23}(\mu^*)G(\theta^*) \\
&\quad + G(\theta^*)'\Omega_{32}(\mu^*) + G(\theta^*)'\Omega_{33}(\mu^*)G(\theta^*) \\
H^* &= G(\theta^*)'WG(\theta^*) + \left(\mu^{*\prime}W \otimes Id_p\right)(\Delta G)(\theta^*).
\end{aligned}$$

Once again, several remarks are in order. First, when the moments are well-specified ($\mu^* = 0$), we find the standard formula for asymptotic variance of a GMM estimator associated with a weighting matrix W. To see that, just note that:

$$\Omega_{ij}(0) = 0, \forall i, j = 1, 2, 3.$$

Second, in the general case, our asymptotic distribution is identical to the asymptotic distribution provided for the GMM estimator $\sqrt{T}\left[\hat{\theta}_T - \theta^*\right]$ by Hall and Inoue (2003). However, it is worth noting that again the use of CLT for U-statistics in the context of jackknife GMM has made the proof much more transparent.

Third, our result is not encompassed by Ai and Chen (2007) because they restrict their attention to a known weighting matrix, chosen for convenience equal to the identity matrix.

7.4 The Time Series Case

We now revisit the analysis of the former subsection when the process $Y_t, t = 1, \ldots, T$ may display some serial dependence, always assumed to be stationary. Our analysis will differ from that of Hall and Inoue (2003) in two respects.

First, Hall and Inoue (2003) put some emphasis on the case where $\{g(Y_t, \theta) - \mu^*\}$ is a martingale difference sequence (see page 370). Of course, this assumption would be very relevant for a well-specified asset pricing model since it would then be, for a (possibly parameter dependent) choice of instruments $Z_t(\theta)$, on moment functions:

$$g(Y_{t+1}, \theta) = Z_t(\theta)[m_{t+1}(\theta)R_{t+1} - 1_n]$$

that are by definition a martingale difference sequence. However when the conditional expectation of $g(Y_{t+1}, \theta)$ given $I(t)$ is not zero, there is no reason to believe that it would coincide with the unconditional expectation μ^*. A pricing error that would be uncorrelated with any function of the past information would be rather easy to fix! Thus, this case is not relevant for our study of pseudo-true values and will not be considered here.

Second, Hall and Inoue (2003) provide a sophisticated analysis of the case of a weighting matrix sequence W_T that is the inverse of the HAC estimator of the long-run variance matrix of the moment conditions. This complicates dramatically the matter since it implies that the estimator W_T is in general converging only at a slow nonparametric rate that may contaminate the estimator θ_T^* of θ^*. This issue is of course very relevant for an "efficient" GMM approach (even though efficiency is never granted in the case of misspecification), but irrelevant for our specific pseudo-true value based on the HJ-distance.

Our extension to time series of Theorems 2 and 3 above is based on Theorem 1.8. of Dehling and Wendler (2010). Their starting point is the function:

$$\varphi(y) = E[V(y, Y_s)],$$

where:

$$\begin{aligned} V(Y_t, Y_s) &= \frac{\partial g(Y_t, \theta^*)'}{\partial \theta} W g(Y_s, \theta^*) + \frac{\partial g(Y_s, \theta^*)'}{\partial \theta} W g(Y_t, \theta^*) \\ \Rightarrow \varphi(y) &= \frac{\partial g(y, \theta^*)'}{\partial \theta} W \mu^* + G(\theta^*)' W g(y, \theta^*). \end{aligned} \tag{7.9}$$

Interestingly enough, the sequence $\varphi(Y_t)$ coincides with the one defined in (7.6), even though it cannot be interpreted anymore as $E[V(Y_t, Y_s)|Y_t]$:

$$\varphi(Y_t) = \frac{\partial g(Y_t, \theta^*)'}{\partial \theta} W\mu^* + G(\theta^*)' W g(Y_t, \theta^*).$$

We are going to show that, as far as the asymptotic distribution of the U -statistic is concerned, this function plays exactly the same role as in the previous subsections. The intuition for that may be found in Lemma 1 of Yoshihara (1976). Under some $\beta-$mixing conditions, the absolute expected error due to the approximation of $E[V(Y_t, Y_s)|Y_t]$ by $\varphi(Y_t)$ can be made arbitrarily small (for large $(s-t)$). We will then show that the time series extension of Theorem 2 holds with $\Omega(\mu) = [\Omega_{i,j}(\mu)]_{1 \le i,j \le 3}$ being defined as a long-term variance matrix. More precisely, we maintain the following assumption that is similar to Assumption (12) in Hall and Inoue (2003):

Assumption LT The vector $Z_T(\mu^*)$ is asymptotically normal with mean zero and variance $\Omega(\mu^*) = [\Omega_{i,j}(\mu^*)]_{1 \le i,j \le 3}$ when $Z_T(\mu)$ is defined as:

$$\begin{aligned}
Z_T(\mu) &= [Z_{1,T}(\mu)', Z_{2,T}(\mu)', Z_{3,T}(\mu)']' \\
Z_{1,T}(\mu) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [g(Y_t, \theta^*) - \mu] \\
Z_{2,T}(\mu) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\frac{\partial g(Y_t, \theta^*)'}{\partial \theta} - G(\theta^*)' \right] W\mu \\
Z_{3,T}(\mu) &= \sqrt{T}[W_T - W]\mu.
\end{aligned}$$

Note that in contrast with Hall and Inoue (2003), we do not need to assume that $\Omega(\mu^*)$ is a positive definite matrix. For instance, nothing precludes that $\Omega_{2,2}(\mu^*) = 0$ or $\Omega_{3,3}(\mu^*) = 0$. However, we also need to maintain the assumptions of Theorem 1.8. of Dehling and Wendler (2010). Since their theorem is written for a scalar U-statistic, we need to apply it for any possible linear combination.

Assumption DW $(Y_t)_{t \in \mathbb{N}}$ is a stationary mixing process such that for all $a \in \mathbb{R}^p, h_a(y_1, y_2) = a' V(y_1, y_2)$ fulfills the assumptions of Theorem 1.8. of Dehling and Wendler (2010).

These assumptions involve the strength of the mixing property jointly with some integrability conditions. Together these assumptions allow Dehling and Wendler (2010) to conclude that the asymptotic distribution of:

$$\frac{\sqrt{T}}{T(T-1)} \sum_{1 \le t \ne s \le T} h_a(Y_t, Y_s) = a' \left(\frac{\sqrt{T}}{T(T-1)} \sum_{1 \le t \ne s \le T} V(Y_t, Y_s) \right)$$

is normal with mean zero and asymptotic variance $4\sigma^2(a)$ with:

$$\sigma^2(a) = \text{Var}[a'\varphi(Y_t)] + 2\sum_{k=1}^{\infty} \text{Cov}[a'\varphi(Y_t), a'\varphi(Y_{t+k})].$$

In other words, for any a in $\mathbb{R}^p$, we get an asymptotic normal distribution that coincides with the asymptotic distribution of $\frac{2}{\sqrt{T}} \sum_{t=1}^{T} a'\varphi(Y_t)$, including the long-term variance matrix. By the Cramér–Wold theorem, the asymptotic distribution of the sequence of random vectors is characterized by the set of asymptotic distributions of all linear combinations. We can then conclude that the asymptotic distribution of:

$$\frac{\sqrt{T}}{T(T-1)}\sum_{1\leq t\neq s\leq T}\left[\frac{\partial g(Y_t,\theta^*)'}{\partial\theta}Wg(Y_s,\theta^*)+\frac{\partial g(Y_s,\theta^*)'}{\partial\theta}Wg(Y_t,\theta^*)\right]$$

coincides with the asymptotic distribution of $\frac{2}{\sqrt{T}}\sum_{t=1}^{T}\varphi(Y_t)$. Therefore, a proof very similar to the one of Theorem 3 can be devised to conclude that:

