Comparing Asset Pricing Models by the Conditional Hansen-Jagannathan Distance

PATRICK GAGLIARDINI and DIEGO RONCHETTI*

First draft: July 2014 This version: May 2015

Please do not circulate without authors' permission†

Abstract

We compare non-nested parametric specifications of the Stochastic Discount Factor (SDF) in terms of their conditional Hansen-Jagannathan (HJ-) distance. This distance is defined as the discrepancy between a parametric SDF family identifying an asset pricing model and the set of admissible SDF's satisfying the conditional no-arbitrage restrictions for a set of traded assets. The conditional HJ-distance accounts for the models' ability to match the dynamic pricing restrictions for any set of managed portfolios, and not just a set of static restrictions for a specific choice of instruments like the often employed (unconditional) HJ-distance. We estimate the conditional HJ-distance by a kernel-based Generalized Method of Moments estimator and establish its large sample properties for model selection purposes. We demonstrate empirically the usefulness of our approach by comparing several SDF models including preference-based specifications, beta-pricing models and recently proposed SDF models that are conditionally linear in the priced risk factors.

JEL classification: C12, C14, G12.

Keywords: Asset pricing model comparison, stochastic discount factor, Hansen-Jagannathan distance, conditional moment restrictions, nonparametric estimation.

---

*Patrick Gagliardini, University of Lugano, Faculty of Economics, Via Buffi 13, CH-6900 Lugano, Switzerland. Phone: +41 58 6664660. Fax: +41 58 6664734. E-mail: patrick.gagliardini@usi.ch. Diego Ronchetti, University of Groningen, Faculty of Economics and Business, PO Box 800, 9700 AV Groningen, The Netherlands. Phone: +31 50 3633460. Fax: +31 50 3637337. E-mail: d.ronchetti@rug.nl.

†We thank the participants at the AMES 2014 in Taipei, the CMES 2014 in Xiamen, the EEA-ESEM 2014 in Toulouse, the SEA meeting 2014 in Zurich and the EEF seminar in Groningen for useful comments.

Modern asset pricing theories can be formulated in terms of the Stochastic Discount Factor (SDF). The SDF accounts for time discount and risk adjustment in the pricing of a future risky payoff. In a setting characterized by a representative investor with time-separable preferences, the Euler equations for the solution of the investment/consumption problem leads to a SDF that is equal to the intertemporal marginal rate of substitution for consumption multiplied by the time discount rate. As several preference specifications are possible in theory, we are confronted with a large set of alternative SDF models, and there is even more latitude in this choice when reduced-form SDF specifications are considered. Therefore, a central question in empirical asset pricing is how to select the most adequate model among a set of competing non-nested parametric SDF families.

In this paper we adopt the point of view that all competing asset pricing models are potentially misspecified and we compare them in terms of their distance from the set of admissible SDF's. A SDF is considered admissible if it matches the dynamic no-arbitrage pricing restrictions implied by a set of test assets. The goals of our paper are (i) the introduction of a conditional version of the Hansen-Jagannathan (HJ-) distance (Hansen and Jagannathan (1997)), which is the distance currently used to compare asset pricing models, (ii) the study of the properties of this new distance and (iii) the illustration of its use in an empirical comparison of some popular models proposed in the equity pricing literature. We then refer in this paper to the distance described in Hansen and Jagannathan (1997) as the *unconditional HJ-distance*, and to the newly introduced distance as the *conditional HJ-distance*. The latter distance captures the conditional pricing errors implied by the SDF model for any managed portfolio of the test assets, and not just for some portfolios like the unconditional HJ-distance. Hence, the conditional HJ-distance fully exploits the conditioning information when comparing the performance of competing asset pricing models.

In order to introduce the framework of our paper, let us consider an economy with N assets traded in discrete time.[1] Let $P_t := [P_{1,t} \ \dots \ P_{N,t}]'$ denote the vector of the trading prices at time t for the N assets, and let $\mathcal{I}_t$ be the information available to investors at time t. The Absence of Arbitrage Opportunities (AAO) in the market is equivalent to the existence of a scalar

[1] In the empirical application discussed in Sec. III six portfolios representative of the market for publicly traded U.S. equities and the 1-month T-Bill are taken as test assets.

stochastic process $\{M_{t,t+1}\}$ such that the random variable $M_{t,t+1}$ is (i) positive, (ii) measurable with respect to (w.r.t.) the information $\mathcal{I}_{t+1}$ and (iii) satisfying the no-arbitrage restriction $P_t = \mathrm{E}\left[M_{t,t+1}P_{t+1}|\,\mathcal{I}_t\right]$, where $\mathrm{E}\left[\cdot|\,\mathcal{I}_t\right]$ denotes the conditional expectation operator under the historical probability measure given information $\mathcal{I}_t$ (see, e.g., Harrison and Kreps (1979), and Hansen and Richard (1987)).[2] As in most of the modern empirical asset pricing literature, we refer to any random variable $M_{t,t+1}$ satisfying properties (ii) and (iii) as an admissible SDF between dates t and $t+1$. Let us assume that the information $\mathcal{I}_t$ is generated by the Markov process of L random state variables collected in vector X_t admitting values in set $\mathcal{X} \subseteq \mathbb{R}^L$. Moreover, let $R_t := [R_{1,t} \ \ldots \ R_{N,t}]'$ be the N-dimensional vector of the assets' gross returns $R_{i,t} = P_{i,t}/P_{i,t-1}$, for any $i = 1, \ldots, N$. Then, the property (iii) for an admissible SDF can be rewritten as

$$\mathrm{E}\left[M_{t,t+1}R_{t+1} - 1_N|\, X_t = x\right] = 0_N, \tag{1.1}$$

for any $x \in \mathcal{X}$, where 1_N and 0_N are N-dimensional vectors of ones and zeros, respectively.

In a parametric asset pricing model the admissible SDF $M_{t,t+1}$ between dates t and $t+1$ is replaced by a candidate SDF, which is a known function m of a random vector Y_{t+1} parameterized by a vector with unknown value in set $\Theta \subseteq \mathbb{R}^p$. The random vector Y_{t+1} collects some priced risk factors contained in X_{t+1} and potentially also some conditioning variables contained in X_t that generate time-varying risk premia. The set of random variables $m(Y_{t+1};\theta)$ for any value $\theta \in \Theta$ constitutes a parametric SDF family, and it identifies the asset pricing model. If at least one admissible SDF belongs to this parametric SDF family we say that the parametric asset pricing model is correctly specified. In this case the AAO assumption in Eq. (1.1) implies the N-dimensional vector conditional moment restriction

$$\mathrm{E}\left[h(Y_{t+1}, R_{t+1};\theta_0)|\, X_t = x\right] = 0_N \tag{1.2}$$

for any $x \in \mathcal{X}$, for the N-dimensional vector

$$h(Y_{t+1}, R_{t+1};\theta) := m(Y_{t+1};\theta)R_{t+1} - 1_N, \tag{1.3}$$

[2] To simplify the exposition we consider assets that do not pay any dividend. If dividends are paid, expected future prices in the AAO are the sum of expected trading prices and dividends.

which represents the conditional moment function of the econometric problem, and for the unknown true value θ_0 of the SDF parameter vector. To ensure the identification of the true parameter value in parameter set Θ, it is customary to assume that the value θ_0 is unique. Differently, if no admissible SDF belongs to the parametric SDF family, we say that the parametric asset pricing model is misspecified.

The estimation of the true parameter value and the testing of the correct model specification are typically addressed in a Generalized Method of Moments (GMM) framework (see, e.g., Hansen (1982), and Hansen and Singleton (1982)). The method is based on the minimization of the GMM criterion, which is a quadratic form of a sample counterpart of a vector unconditional moment restriction derived from the vector conditional moment restriction in Eq. (1.2). To create this unconditional moment restriction, we select a $(q \times N)$-dimensional instrument matrix Z_t for any date t, that is a deterministic function of the components of the state variables vector X_t. Under the hypothesis of correct model specification, using the instrument matrix Z_t and the law of iterated expectations, we derive the following q-dimensional vector unconditional moment restriction:

$$\mathrm{E}\left[Z_t h(Y_{t+1}, R_{t+1}; \theta_0)\right] = 0_q, \tag{1.4}$$

where $\mathrm{E}\left[\cdot\right]$ denotes the unconditional expectation operator under the historical probability measure. The vector $Z_t h(Y_{t+1}, R_{t+1}; \theta_0)$ is interpreted as a collection of q managed portfolios, realized by taking dynamic positions in the N traded assets.[3] More precisely, the rows of instrument matrix Z_t are the weights of these managed portfolios. We will refer to the l.h.s. of Eq. (1.4) valued at a generic value θ of the SDF parameter as the unconditional pricing error vector. It is immediately apparent that the unconditional pricing error vector depends on the chosen instrument matrix Z_t. The value of the minimized GMM criterion multiplied by sample size T, the so-called Hansen's J statistic, can be used to test the correct specification of the parametric asset pricing model.

