Rejoinder:
Pseudo-True SDFs in Conditional Asset Pricing Models

Bertille Antoine[*], Kevin Proulx[†], and Eric Renault[‡]

May 1, 2020[§]

We are extremely grateful to the discussants appointed by the co-Editors of *Journal of Financial Econometrics* for their thorough and constructive comments on our paper. Lars P. Hansen has chosen to set the focus on the population analysis while Patrick Gagliardini and Diego Ronchetti have co-authored an impressive contribution mainly devoted to (statistical) comparison of estimators. Furthermore, Sydney Ludvigson, on the one hand, Raymond Kan and Cesare Robotti, on the other, have shared their discussion between the two dual issues of population analysis and inferential methods. We are also grateful to Rachidi Kotchoni for his discussion at the 2018 CIREQ Econometrics Conference on "*Recent Advances in the Method of Moments*" that was insightful about identification issues.

The four discussions appointed by *Journal of Financial Econometrics* are right to the point, and as a result, share some common observations. For this reason, we have chosen to organize our rejoinder according to the relevant themes of discussion. We first address in section 1 the reviewers' questions about our focus on a misspecified SDF rather than a valid one (the closest to our misspecified one), as well as the status of the positivity constraint. Section 2 considers the related issue of a state dependent pseudo-true value. In our paper, we study a pseudo-true value which, albeit focused on conditional pricing errors, is time invariant, and minimizes a mean squared error

[*]Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A1S6, Canada, email: bertille_antoine@sfu.ca

[†]Analysis Group, Boston, MA, USA, email: kevin.proulx@analysisgroup.com

[‡]Department of Economics, Warwick University, UK, email: Eric.Renault@warwick.ac.uk

[§]Disclaimer: The views expressed in this publication are those of the authors. They do not reflect the opinions or views of Analysis Group.

of conditional pricing. We review possible alternatives based on local GMM (Lewbel, 2007), or Extended Method of Moments (Gagliardini, Gouriéroux and Renault, 2011). Section 3 discusses issues related to estimation of pseudo-true values through a choice of instruments (or managed portfolios), the corresponding estimation dilemma, as well as weak identification issues. Finally, section 4 considers different nonparametric estimation techniques (in particular, kernel smoothing and SMD) for assessing a time invariant pseudo-true value, their relative performances, and the role of conditioning information.

Whenever necessary, the discussants' questions are rewritten with the notations of our paper.

1 Why not a valid SDF?

1.1 Pseudo-true value of structural parameters

L.P. Hansen starts with the following question: "*Why not use the misspecified stochastic discount factor $m(\theta^*)$ to select the valid SDF M^* that is closest to $m(\theta^*)$ and prices vector returns correctly?*". There is no doubt that such a valid SDF is an object of interest, in particular because "*the better the model, the smaller the adjustment*" from $m(\theta^*)$ to M^*. This question is fully sensible when the main goal is to evaluate asset pricing models by comparing their HJ distances. However, as noted by S. Ludvigson, the computation of the HJ distance "*also provides an implicit recommendation on how to choose* θ". This pseudo-true value θ^* of θ is the focus of interest of our paper. It seems to us that there are also economic motivations for choosing a structural parametric model that complement the statistical motivations such as

(i) "*Nonparametric estimators have slower rates of convergence and this may make the practical construction of M^* less reliable*" as acknowledged by L.P. Hansen;

(ii) Models that are not tightly specified with structural content may suffer from over-fitting and poor out of sample performance;

there are also economic motivations. For example, the plausible value of a risk aversion parameter has been the focus of interest for years of research about the equity premium puzzle; disentangling the risk aversion parameter from the elasticity of intertemporal substitution and correctly interpreting the risk aversion parameter when agents are endowed with the recursive preferences of Epstein and Zin (1989) has also been another debate (Garcia, Renault, and Semenov (2006)).

More generally, the specific (pseudo-true) value θ^* of a vector θ of structural parameters is, in our opinion, an object of interest. We refer in the paper to the philosophy promoted by Lettau and Ludvigson (2001), stating that, 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 that are well priced by the pseudo-true SDF $m(\theta^*)$ irrespective of the fact that the corrected SDF M^* prices correctly all the asset returns at hand. This is the reason why we enhance the ability of the pseudo-true SDF $m_{t+1}(\theta_t)$ (resp. $m_{t+1}(\theta^*)$) to price exactly at each point of time t (resp. exactly on average over time) all the local factors.

1.2 Positivity constraint

L.P. Hansen completes his first question with the following one: "*Why not impose that $M > 0$ with probability one in the constraint set for minimization?*". As far as our focus of interest remains the pseudo-true value θ^* (or, more generally the conditional one θ_t) rather than the corrected SDF M^*, the information content of the additional constraint $M > 0$ is unclear. The point is that the essential information comes from the choice of the structural model $\mathcal{L} = \{m_{t+1}(\theta); \theta \in \Theta \subset \mathbb{R}^p\}$. If the structural model is a conditionally affine model (resp. exponentially conditionally affine model), there is no way to make the pseudo-true SDF nonnegative (resp. it is automatically nonnegative) irrespective of maintaining or not the additional constraint $M > 0$.