Theorem 4 *Assume that $Y_t, t=1,\ldots,T$ is a stationary mixing process and the jackknife GMM estimator θ_T^* is defined as:*

$$\arg\min_{\theta\in\Theta}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}g(Y_t,\theta)'W_T g(Y_s,\theta)$$

for a sequence of positive definite matrices W_T converging in probability toward a positive definite matrix W. Then, under assumptions LT and DW, the asymptotic distribution of $\sqrt{T}\left[\theta_T^*-\theta^*\right]$ is normal with zero mean and variance:

$$(H^*)^{-1}\tilde{\Sigma}^*(H^*)^{-1}$$

with:

$$\begin{aligned}
\tilde{\Sigma}^* &= \Sigma^*+\Sigma^{**}\\
\Sigma^* &= G(\theta^*)'W\Omega_{11}(\mu^*)WG(\theta^*)+\Omega_{22}(\mu^*)+\Omega_{21}(\mu^*)WG(\theta^*)+G(\theta^*)'W\Omega_{12}(\mu^*)\\
\Sigma^{**} &= G(\theta^*)'W\Omega_{13}(\mu^*)G(\theta^*)+G(\theta^*)'\Omega_{31}(\mu)WG(\theta^*)+\Omega_{23}(\mu^*)G(\theta^*)\\
&\quad +G(\theta^*)'\Omega_{32}(\mu^*)+G(\theta^*)'\Omega_{33}(\mu^*)G(\theta^*)\\
H^* &= G(\theta^*)'WG(\theta^*)+\left(\mu^{*\prime}W\otimes Id_p\right)(\Delta G)(\theta^*).
\end{aligned}$$

7.5 Bandwidth Dependent Pseudo-True SDF

Following Lavergne and Patilea (2013), we can apply a U-statistic approach similar to the one described above for the estimation of our bandwidth dependent pseudo true SDF $\theta^*(h)$:

$$\begin{aligned}
\hat{\theta}_T(h) &= \arg\min_{\theta\in\Theta}Q_T(\theta)\\
Q_T(\theta) &= \frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\Psi_{s+1}(\theta)'B'(X_s)B(X_t)\Psi_{t+1}(\theta)K\left[\frac{X_s-X_t}{h}\right].
\end{aligned}$$

To derive the asymptotic distribution of $\sqrt{T}\left[\hat{\theta}_T(h)-\theta^*(h)\right]$, the key is the asymptotic distribution of the U-statistic:

$$\begin{aligned}
\frac{\partial Q_T[\theta^*(h)]}{\partial\theta} &= \overline{U}_T=\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}U(Y_t,Y_s)\\
U(X_t,X_s) &= \frac{\partial\Psi_{s+1}(\theta)'}{\partial\theta}B'(X_s)B(X_t)\Psi_{t+1}(\theta)K\left[\frac{X_s-X_t}{h}\right]\\
&\quad +\frac{\partial\Psi_{t+1}(\theta)'}{\partial\theta}B'(X_t)B(X_s)\Psi_{s+1}(\theta)K\left[\frac{X_s-X_t}{h}\right].
\end{aligned}$$

Following a logic similar to the proofs of Theorems 2, 3, and 4 above, we start by noting that under convenient regularity conditions (as well as mixing assumptions in the time series case), $\sqrt{T}U_T$ is asymptotically equivalent to $\frac{2}{\sqrt{T}}\sum_{t=1}^{T}\varphi(X_t)$ with:

$$\begin{aligned}\varphi(x_t) &= E[U(x_t, X_s)] \\ &= E\left\{\frac{\partial\Psi_{s+1}(\theta)'}{\partial\theta}B'(X_s)K\left[\frac{X_s - x_t}{h}\right]\right\}B(x_t)\Psi_{t+1}(\theta) \\ &\quad + \frac{\partial\Psi_{t+1}(\theta)'}{\partial\theta}B'(x_t)E\left\{B(X_s)\Psi_{s+1}(\theta)K\left[\frac{X_s - x_t}{h}\right]\right\}.\end{aligned}$$

By stationarity, we can define time invariant matricial functions $A(.)$ and $C(.)$ such that:

$$\begin{aligned}\varphi(x_t) &= A(x_t)B(x_t)\Psi_{t+1}(\theta) + \frac{\partial\Psi_{t+1}(\theta)'}{\partial\theta}B'(x_t)C(x_t) \\ A(x_t) &= E\left\{\frac{\partial\Psi_{s+1}(\theta)'}{\partial\theta}B'(X_s)K\left[\frac{X_s - x_t}{h}\right]\right\} \\ C(x_t) &= E\left\{B(X_s)\Psi_{s+1}(\theta)K\left[\frac{X_s - x_t}{h}\right]\right\}.\end{aligned}$$

Therefore, $\sqrt{T}U_T$ is asymptotically normal with mean zero and variance $4\Sigma^*$:

$$\begin{aligned}\Sigma^* &= \mathrm{Var}[\varphi(X_t)] \\ &= \mathrm{Var}[A(X_t)B(X_t)\Psi_{t+1}(\theta)] + Var\left[\frac{\partial\Psi_{t+1}(\theta)'}{\partial\theta}B'(x_t)C(x_t)\right] \\ &\quad + \mathrm{Cov}\left[A(x_t)B(x_t)\Psi_{t+1}(\theta), \frac{\partial\Psi_{t+1}(\theta)'}{\partial\theta}B'(x_t)C(x_t)\right] \\ &\quad + \mathrm{Cov}\left[\frac{\partial\Psi_{t+1}(\theta)'}{\partial\theta}B'(x_t)C(x_t), A(x_t)B(x_t)\Psi_{t+1}(\theta)\right].\end{aligned}$$

Then the same argument as the one developed for proving Theorem 2, 3, and 4 allows to prove that, under convenient regularity and mixing conditions, the asymptotic distribution of $\sqrt{T}\left[\hat{\theta}_T(h) - \theta^*(h)\right]$ is normal with zero mean and variance:

$$(H^*)^{-1}\Sigma^*(H^*)^{-1}$$

where:

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

A similar result has been stated by Lavergne and Patilea (see their Theorem 2.3.). Note that we have simplified the analysis by using a fixed known matrix instead of a consistent estimator of a genuine square root of the HJ weighting matrix defined by:

$$\left[E\left[R_{t+1}R'_{t+1}|I(t)\right]\right]^{-1} = B'(X_t)B(X_t).$$

A consistent estimator of $B(X_t)$ would involve a nonparametric rate of convergence that would contaminate the estimator of $\theta^*(h)$, analogous to the case of a HAC estimator pointed out by Hall and Inoue (2003).

8 Numerical Experiments

We investigate in this section the finite sample properties of the SMD estimator relative to the local GMM estimator in a time-series setting. In Section 8.1, we study their performance for estimating an equity risk price parameter in the presence of global misspecification. We then proceed to present evidence in Section 8.2 that the factors commonly used in asset pricing models exhibit semi-strong identification strength. Finally, we illustrate the finite sample properties of the WMD estimator in a linear asset pricing model with semi-strong factors.

8.1 Misspecified Asset Pricing Model

The asset pricing model that we consider in this section is a discrete time option pricing model with an exponentially affine SDF that is driven by two fundamental factors, the return on an underlying asset and a factor driving its stochastic volatility process. We introduce global misspecification by assuming that the econometrician ignores the stochastic volatility factor as a priced source of risk.

8.1.1 Data generating process

The data-generating process (DGP) is a bivariate compound autoregressive process (Car(1)) for the log market return and a stochastic volatility factor $(r_{t+1}, \sigma^2_{t+1})$ defined by the conditional Laplace transform

$$E\left[\exp\left(-v r_{t+1} - u\sigma^2_{t+1}\right)|I_t\right] = \exp\left\{-l(u,v)\sigma^2_t - g(u,v)\right\} \tag{8.1}$$

for complex arguments u and v, and I_t the natural filtration of the state variables. We preclude Granger causality from return to volatility and we assume the volatility factor is Markov of order one. This implies a univariate Car(1) model for the volatility factor

$$E\left[\exp\left(-u\sigma^2_{t+1}\right)|\sigma^2_t\right] = \exp\left\{a(u)\sigma^2_t - b(u)\right\}. \tag{8.2}$$

Furthermore, we assume that stock returns are conditionally serially independent given the volatility path, and that the conditional Laplace transform for log-returns given σ^2_t and σ^2_{t+1} is

$$E\left[\exp(-v r_{t+1})|\sigma^2_t, \sigma^2_{t+1}\right] = \exp\left\{-\alpha(v)\sigma^2_{t+1} - \beta(v)\sigma^2_t - \gamma(v)\right\}. \tag{8.3}$$