Hansen and Jagannathan (1997) introduced a specification test statistic, the unconditional HJ-distance, that is alternative to the Hansen's J statistic for the purpose of testing model specifications. This distance is the minimum L^2-distance of a parametric SDF family from the set of

[3]Being the instrument matrix a function of vector X_t, the dynamic positions are determined by this conditioning variable vector.

all admissible SDF's satisfying the unconditional moment restrictions for the chosen instrument matrix. This distance is defined as

$$d_Z := \min_{\theta \in \Theta} \min_{M_{t,t+1} \in \mathcal{M}_Z} \mathrm{E}\left[(M_{t,t+1} - m(Y_{t+1};\theta))^2\right]^{1/2}, \tag{1.5}$$

where $\mathcal{M}_Z$ is the set of admissible square integrable SDF's for the vector R_t of assets' gross returns and instrument matrix Z_t, i.e.

$$\mathcal{M}_Z := \left\{M_{t,t+1} \in L^2(\mathcal{I}_{t+1}) : \mathrm{E}\left[Z_t\left(M_{t,t+1}R_{t+1} - 1_N\right)\right] = 0_q\right\}, \tag{1.6}$$

where we indicate by $L^2(\mathcal{I}_{t+1})$ the linear space of real random variables with finite second moment and measurable w.r.t. information $\mathcal{I}_{t+1}$. The unconditional HJ-distance d_Z turns out to be the square root of a minimized quadratic form of the unconditional pricing error vector:

$$d_Z = \min_{\theta \in \Theta} \left(\mathrm{E}\left[Z_t h(Y_{t+1}, R_{t+1};\theta)\right]' \Omega_Z \mathrm{E}\left[Z_t h(Y_{t+1}, R_{t+1};\theta)\right]\right)^{1/2}, \tag{1.7}$$

where Ω_Z is the inverse of the $(q \times q)$-dimensional matrix collecting the second unconditional moments of the scaled assets' gross returns, i.e.

$$\Omega_Z := \mathrm{E}\left[Z_t R_{t+1} R'_{t+1} Z'_t\right]^{-1}. \tag{1.8}$$

The HJ-distance is suitable for comparing two possibly non-nested parametric SDF models. Its advantage over the Hansen statistic is that the unconditional pricing error vector of competing models are compared according to the same metric.[4] The large sample properties of the unconditional HJ-distance, which corresponds to a minimized GMM criterion with the non-optimal weighting matrix Ω_Z, are studied in Hansen, Heaton and Luttmer (1995), Hansen and Jagannathan (1997), Parker and Julliard (2005), Kan and Robotti (2009) and Gospodinov, Kan and Robotti (2013). Moreover, the unconditional HJ-distance has been used for asset pricing model selection in empirical work including Hodrick and Zhang (2001) and Kan and Robotti (2009).

[4]The optimal weighting matrix used in the Hansen's J statistic is the inverse of the variance-covariance matrix of the orthogonality vector $Z_t h(Y_{t+1}, R_{t+1};\theta)$, which varies across models. If, for the chosen instrument matrix, a model has very volatile orthogonality vector, this fact may lead artificially to a small J-statistic.

Almeida and Garcia (2012) have considered different discrepancy measures to assess the unconditional distance of the parametric SDF family from the set of admissible SDF's.

More recently, Nagel and Singleton (2011) have proposed to estimate the true (or pseudo-true) value of the SDF parameter vector in a conditionally linear asset pricing model by efficiently exploiting the information contained in the conditional moment restriction in Eq. (1.2). Specifically, Nagel and Singleton (2011) have implemented the optimal instrument matrix by a kernel method (see, e.g., Chamberlain (1987), and Newey (1993)). Besides achieving semiparametric efficiency for estimation, this approach is appealing as it allows for empirical results that do not depend on a specific choice of the instrument matrix. In the working paper Nagel and Singleton (2008) the conditional HJ-distance is defined as the largest unconditional HJ-distance that can be attained with managed portfolios of the assets in the economy (see also Bekaert and Liu (2008) and Chabi-Yo (2008) for the use of a similar scaling approach to derive the conditional HJ-bounds and the unconditional HJ-distance). The specification test of a null conditional HJ-distance is implemented by means of an unconditional HJ-statistic based on this optimal choice of the instrument matrix. However, in Nagel and Singleton (2008) the conditional HJ-distance is neither estimated nor used for model selection among possibly misspecified models.

In this paper we derive the conditional HJ-distance by extending Eqs. (1.5) and (1.6) in a conditional setting. Specifically, we define the conditional HJ-distance δ as the L^2-discrepancy between the candidate parametric SDF family and the set of SDF's satisfying the conditional moment restrictions in Eq (1.2), and not just the unconditional moment restrictions in Eq (1.4) holding for a particular choice of the instrument matrix.

The paper has two main theoretical contributions. First, we study in detail the difference between the conditional and unconditional HJ-distances. We provide upper and lower bounds for the difference $\delta^2 - d_Z^2$ that are valid for general SDF families. In particular, we are able to characterize the difference $\delta^2 - d_Z^2$ explicitly for families of SDF that are conditionally linear in the priced risk factors. We show how this difference is related to the component of the conditional pricing error vector which is unspanned by the instrument matrix. We demonstrate that the difference between conditional and unconditional HJ-distances can be arbitrarily large, and that the ranking for the degree of model misspecification between two misspecified SDF families can be reversed, depending on which distance is used for the comparison. The second theoretical contribution is the definition of a sample analogue of the conditional HJ-distance and the description of its large sample properties, for both correctly specified and misspecified models. In constructing the sample conditional HJ-distance we estimate the conditional expectation of the moment vector $h(Y_{t+1}, R_{t+1}; \theta)$ given the state variable vector X_t by kernel regression methods. The large sample results allow us to develop a model selection procedure based on the conditional HJ-distance.

Our empirical contribution consists in the comparison of fourteen parametric SDF specifications for the U.S. equity and short term T-Bill markets in terms of the conditional HJ-distance. Two specifications are preference-based SDF models with the time-separable CRRA utility and time-nonseparable preferences of Epstein and Zin (1989, 1991) (see, e.g., Stock and Wright (2000) for the GMM estimation of these models). A specification corresponds to the linearization of the preference-based SDF model with time-separable CRRA utility for small values of the logarithmic consumption growth. Three SDF specifications, considered also in Nagel and Singleton (2011), are conditionally linear in logarithmic consumption growth and correspond to dynamic versions of the linearized Consumption-based Capital Asset Pricing Model (CCAPM) with time-varying risk premia. This variation of the coefficients over time is due to either consumption to wealth ratio (Lettau and Ludvigson (2001)), or corporate bond spread (Jagannathan and Wang (1996)) or labor income to consumption ratio (Santos and Veronesi (2006)). Another SDF specification is the linearization of the preference-based SDF model with time-nonseparable preferences of Epstein and Zin (1989, 1991) including personal consumption of nondurables, services and durables. The last five SDF specifications considered in our empirical analysis correspond to the Capital Asset Pricing Model (CAPM, Treynor (1962), Sharpe (1964), Lintner (1965) and Mossin (1966)) and some of its extensions largely used in empirical studies. Specifically, we consider the three-factor Fama and French (1993) model (see also Fama and French (1998)) and its extensions including either the maturity risk and the default risk factors, or a momentum factor (Carhart (1997)), or a liquidity factor (Pastor and Stambaugh (2003)), or the profitability and the investment factors (Fama and French (2014) and Hou, Xue and Zhang (2014)). We also include the four-factor model introduced in Novy-Marx (2013). To highlight the differences between the results obtained by relying on the unconditional and conditional HJ-distances, we report the results of empirical analyses based on the two distances.

The paper is organized as follows. In Sec. I we introduce the conditional HJ-distance and characterize the difference w.r.t. its unconditional counterpart. We also introduce a kernel-based GMM estimator of the conditional HJ-distance. In Sec. II we establish the large sample properties of the sample conditional HJ-distance, for both correctly specified and misspecified models, and develop a model selection procedure. In Sec. III we report the results of an empirical comparison of several misspecified models for the market of publicly traded U.S. equities and short term T-bills. Sec. IV concludes. The regularity conditions and the proofs of theoretical results are given in the appendices. The proofs of auxiliary lemmas are given in the supplementary material available on request.

I. Conditional HJ-distance

We introduce in Sec. I.A the conditional HJ-distance. In Sec. I.B we describe its theoretical properties and discuss the difference with the unconditional HJ-distance. In Sec. I.C we consider the estimation of the conditional HJ-distance.

A. Definition of conditional HJ-distance

In this section we derive the expression of the conditional HJ-distance along the lines of its unconditional counterpart given in the introductory section. We start by defining the set $\mathcal{M}$ of admissible SDF's for the chosen test assets' gross returns:

$$\mathcal{M} := \left\{ M_{t,t+1} \in L^2(\mathcal{I}_{t+1}) : \mathrm{E}\left[M_{t,t+1} R_{t+1} - 1_N \right| X_t = x] = 0_N \text{ for any } x \in \mathcal{X} \right\}. \quad (2.1)$$

Let $\{m(\cdot;\theta) : \theta \in \Theta\}$ be a parametric SDF family. Adapting the definition of unconditional HJ-distance given in Eq. (1.5) to the setting with dynamic pricing restrictions, we define the conditional HJ-distance δ as

$$\delta := \min_{\theta \in \Theta} \min_{M_{t,t+1} \in \mathcal{M}} \mathrm{E}\left[(M_{t,t+1} - m(Y_{t+1};\theta))^2 \right]^{1/2}. \quad (2.2)$$

Differently than set $\mathcal{M}_Z$ and unconditional HJ-distance d_Z, the set $\mathcal{M}$ and the conditional HJ-distance δ do not depend on any instrument matrix Z_t.

Let us consider the minimization problem that defines the squared distance δ^2 in Eq. (2.2). For a given value of parameter $\theta \in \Theta$, the Lagrangian function $\mathcal{L}$ for the constrained minimization w.r.t. the admissible SDF $M_{t,t+1} \in \mathcal{M}$ is given by

$$\begin{aligned}\mathcal{L}(\theta) &= \mathrm{E}\left[(m(Y_{t+1};\theta) - M_{t,t+1})^2\right] + 2\int_{\mathcal{X}} \lambda(x)'\mathrm{E}\left[M_{t,t+1}R_{t+1} - 1_N | X_t = x\right] f_X(x)dx \\ &= \mathrm{E}\left[(m(Y_{t+1};\theta) - M_{t,t+1})^2\right] + 2\mathrm{E}\left[\lambda(X_t)'\left(M_{t,t+1}R_{t+1} - 1_N\right)\right],\end{aligned}$$

where $\lambda(\cdot)$ is a N-dimensional functional Lagrange multipliers vector, function f_X denotes the stationary probability density function (pdf) of process $\{X_t\}$, and we use the law of iterated expectations. By rearranging terms, the Lagrangian function can be written as the sum of

$$\mathrm{E}\left[\left(M_{t,t+1} - m(Y_{t+1};\theta) + \lambda(X_t)'R_{t+1}\right)^2\right]$$

and a term that is independent of the admissible SDF $M_{t,t+1}$. Then, the first-order condition for optimizing the Lagrangian w.r.t. $M_{t,t+1}$ yields the condition

$$M_{t,t+1} = m(Y_{t+1};\theta) - \lambda(X_t)'R_{t+1}. \tag{2.3}$$

The SDF in Eq. (2.3) has to satisfy the conditional no-arbitrage restriction in Eq. (1.1) and thus

$$\lambda(X_t) = \mathrm{E}\left[R_{t+1}R_{t+1}'|X_t\right]^{-1}\mathrm{E}\left[m(Y_{t+1};\theta)R_{t+1} - 1_N|X_t\right] = \Omega(X_t)e(X_t;\theta), \tag{2.4}$$

where $\Omega(X_t)$ is the inverse of the $(N \times N)$-dimensional conditional second-order moment matrix of the assets' gross returns vector R_{t+1} given X_t, i.e.