Again, it seems to us that the only potential benefits of imposing a nonnegativity constraint are statistical ones when considering inference on the modified SDF M^* rather than on the structural model. In their discussion, R. Kan and C. Robotti dub "*unconstrained HJ distance*" (resp. "*constrained HJ distance*") the distance between the proposed SDF $m_{t+1}(\theta)$ and the set of admissible SDFs (resp. nonnegative admissible SDFs). They recall the in-depth analysis of Gospodinov, Kan and Robotti (2016) where "*the equivalence of the asymptotic distributions of the sample constrained and constrained HJ-distance tests*" is shown. Therefore, there is no clear benefit in taking the constraint into account for specification testing. As far as estimation is concerned, it is intriguing to think about the potential use of the analysis of Moon and Schorfheide (2009) regarding empirical likelihood estimators where inequality moment conditions provide over-identifying information. While they show that the use of this information leads to a reduction of the asymptotic mean-squared estimation error, one may wonder if over-identifying information like

$$E[MZ] \geq 0, \forall Z \geq 0$$

may be relevant for moment-based estimation of a valid SDF M.

2 Why not a State-Dependent Pseudo-True Value?

L.P. Hansen's third question asks: "*Why not let the "parameter" θ be scaling of the underlying factors for conditional factor models or their exponential counterparts (...) and solve conditional minimization (...)?*". As explained below, it is arguably the most important question since it is at the core of why one may be interested in pseudo-true values.

As already mentioned, the evaluation of asset pricing models through their HJ distances is not our main interest. That being said, and as noticed by R. Kan and C. Robotti, by setting the focus on the average HJ distance $\delta^2(\theta)$ (and not on the conditional one $\delta^2(\theta)[I(t)]$), we overlook the "*potentially interesting task of determining how different models perform relative to each other over time*". We actually share an alternative viewpoint where an excessively time-varying minimum pricing error is a shortcoming for the practical use of an asset pricing model; accordingly, it then makes sense to minimize the mean squared pricing error over time.

However, the main question remains: why not perform "conditional parameter estimation" of a state dependent pseudo-true value θ_t? To address this important issue, it is worth adopting the point of view put forward by Lewbel (2007).

2.1 Estimation of a state dependent pseudo-true value

We are working on the n conditional moment restrictions:

$$E[\psi_{t+1}(\theta)\,|I(t)] \;=\; 0 \quad \textit{with} \quad \psi_{t+1}(\theta) = m\,[I(t+1), \theta]\,R_{t+1} - 1_n \tag{1}$$

When setting the focus on a state-dependent pseudo-true θ_t, we choose to see θ_t as an additional random vector of parameters in the estimating functions $\psi_{t+1}(\theta_t)$. As noted by Lewbel (2007), it may be worth simplifying the general theory of Ai and Chen (2003) regarding conditional moment restrictions with unknown functions by imposing that the unknown function depends on the state of nature only through the conditioning information:

$$\theta_t = \tilde{\theta}[I(t)] \tag{2}$$

where $\tilde{\theta}[.]$ is a nonparametric deterministic function to estimate. The extension of Gagliardini, Gouriéroux and Renault (2011) (GGR hereafter) of Lewbel's (2007) i.i.d. setting to time series is especially convenient for us since the interpretation of a state-dependent pseudo-true value is fully consistent with (2). Such a framework puts on the table two alternative interpretations of the concept of misspecification of an asset pricing model:

- Either we consider that the asset pricing model (1) is misspecified because it assumes a constant vector θ of structural parameters, but we do not question the existence of a state dependent solution $\tilde{\theta}[I(t)]$ of:

$$E[\psi_{t+1}(\tilde{\theta}[I(t)])\,|I(t)] = 0 \tag{3}$$

- Or, we consider that the relaxation (3) of model (1) is still misspecified, meaning that $\tilde{\theta}[I(t)]$ is defined as a pseudo-true value (denoted by θ_t in the paper) that minimizes the conditional HJ distance $\delta^2(\theta)\,[I(t)]$ without putting it at zero.

Lewbel (2007) develops a local GMM methodology under the maintained assumption (3) with the conditioning information $I(t)$ summarized by some state variable vector X_t,

$$\tilde{\theta}[I(t)] = \theta(x)$$

for all possible value x of the random vector X_t. Up to replacing expectations and variances by conditional ones given $X_t = x_t$, the pattern of local GMM formulas (objective function, asymptotic distribution of GMM estimators) looks similar to standard GMM based on unconditional expectations. However, while we end up with a pointwise asymptotic normal distribution of the estimator $\hat{\theta}^{(T)}(x)$ for any possible value x of X_t, the rate of convergence of this estimator is nonparametric, and determined by the nonparametric estimator $\hat{E}^{(T)}[\psi_{t+1}(\theta)\,|X_t]$ of the conditional expectation operator $E[\psi_{t+1}(\theta)\,|X_t]$. Even though this asymptotic theory has not been developed in the literature, it should characterize the asymptotic distribution of $\hat{E}^{(T)}[\psi_{t+1}(\hat{\theta}^{(T)}(x))\,|X_t = x_t]$ and the associated J-test of overidentification of the null hypothesis:

$$\exists \theta_t : E[\psi_{t+1}(\theta_t)\,|X_t = x_t] = 0 \tag{4}$$

The null hypothesis associated with (4) is defined at each date $t_0 = 1, ..., T$. For the test of a given $H_0(t_0)$, the complete dataset at dates $t \neq t_0$ is used to define the nonparametric estimator of the conditional expectation operator $E[\psi_{t+1}(\theta)\,|X_t]$. Note that our discussion is simplified by only contemplating individual tests of the hypotheses $H_0(t_0)$, $t_0 = 1, ..., T$; of course, joint tests could also be performed. In any event, there are two cases of interest:

- Either we fail, in general, to reject the null hypotheses $H_0(t_0)$, $t_0 = 1, ..., T$, and so we have estimated a sequence $\theta_t, t = 1, ..., T$ of conditional true values. The model is misspecified because the true values are time varying. There is still room to make inference on our constant pseudo-true value θ^* that minimizes the average HJ distance $\delta^2(\theta)$ while the conditional ones $\delta^2(\theta)[I(t)]$ actually define state-dependent true values. For inference on these conditional true values, a dynamic extension of the results of Lewbel (2007) (see GGR) is available.