The univariate conditional Laplace transforms (8.2) and (8.3) impose the following constraint on the bivariate conditional Laplace transform (8.1):

$$\begin{aligned} l(u,v) &= a[u + \alpha(v)] + \beta(v) \\ g(u,v) &= b[u + \alpha(v)] + \gamma(v). \end{aligned} \tag{8.4}$$

The functions $a(.)$, $b(.)$, $\alpha(.)$, $\beta(.)$, and $\gamma(.)$ are all defined on neighborhood of zero in the complex plane, and we have by definition

$$a(0) = b(0) = \alpha(0) = \beta(0) = \gamma(0) = 0.$$

A structure preserving exponential affine pricing kernel is specified as:

$$M_{t,t+1}(\theta) = \exp\left(-r_{f,t}\right)\exp\left\{m_0(\theta) + m_1(\theta)\sigma^2_t - \theta_1\sigma^2_{t+1} - \theta_2 r_{t+1}\right\}, \tag{8.5}$$

where θ_1 and θ_2 are preference parameters which characterize the prices of volatility risk and equity risk, respectively, and $m_0(\theta)$ and $m_1(\theta)$ with $\theta = (\theta_1, \theta_2)$ are defined in order to match the exogenously specified dynamics of the risk-free rate, $r_{f,t}$, which we assume to be zero. This is akin to the restriction

$$E\left[\exp\left\{m_0(\theta) + m_1(\theta)\sigma_t^2 - \theta_1\sigma_{t+1}^2 - \theta_2 r_{t+1}\right\}|I_t\right] = 1$$

from which it follows that

$$m_0(\theta) = b[\theta_1 + \alpha(\theta_2)] + \gamma(\theta_2) \tag{8.6}$$

and

$$m_1(\theta) = a[\theta_1 + \alpha(\theta_2)] + \beta(\theta_2). \tag{8.7}$$

The exponential affine pricing kernel is structure preserving in the sense that the risk-neutral dynamics are defined by the following bivariate Car(1):

$$E^*\left[\exp\left(-u\sigma_{t+1}^2 - v r_{t+1}\right)|I(t)\right] = \exp\left\{-l^*(u,v)\sigma_t^2 - g^*(u,v)\right\}$$
$$l^*(u,v) = a^*[u + \alpha^*(v)] + \beta^*(v)$$
$$g^*(u,v) = b^*[u + \alpha^*(v)] + \gamma^*(v),$$

where

$$\alpha^*(v) = \alpha(\theta_2 + v) - \alpha(\theta_2)$$
$$\beta^*(v) = \beta(\theta_2 + v) - \beta(\theta_2)$$
$$\gamma^*(v) = \gamma(\theta_2 + v) - \gamma(\theta_2)$$

and

$$a^*(u) = a[u + \theta_1 + \alpha(\theta_2)] - a[\theta_1 + \alpha(\theta_2)]$$
$$b^*(u) = b[u + \theta_1 + \alpha(\theta_2)] - b[\theta_1 + \alpha(\theta_2)].$$

The particular bivariate Car(1) process that we simulate from is the ARG(1)-Normal specification introduced by Khrapov and Renault (2016) as an extension (with leverage effect) of the discrete time stochastic volatility model of Darolles, Gourieroux, and Jasiak (2006). In the ARG(1)-Normal model, the volatility factor follows an autoregressive gamma process of order one (ARG(1)) defined by:

$$a(u) = \frac{\rho u}{1 + cu}, \qquad b(u) = \delta \log(1 + cu) \tag{8.8}$$

with

$$\rho \in [0, 1), c > 0, \delta > 0$$

so that the affine volatility dynamics are parametrized as:

$$\begin{aligned} E\left[\sigma_{t+1}^2|I_t\right] &= \rho\sigma_t^2 + \delta c \\ V\left[\sigma_{t+1}^2|I_t\right] &= 2c\rho\sigma_t^2 + \delta c^2. \end{aligned} \tag{8.9}$$

Furthermore, the log market return r_{t+1} is conditionally Gaussian with conditional Laplace transform characterized by:

$$\alpha(v) = \psi v - \frac{1}{2}v^2\left(1-\phi^2\right), \beta(v) = va^*(-\phi k), \gamma(v) = vb^*(-\phi k), \tag{8.10}$$

where:

$$k = k(c,\rho) = \frac{1}{\sqrt{c(1+\rho)}}$$

$$\psi = \psi(c,\rho,\phi,\theta_1,\theta_2) = k\phi + \left(1-\phi^2\right)\left(\theta_2 - \frac{1}{2}\right)$$

and:

$$a^*(u) = \frac{\rho^* u}{1+c^* u}, b^*(u) = \delta \log(1+c^* u)$$

$$\rho^*(u) = \frac{\rho}{\chi^2(\theta)}, c^*(u) = \frac{c}{\chi(\theta)}, \text{ and } \chi(\theta) = 1 + c[\theta_1 + \alpha(\theta_2)].$$

In addition to the three stochastic volatility parameters (c, ρ, δ), the conditional distribution of the log market return is characterized by three additional parameters (ϕ, θ_1, θ_2) which capture leverage effect, volatility risk premium, and equity risk premium, respectively. Identification of the volatility risk premium parameter θ_1 is fragile as it only occurs in the function $\chi(\theta)$, whereas the equity risk premium parameter θ_2 is prevalent throughout. To keep the exposition brief, we refer the reader to Khrapov and Renault (2016) for further details on the above parameterization of the log market return.

8.1.2 Simulation design

We simulate 1000 replications of sample size T = 2500 from the ARG(1)-Normal model with parameter values:

$$(c,\rho,\delta,\phi,\theta_1,\theta_2) = (0.10, 0.60, 1.25, 0.00, -0.70, 0.50).$$

Although the model allows for leverage effect as captured by the parameter ϕ, we assume that it is zero in order to be consistent with our state variable framework introduced in Section 6.2.

We initialize the volatility process at its unconditional mean, and we simulate it using the following well known procedure of Gouriéroux and Jasiak (2006):

1. Draw an integer valued latent variable U_t from the Poisson distribution with intensity $\left(\rho\sigma_t^2/c\right)$.
2. Draw (σ_{t+1}^2/c) from a gamma distribution with degree of freedom $\delta + U_t$.

We simulate the log market return based on its first two conditional moments:

$$E\left[r_{t+1}|\sigma_t^2,\sigma_{t+1}^2\right] = \alpha'(0)\sigma_{t+1}^2 + \beta'(0)\sigma_t^2 + \gamma'(0)$$

$$V\left[r_{t+1}|\sigma_t^2,\sigma_{t+1}^2\right] = -\alpha''(0)\sigma_{t+1}^2 - \beta''(0)\sigma_t^2 - \gamma''(0)$$

and the affine decomposition:

$$r_{t+1} = E\left[r_{t+1}|\sigma_t^2,\sigma_{t+1}^2\right] + \sqrt{V\left[r_{t+1}|\sigma_t^2,\sigma_{t+1}^2\right]}\epsilon_{t+1},$$

where ϵ_{t+1} is a standard normal random variable. The above parameterization of the model implies

$$\alpha'(0) = \psi, \beta'(0) = a^*(-\phi k), \gamma'(0) = b^*(-\phi k)$$
$$\alpha''(0) = -(1-\phi^2), \text{ and } \beta''(0) = \gamma''(0) = 0.$$

8.1.3 Estimation

We base our estimation of a well-specified model on the conditional moment restriction:

$$E\left[\exp\{m_0(\theta) + m_1(\theta)\sigma_t^2 - \theta_1\sigma_{t+1}^2 - \theta_2 r_{t+1}\}\exp\{r_{t+1}\} - 1|\sigma_t^2\right] = 0, \quad (8.11)$$

where $m_0(\theta)$ and $m_1(\theta)$ are defined in Equations (8.6) and (8.7), respectively. For the globally misspecified case, we instead base our estimation on the conditional moment restriction:

$$E\left[\exp\{m_0(\theta) - \theta_2 r_{t+1}\}\exp\{r_{t+1}\} - 1|\sigma_t^2\right] = 0. \quad (8.12)$$

Even though the volatility factor is recognized to be the proper conditioning variable in each case, in the globally misspecified case the econometrician incorrectly assumes that the it is not a priced risk factor. In order to further simplify matters, we only attempt to estimate the equity risk price, θ_2. This is consistent with estimating the parameters which govern the state variable dynamics (c, ρ, δ, ϕ) separately from the preference parameters (θ_1, θ_2), while also using option prices to (correctly) calibrate the volatility risk price, θ_1, for the well-specified case.