$$\Omega(X_t) := \mathrm{E}\left[R_{t+1}R_{t+1}'|X_t\right]^{-1}, \tag{2.5}$$

and $e(X_t;\theta)$ is the N-dimensional vector of the conditional pricing errors, i.e.

$$e(X_t;\theta) := \mathrm{E}\left[h(Y_{t+1}, R_{t+1};\theta)|\, X_t\right] = \mathrm{E}\left[m(Y_{t+1};\theta)R_{t+1} - 1_N|X_t\right], \tag{2.6}$$

for any $\theta \in \Theta$. By replacing Eqs. (2.3) and (2.4) into the objective function in Eq. (2.2) we get

$$\delta = \min_{\theta\,\in\,\Theta} \mathrm{E}\left[\lambda(X_t)'\mathrm{E}\left[R_{t+1}R'_{t+1}|X_t\right]\lambda(X_t)\right]^{1/2} = \min_{\theta\,\in\,\Theta} \mathrm{E}\left[e(X_t;\theta)'\Omega(X_t)e(X_t;\theta)\right]^{1/2}. \tag{2.7}$$

The last expression corresponds to the conditional HJ-distance proposed in Nagel and Singleton (2008).[5] This distance is null if, and only if, the conditional moment restriction in Eq. (1.2) is satisfied, that is, the parametric SDF family $\{m(\cdot;\theta) : \theta \in \Theta\}$ is correctly specified. Differently, when the asset pricing model is misspecified, the minimization problem in Eq. (2.2) (or equivalently in Eq. (2.7)) defines the pseudo-true parameter value $\theta_* \in \Theta$. The pseudo-true parameter value θ_* coincide with the true parameter value θ_0 if the model is correctly specified. Moreover, the minimized criterion gives the L^2-distance between the candidate SDF $m(Y_{t+1};\theta_*)$ and the closest admissible SDF in set $\mathcal{M}$.

Hansen and Jagannathan (1997) suggest to consider the dual problem associated to the definition of unconditional HJ-distance to study the same distance. This distance is null if, and only if, the Lagrange multiplier in the dual representation is null. This observation is at the basis of the Lagrange multiplier test proposed by Gospodinov, Kan and Robotti (2013). Similarly, as we can see from the first equality in Eqs. (2.7) we have $\delta = 0$ if, and only if, $\lambda(x) = 0$ for almost all $x \in \mathcal{X}$.

B. Comparison of conditional and unconditional HJ-distances

If a model is correctly specified, both conditional and unconditional HJ-distances are null. However, if the model is misspecified these distances can differ, and in this section we study the discrepancy between them. Before describing in detail this discrepancy, let us first explain why the conditional HJ-distance is larger than the unconditional HJ-distance. Let us consider

[5] Differently from our definition, some authors refer to the argument of the unconditional expectation operator in Eq. (2.7) as squared conditional HJ-distance (see, e.g., Balduzzi and Robotti (2010), and Fang, Ren and Yuan (2011)).

the definition of set $\mathcal{M}$ of admissible SDF's satisfying the conditional moment restrictions in Eq. (2.1) and the definition of set $\mathcal{M}_Z$ of admissible SDF's satisfying the unconditional moment restrictions derived from the chosen instrument matrix in Eq. (1.6). By the law of iterated expectations and considering the definitions of the unconditional and conditional HJ-distances in Eqs. (1.5) and (2.2) we deduce that $\mathcal{M} \subseteq \mathcal{M}_Z$. Therefore, the conditional HJ-distance, which is a measure minimized over just the subset $\mathcal{M}$, is not smaller than the unconditional HJ-distance, which is the corresponding measure minimized over the entire set $\mathcal{M}_Z$:

$$\delta \geqslant d_Z. \tag{2.8}$$

In this section we study the difference between the two distances. We characterize the situations in which this discrepancy is large, and leading to different rankings of competing SDF families, and situations in which it is null. We introduce in Subsec. B.1 a representation of the two HJ-distances as vector norms in a suitable Hilbert space. We then provide in Subsec. B.2 an upper and lower bound for the difference between these two norms, for any parametric SDF family. We particularize the results in Subsec. B.3 to the case of SDF families that are conditionally linear w.r.t. the priced risk factors, providing exact expressions for the difference between the two norms.

B.1 Interpreting the HJ-distances as weighted L^2-norms

We introduce the Hilbert space $L^2_\Omega(\mathcal{X})$ of q-dimensional vector functions $\varphi(X_t)$ with real-valued square integrable elements endowed with the inner product

$$\langle \varphi, \psi \rangle_{L^2_\Omega(\mathcal{X})} := \mathrm{E}\left[\varphi(X_t)'\Omega(X_t)\psi(X_t)\right], \tag{2.9}$$

and associated norm $\|\varphi\|_{L^2_\Omega(\mathcal{X})} = \langle \varphi, \varphi \rangle^{1/2}_{L^2_\Omega(\mathcal{X})}$, for any $\varphi, \varphi \in L^2_\Omega(\mathcal{X})$. We refer to them as the $L^2_\Omega(\mathcal{X})$-inner product and the $L^2_\Omega(\mathcal{X})$-norm, respectively. From Eq. (2.7) the conditional HJ-distance can be written as the minimized $L^2_\Omega(\mathcal{X})$-norm of the conditional pricing error vector w.r.t. parameter θ:

$$\delta = \min_{\theta \,\in\, \Theta} \|e(\cdot;\theta)\|_{L^2_\Omega(\mathcal{X})} = \|e(\cdot;\theta_*)\|_{L^2_\Omega(\mathcal{X})}. \tag{2.10}$$

In order to represent also the unconditional HJ-distance as a $L^2_\Omega(\mathcal{X})$-norm, let us parameterize without loss of generality the instrument matrix Z_t by means of a $(N \times q)$-dimensional matrix function A of vector X_t as follows:[6]

$$Z_t = A(X_t)'\Omega(X_t). \tag{2.11}$$

Using this parameterization and the law of iterated expectations the inverse of matrix Ω_Z and the unconditional pricing error vector become

$$\Omega_Z^{-1} = \mathrm{E}\left[A(X_t)'\Omega(X_t)A(X_t)\right], \text{ and } \mathrm{E}\left[Z_t h\left(X_{t+1};\theta\right)\right] = \mathrm{E}\left[A(X_t)'\Omega(X_t)e\left(X_t;\theta\right)\right], \tag{2.12}$$

for any $\theta \in \Theta$. The elements of matrix Ω_Z^{-1} are the $L^2_\Omega(\mathcal{X})$-inner products between the columns of the matrix function $A(X_t)$ in the Hilbert space $L^2_\Omega(\mathcal{X})$. The elements of the unconditional pricing error vector are the $L^2_\Omega(\mathcal{X})$-inner products between the columns of the matrix function $A(X_t)$ and the conditional pricing errors vector function $e\left(X_t;\theta\right)$. Therefore, from Eq. (1.7) the unconditional HJ-distance under the hypothesis of model misspecification is

$$d_Z = \min_{\theta\,\in\,\Theta} \|\mathcal{P}_A[e(\cdot;\theta)]\|_{L^2_\Omega(\mathcal{X})} = \|\mathcal{P}_A[e(\cdot;\theta_Z)]\|_{L^2_\Omega(\mathcal{X})}, \tag{2.13}$$

where $\mathcal{P}_A$ denotes the orthogonal projection operator onto the linear subspace of $L^2_\Omega(\mathcal{X})$ spanned by the columns of the matrix function $A(X_t)$, and $\theta_Z \in \Theta$ is the minimizer of the criterion in Eq. (1.5), or equivalently in Eq. (1.7) (see App. A.1 for the proof). Both the operator $\mathcal{P}_A$ and the unconditional HJ-distance d_Z depend on the instrument matrix Z_t through the matrix function $A(X_t)$ introduced in Eq. (2.11). As stressed in the introductory section, values θ_* and θ_Z coincide only if the identification of the model provided by the conditional moment restriction in Eq. (1.2) and the unconditional moment restriction in Eq. (1.4) coincide.

B.2 Results for general parametric SDF families

[6]The columns of the matrix function A are interpreted as the portfolio weights in the instrument matrix Z_t scaled by the conditional second moment matrix of the assets' gross returns, i.e.

$$A(X_t)' = Z_t\Omega(X_t)^{-1}.$$

By means of the representations introduced in the previous section, we refine the relation between the two HJ-distances given by Ineq. (2.8).

PROPOSITION 1: *The conditional HJ-distance* δ *and the unconditional HJ-distance* d_Z *are such that*

$$\|\mathcal{P}_A^{\perp}[e(\cdot;\theta_Z)]\|^2_{L^2_\Omega(\mathcal{X})} \geqslant \delta^2 - d_Z^2 \geqslant \|\mathcal{P}_A^{\perp}[e(\cdot;\theta_*)]\|^2_{L^2_\Omega(\mathcal{X})},$$

where $\mathcal{P}_A^{\perp}[\cdot] := \mathbf{I}_N - \mathcal{P}_A[\cdot]$ *is the projection operator onto the linear subspace of* $L^2_\Omega(\mathcal{X})$ *that is orthogonal to the space spanned by the columns of the matrix function* $A(X_t)$ *and* $\mathbf{I}_N$ *is the* $(N \times N)$*-dimensional identity matrix.*

Proof. See App. A.2. □

The upper and lower bounds for the difference between the two squared HJ-distances given in Prop. 1 depend (among other quantities) on the instrument matrix Z_t through matrix $A(X_t)$. Let us focus on the lower bound $\|\mathcal{P}_A^{\perp}[e(\cdot;\theta_*)]\|^2_{L^2_\Omega(\mathcal{X})}$. This bound implies that the difference $\delta^2 - d_Z^2$ cannot be small when the $L^2_\Omega(\mathcal{X})$-norm of the orthogonal projection of the conditional pricing error vector $e(X_t;\theta_*)$ onto the space that is unspanned by the columns of matrix $A(X_t)$ is large. The intuition behind this finding is as follows. When the instrument matrix Z_t does not appropriately describe the variation of the conditional pricing error vector $e(X_t;\theta_*)$, a large part of the variability of this vector is not captured by the unconditional HJ-distance. Stated differently, if the cross-moments between the weights of the portfolio chosen by means of the instrument matrix Z_t and the conditional pricing error vector $e(X_t;\theta_*)$ are close to zero, the unconditional HJ-distance cannot measure accurately the variation of the conditional pricing error vector. We can then note a consequence of Prop. 1.