- Or, the null hypotheses $H_0(t_0)$, $t_0 = 1, \cdots, T$, are generally rejected, which means that the asset pricing model is misspecified date by date. Then, the results of Lewbel (2007) do not apply anymore. The asymptotic theory of Lewbel (2007) should be revisited to account for misspecification, as thoroughly as Hall and Inoue (2003) had to revisit the GMM theory of Hansen (1982) to accommodate misspecification. One may wonder whether such a study is worthwhile. After all, what is the point to resort to a structural asset pricing model if the model does not maintain any structural constraint? This is the reason why we have rather chosen to set the focus on the constant pseudo-true value θ^*.

2.2 Additional structural restrictions

In our opinion, working on a sequence of conditional pseudo-true values is worth doing only if one introduces some additional restrictions to keep some structural features for our misspecified asset pricing model. We know at least two examples of this strategy in the extant literature. In both cases, some additional structural restrictions are reckoned to recover estimators with root T convergence, at least in some directions of the parameter space.

GGR have extended Lewbel (2007) by considering that, besides some local moment restrictions,

$$E[\psi_{t_0+1}(\theta)\,|X_{t_0} = x_{t_0}] = 0 \tag{5}$$

there are some global moment restrictions,

$$E[\varphi_t(\theta)] = 0, \quad \forall t = 1, ..., T \tag{6}$$

Note that the local moment restrictions (5) mean that the solution θ may depend on the state at date t_0. Since we set the focus on this particular date (rather than considering local restrictions at other dates), our notation does not need to explicitly account for the fact that θ may be time-varying. Moreover, the global moment restrictions (6) may be not informative enough to fully identify a constant vector θ; they may only define a few components of the parameter vector (or some linear combinations) that are time invariant and may be estimated at a parametric rate of convergence. GGR deliver the comprehensive asymptotic theory for inference on θ under the joint set of moment restrictions (5) and (6). Some linear combinations of θ have root T consistent estimators while others are only consistently estimated at nonparametric rates. GGR had in mind option pricing when the underlying asset and the short term bond price properly discounted by a SDF $m_{t+1}(\theta)$ and relevant instruments have dynamics conformable to global moment restrictions (6). By contrast, the option prices are only considered at time t_0 (with various strike prices and maturities) and only provide state dependent conditional information. Typically, the price of volatility risk is only identified through option price data, is estimated at a nonparametric rate, and depends on the special state at time t_0.

The bottom line is that we never question that the state dependent directions in the parameter space are about true values, and not pseudo-true values. It is also the case of an alternative approach recently put forward by Creal et al. (2020) who, similarly to GGR, allow some directions in the parameter space to be state dependent. However, by contrast with GGR, they do not give up parametric rate of convergence to estimate these directions, because they add some information about their dynamics and the possibility to filter them. The key idea is to define some "observation driven filtering" that generalizes the former literature (see e.g. Creal, Koopman and Lucas (2013) on score driven generalized autoregressive models with time varying parameters). Such additional filtering information is arguably structural, even though not explicitly based on economic theory. Once again we are talking about state dependent true values, as there is no such thing as state dependent pseudo-true values.

3 Why not Managed Portfolios?

3.1 An identification issue

As explained in the paper, a fixed pseudo-true value $\theta^* \in \Theta \subset \mathbb{R}^p$ should be characterized by the p first-order conditions,

$$E\left\{E\left[\frac{\partial m_{t+1}(\theta^*)}{\partial \theta} R'_{t+1} \left| I(t)\right.\right] \Omega^{-1}\left[I(t)\right] E[m_{t+1}(\theta^*) R_{t+1} - 1_n \left| I(t)\right.]\right\} = 0 \qquad (7)$$

These first-order conditions average out over the conditioning information $I(t)$ the local conditions that define a state-dependent (pseudo) true value θ_t as solution of

$$E\left[\frac{\partial m_{t+1}(\theta_t)}{\partial\theta} R'_{t+1}\,|I(t)]\right]\Omega^{-1}[I(t)]\,E[m_{t+1}(\theta_t)\,R_{t+1}-1_n\,|I(t)]=0 \tag{8}$$

We have proposed to interpret these definitions through the concept of local factors,

$$F^L_{t+1}(\theta)=\frac{\partial m_{t+1}(\theta)}{\partial\theta}$$

For the case 1 of conditionally affine SDF, the local factors are:

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

so that with an appropriate choice of factors, we can assume that

$$\texttt{Rank}\left\{E\left[F^L_{t+1}(\theta)F^L_{t+1}(\theta)'\,|I(t)]\right]\right\}=p \tag{9}$$

This rank condition should also be fulfilled in the case of an exponentially affine model. With the notations of the paper:

$$m_{t+1}(\theta)=\exp\left[\tilde{m}_{t+1}(\theta)\right]\;\Rightarrow\;F^L_{t+1}(\theta)=m_{t+1}(\theta)\begin{bmatrix}1\\ F_{t+1}\end{bmatrix}$$