8.1.4 Results

Our baseline results for the well-specified model are reported in Table A.1. We report mean bias, median bias, standard deviation, root mean squared error, and the sampling distribution skewness and kurtosis for our estimates of the equity risk price θ_2. Both estimators, local GMM and SMD, are specified with a Gaussian kernel and the identity weighting matrix. We provide results for bandwidths, h, in the set $\{10, 1, 0.10, 0.01\}$.

In the well-specified case we see that both estimators perform comparably well over a range of bandwidths, but that both exhibit identification issues when we undersmooth. Histograms (not provided here) show a bimodal sampling distribution for the SMD estimator with modes of 0.50 and 1.00, but a unimodal distribution for local GMM centered around 1.00. A bimodal distribution also appears for the local GMM estimator for h = 0.10. This suggests that the extra degree of smoothing in local GMM compared to SMD leads to less stable behavior for small bandwidth choices.

Our baseline results for the globally misspecified case are reported in Table A.2. The reported statistics are the same as in Table A.1, but the estimators are now specified using the conditional HJ weighting matrix. For our baseline results, the bandwidth for the weighting matrix, b, coincides with the bandwidth for the conditional moment restrictions, h, across all specifications. Finally, due to a lack of a closed-form solutions for the pseudo-true values, we first simulated them using a much larger sample size of T = 7, 500. These are also reported in Table A.2 as $\theta_2^*(h)$.

In the misspecified case, we see that both estimators perform comparably across all bandwidths, but we hesitate to make any strict comparisons between the two estimators as they are estimating different pseudo-true values. Encouragingly, we also see that (at least for the model in question) the pseudo-true values do not vary drastically with the choice of bandwidth. Furthermore, we note that the pseudo-true for local GMM appears to be less sensitive to the bandwidth choice. As expected based on our asymptotic distributional theory, we also see that the sampling distributions exhibit normal behavior with near zero skewness, and kurtosis near three.

In Tables A.3 and A.4 we report results for the misspecified case again, but with a bandwidth for the conditional HJ weighting matrix estimator that is fixed across specifications at b = 1.0 and b = 0.10, respectively. We do not find considerable differences between the results in Tables A.2, A.3, and A.4.

As we have argued previously, we may not want to be overly parsimonious with respect to the conditioning information set in the misspecified case. For this reason, we revisit the baseline results with additional conditioning information by including one-period lagged returns as a conditioning variable alongside the stochastic volatility factor.

For the well-specified case with additional conditioning information, the results in Table A.5 show similar patterns to those in Table A.1. The most striking difference is the change in the sampling distribution statistics for local GMM at a relatively large bandwidth of h = 1; local GMM starts to exhibit nonnormal behavior, and once again we have identification issues for the smallest bandwidth choices.

For the misspecified case with additional conditioning information, the results in Table A.6 confirm that the pseudo-true value is a function of the conditioning information for both estimators. Moreover, we find that the pseudo-true value for SMD is more sensitive to the choice of conditioning information for large bandwidths, but that it is more stable across the entire range of bandwidths than that of local GMM.

Finally, we reconsider the well-specified case in the presence of leverage effect in order to confirm whether or not the assumption of no leverage effect introduced in Section 6.2 is necessary. The results in Table A.7 clearly show that SMD and local GMM are both biased in the presence of leverage effect.

8.2 Estimation with Identification Weakness

In this section, we present evidence that the factors commonly used in asset pricing models exhibit semi-strong identification strength. We then illustrate the finite sample properties of our new inference procedures in a linear asset pricing model with semi-strong factors.

8.2.1 Factors display different identification strengths

We start with a covariance study of the factors commonly used in asset pricing models to illustrate their identification strengths. Consider the usual case where a CLT holds for the empirical covariance between the market return r_t and a given (observable) factor f_t under standard regularity assumptions, that is:

$$\sqrt{T}\left(\widehat{\mathrm{Cov}}_T(r_t, f_t) - \mathrm{Cov}(r_t, f_t)\right) \xrightarrow{d} z$$

with z random variable normally distributed with mean 0 and given variance σ^2.

It is useful to consider the implication of the above result for T large enough,

$$\widehat{\mathrm{Cov}}(r_t, f_t) = \frac{z}{\sqrt{T}} + \mathrm{Cov}(r_t, f_t)$$

Let us now consider the following cases:

1. Strong factor, $\text{Cov}(r_t, f_t) = O(1)$: for T large, $\widehat{\text{Cov}}(r_t, f_t) = O(1)$;
2. Semi-strong/Near-weak factor, $\text{Cov}(r_t, f_t) = O\big(1/T^{\lambda}\big)$ and $0 < \lambda < 1/2$: for T large, $\widehat{\text{Cov}}(r_t, f_t) = \mathcal{O}\big(1/T^{\lambda}\big)$;
3. Weak factor, $\text{Cov}(r_t, f_t) = O\big(1/T^{\lambda}\big)$ and $\lambda = 1/2$: for T large, $\widehat{\text{Cov}}(r_t, f_t) = \mathcal{O}_P\Big(1/\sqrt{T}\Big) + \mathcal{O}\Big(1/\sqrt{T}\Big)$;
4. Super-weak factor, $\text{Cov}(r_t, f_t) = O\big(1/T^{\lambda}\big)$ and $\lambda > 1/2$: for T large, $\widehat{\text{Cov}}(r_t, f_t) = \mathcal{O}_p\Big(1/\sqrt{T}\Big)$.

Regardless of the properties of the factors, assuming the above CLT applies, the empirical covariance is always of order $1/T^{\nu}$ with $0 \le \nu \le 1/2$.

In order to gain intuition on the properties of macro-finance factors commonly used in empirical studies, we use monthly data and consider n different sample sizes T_i (with $i = 1, \cdots, n$) to compute the associated sample covariances c_i between r_t and f_t. We then regress $\log(c_i)$ on $\log(T_i)$,

$$\log(c_i) = \alpha - \beta \log(T_i) + u_i,$$

and estimate the associated slope coefficients by OLS. We consider the following three popular factors: *cay* (the factor developed by Lettau and Ludvigson (2001a)), *cg* (the growth rate in real per capita nondurable consumption (seasonally adjusted at annual rates) from the Bureau of Economic Analysis), and *smb* (the second factor of Fama and French, "small minus big"). The data are from February 1959 until December 2012. The different sample sizes are 196, 246, 296,…, 646, and the associated estimated slopes are as follows, all between 0 and 0.5:

Factor	cay	cg	smb
Slope	0.0824	0.2936	0.4573

To conclude, our empirical results reveal that the three factors are semi-strong, with *cay* stronger than *cg* stronger than *smb*.

8.2.2 Pricing model with semi-strong factors

In this section, we illustrate the finite sample properties of our new inference procedures. We focus on a linear asset pricing model where the SDF is motivated by Epstein–Zin preferences. More specifically, we consider the following pricing equations:

$$E[R_{t+1} m_{t+1}(\eta) - \mathbf{1} | \mathcal{F}_t] = \mathbf{0}$$

with

$$m_{t+1}(\eta) = m_{t+1}(\delta, \gamma, \psi) = \left[\delta (cg_{t+1})^{-1/\psi}\right]^{\gamma} \left[\frac{1}{R_{m,t+1}}\right]^{1-\gamma},$$

and

$$\mathcal{F}_t = \{Z_\tau, \tau \leq t\} \qquad \text{with} \qquad Z_t = \{cay_t, cg_t, smb_t, R_{m,t}\}.$$

The linear approximation of the above pricing model yields the following conditional moment equations:

$$E[R_{t+1}(1 + \log\,(m_{t+1}(\theta))) - \mathbf{1}|\mathcal{F}_t] = \mathbf{0}$$

with three unknown parameters $\theta = (a, b, c)$ and

$$\begin{aligned} 1 + \log(m_{t+1}(\theta)) = 1 + \log(m_{t+1}(\delta, \gamma, \psi)) &= 1 + \gamma\log\delta - \frac{\gamma}{\psi}\log(cg_{t+1}) - (1-\gamma)\log\left(R_{m,t+1}\right) \\ &= a + b\log\,(cg_{t+1}) + c\log\,\left(R_{m,t+1}\right), \end{aligned}$$

and

$$\mathcal{F}_t = \{Z_\tau, \tau \leq t\} \qquad \text{with} \qquad Z_\tau = \{cay_\tau, cg_\tau, smb_\tau, R_{m,\tau}\}.$$

We are pricing the seventeen industry portfolios; definitions and data can be found on K. French's webpage at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.First, we investigate the informational content of the different instruments featured in $\mathcal{F}_t$. More specifically, we compute confidence intervals for each parameter using the estimators of Antoine and Lavergne (2014)[1] and the Local-GMM (computed with the data-dependent optimal bandwidth) which rely on conditional moment equations based on different combinations of the conditioning variables *cay*, *cg*, *smb*, and R_m. As a benchmark, we also compute confidence regions based on the efficient GMM estimator obtained using five unconditional moments using the constant and the first lag of each conditioning variable, as well as nine and thirteen unconditional moments, respectively with the first two and three lags of each conditioning variable. Results are displayed in Table A.8.