REMARK 1: *Given an upper bound on the value of the unconditional HJ-distance, there exist parametric SDF families and instrument matrix choices for which the difference between the conditional and unconditional HJ-distances is arbitrarily large.*

As a simple illustration of Rmk. 1, let us consider the case of constant instrument matrix, i.e. $Z_t = \mathbf{I}_N$, and test asset returns with constant second moment, i.e. $\Omega = \Omega_Z = \mathrm{E}[R_{t+1}R_{t+1}']^{-1}$. First, from Eq. (1.7) we deduce that $\mathrm{E}[e(X_t;\theta_*)]'\Omega\mathrm{E}[e(X_t;\theta_*)]$ is an upper bound for d_Z^2. Second, from Prop. 1 and the invariance of the trace operator under cyclical permutations we can write the lower bound for the difference $\delta^2 - d_Z^2$ as

$$\begin{aligned} \|\mathcal{P}_A^{\perp}[e(\cdot;\theta_*)]\|^2_{L^2_\Omega(\mathcal{X})} &= \mathrm{E}\left[(e(X_t;\theta_*) - \mathrm{E}[e(X_t;\theta_*)])' \,\Omega\, (e(X_t;\theta_*) - \mathrm{E}[e(X_t;\theta_*)])\right] \\ &= \mathrm{Tr}\left[\Omega \mathrm{V}[e(X_t;\theta_*)]\right]. \end{aligned}$$

Thus, if the conditional pricing error vector $e(X_t;\theta_*)$ has null unconditional mean and it is sufficiently volatile, the unconditional HJ-distance d_Z is null while the conditional HJ-distance δ can be arbitrarily large.

In the following Cor. 1 we give sufficient and necessary conditions for the two HJ-distances to coincide for a given parametric SDF family.

COROLLARY 1: *The conditional and unconditional HJ-distances coincide if the vector function $e(X_t;\theta_Z)$ is spanned by the columns of matrix function $A(X_t)$. Conversely, if the two distances coincide, then the vector function $e(X_t;\theta_*)$ is spanned by the columns of matrix function $A(X_t)$.*

On the one side, if vector function $e(X_t;\theta_Z)$ is spanned by the columns of matrix function $A(X_t)$, the vector $\mathcal{P}_A^{\perp}[e(\cdot;\theta_Z)](X_t)$ is null by the same definition of projection operator. Then, from Ineq. (2.8) and Prop. 1, the conditional and unconditional HJ-distances coincide. Hence, the conditional HJ-distance equals the unconditional HJ-distance when the information in the instrument matrix is rich enough for $A(X_t)$ to span the conditional pricing error vector for the value θ_Z of the SDF parameter vector. On the other side, if the two distances coincide, from Prop. 1 the vector $\mathcal{P}_A^{\perp}[e(\cdot;\theta_*)]$ is null, which means that the vector function $e(X_t;\theta_*)$ is spanned by the columns of matrix function $A(X_t)$.

Let us now illustrate the implications of Prop. 1 when comparing the values of the conditional and unconditional HJ-distances for distinct parametric SDF families.

REMARK 2: *The unconditional and conditional HJ-distances can yield different rankings for the degree of misspecification of competing parametric SDF families.*

Let us explain Rmk. 2 by considering the rankings for the degree of misspecification of two competing parametric SDF families $\mathcal{F} := \{m(\cdot;\theta); \theta \in \Theta\}$ and $\tilde{\mathcal{F}} := \{\tilde{m}(\cdot;\tilde{\theta}); \tilde{\theta} \in \tilde{\Theta}\}$. Every quantity that refers to the latter family is denoted as for the former but with a superscript tilde.

Let us assume we have $\tilde{d}_Z < d_Z$, that is, the SDF family $\tilde{\mathcal{F}}$ has a lower degree of misspecification than the SDF family $\mathcal{F}$ in terms of the unconditional HJ-distance computed using the instrument matrix Z_t. By Prop. 1 applied to the SDF family $\tilde{\mathcal{F}}$ the difference $\tilde{\delta}^2 - \tilde{d}_Z^2$ can be arbitrarily large if the $L^2_\Omega(\mathcal{X})$-norm of the orthogonal projection of the conditional pricing error vector $\tilde{e}(\cdot;\tilde{\theta}_*)$ onto the space unspanned by the columns of the function matrix $A(X_t)$ is sufficiently large. In particular we can have $\tilde{\delta}^2 - \tilde{d}_Z^2 \geq \delta^2 - \tilde{d}_Z^2$, which implies that $\tilde{\delta} \geq \delta$.[7] Therefore, the parametric SDF family $\tilde{\mathcal{F}}$ has a higher degree of misspecification than the SDF family $\mathcal{F}$ in terms of the conditional HJ-distance. Summarizing out results, using to the two HJ-distances we obtain different rankings for the degree of model misspecification.

B.3 Results for linear SDF families

We obtain sharper results for conditionally linear SDF's, which correspond to the following specification:

$$m(Y_{t+1};\theta) = \tilde{Y}'_{t+1}\theta, \tag{2.14}$$

where the elements of the p-dimensional vector $\tilde{Y}_{t+1}$ are functions of the priced factors and the conditioning information. Examples of such linear SDF specifications are considered by Nagel and Singleton (2011), and are among the models included in our empirical analysis in Sec. III. The conditional pricing error vector at time t is

$$e(X_t;\theta) = B(X_t)\theta - 1_N, \tag{2.15}$$

for any $\theta \in \Theta$, and the $(N \times p)$-dimensional matrix function

$$B(X_t) := \mathrm{E}\left[R_{t+1}\tilde{Y}'_{t+1} \middle| X_t \right],$$

which consists of the conditional cross-moments of assets' gross returns and SDF factors given X_t. Adapting the expression for the conditional HJ-distance δ given in Eq. (2.10) to this case, we see that this distance is the $L^2_\Omega(\mathcal{X})$-norm of the residual of the orthogonal projection of the

[7] From Eq. (2.13) the quantity $\delta^2 - \tilde{d}_Z^2$ depends on the SDF family $\tilde{\mathcal{F}}$ only via $\mathcal{P}_A[\tilde{e}(\cdot;\tilde{\theta}_Z)]$, i.e. the part of the conditional pricing error vector $\tilde{e}(X_t;\tilde{\theta}_Z)$ that is spanned by the instrument matrix. Thus, it is possible to have a large $L^2_\Omega(\mathcal{X})$-norm of vector $\mathcal{P}_A^\perp[\tilde{e}(\cdot;\tilde{\theta}_*)]$ without implications for the value of quantity $\delta^2 - \tilde{d}_Z^2$.

constant vector 1_N onto the space spanned by the columns of matrix function $B(X_t)$.

The following Prop. 2 gives explicit expressions for the difference of the squared conditional and unconditional HJ-distances in terms of the conditional pricing error vector $e(X_t;\cdot)$ valued at θ_* and θ_Z. It also provides two sufficient and necessary conditions for the conditional and unconditional HJ-distances to coincide.

PROPOSITION 2: *For the linear SDF specification given in Eq.* (2.14) *and the parameterization of the instrument matrix given in Eq.* (2.11)*, we have*

$$\delta^2 - d_Z^2 = \|\mathcal{P}_A^{\perp}[e(\cdot;\theta_*)]\|^2_{L^2_\Omega(\mathcal{X})} + \|\mathcal{P}_{\mathcal{P}_A[B]}[e(\cdot;\theta_*)]\|^2_{L^2_\Omega(\mathcal{X})} \tag{2.16}$$

where $\mathcal{P}_{\mathcal{P}_A[B]}$ denotes the orthogonal projection operator onto the linear subspace of $L^2_\Omega(\mathcal{X})$ spanned by the columns of the matrix $\mathcal{P}_A[B]$, and

$$\delta^2 - d_Z^2 = \|\mathcal{P}_A^{\perp}[e(\cdot;\theta_Z)]\|^2_{L^2_\Omega(\mathcal{X})} - \|e(\cdot;\theta_Z) - e(\cdot;\theta_*)\|^2_{L^2_\Omega(\mathcal{X})}. \tag{2.17}$$

Moreover we have

$$\delta = d_Z \quad \Leftrightarrow \quad \mathcal{P}_A^{\perp}[e(\cdot;\theta_*)](X_t) = 0_N \quad \Leftrightarrow \quad \mathcal{P}_A^{\perp}[e(\cdot;\theta_Z)](X_t) = 0_N. \tag{2.18}$$

Proof. See App. A.3. □

In Eq. (2.16), the difference between the squared conditional and unconditional HJ-distances is written as the sum of the squared $L^2_\Omega(\mathcal{X})$-norms of the projections of the conditional pricing error vector $e(\cdot;\theta_*)$ on two mutually orthogonal spaces. The first one is the linear space that is unspanned by the columns of matrix function $A(X_t)$, while the second one is the linear space that is generated by the columns of matrix function $\mathcal{P}_A[B](X_t)$, i.e. the component of the column space of matrix function $B(X_t)$ that is spanned by the columns of matrix function $A(X_t)$. We offer an alternative expression for the difference between the squared conditional and unconditional HJ-distances in Eq. (2.17), where this quantity is written as the difference of the squared $L^2_\Omega(\mathcal{X})$-norms of two vector functions. The first one is $\mathcal{P}_A^{\perp}[e(\cdot;\theta_Z)](X_t)$, that is the residual of the orthogonal projection of the conditional pricing error vector $e(\cdot;\theta_Z)$ onto the columns of matrix function $A(X_t)$. The second one is the difference $e(X_t;\theta_Z) - e(X_t;\theta_*)$. Note that the first terms in the right hand sides (r.h.s.) of Eqs. (2.16) and (2.17) correspond to the lower and upper bounds, respectively, for the difference $\delta^2 - d_Z^2$ given in Prop. 1. Thus, Prop. 2 explains how far the difference $\delta^2 - d_Z^2$ is from these lower and upper bounds in case of linear SDF families.