By contrast, for the case 2 of conditionally affine SDF ("scaled multifactor model" as in Lettau and Ludvigson (2001)), the local factors are:

$$F^L_{t+1}(\theta)=W_t\otimes F_{t+1}$$

and, without additional equality constraints on the parameter θ, it is obviously associated with some rank deficiency:

$$\texttt{Rank}\left\{E\left[F^L_{t+1}(\theta)F^L_{t+1}(\theta)'\,|I(t)]\right]\right\}=q<p$$

As shown in the paper, one can always define a subvector $\tilde{F}^L_{t+1}(\theta)$ of $F^L_{t+1}(\theta)$ of size q, such that the square matrix $E\left[\tilde{F}^L_{t+1}(\theta)\tilde{F}^L_{t+1}(\theta)'\,|I(t)]\right]$ is nonsingular. On the one hand, (7) is equivalent to exact unconditional pricing of $\tilde{F}^L_{t+1}(\theta^*)$:

$$E\left[m_{t+1}(\theta^*)\,\tilde{F}^L_{t+1}(\theta^*)\right]=1_q \tag{10}$$

On the other hand, (8) is equivalent to exact conditional pricing of $\tilde{F}^L_{t+1}(\theta_t)$:

$$E[m_{t+1}(\theta_t)\,\tilde{F}^L_{t+1}(\theta_t)\,|I(t)]=1_q \tag{11}$$

Obviously, when $q < p$, θ^* is not identified and cannot be consistently estimated, irrespective of the estimation method: local GMM as in Gagliardini and Ronchetti's work, or SMD as in our paper. A solution would be an XMM-type approach as put forward by GGR. In our context, it would amount to assume that there is a specific date t_0 such that the conditional pricing equations (11) are fulfilled by $\theta_t = \theta_{t_0}$ at all dates $t = 1, ..., T$ for at least a $\breve{q}$-dimensional ($\breve{q} \leq q$) subset $\breve{F}^L_{t+1}(\theta)$ of factors $\tilde{F}^L_{t+1}(\theta_t)$:

$$E[m_{t+1}(\theta_{t_0})\, \breve{F}^L_{t+1}(\theta_{t_0})\, |I(t)] = 1_{\breve{q}}, \quad \forall t = 1, ..., T$$

This would imply that some linear combinations of θ_{t_0} are not state dependent and would provide additional identifying information about θ^*. Such an extension from local GMM to XMM is beyond the scope of this paper, and we maintain the rank condition (9).

An alternative approach would be to live with lack of identification. Korsaye, Quaini and Trojani (2020) define an "APT-consistent" SDF such that, it first exactly prices the traded factors (our condition (10)), and, second, it implies pricing errors that satisfy some APT bound. In other words, the pseudo-true SDF is not unique, but the focus is set on those SDFs which, albeit not valid, satisfy some misspecification threshold. This is an interesting illustration of the general strategy suggested by L.P. Hansen in his discussion: "*Why the focus on inference about a pseudo-true parameter vector instead of on the more general construction of a set of parameters that satisfies a pre-specified misspecification bound?*".

This question paves the way for an interesting extension of partial identification in asset pricing. In the first *Journal of Financial Econometrics* Invited Lecture (see also Kaido and White (2009)), Halbert White[1] showed that "*under incomplete markets, the market price of risk is not point-identified but is instead identified as a bounded subset of an affine subspace*". For us, the pseudo-true SDF is only required to price a set of q local factors which may not be sufficient (when $q < p$) to uniquely identify the pseudo-true value of the SDF $m_{t+1}(\theta^*)$ - just like with incomplete markets. This would lead to define as target of inference an identified set as suggested above by L.P. Hansen.

[1]We are pleased that this Halbert White Jr. Memorial JFEC Invited Lecture gives us an opportunity to revisit H. White's seminal work.

3.2 The estimation dilemma

While maintaining the assumption of point identification of the pseudo-true value θ^*, we fully concur with S. Ludvigson's statement that estimation of a pseudo-true value "*poses a fundamental dilemma*". As mentioned in the paper, the first-order conditions (7) seem to define a vector of "optimal instruments":

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

These instruments, dubbed "HJ-optimal instrument matrix" by Gagliardini and Ronchetti in their discussion, may look a bit simpler than the genuine optimal instruments for efficient GMM since, in line with the philosophy of the HJ distance, the weighting matrix $\Omega[I(t)]$ does not depend on unknown parameters (that would call for a first step estimator). Moreover, as it is well-known, the first-order conditions can be rewritten and reinterpreted as pricing equations for managed portfolios with payoffs $Z_t(\theta^*) R_{t+1}$:

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

However, as rightly noted by S. Ludvigson, when the SDF $m_{t+1}(\theta)$ is a nonlinear function of unknown θ - that is, when we go beyond conditional affine SDFs - these optimal instruments do not "*in general present a feasible estimation strategy*". This is in sharp contrast with the case of efficient GMM where one faces a seemingly analog problem. Following Newey (1990), "a problem in constructing efficient instrumental variables estimators for such models is that the optimal instruments involve a conditional expectation (...) Nonparametric methods provide a way of avoiding this difficulty". Newey (1990) promotes in particular series approximation which may be applied in two ways:

- Either using a countable basis of the space of $I(t)$-measurable functions, with the hope that, by picking a large number J_T of instruments $Z_{j,t}, j = 1, \cdots, J_T$ in this base, one will span asymptotically the optimal instruments $Z_t(\theta^*)$. For J_T going to infinity at a proper rate with T, the GMM estimator based on managed portfolios with payoffs $Z'_{j,t} R_{t+1}, j = 1, \cdots, J_T$ will be asymptotically equivalent to efficient GMM.