When conditioning variables are added, the associated confidence intervals of Antoine and Renault (2009) are always narrower, and always a strict subset of the original confidence intervals. This is in contrast with the estimation based on local GMM and GMM: when conditioning variables are added, the new local GMM-based confidence interval is often much wider, indicating issues related to the curse of dimensionality; when instruments are added, the new GMM-based confidence interval often has a null intersection with the original interval. Since it is always unclear which instruments should be used in practice, one may obtain very different (and misleading) inference results.

All estimation techniques reveal that a is positive and slightly smaller than one; however, they disagree for b and c: AL inference suggests that both b and c are negative and both between -1 and zero, which conforms to the economic interpretation of the coefficients, whereas local GMM and GMM inference suggest conflicting results depending on the set of conditioning variables/instruments used.

Second, we investigate the convergence rates of the estimators of Antoine and Lavergne (2014) obtained with different conditioning variables. We follow the procedure highlighted in the previous section by considering different sample sizes from 196 to 646 and computing the associated estimators and their variances. More specifically, for a given sample size

1 The estimator of Antoine and Lavergne (2014) is a special case of SMD where smoothing is not used.

T_i, we compute the ratios of the standard deviations of the estimators of a and b, as well as a and c as follows,

$$\frac{\sqrt{\text{Var}\left(\hat{b}_i\right)}}{\sqrt{\text{Var}(\hat{a}_i)}} \quad \text{and} \quad \frac{\sqrt{\text{Var}(\hat{c}_i)}}{\sqrt{\text{Var}(\hat{a}_i)}}$$

and regress the logarithm of such ratio on the logarithm of the sample size. According to asymptotic results developed in Antoine and Renault (2009) and Antoine and Lavergne (2014), the estimator of a is strongly identified with a standard rate of convergence $\sqrt{T}$, whereas estimators of b and c may not be as strongly identified. As a result, we expect the slope coefficients of the above regression to be positive coefficients between 0 and 0.5: the closer to zero, the stronger the identification. Our results are displayed in Table A.9.

Some of the results for the AL estimators obtained with one conditioning variable are hard to reconcile with the asymptotic theory mentioned above, especially the negative slopes. However, when computing the AL estimators based on the four conditioning variables, our results suggest that b and c are semi-strongly identified with a similar identification strength: both slope coefficients for the estimators using the four conditioning variables are approximately 0.14 which means that they are associated with a rate of convergence of $T^{0.50-0.14} = T^{0.36}$. This also supports the confidence intervals obtained in Table A.8: the length of the interval for a is much narrower than that of b or c.

The results for the local GMM estimators with four conditioning variables and for the GMM estimators with thirteen IV also suggest that b and c are semi-strongly identified: corresponding rates of convergence are $T^{0.09}$ and $T^{0.37}$ for Local GMM estimators, and $T^{0.09}$ and $T^{0.19}$ for efficient GMM; all GMM-based estimators are somewhat weaker than corresponding AL estimators.

9 Concluding Remarks

In this article, we pursue the research agenda put forward by H. White's (1994) Econometric Society monograph to examine the consequences of misspecification in econometrics for the interpretation and properties of parameters estimated from a misspecified model. However, since our focus of interest is inference on asset pricing models, the maximum likelihood approach developed by White (1994) is hardly relevant. As clearly characterized by Hansen and Richard (1987), the empirical content of an asset pricing model is encapsulated in the definition of a SDF, or in practice, a parametric family of SDFs. A valid SDF is supposed to deliver asset prices as solutions of some moment conditions. Therefore, a SDF model is misspecified insofar as there is no value of the vector θ of unknown parameters that fulfills all of the moment conditions for all observed asset prices.

The main motivation of this article is the observation that extant literatures for inference on misspecified models have developed in very different ways for the methodology of (quasi) maximum likelihood and GMM-type approaches, respectively.

In the maximum likelihood literature, there is little doubt that the relevant measure of misspecification is the Kullback–Leibler distance between the DGP and the parametric family of probability distributions that characterize the misspecified model of interest. The probability limit of the quasi maximum likelihood estimator is indeed the value of the parameters that minimizes this distance, and the focus of interest is inference about this so-called pseudo-true value of θ.

In the asset pricing literature, the focus has been mainly on comparison of the degree of misspecification of different competing models as exemplified for instance in the list of papers mentioned in the conclusion of Hansen and Jagannathan (1997). In contrast, little has been done on interpretation and estimation of the pseudo-true value. The ambition of this article has been to shed more light on these issues, in particular through the following statements.

First, we follow Lettau and Ludvigson (2001b) to note that the approach "of scaling factors with information available in the current period leads to a multifactor unconditional model in place of a single-factor, conditional model. This approach therefore provides a justification for requiring more than one factor to explain the behavior of expected returns" even if one knows that a well-specified SDF should always provide a one factor model. More generally, we have argued that the ability to price accurately on average the local factors is the main feature of interest of our pseudo true SDF.

Second, it is interesting to notice the analogy between linear exponential families for the study of maximum likelihood with misspecification and the conditionally affine (or exponentially conditionally affine) pricing factor models. For quasi maximum likelihood the linear exponential family gives a setting where the pseudo-true value may coincide with the true unknown value of the parameters because there are cases where "the consequences for the QMLE of ignoring or misspecifying features of the conditional density that are not of direct interest" (White, 1994, page 62) is not harmful for the characterization of the true unknown value of θ (see also Gourieroux, Monfort, and Trognon, 1984). Similarly, we have stressed that our HJ pseudo-true value delivers exact pricing on average of the factors (or at least of the discounted factors) in the context of conditionally affine (or exponentially conditionally affine) pricing factor models.

Third, White (1994) (see chapter 8) emphasizes "the crucial role played by the martingale difference sequence requirement" on the score vector, while "this martingale difference assumption generally fails in the presence of dynamic misspecification." A similar failure has led us to question the traditional GMM approaches, including local GMM, even when they try to accommodate misspecification. We have shown that an approach in terms of jackknife GMM (including SMD with fixed bandwidth) and Central Limit Theorems for U-statistics is much more transparent and flexible. However, the ability to incorporate in the estimating equations estimated quantities that come with a nonparametric rate of convergence (as a kernel estimator of a conditional HJ weighting matrix) is still work in progress. More generally, quadratic forms of Markov chains arise in a variety of estimation problems with misspecified moment restrictions and a Markov setting of state variables. A martingale decomposition as put forward by Atchadé and Cattaneo (2014) that "resembles the well-known Hoeffding decomposition of U-statistics" should be very helpful in this respect, beyond the few examples provided in this article.

Finally, we have provided in this article a few elements to bridge the gap between two important issues for inference in conditional pricing models, namely misspecification and weak identification. The development of methods able to tackle simultaneously these two issues is still in its infancy with the notable exception of Gospodinov, Kan, and Robotti (2014, 2018). We consider that the SMD approaches put forward in this article, after Lavergne and Patilea (2013) as well as Antoine and Lavergne (2014), are promising in this respect.