Eqs. (2.18) provide two conditions for equality between conditional and unconditional HJ-distances in a linear SDF family. These conditions amount to the spanning of the conditional pricing error vectors $e(X_t;\theta_Z)$ and $e(X_t;\theta_*)$ by the columns of matrix function $A(X_t)$. These two conditions are both sufficient and necessary for $\delta = d_Z$, and are therefore equivalent. In this respect, Prop. 2 is a stronger result than Cor. 1, albeit limited to the case of linear SDF families. Let us now relate the conditions in Eqs. (2.18) to the optimal instruments. The choice $A(X_t) = B(X_t)$ in Eq. (2.11) corresponds to the adoption of the optimal instrument matrix for the estimation of the value θ_0 of the SDF parameter vector in a correctly specified linear SDF family (see, e.g., Chamberlain (1987), Newey (1993), and Nagel and Singleton (2011)). For this choice of instruments we have $q = p$, i.e. the set of unconditional moment restrictions is exactly identified, and the unconditional HJ-distance is null by construction whenever the SDF family is correctly specified or not. In the former case the conditional HJ-distance vanishes as well, and the two conditions for $\delta = d_Z$ given in Prop. 2 are satisfied. Indeed, in such a case we have $\theta_Z = \theta_0$ and $e(\cdot;\theta_Z) = e(\cdot;\theta_0) = 0$. Differently, if the SDF family is misspecified, we have $\delta > 0$. More precisely, for the optimal instruments, vector $B(X_t)\theta$ is spanned by the columns of matrix function $A(X_t)$, for any θ, and $\mathcal{P}_A^{\perp}[e(\cdot;\theta_*)] = \mathcal{P}_A^{\perp}[e(\cdot;\theta_Z)] = -\mathcal{P}_A^{\perp}[1_N]$. Therefore, the sufficient and necessary conditions for obtaining $d_Z = \delta$ in Prop. 2 is adding instruments to the optimal ones so that the columns of matrix function $A(X_t)$ span the constant vector 1_N.

C. *Sample conditional HJ-distance*

The conditional HJ-distance δ, and the value θ_0 of the SDF parameter vector if the model is correctly specified, or the value θ_* in case of model misspecification, are unobservable characteristics of the data generating process. In this section we describe an estimation methodology for these characteristics based on a sample of T time series observations on the stationary state variables joint process for vectors of state variables and risk factors X_t and Y_{t+1}. The methodology is particularly useful when the conditioning information vector X_t is low dimensional. The conditional HJ-distance can be estimated by replacing the unconditional expectation in the criterion in Eq. (2.7) by a sample average, and the conditional expectation in the definition of the conditional pricing error vector in Eq. (2.6) by a nonparametric regression. We consider kernel smoothing and denote by K a kernel function on set $\mathbb{R}^L$ and by H_T a symmetric and positive definite $(L \times L)$-dimensional bandwidth matrix, which depends on the sample size T and converges to the null matrix as T tends to infinity.[8] The conditional pricing error vector $e(X_t; \theta)$ is estimated by the Nadaraya-Watson kernel regression estimator computed on the sample of $T-1$ observations of the pair of vectors Y_{i+1} and X_i:

$$\hat{e}_T(X_t; \theta, H) := \sum_{i=1}^{T-1} w(X_t, X_i; H) h(Y_{i+1}, R_{i+1}; \theta), \tag{2.19}$$

where the kernel weighting function w is defined as

$$w(x, \tilde{x}; H) := K\left(H_T^{-1/2}(x - \tilde{x})\right) \Big/ \sum_{j=1}^{T} K\left(H_T^{-1/2}(x - X_j)\right), \tag{2.20}$$

for any $x, \tilde{x} \in \mathcal{X}$ and any $(L \times L)$-dimensional real symmetric positive definite matrix H_T. Thus, vector $\hat{e}_T(X_t; \theta, H_T)$ is a weighted sample average of the vector h, such that the closer is the value X_i of the state variables vector at date i to value X_t, the larger is the weight for that date. Then, the HJ-distance δ is estimated by the sample conditional HJ-distance $\hat{\delta}_T$ defined by

$$\hat{\delta}_T^2 := \min_{\theta \in \Theta} \mathcal{Q}_T(\theta), \qquad \mathcal{Q}_T(\theta) := \frac{1}{T} \sum_{t=1}^{T} \mathbf{I}(X_t) \hat{e}_T(X_t; \theta, H_{OPT})' \hat{\Omega}_T(X_t) \hat{e}_T(X_t; \theta, H_{OPT}), \tag{2.21}$$

where

$$\hat{\Omega}_T(X_t) := \left(\sum_{i=1}^{T-1} w(X_t, X_i) R_{i+1} R_{i+1}' \right)^{-1} \tag{2.22}$$

is the inverse of the kernel regression estimator of matrix $\mathrm{E}\left[R_{t+1} R_{t+1}' | X_t\right]$ taken as a consistent estimator of the matrix $\Omega(X_t)$, and H_{OPT} is the bandwidth matrix selected on the basis of a criterion, such as a cross-validation criterion. The indicator variable $\mathbf{I}(X_t)$ is equal to 1 when

[8] In the empirical application in Sec. III we consider regression functions given a bivariate conditioning information vector X_t and bandwidth matrices.

X_t is in a given compact subset $\mathcal{X}_\star$ of the state variables support $\mathcal{X}$ independent on T, and 0 otherwise. This indicator variable is used as a trimming factor to control boundary effects in the kernel regressions.[9] We denote by $\hat{\theta}_T$ the minimizer of the criterion function $\mathcal{Q}_T(\theta)$, that is an estimator of value θ_0 in case of correct model specification, and of value θ_* in case of model misspecification.

The sample conditional HJ-distance $\hat{\delta}_T$ and the estimator $\hat{\theta}_T$ are closely related to some test statistics and estimators recently proposed in the econometric literature on conditional moment restrictions models. Specifically, estimator $\hat{\theta}_T$ corresponds to a Minimum Distance estimator of Ai and Chen (2003) for an unknown finite dimensional parameter, and to an Euclidean Empirical Likelihood (EEL) estimator of Antoine, Bonnal and Renault (2007) with a non-optimal weighting matrix. The EEL estimator is a member of the class of information-based GMM estimators. These estimators are defined by searching for the value of the structural parameter θ and the distribution of the data that minimize the discrepancy between this distribution and the empirical one, subject to the conditional moment restrictions implied by the model. The EEL estimator relies on a quadratic distance to measure the discrepancy between distributions. Other choices lead to different information-based GMM estimators, such as the Smoothed Empirical Likelihood estimator based on the Kullback-Leibler divergence in Kitamura, Tripathi and Ahn (2004). For correctly specified models, the information-based GMM estimators are asymptotically equivalent to standard GMM estimators based on the optimal choice of instruments and weighting matrix. Thus, they attain the semi-parametric efficiency bound for estimating the true value of parameter θ_0 from the conditional moment restriction $\mathrm{E}\left[h(Y_{t+1}, R_{t+1}; \theta_0)|X_t\right] = 0_N$. The squared sample conditional HJ-distance $\hat{\delta}_T^2$ is asymptotically equivalent to the statistic proposed by Tripathi and Kitamura (2003) to test the conditional moment restriction $\mathrm{E}\left[h(Y_{t+1}, R_{t+1}; \theta)|X_t\right] = 0_N$, for some $\theta \in \Theta$, in a setting with i.i.d. data. In particular, the squared sample conditional HJ-distance $\hat{\delta}_T^2$ has matrix $\hat{\Omega}_T$ in place of the optimal weighting matrix used in Tripathi and Kitamura (2003). The econometric literature on conditional moment restriction models has focused on the large sample properties of the estimators under correct model specification, and on the

[9] See, e.g., Tripathi and Kitamura (2003) and Su and White (2013) for the introduction of trimming factors in test statistics to account for the inaccuracy of nonparametric estimation in tail regions. Moreover, as noted by Aït-Sahalia, Bickel and Stoker (2001), considering a subset of the state variables support allows to focus on those states which are more relevant for the analysis.

consistency of the specification tests under the alternative hypothesis of model misspecification, mostly in a i.i.d. data framework.[10] On the other hand, Hall and Inoue (2003) investigate in depth the asymptotic properties of GMM estimators in misspecified unconditional moment restrictions models. Instead, the large sample behavior of the estimator $\hat{\theta}_T$ of the value θ_* for misspecified models, and the large sample behavior of the statistic $\hat{\delta}_T$ as an estimator of the conditional HJ-distance in misspecified SDF models are unexplored issues. These issues are the topics of the next section.

The sample conditional HJ-distance $\hat{\delta}_T$ is affected by the curse of dimensionality, so that $\hat{\delta}_T$ defined in Eq. (2.21) is a reliable estimator of the conditional HJ-distance δ only if the dimension L of the conditioning information vector X_t is low, say $L \leq 5$. However, for sake of clarity, let us stress that even in the case of a low dimensional vector X_t, the econometrician has to choose a finite number q of deterministic functions of vector X_t to create the instrument matrix Z_t. This instrument matrix may include just a part of the entire information that is necessary to fully describe the behavior of the conditional pricing error vector, as it has been explained in Sec. I.B.

II. Model selection using the conditional HJ-distance

The use of the conditional HJ-distance for specification testing and model selection requires the knowledge of its distribution under both the hypotheses of correct model specification and model misspecification. In both cases the distribution in finite samples is unknown, and we rely on large sample approximations. We consider in Secs. II.A and II.B these approximations, for correct model specification and model misspecification, respectively. In Sec. II.C we describe a model selection among two competing potentially misspecified parametric SDF families on the basis of their sample conditional HJ-distances.