- Or, computing a nonparametric estimator of $E\left[\frac{\partial m_{t+1}(\theta)}{\partial \theta} R'_{t+1} \left| I(t)\right.\right]$ where θ is replaced by a first-step consistent estimator $\tilde{\theta}_T$; possibly with a series estimator, or a nearest neighbor estimator as in Newey (1990).

Unfortunately, none of these classical strategies will work in our context - hence the aforementioned estimation dilemma. Since there is no such thing as a true value of θ, each choice of a set $(Z_{j,t})_{1\leq j\leq J}$ of instruments will define a different pseudo-true value, that may not even exist; and there is no reason to believe that we will capture the optimal instruments $Z_t(\theta^*)$ when J goes to infinity. Moreover, a first-stage consistent estimator of θ^* is not available. Therefore, any choice of an initial value θ^1 to compute instruments $Z_t(\theta^1)$ will also lead to different pseudo-true values. This is a problem similar to the issue met by Hall and Inoue (2003) for GMM with misspecification when any value of θ chosen to compute a weighting matrix will lead to a different pseudo-true value (they consider unconditional moment restrictions, so that there is no issue of optimal instruments). As explained in the paper, the only hope would be the convergence of an iterative strategy, thanks to some contraction mapping property whose validity is not easy to check.

The bottom line is well summarized by S. Ludvigson: "*it seems clear from the above discussion that any resolution of this dilemma must (in general) abandon the use of some finite number Z_ts (or choice of arbitrary managed portfolios)*" to give some clear empirical content to the pseudo-true value. This provides the motivation for nonparametric approaches described in section 4.

3.3 Weak Identification

As recalled by Kan and Robotti, "*Dominguez and Lobato (2004) argue that, despite its computational attractiveness, the standard GMM approach of Hansen (1982) based on unconditional moment restrictions may result in efficiency losses and inconsistencies that arise from possible nonidentifiability of the parameters of interest by the unconditional moment restrictions even when the conditional moment restrictions identify the parameters*". We much appreciate that Kan and Robotti interpret our work as aiming "*at shedding some light on these issues*" for the conditional HJ-distance. We have initiated above a discussion on the tension between identification from the conditional moment restrictions versus unconditional ones. However, we acknowledge that "*a rigorous treatment of the HJ distance metric in the presence of conditional moment restrictions is still in its infancy*" and much work remains to be done.

That being said, it is worth contemplating some extension of the Dominguez and Lobato's (2004) contribution, by considering in our context the possibility that the so-called optimal instruments $Z_t(\theta^*)$ may not be the most efficient ones to protect us against identification issues. Let us consider more generally the estimation of the pseudo-true value of interest from the moment conditions:

$$E[m_{t+1}(\theta) Z_t R_{t+1} - Z_t 1_n] = 0 \tag{12}$$

where Z_t is a given vector of H ($\geq p$) managed portfolios (with payoffs $Z_t R_{t+1}$). When considering these moment conditions, it is important to keep in mind that:

- First, if we do not maintain the assumption that the state dependent parameters θ_t do not depend on the state $I(t)$, we do not know whether equations (12) admit at least a solution θ. In other words, we may have to follow Hall and Inoue (2003) to define a pseudo-true value as minimizing a norm of the expected vector in (12).

- Second, irrespective of how the moment conditions (12) may define a pseudo-true value, it will, in general, depend on the choice of the managed portfolios. We will denote it below by θ^*, without explicit accounting for its dependence on the choice of the managed portfolios used in (12) and, possibly, on a weighting matrix.

In spite of the dependence of the pseudo-true value θ^* on the choice of managed portfolios Z_t, it is worth studying to what extent different choices of managed portfolios Z_t may lead to different levels of identification strength. Antoine and Renault (2020) define identification strength through a sequence of deterministic nonsingular matrices M_T of size p such that

$$\lim_{T=\infty} \frac{\partial \rho_T}{\partial \theta'}(\theta^*) M_T = \Gamma(\theta^*)$$

where the limit matrix $\Gamma(\theta^*)$ exists, is full-column rank, and the sequence of functions $\rho_T(\theta)$ define the moment conditions

$$\rho_T(\theta) = E[m_{t+1}(\theta) Z_t R_{t+1} - Z_t 1_n].$$

These moments may depend on the sample size T since we consider a drifting DGP to accommodate weak moment asymptotics *à la* Stock and Wright (2000). Typically, standard (strong) identification is associated with a sequence of matrices M_T that can be chosen constant (such as the identity matrix without loss of generality), while weak identification corresponds to a sequence of matrices that blows up to infinity, possibly as fast as $\sqrt{T}$. Antoine and Renault (2020) set the focus (and test for) intermediate cases of nearly-strong/nearly-weak identification where $\|M_T\| = o(\sqrt{T})$, but where the sequence M_T may diverge to infinity in different directions, albeit slower than $\sqrt{T}$:

the slower the divergence of M_T, the stronger the identification of θ^* in corresponding directions. It is worth interpreting this property in terms of managed portfolios. Since

$$\frac{\partial \rho_T}{\partial \theta'}(\theta^*) = E\left[Z_t R_{t+1} \frac{\partial m_{t+1}}{\partial \theta'}(\theta^*)\right] = E\left[Z_t R_{t+1} F_{t+1}^L(\theta^*)'\right]$$

we see that

$$\frac{\partial \rho_T}{\partial \theta'}(\theta^*) M_T = E\left[Z_t R_{t+1}\left(M_T' F_{t+1}^L(\theta^*)\right)'\right]$$

Therefore, up to a risk free asset, the identification strength is characterized by the amount of rescaling of local factors that is called for to (asymptotically) get a full rank covariance matrix with the returns on managed portfolios under test. While this amount of rescaling has been documented in the paper for some popular models, it is worth recalling that variants of SMD have been put on the table by Antoine and Lavergne (2014, 2020) for estimation and inference robust to weak identification. A key idea is precisely to give up projections on instruments to avoid the perverse interplay between the number of instruments and their strength.