APPENDIX

A.1 Proofs

Theorem 1

$$\begin{aligned} e_Z[\theta^1,\theta^2] &= E\{Z_t(\theta^1)\Psi_{t+1}(\theta^2)\} \\ &= E\{Z_t(\theta^1)E[\Psi_{t+1}(\theta^2)|I(t)]\} \\ &\Rightarrow \frac{\partial e_Z(\theta^1,\theta^2)}{\partial\theta^{2\prime}} = E\left\{Z_t(\theta^1)E\left[\frac{\partial\Psi_{t+1}(\theta^2)}{\partial\theta'}|I(t)\right]\right\} \\ &= E\left\{E\left[\frac{\partial\Psi'_{t+1}(\theta^2)}{\partial\theta}|I(t)\right]\Omega^{-1}[I(t)]E\left[\frac{\partial\Psi_{t+1}(\theta^2)}{\partial\theta'}|I(t)\right]\right\}. \end{aligned}$$

Hence:

$$\frac{\partial e_Z(\theta,\theta)}{\partial\theta^{2\prime}} = \Omega_Z(\theta)$$

is a positive definite matrix, and in particular a nonsingular matrix. Moreover, since we have by definition:

$$\theta^1 = \theta^2 = \theta^* \Rightarrow e_Z[\theta^1,\theta^2] = 0$$

we can apply the implicit function theorem to define the function $\overline{\theta}(.)$ as announced in Theorem 1. The differentiation of the identity (4.8) gives:

$$\frac{\partial e_Z(\theta,\overline{\theta}(\theta))}{\partial\theta^1} + \frac{\partial e_Z(\theta,\overline{\theta}(\theta))}{\partial\theta^{2\prime}}\frac{\partial\overline{\theta}(\theta)}{\partial\theta} = 0$$

leading to the formula of Theorem 1 in the case $\theta = \theta^*$. QED

Proposition 1. By definition of the scalar product:

$$P_C e[.,\theta^*] = C[.]b$$

with:

$$b = \left[E\{C'[I(t)]\Omega^{-1}[I(t)]C[I(t)]\}\right]^{-1}E\{C'[I(t)]\Omega^{-1}[I(t)]e[I(t),\theta^*]\}.$$

Therefore:

$$\delta_Z^2 = \delta_Z^2(\theta^*) = e_Z[\theta^*]'\Omega_Z^{-1}e_Z[\theta^*] = ||P_C e[.,\theta^*]||^2$$

since:

$$\begin{aligned} e_Z[\theta] &= E\{Z_t[m[I(t+1),\theta]R_{t+1} - 1_n]\} \\ &= E\{C'[I(t)]\Omega^{-1}[I(t)]e[I(t),\theta]\} \end{aligned}$$

and:

$$\begin{aligned}\Omega_Z &= E\{Z_t\Omega[I(t)]Z_t'\}\\ &= E\{C'[I(t)]\Omega^{-1}[I(t)]C[I(t)]\}.\end{aligned}$$

QED

A.2 Results for Numerical Experiments

A.2.1 Misspecified asset pricing model

Table A.1. Finite sample properties of the SMD and Local GMM estimators for a well-specified asset pricing model

	h = 10		h = 1		h = 0.10		h = 0.01	
	SMD	LGMM	SMD	LGMM	SMD	LGMM	SMD	LGMM
Bias	−0.0154	−0.0114	−0.0135	−0.0111	−0.0003	−0.0017	0.2384	0.4511
Med	−0.0158	−0.0123	−0.0152	−0.0119	−0.0019	−0.0044	−0.0001	0.4926
Std	0.0460	0.0465	0.0463	0.0466	0.0528	0.0593	4.4000	0.1353
Rmse	0.0485	0.0479	0.0482	0.0478	0.0528	0.0593	4.4043	0.4709
Skew	0.078	0.093	0.103	0.095	0.836	2.063	21.433	−2.694
Kurt	3.143	3.170	3.112	3.164	9.164	17.933	473.210	8.634

We report mean bias (Bias), median bias (Med), standard deviation (Std), root mean squared error (RMSE), skewness, and kurtosis for two different estimators: SMD and Local GMM (LGMM). Both estimators are specified with a Gaussian Kernel and the identity weighting matrix.

Table A.2. Finite sample properties of the SMD and Local GMM estimators for a misspecified asset pricing model

	h = 10		h = 1		h = 0.10		h = 0.01	
	SMD	LGMM	SMD	LGMM	SMD	LGMM	SMD	LGMM
$\theta_2^*(h)$	0.2088	0.2098	0.2059	0.2090	0.1660	0.1998	0.1435	0.2055
Bias	−0.0058	−0.0040	−0.0057	−0.0040	−0.0071	−0.0023	−0.0238	0.0152
Med	−0.0049	−0.0032	−0.0051	−0.0033	−0.0080	−0.0026	−0.0240	0.0155
Std	0.0384	0.0385	0.0380	0.0383	0.0392	0.0371	0.0411	0.0361
Rmse	0.0388	0.0387	0.0384	0.0385	0.0398	0.0372	0.0475	0.0391
Skew	−0.166	−0.154	−0.160	−0.158	−0.103	−0.149	−0.115	−0.144
Kurt	3.174	3.210	3.152	3.197	2.962	3.130	2.857	3.024

We report the pseudo-true equity risk price $\theta_2^*(h)$, mean bias (Bias), median bias (Med), standard deviation (Std), root mean squared error (RMSE), skewness, and kurtosis for two different estimators: SMD and Local GMM (LGMM). Both estimators are specified with a Gaussian Kernel and the Hansen–Jagannathan weighting matrix. The bandwidth for the weighting matrix, b, coincides with the bandwidth for the conditional moment restrictions, h, across all specifications.

Table A.3. Finite sample properties of the SMD and Local GMM estimators for a misspecified asset pricing model

	h = 10		h = 1		h = 0.10		h = 0.01	
	SMD	LGMM	SMD	LGMM	SMD	LGMM	SMD	LGMM
$\theta_2^*(h)$	0.2108	0.2098	0.2059	0.2090	0.1773	0.2179	0.1682	0.2543
Bias	−0.0064	−0.0040	−0.0057	−0.0040	−0.0054	0.0003	−0.0059	0.0511
Med	−0.0054	−0.0032	−0.0051	−0.0033	−0.0069	0.0015	−0.0057	0.0518
Std	0.0382	0.0385	0.0380	0.0383	0.0395	0.0429	0.0430	0.0546
Rmse	0.0388	0.0387	0.0384	0.0385	0.0398	0.0429	0.0433	0.0748
Skew	−0.191	−0.154	−0.160	−0.158	−0.100	−0.017	−0.041	0.239
Kurt	3.139	3.210	3.152	3.197	2.996	3.386	2.902	3.168

We report the pseudo-true equity risk price $\theta_2^*(h)$, mean bias (Bias), median bias (Med), standard deviation (Std), root mean squared error (RMSE), skewness, and kurtosis for two different estimators: SMD and Local GMM (LGMM). Both estimators are specified with a Gaussian Kernel and the Hansen–Jagannathan weighting matrix. The bandwidth for the weighting matrix is fixed at b = 1.0 across all specifications.

Table A.4. Finite sample properties of the SMD and Local GMM estimators for a misspecified asset pricing model

	h = 10		h = 1		h = 0.10		h = 0.01	
	SMD	LGMM	SMD	LGMM	SMD	LGMM	SMD	LGMM
$\theta_2^*(h)$	0.1935	0.2098	0.1910	0.2079	0.1659	0.1998	0.1556	0.2264
Bias	−0.0090	−0.0041	−0.0077	−0.0039	−0.0070	−0.0023	−0.0075	0.0400
Med	−0.0082	−0.0032	−0.0079	−0.0033	−0.0079	−0.0026	−0.0082	0.0418
Std	0.0369	0.0385	0.0369	0.0382	0.0392	0.0371	0.0422	0.0380
Rmse	0.0380	0.0387	0.0377	0.0383	0.0398	0.0372	0.0429	0.0552
Skew	−0.144	−0.154	−0.131	−0.159	−0.104	−0.149	−0.079	−0.167
Kurt	2.955	3.210	2.961	3.190	2.962	3.130	2.832	2.989

We report the pseudo-true equity risk price $\theta_2^*(h)$, mean bias (Bias), median bias (Med), standard deviation (Std), root mean squared error (RMSE), skewness, and kurtosis for two different estimators: SMD and Local GMM (LGMM). Both estimators are specified with a Gaussian Kernel and the Hansen–Jagannathan weighting matrix. The bandwidth for the weighting matrix is fixed at b = 0.10 across all specifications.