[10]An exception is Gospodinov and Otsu (2012) who allow for serial dependence in the sample.

A. *Large sample properties of the sample conditional HJ-distance under correct model specification*

Let $\mathcal{F} := \{m(\cdot;\theta) : \theta \in \Theta\}$ be a correctly specified parametric SDF family, and let $\hat{\delta}_T$ be the sample conditional HJ-distance defined in Eq. (2.21). The asymptotic distribution of statistic $\hat{\delta}_T^2$ as $T \to \infty$, under a set of regularity conditions collected in App. C, is given in the following Prop. 3. The set of regularity conditions includes the restrictions on the rate of convergence to 0 of the bandwidth b_T as $T \to \infty$ and the conditions on the time series dependence of process $\{X_t\}$.

PROPOSITION 3: *If the parametric SDF family is correctly specified, as $T \to \infty$ the sample conditional HJ-distance $\hat{\delta}_T$ is such that*

$$T b_T^{L/2} \left(\hat{\delta}_T^2 - a_T \right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_0^2\right),$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, the centering variable a_T is defined as

$$a_T := \text{tr}\left[\frac{1}{T} \sum_{t=1}^{T} \mathbf{I}(X_t) \hat{\Omega}_T(X_t) \left(\sum_{\substack{i=1 \\ i \neq t}}^{T-1} w(X_t, X_i)^2 h(Y_{i+1}, R_{i+1}; \hat{\theta}_T) h(Y_{i+1}, R_{i+1}; \hat{\theta}_T)' \right) \right]$$

and the asymptotic variance σ_0^2 is defined as

$$\sigma_0^2 := 2 \left(\int_{\mathbb{R}^L} \mathcal{K}(u)^2 du \right) \int_{\mathcal{X}_\star} \text{Tr}\Big[V_0(x) \Omega(x) V_0(x) \Omega(x) \Big] dx,$$

for the kernel convolution $\mathcal{K}(u) := \int_{\mathbb{R}^L} K(w) K(w-u) dw$ and the conditional variance-covariance matrix $V_0(x) := V[h(Y_{t+1}, R_{t+1}; \theta_0) | X_t = x]$ of the moment function at SDF parameter value θ_0 given the value x of the conditioning state variable vector.

Proof. See App. D.3. □

The proof of Prop. 3 follows closely the proof of Th. 4.1 in Tripathi and Kitamura (2003) by extending their results to a setting with serially dependent data. The asymptotic distribution in Prop. 3 of the squared HJ-distance sharply differs from the asymptotic distribution of the sample squared unconditional HJ-distance (see, e.g., Hansen, Heaton and Luttmer (1995), Hansen and Jagannathan (1997), and Kan and Robotti (2009)). Indeed, the asymptotic distribution of the sample squared conditional HJ-distance $\hat{\delta}_T^2$ is Gaussian (instead of a linear combination of independent chi-square variables), it involves a centering term a_T, and the normalizing factor is $Tb_T^{L/2}$ (instead of T). These differences are due to the fact that the conditional moment restriction in Eq. (1.2) corresponds to an infinity of unconditional moment restrictions.[11] Asymptotic normality is derived by applying a Central Limit Theorem for quadratic forms derived in Yoshihara (1976, 1989).

The centering term a_T for the asymptotic distribution of $\hat{\delta}_T^2$ is nonnegative, since it is a weighted sum of positive definite quadratic forms in the conditional pricing error vector. Then, since also $\hat{\delta}_T^2$ is nonnegative, the test of the null hypothesis $\delta = 0$ is unilateral. The rejection rule for a test of the correct specification of the SDF family with $\alpha\%$ asymptotic level is

$$\hat{\delta}_T^2 \geq a_T + c_{\alpha/2} \frac{\hat{\sigma}_0}{\sqrt{Tb_T^{L/2}}}, \tag{3.1}$$

where $c_{\alpha/2}$ is the $(1-\alpha)$-quantile of the standard normal probability distribution, and $\hat{\sigma}_0$ is a consistent estimator of σ_0 obtained by replacing in its definition the matrices $V_0(x)$ and $\Omega(x)$ by kernel estimators.

B. *Large sample properties of the sample conditional HJ-distance under model misspecification*

We consider in this section the large sample properties of the sample conditional HJ-distance $\hat{\delta}_T^2$ for misspecified models. However, let us first make a general remark about the criterion function $\mathcal{Q}_T$ in Eq. (2.21) that applies to both correctly specified and misspecified models. Since the criterion function $\mathcal{Q}_T$ involves the indicator variable $\mathbf{I}(X_t)$ for set $\mathcal{X}_\star$, the sample conditional HJ-distance $\hat{\delta}_T$ is not a consistent estimator of δ but of the quantity

$$\delta_\star := \min_{\theta \in \Theta} \mathrm{E}\left[\mathbf{I}(X_t) e(X_t;\theta)' \Omega(X_t) e(X_t;\theta)\right]^{1/2} \tag{3.2}$$

[11]See also the introduction of Domínguez and Lobato (2004) for a discussion about infinite sets of restrictions implied by conditional moment restrictions.

instead. The quantity $\delta_\star$ is the version of the conditional HJ-distance when (i) the discrepancy between the parametric SDF family and the set of admissible SDF's is measured by the L^2-distance restricted on set $\mathcal{X}_\star$, and (ii) the no-arbitrage restrictions are imposed just for the values of the conditioning state variables in set $\mathcal{X}_\star$. We can then express the distance $\delta_\star$ as

$$\delta_\star = \min_{\theta \in \Theta} \min_{M_{t,t+1} \in \mathcal{M}_\star} \mathrm{E}\left[\mathbf{I}(X_t)\left(M_{t,t+1} - m(Y_{t+1};\theta)\right)^2\right]^{1/2} = \|\mathbf{I}(\cdot)e(\cdot;\theta_\star)\|_{L^2_\Omega(\mathcal{X})},$$

where the set $\mathcal{M}_\star$ is defined as in Eq. (2.1) with set $\mathcal{X}_\star$ in place of set $\mathcal{X}$, and $\theta_\star$ is the minimizer of the criterion in Eq. (3.2). The distance $\delta_\star$ generally differs from δ, and similarly $\theta_\star$ generally differs from θ_*. However, if the set $\mathcal{X}_\star$ is chosen sufficiently large, the quantities $\delta_\star$ and $\theta_\star$ can be made arbitrarily close to δ and θ_*, respectively.

Let us assume that family $\mathcal{F}$ is misspecified and that the conditional pricing error $e(X_t;\theta_\star)$ does not vanish almost surely on the subset $\mathcal{X}_\star$ of the state variable support, i.e.

$$\delta_\star > 0. \tag{3.3}$$

The large sample behaviour of the squared sample conditional HJ-distance $\hat{\delta}_T^2$ is given in the next Prop. 4.[12]

PROPOSITION 4: *If the parametric SDF family is misspecified, and Ineq.* (3.3) *holds, as* $T \to \infty$ *the sample conditional HJ-distance* $\hat{\delta}_T$ *is such that*

$$\sqrt{T}\left(\hat{\delta}_T^2 - \delta_\star^2\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_\star^2\right),$$

where the asymptotic variance $\sigma_\star^2$ *is the long-run variance*

$$\sigma_\star^2 := \sum_{l=-\infty}^{\infty} \mathrm{E}\left[\phi_t, \phi_{t-l}\right]$$

for the process of the scalar random variable ϕ_t *defined as*

$$\phi_t := \varepsilon_t - \mathrm{E}[\varepsilon_t] + \eta_{t+1} - \mathrm{E}[\eta_{t+1}|X_t], \tag{3.4}$$

[12]The large sample distribution of estimator $\hat{\theta}_T$ for a misspecified SDF family is given in App. E.1.

where

$$\varepsilon_{t+1} := \mathbf{I}(X_t)e(X_t;\theta_\star)'\Omega(X_t)e(X_t;\theta_\star),$$

and

$$\eta_{t+1} := 2\mathbf{I}(X_t)e(X_t;\theta_\star)'\Omega(X_t)h(Y_{t+1},R_{t+1};\theta_\star) - \mathbf{I}(X_t)\left(e(X_t;\theta_\star)'\Omega(X_t)R_{t+1}\right)^2.$$

Proof. See App. E.2. □

Prop. 4 shows that the square sample conditional HJ-distance is a consistent and asymptotically normal estimator of $\delta^2_\star$. The convergence rate is the standard parametric rate $\sqrt{T}$ and differs from the convergence rate $Tb_T^{L/2}$ under correct model specification given in Prop. 3. The asymptotic variance $\sigma^2_\star$ involves the variance and the autocovariances of process $\{\phi_t\}$. To interpret the terms in $\sigma^2_\star$, let us note from Eq. (2.21) that the squared sample conditional HJ-distance is the sample average of quadratic forms of kernel regression estimators. The zero-mean process $\{\varepsilon_t - \mathrm{E}[\varepsilon_t]\}$ is an error term induced by sample averaging, and the process $\{\eta_{t+1} - \mathrm{E}[\eta_{t+1}|X_t]\}$ is induced by the kernel smoothing of the pricing errors and squared gross returns.