4 Inference based on nonparametric estimation of conditional moments

4.1 Different kernel-based nonparametric strategies

As explained above, the use of managed portfolios is problematic in the context of misspecification, and motivated us to resort to kernel smoothing to capture the conditional moments of interest. As summarized in Gagliardini and Ronchetti's discusssion (referred to as GR hereafter), there are basically three possible approaches:

(i) The local GMM estimator;

(ii) The GMM estimator with HJ-optimal instruments;

(iii) The smooth minimum distance (SMD) estimator with fixed kernel bandwidth h which is the focus of sections 6 and 7 in our paper.

The terminology "local GMM" follows from Gospodinov and Otsu (2012) and should not be confused with Lewbel's (2007) method that was also dubbed "local GMM". Lewbel (2007) sets the focus on conditional moment restrictions for only one value of the conditioning variable, so that the parameters of interest are estimated at nonparametric rates of convergence as function of this conditioning value. By contrast, Gospodinov and Otsu's (2012) method belong to the class of localized versions of the generalized empirical likelihood as proposed by Kitamura, Tripathi and Ahn (2004), Antoine, Bonnal and Renault (2007), and Smith (2007). In this case, different conditioning values are considered as uniformly defining the same values of unknown parameters that are then estimated with root-T rates of convergence. Typically, all these localized versions of the generalized empirical likelihood deliver root-T consistent asymptotically normal estimators that reach the semiparametric efficiency bound (with valid conditional moment restrictions). GGR is a mixture of these two approaches, by considering that uniformity (and associated parametric rates of convergence for estimators) is only achieved for some directions in the parameter space.

Gagliardini and Ronchetti's (2016) local GMM estimator is defined as:

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \left[\frac{1}{T} \sum_{t=1}^{T} 1(x_t) \hat{E}[\psi_{t+1}(\theta) \,| x_t]' \, \hat{\Omega}^{-1} (x_t) \, \hat{E}[\psi_{t+1}(\theta) \,| x_t] \right] \tag{13}$$

where $1(x_t)$ is a trimming factor, $\hat{\Omega}(x_t) = \hat{E}[R_{t+1}R'_{t+1} \,| x_t]$, and $\hat{E}[. \,| x_t]$ denotes the Nadaraya-Watson kernel regression estimator. Up to the trimming factor, the criterion function corresponds to the sample counterpart of the population expectation that defines our conditional average HJ distance $\delta^2(\theta)$. In the conditional extension of (efficient) continuously updated GMM (where $\hat{\Omega}(x_t)$ is replaced by a Nadaraya-Watson estimator of the conditional variance of $\psi_{t+1}(\theta)$), Antoine, Bonnal and Renault (2007) showed that such an estimator was asymptotically equivalent to efficient GMM based on a non-parametric estimator of optimal instruments. In their discussion, GR extend this result to the case of HJ-optimal instruments with possible misspecification; see their Appendix A1.

S. Ludvigson highlights as an alternative the methodology of Ai and Chen (2007) which amounts to replacing the Nadaraya-Watson estimator $\hat{E}[\psi_{t+1}(\theta) \,| x_t]$ in (13) by a sieve estimator. Ai and Chen (2007) recognize that, in case of misspecification, the nonparametric estimator of the weighting matrix $\hat{\Omega}^{-1}(x_t)$ may contaminate the estimators of the parameters of interest (including their rate of convergence), and they resort to using the identity weighting matrix (instead of the HJ one). By doing so, they implicitly point out a practical difficulty in the application of local GMM as in Gagliardini and Ronchetti (2016) when, due to misspecification, a long-term variance matrix needs to be estimated (see also below). Up to this difficulty, one should acknowledge how seminal the work of Ai and Chen (2007) was in the context of possibly misspecified conditional moment restrictions. Moreover, the fact that different sets of conditioning variables may be considered in different moment restrictions should allow for some flexibility in the crucial choice of conditioning information for the definition of pseudo-true values. More generally, when nonparametric estimation of conditional expectations is concerned, there are well-documented pros and cons of the respective approaches of kernel smoothing and sieve estimation.

In addition, GR provide an insightful interpretation of our estimation strategy inspired by Lavergne and Patilea's (2013) SMD estimator with fixed kernel bandwidth h. We know that the corresponding pseudo-true value $\theta^*(h)$ depends on the chosen bandwidth h, which can be seen as a calibration parameter. Several comments are in order:

(i) GR show that when $h \to \infty$, the vector $\theta^*(h)$ converges to the unconditional pseudo-true value. This is conformable to intuition: the less we smooth (or the smaller the bandwidth h), the more all relevant conditioning information is taken into account. Even though our pseudo-true value is not state dependent (see section 2 above), for finite h, it does set the focus on conditional pricing errors. GR interestingly show that "*when h increases, the criterion $Q_\infty(\theta, h)$ gives less weight to the unconditional Fourier transforms of the pricing errors for large frequencies of the process $\{x_t\}$*". In other words, there is arguably a trade-off between taking h small (with associated curse of dimensionality) and introducing more state variables in the conditioning information. As we argued, "we are not keen on restricting the information set $I(t)$ since this would also modify the definition of the pseudo-true value". In this respect, we see the fixed bandwidth h as a price to pay for the flexibility to incorporate more state variables.