Table A.5. Finite sample properties of the SMD and Local GMM estimators for a well-specified asset pricing model with additional conditioning information

	h = 10		h = 1		h = 0.10		h = 0.01	
	SMD	LGMM	SMD	LGMM	SMD	LGMM	SMD	LGMM
Bias	−0.0165	−0.0114	−0.0665	−0.0131	−0.1385	0.4715	−0.0919	0.4825
Med	−0.0170	−0.0123	−0.0648	−0.0148	−0.1384	0.4768	−0.1487	0.4826
Std	0.0458	0.0465	0.0408	0.0548	0.0441	0.0446	0.3384	0.0141
Rmse	0.0487	0.0479	0.0780	0.0563	0.1454	0.4737	0.3505	0.4827
Skew	0.072	0.093	−0.027	3.001	−0.099	−2.917	−6.446	0.056
Kurt	3.136	3.170	2.986	32.080	2.791	24.721	113.684	3.271

We report mean bias (Bias), median bias (Med), standard deviation (Std), root mean squared error (RMSE), skewness, and kurtosis for two different estimators: SMD and Local GMM (LGMM). Both estimators are specified with a product of Gaussian Kernels and the identity weighting matrix.

Table A.6. Finite sample properties of the SMD and Local GMM estimators for a misspecified asset pricing model with additional conditioning information

	h = 10		h = 1		h = 0.10		h = 0.01	
	SMD	LGMM	SMD	LGMM	SMD	LGMM	SMD	LGMM
$\theta_2^*(h)$	0.161	0.2098	0.1608	0.2069	0.1199	0.2486	0.1035	1.1176
Bias	−0.0405	−0.0041	−0.0280	−0.0034	−0.0169	0.1258	−0.0130	−0.0465
Med	−0.0398	−0.0032	−0.0286	−0.0028	−0.0156	0.0171	−0.0191	−0.0457
Std	0.0395	0.0385	0.0377	0.0383	0.0434	0.2987	0.1017	0.0101
Rmse	0.0566	0.0387	0.0469	0.0384	0.0466	0.3240	0.1025	0.0476
Skew	−0.132	−0.154	−0.086	−0.148	−0.165	2.099	0.210	−0.698
Kurt	2.947	3.211	2.875	3.198	2.958	5.632	3.411	4.360

We report the pseudo-true equity risk price $\theta_2^*(h)$, mean bias (Bias), median bias (Med), standard deviation (Std), root mean squared error (RMSE), skewness, and kurtosis for two different estimators: SMD and Local GMM (LGMM). Both estimators are specified with a product of Gaussian Kernels and the Hansen–Jagannathan weighting matrix. The bandwidth for the weighting matrix is fixed at b = 0.10 across all specifications.

Table A.7. Finite sample properties of the SMD and Local GMM estimators for an asset pricing model with leverage effect

	h = 10		h = 1		h = 0.10		h = 0.01	
	SMD	LGMM	SMD	LGMM	SMD	LGMM	SMD	LGMM
Bias	0.0492	0.0531	0.0512	0.0535	0.0660	0.0661	0.0951	0.4808
Med	0.0491	0.0528	0.0501	0.0532	0.0633	0.0599	0.0684	0.4959
Std	0.0430	0.0436	0.0434	0.0437	0.0504	0.0682	0.1192	0.0808
Rmse	0.0653	0.0687	0.0671	0.0690	0.0830	0.0950	0.1525	0.2390
Skew	0.137	0.161	0.160	0.162	0.069	2.891	2.424	–4.341
Kurt	3.157	3.200	3.133	3.196	6.912	18.058	8.697	22.097

We report mean bias (Bias), median bias (Med), standard deviation (Std), root mean squared error (RMSE), skewness, and kurtosis for two different estimators: SMD and Local GMM (LGMM). Both estimators are specified with a Gaussian Kernel and the identity weighting matrix.

A.2.2 Estimation with identification weakness

Table A.8. Confidence intervals for each parameter using the SMD estimator with $h=1$ with different conditioning variables, as well as the local GMM estimator with different conditioning variables, and the efficient GMM estimator based on 4, 8, and 12 unconditional moments

	SMD with $h=1$			
1 conditioning variable	*cay*	*cg*	*smb*	R_m
Confidence interval for *a*	[0.9768; 1.0049]	[0.9734; 1.0054]	[0.9691; 1.0049]	[0.9825; 1.0101]
Confidence interval for *b*	[−16.6987; 7.2351]	[−17.6110; 8.4584]	[−14.7980; 10.6002]	[−24.4754; 4.9896]
Confidence interval for *c*	[−0.6809; 2.5114]	[−0.8129; 3.0894]	[−0.6992; 2.8511]	[− 1.0102; 2.8754]
4 conditioning variables	(cay, cg, smb, R_m)			
Confidence interval for *a*	[0.9921; 0.9924]			
Confidence interval for *b*	[−0.9513;−0.8472]			
Confidence interval for *c*	[−0.1273;−0.0730]			

	Local-GMM			
1 conditioning variable	*cay*	*cg*	*smb*	R_m
Confidence interval for *a*	[0.9861; 1.0069]	[0.9488; 1.0507]	[5.3; 35.3]	[0.9228; 0.9430]
Confidence interval for *b*	[−10.6581; 7.9169]	[−10.6606; 11.3798]	[933.3; 7459.5]	[−32.0638; −11.6121]
Confidence interval for *c*	[−1.6724; 0.1225]	[−3.2114; 0.6840]	[119.4; 938.3]	[−1.7861; 0.2510]
4 conditioning variables	(cay, cg, smb, R_m)			
Confidence interval for *a*	[0.8876; 1.1110]			
Confidence interval for *b*	[−46.9301; 44.4371]			
Confidence interval for *c*	[−7.2303; 5.4776]			

	Efficient GMM		
	4 IV	8 IV	12 IV
Confidence interval for *a*	[0.9904; 0.9907]	[0.9907; 0.9908]	[0.9918; 0.9919]
Confidence interval for *b*	[−0.3168;−0.2324]	[−0.1142;−0.0661]	[0.3433; 0.3856]
Confidence interval for *c*	[0.1855; 0.2220]	[−0.0255;−0.0044]	[−0.4598;−0.4434]

Note: We price seventeen industry portfolios.

Table A.9. Estimated slopes of the regression of the log ratio of standard deviations on the log sample size for parameter estimated by SMD ($h = 1$) and local GMM with different conditioning variables, as well as efficient GMM with 4, 8, and 12 instruments

	SMD with $h = 1$			
1 conditioning variable	*cay*	*cg*	*smb*	R_m
Slope for log $(\sqrt{\text{Var}(\hat{b})}/\sqrt{\text{Var}(\hat{a})})$	0.49	0.53	0.44	0.64
Slope for log $(\sqrt{\text{Var}(\hat{c})}/\sqrt{\text{Var}(\hat{a})})$	−0.79	−0.79	−0.92	−0.58
4 conditioning variables	(cay, cg, smb, R_m)			
Slope for log $(\sqrt{\text{Var}(\hat{b})}/\sqrt{\text{Var}(\hat{a})})$	0.141			
Slope for log $(\sqrt{\text{Var}(\hat{c})}/\sqrt{\text{Var}(\hat{a})})$	0.145			

	Local GMM			
1 conditioning variable	*cay*	*cg*	*smb*	R_m
Slope for log $(\sqrt{\text{Var}(\hat{b})}/\sqrt{\text{Var}(\hat{a})})$	0.584	1.072	−0.019	0.293
Slope for log $(\sqrt{\text{Var}(\hat{c})}/\sqrt{\text{Var}(\hat{a})})$	0.301	0.205	−0.501	0.215
4 conditioning variables	(cay, cg, smb, R_m)			
Slope for log $(\sqrt{\text{Var}(\hat{b})}/\sqrt{\text{Var}(\hat{a})})$	0.406			
Slope for log $(\sqrt{\text{Var}(\hat{c})}/\sqrt{\text{Var}(\hat{a})})$	0.125			

	Efficient GMM		
	4 IV	8 IV	12 IV
Slope for log $(\sqrt{\text{Var}(\hat{b})}/\sqrt{\text{Var}(\hat{a})})$	0.01	0.10	0.41
Slope for log $(\sqrt{\text{Var}(\hat{c})}/\sqrt{\text{Var}(\hat{a})})$	0.12	0.07	0.31

Notes: We price seventeen industry portfolios. Note that with 12 IV, the smallest sample has size $T = 244$ to avoid numerical instabilities.