A consistent estimator $\hat{\sigma}^2_\star$ of the asymptotic variance $\sigma^2_\star$ is obtained by using a kernel regression estimator for the conditional expectations, and an Heteroskedasticity and Autocorrelation (HAC) robust estimator for the sum of the autocovariances of process ε_t (see e.g. Andrews and Monahan (1992) and Newey and West (1994)). A confidence interval for $\delta^2_\star$ is obtained from the asymptotic distribution of $\log(\delta^2_\star)$ inverting the logarithmic transformation. The asymptotic distribution of $\log(\delta^2_\star)$ is obtained by the applying the delta method with the logarithmic transformation to the following Prop. 5. The lower bounds L and the upper bound U of a confidence interval at the asymptotic $\alpha\%$-confidence level are defined as

$$L := \hat{\delta}^2_T \exp\left(-c_{\alpha/2}\frac{\hat{\sigma}_\star}{\sqrt{T}\hat{\delta}^2_T}\right), \qquad U := \hat{\delta}^2_T \exp\left(c_{\alpha/2}\frac{\hat{\sigma}_\star}{\sqrt{T}\hat{\delta}^2_T}\right). \tag{3.5}$$

C. *Model selection*

In this section we study the pairwise comparison of competing no-arbitrage asset pricing models. Let $\mathcal{F} = \{m(\cdot;\theta);\theta\in\Theta\}$ and $\tilde{\mathcal{F}} = \{\tilde{m}(\cdot;\tilde{\theta});\tilde{\theta}\in\tilde{\Theta}\}$ be two competing parametric

SDF families, with conditional HJ-distances δ and $\tilde{\delta}$, respectively. Under the null hypothesis $\mathcal{H}_0$ the two HJ-distances coincide:

$$\mathcal{H}_0 : \delta_\star = \tilde{\delta}_\star.$$

We test this null hypothesis against the alternative hypothesis

$$\mathcal{H}_A : \delta_\star > \tilde{\delta}_\star,$$

that is, the parametric SDF family $\tilde{\mathcal{F}}$ is preferred to $\mathcal{F}$ in terms of the conditional HJ-distance. The test statistic we use is based on the difference between the squared sample conditional distances $\hat{\delta}_T^2 - \hat{\tilde{\delta}}_T^2$, which is a consistent estimator of the difference $\delta_\star^2 - \tilde{\delta}_\star^2$. The null hypothesis $\mathcal{H}_0$ is rejected in favor of the alternative hypothesis $\mathcal{H}_A$ if the realized value of the statistic $\hat{\delta}_T^2 - \hat{\tilde{\delta}}_T^2$ is larger than a critical value dependent on the chosen asymptotic $\alpha\%$-confidence level. The asymptotic distribution of $\hat{\delta}_T^2 - \hat{\tilde{\delta}}_T^2$ under the null hypothesis is given in the next Prop. 5

PROPOSITION 5: *Let $\Delta\phi_t := \phi_t - \tilde{\phi}_t$, where ϕ_t and $\tilde{\phi}_t$ are scalar random variables defined as in Eq.* (3.4) *for families $\mathcal{F}$ and $\tilde{\mathcal{F}}$, respectively, and*

$$\sigma_\Delta^2 = \sum_{j=-\infty}^{\infty} \mathrm{E}[\Delta\phi_t \Delta\phi_{t-j}] > 0.$$

Then, under the null hypothesis $\mathcal{H}_0$, as $T \to \infty$, we have

$$\sqrt{T}\left(\hat{\delta}_T^2 - \hat{\tilde{\delta}}_T^2\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_\Delta^2\right).$$

Under the conditions of Prop. 5 the rejection region for a test of $\mathcal{H}_0$ at the $\alpha\%$ asymptotic level is

$$\hat{\delta}_T^2 - \hat{\tilde{\delta}}_T^2 \geq c_{\alpha/2} \frac{\hat{\sigma}_\Delta}{\sqrt{T}}, \tag{3.6}$$

where $\hat{\sigma}_\Delta$ is a consistent estimator of σ_Δ. Prop. 5 does not cover the case when $\sigma_\Delta = 0$. This case occurs if, and only if, the stochastic process $\{\phi_t\}$ vanishes (MEGLIO $\phi_t = 0$, for any t ?). The first possibility for this case to happen is that the conditional pricing errors at the pseudo-true parameter value are uniformly zero over the entire support $\mathcal{X}$ of the state variable vector for both models, that is $e(\cdot;\theta_\star) = \tilde{e}(\cdot;\tilde{\theta}_\star) = 0$. In this case, all the terms in the r.h.s. of Eq. (3.4) are zero. This condition is equivalent to $\delta = \tilde{\delta} = 0$, i.e. both models are correctly specified. The second possibility is that the pseudo-true SDF's coincide, i.e. $m(\cdot;\theta_\star) = \tilde{m}(\cdot;\tilde{\theta}_\star) = 0$. In this case, the r.h.s. of Eq. (3.4) for families $\mathcal{F}$ and $\tilde{\mathcal{F}}$ coincide term by term.[13] We can then consider separately these events and write the null hypothesis $\mathcal{H}_0$ as a union:

$$\mathcal{H}_0 = \mathcal{H}_{0,1} \cup \mathcal{H}_{0,2} \cup \mathcal{H}_{0,3}, \tag{3.7}$$

where the three hypotheses $\mathcal{H}_{0,1}, \mathcal{H}_{0,2}$ and $\mathcal{H}_{0,3}$ are defined as

$$\mathcal{H}_{0,1} := \left\{\delta = \tilde{\delta} = 0\right\}, \mathcal{H}_{0,2} := \left\{m(\cdot;\theta_\star) = \tilde{m}(\cdot;\tilde{\theta}_\star)\right\}, \mathcal{H}_{0,3} := \left\{m(\cdot;\theta_\star) \neq \tilde{m}(\cdot;\tilde{\theta}_\star), \delta = \tilde{\delta} > 0\right\}.$$

The asymptotic distribution in Prop. 5 is valid only under hypothesis $\mathcal{H}_{0,3}$, that is when variable ϕ_t is not identically null. To address this issue, we follow Gospodinov, Kan and Robotti (2013) in the unconditional framework, and distinguish the three cases of nested, strictly non-nested and overlapping models.

C.1 Nested models

Let us first assume that one of the two SDF families is included in the other one. Without loss of generality, let family $\tilde{\mathcal{F}}$ be nested into family $\mathcal{F}$, i.e. $\tilde{\mathcal{F}} \subset \mathcal{F}$. In this case $\mathcal{H}_0 = \mathcal{H}_{0,2}$. Indeed, if the pseudo-true SDF's differ, the conditional HJ-distance of the family $\tilde{\mathcal{F}}$ is larger than the conditional HJ-distance of the family $\mathcal{F}$. The equality of the pseudo SDF's $m(\cdot;\theta_\star) = \tilde{m}(\cdot;\tilde{\theta}_\star)$ occurs when a specific set of r transformations of the parameter of the nesting family equates zero, i.e.

$$\psi(\theta_\star) = 0_r,$$

where ψ is the r-dimensional vector of restrictions on parameter vector $\theta_\star$. Therefore, we can test the hypothesis $\mathcal{H}_0$ by means of the following Wald test statistic:

$$\xi := T\psi(\hat{\theta}_T)' \left(\left(\frac{\partial \psi(\hat{\theta}_T)}{\partial \theta'} \right) \hat{\Sigma}_\star \left(\frac{\partial \psi(\hat{\theta}_T)'}{\partial \theta} \right) \right)^{-1} \psi(\hat{\theta}_T)$$

[13]If the conditional pricing errors are non-zero and $m(\cdot;\theta_\star) \neq \tilde{m}(\cdot;\tilde{\theta}_\star)$, the uncertainty sources at time t in ϕ_t are $m(Y_t;\theta_\star)$, $\tilde{m}(Y_t;\tilde{\theta}_\star)$ and $R_t R_t'$, and these sources cannot compensate each other to yield $\phi_t = 0$ at all dates.

for the $(r \times p)$-dimensional Jacobian matrix $\frac{\partial \psi(\hat{\theta}_T)}{\partial \theta'}$ and the $(p \times p)$-dimensional matrix $\hat{\Sigma}_\star$, which is a consistent estimator of the asymptotic variance $\Sigma_\star$ of estimator $\hat{\theta}_T$ defined in Lemma 5 (see App. E.1). Under $\mathcal{H}_{0,2}$, statistic ξ is asymptotically χ^2_r-distributed.

C.2 Strictly non-nested models

When the two parametric SDF families are strictly non-nested, i.e. $\mathcal{F} \cap \tilde{\mathcal{F}} = \emptyset$, the pseudo-true SDF's cannot coincide.[14] Therefore, in this case $\mathcal{H}_0 = \mathcal{H}_{0,1} \cup \mathcal{H}_{0,3}$, with $\mathcal{H}_{0,3} := \left\{\delta = \tilde{\delta} > 0\right\}$. To distinguish between the hypotheses $\mathcal{H}_{0,1}$ and $\mathcal{H}_{0,3}$ we propose a pre-testing strategy. First, we test $\mathcal{H}_{0,1}$. Then, if the hypothesis $\mathcal{H}_{0,1}$ is rejected, we test the hypothesis $\mathcal{H}_{0,3}$ using the rejection rule in Ineq. (3.1). We neglect the effect of pre-testing on the level of the test.

We can test $\mathcal{H}_{0,1}$ by means of the statistic $\hat{\delta}^2 + \hat{\tilde{\delta}}^2$ (see App. N for the asymptotic distribution of this test statistic). Alternatively, we can test individually the hypotheses $\delta = 0$ and $\tilde{\delta} = 0$, using Prop. 3.

C.3 Overlapping models

The families $\mathcal{F}$ and $\tilde{\mathcal{F}}$ are overlapping if they contain common SDF's and are not nested. In this case, any of the subsets $\mathcal{H}_{0,1}$, $\mathcal{H}_{0,2}$ and $\mathcal{H}_{0,3}$ of the null hypothesis $\mathcal{H}_0$ can occur and we cannot simplify Eq. (3.7). We propose a pre-testing strategy. First, we test individually hypotheses $\mathcal{H}_{0,1}$ and $\mathcal{H}_{0,2}$. Second, if the hypotheses $\mathcal{H}_{0,1}$ and $\mathcal{H}_{0,2}$ are both rejected, we test the null hypothesis $\mathcal{H}_{0,3}$ using the rejection rule in Eq. (3.1). The test of hypothesis $\mathcal{H}_{0,1}$ is as in Sec. C.2. To test hypothesis $\mathcal{H}_{0,2}$ we use that the equality of the pseudo-true SDF's occurs when the pseudo-true parameters of the two families satisfy a set of restrictions

$$\mathcal{H}_{0,2} = \left\{\psi(\theta_\star) = 0_r, \tilde{\psi}(\tilde{\theta}_\star) = 0_{\tilde{r}}\right\}$$

where ψ and $\tilde{\psi}$ are r- and $\tilde{r}$-dimensional vectors of restrictions on parameter vectors $\theta_\star$ and $\tilde{\theta}_\star$,

[14] The numerical value of pseudo-true values of strictly non-nested parametric SDF families may be the same. However the interpretation of the parameters differ in the two models.

respectively. The test of hypothesis $\mathcal{H}_{0,2}$ can be based on the Wald statistic

$$\xi := T\hat{\psi}' \left(\hat{H}\hat{V}\hat{H}'\right)^{-1} \hat{\psi}, \qquad \hat{\psi} := [\psi(\hat{\theta}_T)' \; \tilde{\psi}(\hat{\tilde{\theta}}_T)']', \qquad \hat{H} := \begin{bmatrix} \frac{\partial \psi(\hat{\theta}_T)}{\partial \theta'} & \mathbf{0}_{r\times\tilde{p}} \\ \mathbf{0}_{\tilde{r}\times p} & \frac{\partial \tilde{\psi}(\hat{\tilde{\theta}}_T)'}{\partial \theta} \end{bmatrix}.$$

The statistic ξ is asymptotically $\chi^2_{r+\tilde{r}}$-distributed under $\mathcal{H}_{0,2}$.