(ii) As recalled by S. Ludvigson "*in implementation, fixing h (or the number of sieves) is exactly what a researcher faced with a finite amount of data would do*". This is of course our rationale for considering SMD with fixed bandwidth h. Our paper also puts forward a state variables framework (see section 6.2.) that allows us to interpret the SMD estimator similarly to the i.i.d. case so that, following Lavergne and Patilea's (2013) argument, we believe that, even in large samples, there is some rationale for fixed bandwidth h. This is also illustrated by GR: "*as found in APR and confirmed in our numerical experiments in Section 3, there are empirically relevant frameworks in which some components of vector $\theta^*(h)$ vary slowly over relative small values of the parameter h*".

(iii) We expect that for a bandwidth h converging to zero, our pseudo-true value $\theta^*(h)$ will converge towards the genuine conditional pseudo-true value θ^* as estimated by Gagliardini and Ronchetti (2016) with local GMM (or equivalently with HJ optimal instruments). GR confirm our intuition by a formal proof that goes one step further. It is worth keeping in mind that, even when two sets of moment conditions define the same vector of parameters, and by definition two alternative consistent estimators of the same vector, the asymptotic distributions of the two estimators may differ.

What is arguably the most impressive contribution of GR's thorough discussion is the fact that they are able to question the asymptotic equivalence of local GMM and SMD (with bandwidth h converging to zero). In our paper (see page 23), we present the loss of a martingale difference property due to misspecification as "one of the main motivations for the introduction (...) of the alternative SMD approach that sets the focus on the misspecified model conditional moment restrictions for resorting to a central limit theorem for U-statistics." The asymptotic theory developed by GR confirms that it is precisely this loss of martingale difference sequence property that will drive a wedge between these two methodologies. Of course, this is tightly related to the treatment of relevant conditioning information through state variables.

4.2 The Role of Conditioning Information

In section 6.2., our paper deploys three assumptions (A1, A2, A3) about the role of state variables in summarizing the relevant information about returns and SDF dynamics. Similarly, GR introduce an assumption S that is "*more general than Assumptions A1, A2, A3 (...) because it accommodates for the fact that the conditioning variable x_t used by the econometrician can be a subcomponent of the state vector s_t*". Even though there is no general implication law, it is obviously Assumption A1 in our paper that requires a large enough set of state variables s_t. It basically assumes that the state variables s_t are a sufficient statistic to forecast the next net discounted return:

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

Formula (16) in GR shows that forecasting this vector (and the corresponding Jacobian matrix) is tantamount to forecast the score vector $\varphi_{t+1}(\theta^*)$ of the local GMM estimator. In other words, our assumptions A1, A2, A3 allow the perverse term $\delta(s_t)$ to disappear in the decomposition (18) of the score vector stated in GR. Our assumptions are suggested by a common practice in option pricing (see e.g. Garcia, Luger an Renault

(2003) and references therein). For instance, option pricing models with stochastic volatility characterize option prices as one-to-one functions of latent volatility, while, given the path of the volatility process, consecutive asset returns remain serially independent. In other words, if the econometrician "observes" volatility through option prices data, assumption A1 is fulfilled. The consequences of this observation for econometric inference are extensively discussed in Pastorello, Patilea and Renault (2003). While the exogeneity assumption A2 is rather innocuous for elicited state variables, assumption A3 maintains the absence of leverage effect which is more problematic for option pricing, but also maintained by GR.

For sake of expositional simplicity, we only comment on GR's asymptotic theory under the maintained assumptions A1, A2, A3. In this case, their decomposition (18) of the score vector for local GMM can be written as

$$\begin{aligned} \varphi_{t+1}(\theta^*) &= \xi_{t+1} + g(x_t) \qquad \text{with} \qquad \xi_{t+1} = \varphi_{t+1}(\theta^*) - E[\varphi_{t+1}(\theta^*)\,|I(t)] \\ \text{and} \quad g(x_t) &= E[\varphi_{t+1}(\theta^*)\,|I(t)] = 1(x_t)\, J(x_t,\theta^*)\,\Omega^{-1}(x_t)\, e(x_t,\theta^*) \\ \text{with} \qquad e(x_t,\theta) &= E[\Psi_{t+1}(\theta)\,|I(t)]\,, \quad J(x_t,\theta) = E[\frac{\partial \Psi_{t+1}(\theta)}{\partial \theta'}\,|I(t)] \end{aligned}$$

When the asset pricing model is correctly specified, θ^* is the true unknown value of θ, $e(x_t,\theta^*) = 0$, and the score $\varphi_{t+1}(\theta^*) = \xi_{t+1}$ is a martingale difference sequence (mds hereafter). In this case, we have the standard semiparametric efficiency bound for GMM with conditional moment restrictions (as in the iid case, or as in Gospodinov and Otsu (2012) for time series). Then, local GMM and SMD with bandwidth h going to zero are asymptotically equivalent and miss the efficiency bound only because they do not use the optimal weighting matrix.
When the asset pricing model is misspecified, the additional term $g(x_t)$ is different from zero: it causes the mds property for the score of local GMM to break down, and drives a wedge between local GMM and SMD. According to GR, when the bandwidth h is pushed to zero, the SMD method leads to a score vector:

$$\lim_{h\to 0} \varphi_{t+1}(\theta^*(h), h) = \xi_{t+1} + 2g(x_t)\,.$$

We would guess that the symmetrization of the objective function needed to resort to the asymptotic theory of U-statistics is responsible for this duplication of the scaled pricing error term $g(x_t)$. Since, by definition, the two terms ξ_{t+1} and $g(x_t)$ are not correlated, the duplication of the term $g(x_t)$ can only be responsible for an increase of asymptotic variance of the estimators. Basically, in addition to the variance of the mds part, the local GMM (resp. the SMD with $h \to 0$) will involve once the long term variance (resp. twice the long term variance) of $g(x_t)$. Note that it is indeed a long term variance since $g(x_t)$ is not a mds.