References

Ai, C., and X. Chen. 2003. Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions. *Econometrica* 71: 1795–1843.

Ai, C., and X. Chen. 2007. Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models with Different Conditioning Variables. *Journal of Econometrics* 141: 5–43.

Angrist, J., G. Imbens, and A. Krueger. 1999. Jackknife Instrumental Variables Estimation. *Journal of Applied Econometrics* 14: 57–67.

Antoine, B., H. Bonnal, and E. Renault. 2007. On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Euclidean Empirical Likelihood. *Journal of Econometrics* 138: 461–487.

Antoine, B., and P. Lavergne. 2014. Conditional Moment Models under Semi-Strong Identification. *Journal of Econometrics* 182: 59–69.

Antoine, B., and P. Lavergne. 2018. "Identification-Robust Non-Parametric Inference in a Linear IV Model." SFU Working Paper.

Antoine, B., and E. Renault. 2009. Efficient GMM with Nearly-Weak Instruments. *Econometrics Journal* 12: S135–S171.

Atchadé, Y. F., and M. D. Cattaneo. 2014. A Martingale Decomposition for Quadratic Forms of Markov Chains (with Applications). *Stochastic Processes and Their Applications* 124: 646–677.

Bierens, H. J. 1982. Consistent Model Specification Tests. *Journal of Econometrics* 20: 105–134.

Chao, J. C., and N. R. Swanson. 2005. Consistent Estimation with a Large Number of Weak Instruments. *Econometrica* 73: 1673–1692.

Darolles, S., C. Gourieroux, and J. Jasiak. 2006. Structural Laplace Transform and Compound Autoregressive Models. *Journal of Time Series Analysis* 27: 477–503.

Dehling, H., and M. Wendler. 2010. Central Limit Theorem and the Bootstrap for U-Statistics of Strongly Mixing Data. *Journal of Multivariate Analysis* 101: 126–137.

Fama, E. F., and K. R. French. 1995. Size and Book-to-Market Factors in Earnings and Returns. *The Journal of Finance* 50: 131–155.

Fama, E. F., and K. R. French. 1996. Multifactor Explanations of Asset Pricing Anomalies. *The Journal of Finance* 51: 55–84.

Fan, Y., and Q. Li. 1996. Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms. *Econometrica* 865–890.

Fan, Y., and Q. Li. 2000. Consistent Model Specification Tests. *Econometric Theory* 16: 1016–1041.

Fang, Y., Y. Ren, and Y. Yuan. 2011. Nonparametric Estimation and Testing of Stochastic Discount Factor. *Finance Research Letters* 8: 196–205.

Ferson, W. E., and C. R. Harvey. 1999. Conditioning Variables and the Cross Section of Stock Returns. *The Journal of Finance* 54: 1325–1360.

Frazier, D. T. 2018. A Simple Iterative Z-Estimator for Semiparametric Models. *Econometric Theory (Available Online)* 1–31.

Gagliardini, P., C. Gourieroux, and E. Renault. 2011. Efficient Derivative Pricing by the Extended Method of Moments. *Econometrica* 79: 1181–1232.

Gagliardini, P., and D. Ronchetti. 2016. "Comparing Asset Pricing Models by the Conditional Hansen-Jagannathan Distance." Working Paper.

Gallant, A. R., and H. White. 1988. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. New York: Blackwell.

Garcia, R., R. Luger, and E. Renault. 2003. Empirical Assessment of an Intertemporal Option Pricing Model with Latent Variables. *Journal of Econometrics* 116: 49–83.

Gospodinov, N., R. Kan, and C. Robotti. 2014. Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Risk Factors. *Review of Financial Studies* 27: 2139–2170.

Gospodinov, N., R. Kan, and C. Robotti. 2018. Too Good to Be True? Fallacies in Evaluation Risk Factor Models. *Journal of Financial Economics* (Forthcoming).

Gospodinov, N., and T. Otsu. 2012. Local GMM Estimation of Time Series Models with Conditional Moment Restrictions. *Journal of Econometrics* 170: 476–490.

Gouriéroux, C., and J. Jasiak. 2006. Autoregressive Gamma Processes. *Journal of Forecasting* 25: 129–152.

Gourieroux, C., A. Monfort, and A. Trognon. 1984. Pseudo Maximum Likelihood Methods: Theory. *Econometrica* 52: 681–700.

Hall, A. R. 2005. *Generalized Method of Moments*. New York: Oxford University Press.

Hall, A. R., and A. Inoue. 2003. The Large Sample Behaviour of the Generalized Method of Moments Estimator in Misspecified Models. *Journal of Econometrics* 114: 361–394.

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

Hansen, L. P., and S. F. Richard. 1987. The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models. *Econometrica* 587–613.

Hausman, J. A., W. K. Newey, T. Woutersen, J. C. Chao, and N. R. Swanson. 2012. Instrumental Variable Estimation with Heteroskedasticity and Many Instruments. *Quantitative Economics* 3: 211–255.

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

Jones, M., S. Davies, and B. Park. 1994. Versions of Kernel-Type Regression Estimators. *Journal of the American Statistical Association* 89: 825–832.

Khrapov, S., and E. Renault. 2016. "Affine Option Pricing Model in Discrete Time." Working Paper.

Kitamura, Y., G. Tripathi, and H. Ahn. 2004. Empirical Likelihood-Based Inference in Conditional Moment Restriction Models. *Econometrica* 72: 1667–1714.

Lavergne, P., and V. Patilea. 2013. Smooth Minimum Distance Estimation and Testing with Conditional Estimating Equations: Uniform in Bandwidth Theory. *Journal of Econometrics* 177: 47–59.

Lettau, M., and S. Ludvigson. 2001a. Consumption, Aggregate Wealth, and Expected Stock Returns. *The Journal of Finance* 56: 815–849.

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

Lewbel, A. 2007. A Local Generalized Method of Moments Estimator. *Economics Letters* 94: 124–128.

Ludvigson, S. C. 2013. "Advances in Consumption-Based Asset Pricing: Empirical Tests." In George M. Constantinides, Milton Harris and Rene M. Stulz (eds.), *Handbook of the Economics of Finance*, Vol. 2, Part B, pp. 799–1612. Amsterdam: Elsevier.

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

Nagel, S., and K. J. Singleton. 2011. Estimation and Evaluation of Conditional Asset Pricing Models. *The Journal of Finance* 66: 873–909.

Newey, W. K., and F. Windmeijer. 2009. Generalized Method of Moments with Many Weak Moment Conditions. *Econometrica* 77: 687–719.

Pastorello, S., V. Patilea, and E. Renault. 2003. Iterative and Recursive Estimation in Structural Nonadaptive Models. *Journal of Business & Economic Statistics* 21: 449–509.

Peñaranda, F., J. M. Rodríguez-Poo, and S. Sperlich. 2017. "Nonparametric Specification Testing of Conditional Asset Pricing Models." Working Paper.

Peñaranda, F., and E. Sentana. 2015. A Unifying Approach to the Empirical Evaluation of Asset Pricing Models. *Review of Economics and Statistics* 97: 412–435.

Powell, J. L., J. H. Stock, and T. M. Stoker. 1989. Semiparametric Estimation of Index Coefficients. *Econometrica* 1403–1430.

Roussanov, N. 2014. Composition of Wealth, Conditioning Information, and the Cross-Section of Stock Returns. *Journal of Financial Economics* 111: 352–380.

Smith, R. J. 2007. Efficient Information Theoretic Inference for Conditional Moment Restrictions. *Journal of Econometrics* 138: 430–460.

Wang, K. Q. 2003. Asset Pricing with Conditioning Information: A New Test. *The Journal of Finance* 58: 161–196.

White, H. 1994. *Estimation, Inference and Specification Analysis*. Econometric Society Monographs, No. 22. Cambridge, UK: Cambridge University Press.

Yoshihara, K-i. 1976. Limiting Behavior of U-Statistics for Stationary, Absolutely Regular Processes. *Zeitschrift für Wahrscheinlichkeitstheorie Und Verwandte Gebiete* 35: 237–252.

Zheng, J. X. 1996. A Consistent Test of Functional Form via Nonparametric Estimation Techniques. *Journal of Econometrics* 75: 263–289.