If local GMM appears to deliver an asymptotically more accurate estimator of the pseudo-true value than SMD with bandwidth pushed to zero, GR point out that this ranking is not always guaranteed: for example, in more general situations where the observed set of state variables may not be sufficient to maintain assumption A1. Moreover, as already explained, one may actually prefer to work with SMD and a fixed bandwidth: e.g. to ensure a relevant definition of the pseudo-true value of interest by incorporating many state variables. An alternative (or complementary) strategy may be the use of modern techniques of dimensionality reduction to incorporate all relevant conditioning information through some well chosen statistical summaries. A typical example is the work of Ludvigson and Ng (2009) who resort to dynamic factor analysis to retrieve the relevant conditioning information encapsulated in some macro factors.

References

1. Ai, C. and X. Chen (2003), "Efficient estimation of models with conditional moment restrictions containing unknown functions", *Econometrica*, Vol.71, p1795-1843.

2. Ai, C., and X. Chen (2007), "Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models with Different Conditioning Variables", *Journal of Econometrics*, Vol.141, p5-43.

3. 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*, Vol.138, p461-487.

4. Antoine, B. and P. Lavergne (2014), "Conditional Moment Models under Semi-Strong Identification", *Journal of Econometrics*, Vol.182, p59-69.

5. Antoine, B. and P. Lavergne (2020), "Identification-robust non-parametric inference in a linear IV model", Discussion Papers, Department of Economics, Simon Fraser University, `https://ideas.repec.org/p/sfu/sfudps/dp20-03.html`

6. Antoine, B., and E. Renault (2020), "Testing identification strength", forthcoming at *Journal of Econometrics*.

7. Creal, D., S.J. Koopman, and A. Lucas (2013), "Generalized Autoregressive Score Models with Applications", *Journal of Applied Econometrics*, Vol.28, p777-795

8. Creal, D., S.J. Koopman, A. Lucas, and M. Zamojski (2020), "Observation-Driven Filtering of Time-Varying Parameters Based on Moment Conditions", *Working Paper* available at `https://sites.google.com/view/drewcreal/home`.

9. Dominguez, M., and I. Lobato (2004), "Consistent estimation of models defined by conditional moment restrictions", *Econometrica*, Vol.72, p1601-1615.

10. Epstein L.G., and S.E. Zin (1989), "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework", *Econometrica*, Vol.57, p937-969.

11. Gagliardini, P., C. Gouriéroux, and E. Renault (2011), "Efficient derivative pricing by the extended method of moments", *Econometrica*, Vol.79, p1181-1232.

12. Garcia, R., R. Luger, and E. Renault (2003), "Empirical assessment of an intertemporal option pricing model with latent variables", *Journal of Econometrics*, Vol.116, p49-83.

13. Garcia, R., E. Renault, and A. Semenov (2006), "Disentangling Risk Aversion and Intertemporal Substitution through a Reference Level", *Finance Research Letters*, Vol.3, p181-193.

14. Gospodinov, N., and T. Otsu (2012), "Local GMM estimation of time series models with conditional moment restrictions", *Journal of Econometrics*, Vol.170, p476-490.

15. Hall, A. R., and A. Inoue (2003), "The Large Sample Behaviour of the Generalized Method of Moments Estimator in Misspecified Models", *Journal of Econometrics*, Vol.114, p361-394.

16. Kaido, H., and H. White (2009), "Inference on Risk-Neutral Measures for Incomplete Markets", *Journal of Financial Econometrics*, Vol.7, p199-246.

17. Kitamura, Y., G. Tripathi, and H. Ahn (2004), "Empirical Likelihood-Based Inference in Conditional Moment Restriction Models", *Econometrica*, Vol.72, p1667-1714.

18. Korsaye, S., A. Quaini, and F. Trojani (2019), "Smart SDFs", *Working paper*.

19. Lavergne, P., and V. Patilea, (2013), "Smooth minimum distance estimation and testing with conditional estimating equations: uniform in bandwidth theory", *Journal of Econometrics*, Vol.177, p47-59.

20. Lettau, M., and S. Ludvigson (2001), "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying", *Journal of Political Economy*, Vol.109, p1238-1287.

21. Ludvigson, S., and S. Ng (2009), "Macro Factors in Bond Risk Premia", *Review of Financial Studies*, Vol.22, p5027-5067.

22. Newey, W. (1990), "Efficient Instrumental Variables Estimation of Nonlinear Models", *Econometrica*, Vol.58, p809-837.

23. Pastorello, S., V. Patilea, and E. Renault (2003), "Iterative and Recursive Estimation in Structural Nonadaptive Models", *Journal of Business & Economic Statistics*, Vol.21, p449-482.

24. Smith, R.J. (2007), "Efficient information theoretic inference for conditional moment restrictions", *Journal of Econometrics*, Vol.138, p430-460.