III. Empirical comparison of parametric SDF specifications for portfolios of publicly traded U.S. equities and short term T-bills

In this section we compare 16 no arbitrage asset pricing models for managed portfolios of publicly traded U.S. equities and short term T-bills during the last 50 years, on the basis of both unconditional and conditional HJ-distances. The managed portfolios are often considered in the modern asset pricing literature as representative of all the traded U.S. equities.[15] However, the chosen test assets may represent just the part of variability of equity returns that is well explained by a model, and ignore another part of this variability that is not equivalently well explained by the same model. Therefore, the results of the analysis are only limited to the test portfolios, and we do not claim them to hold for the entire set of U.S. equities. In Sec. III.A we describe the SDF specifications for the competing models. In Sec. III.B we describe the data used for the analysis. In Sec. III.C we compare the models on the basis of both unconditional and conditional HJ-distances and discuss the differences.

A. *Competing asset pricing models*

The competing asset pricing models are (i) three structural preference-based asset pricing models, (ii) three of their linear approximations, (iii) three extensions of the structural models to include additional priced factors and time-varying risk premia, and (iv) seven reduced form spec-

[15] As stressed in Nagel and Singleton (2011), considering a low number of assets accounting for "size" and "value" effects is enough to capture most of the cross-sectional return variation in the market for publicly traded U.S. equities (see also Fama and French (1993) and Lewellen, Nagel and Shanken (2010)).

ifications inspired by linear factor models. In the three structural preference-based asset pricing models the SDF is implied by the behavior of a representative agent who invests, saves and consumes. We refer to these SDF specifications as CRRA, EZ and DCRRA specifications (see, e.g., Stock and Wright (2000) for the GMM estimation of the first two models). In empirical studies of structural preference-based asset pricing models the SDF implied by the behavior of a representative agent is often linearized, so that we consider three linearizations of the CRRA, EZ and DCRRA specifications for small consumption growth and return on the market portfolio in the LCRRA, LEZ and LDCRRA specification, respectively. The coefficients of the LCRRA specification become time-varying in the DEF, CAY, YC specifications, which allow for time-varying risk premia. Finally, the reduced form specifications are "beta pricing models", which move from the CAPM and account for stylized facts in the cross-section of equity returns. We refer to these last specifications as the CAPM, FF3, FFREV, FF5, FFM, FFL and NM specifications. In the remaining part of the section we describe more precisely all the SDF specifications.

CRRA The CRRA specification is associated to the Consumption-based CAPM with time-separable preferences and constant relative risk-aversion (CRRA) utility for the representative agent (see, e.g., Hansen and Singleton (1982)):

$$m(Y_{t+1};\theta) = \theta_1 (C_{t+1}/C_t)^{-\theta_2},$$

where C_t is the personal consumption at time t of nondurables and services, and $Y_{t+1} = C_{t+1}/C_t$ is the consumption growth. The parameter $\theta = [\theta_1\ \theta_2]'$ includes the time-discount rate θ_1 and the risk-aversion parameter θ_2.

EZ The EZ specification is implied by the time-nonseparable Epstein and Zin (1989, 1991) preferences of the representative agent:

$$m(Y_{t+1};\theta) = \theta_1^{\theta_3} (C_{t+1}/C_t)^{-\theta_2\theta_3} \left(1 + 0.01 \times MKT_t\right)^{\theta_3 - 1},$$

where MKT_t is the gross return on the market portfolio in excess of the risk-free rate measured in percentage points, and $Y_{t+1} = [C_{t+1}/C_t\ MKT_{t+1}]'$. The parameter vector $\theta = [\theta_1\ \theta_2\ \theta_3]'$ includes the time discount rate θ_1 and the two additional parameters θ_2 and

θ_3. The parameterization is such that the risk-aversion is $1-\theta_3(1-\theta_2)$ and the elasticity of intertemporal substitution is $1/\theta_2$. The EZ specification reduces to the CRRA specification if $\theta_3 = 1$.

DCRRA The DCRRA specification corresponds to the Durable CCAPM (Yogo (2006)). This specification is an extension of the EZ specification to account for the consumption of durables. The SDF has the expression

$$m(Y_{t+1};\theta) = \theta_1^{\theta_3}(C_{t+1}/C_t)^{-\theta_2\theta_3}\left(\frac{v\left(C_{D,t+1}/C_{t+1}\right)}{v\left(C_{D,t}/C_t\right)}\right)^{\theta_3(\theta_4-\theta_2)}(1+0.01\times MKT_{t+1})^{\theta_3-1},$$

where $C_{D,t+1}$ is the personal consumption at time t of durables, and the function $v(\cdot)$ is defined as

$$v\left(\cdot\right) := \left(1-\theta_5+\theta_5\left(\cdot\right)^{1-\theta_4}\right)^{1/(1-\theta_4)}$$

with $\theta_5 \in (0,1)$. In this specification $Y_{t+1} = [C_{t+1}/C_t\ C_{D,t+1}/C_{t+1}\ C_{D,t}/C_t\ MKT_{t+1}]'$ and $\theta = [\theta_1\ \theta_2\ \theta_3\ \theta_4\ \theta_5]'$. The DCRRA specification reduces to the EZ specification if $\theta_4 = \theta_2$.

LCRRA The LCRRA specification corresponds to the unconditional (or "static") version of the linear Consumption CAPM (CCAPM). It is a linearization of the CRRA specification for small values of the logarithmic growth of personal consumption of nondurables and services $\log\left(C_{t+1}/C_t\right)$:

$$m(Y_{t+1};\theta) = \theta_1 + \theta_2\log\left(C_{t+1}/C_t\right),$$

where $Y_{t+1} = C_{t+1}/C_t$ and $\theta = [\theta_1\ \theta_2]'$.[16]

LEZ The LEZ specification corresponds to a linearization of the EZ specification for small values of the logarithmic growth of personal consumption of nondurables and services $\log\left(C_{t+1}/C_t\right)$ and small values of the logarithm of the gross return on the market portfolio $R_{m,t}$. The SDF has the expression

$$m(Y_{t+1};\theta) = \theta_1 + \theta_2\log\left(C_{t+1}/C_t\right) + \theta_3 MKT_{t+1}.$$

[16] The parameters in the LCRRA specification are renamed after linearizing the SDF in the CRRA specification.

In this specification $Y_{t+1} = [C_{t+1}/C_t \; MKT_{t+1}]'$ and $\theta = [\theta_1 \; \theta_2 \; \theta_3]'$. The LEZ specification reduces to the LCRRA specification for $\theta_3 = 0$.

LDCRRA The LDCRRA specification corresponds to the linearization of the Durable CCAPM for small values of the logarithmic growth of personal consumption of nondurables and services $\log(C_{t+1}/C_t)$, small values of the logarithmic ratio of personal consumption of nondurables and services to consumption of durables $\log(C_{D,t}/C_t)$ and small values of the logarithm of the gross return on the market portfolio $R_{m,t}$. The SDF has the expression

$$m(Y_{t+1};\theta) = \theta_1 + \theta_2 \log(C_{t+1}/C_t) + \theta_3 \log(C_{D,t+1}/C_{t+1}) + \theta_4 \log(C_{D,t}/C_t) + \theta_5 MKT_{t+1}.$$

In this specification $Y_{t+1} = [C_{t+1}/C_t \; C_{D,t+1}/C_{t+1} \; C_{D,t}/C_t \; MKT_{t+1}]'$ and $\theta = [\theta_1 \; \theta_2 \; \theta_3 \; \theta_4 \; \theta_5]'$.[17] The LDCRRA specification reduces to the LEZ specification for $\theta_3 = \theta_4 = 0$.

DEF, CAY, YC The three specifications DEF, CAY, YC of the SDF correspond to conditional (or "dynamic") versions of the linear CCAPM. Each specification is affine in logarithmic consumption growth with time-varying coefficients. In particular, the coefficients themselves are affine functions of some conditioning state variables. The three specifications admit the common SDF expression

$$m(Y_{t+1};\theta) = (\theta_1 + \theta_3 u_t) + (\theta_2 + \theta_4 u_t) \log(C_{t+1}/C_t),$$

where $Y_{t+1} = [C_{t+1}/C_t \; u_t]'$ and $\theta = [\theta_1 \; \theta_2 \; \theta_3 \; \theta_4]'$. The conditioning state variable u_t allows for time-varying risk premia. It is the corporate bond spread DEF_t in the DEF specification (see, e.g., Jagannathan and Wang (1996)), the consumption to wealth ratio CAY_t in the CAY specification (see, e.g., Lettau and Ludvigson (2001)), and the labor income to consumption ratio YC_t in the YC specification (see, e.g., Santos and Veronesi (2006)).[18] The three specifications DEF, CAY, YC reduce to the LCRRA specification for $\theta_3 = \theta_4 = 0$.

CAPM, FF3, FFREV, FF5, FFM, FFL, NM The seven SDF specifications CAPM, FF3, FFREV,

[17] The LDCRRA specification is an example of the illustration in Eq. (32) of Nagel and Singleton (2011).

[18] The DEF, CAY and YC specifications are examples of the illustration in Eq. (29) and (41) of Nagel and Singleton (2011).

FF5, FFM, FFL, NM are used in empirical asset pricing studies to account for some stylized facts in the cross-section of equity returns. They admit the common SDF expression

$$m(Y_{t+1};\theta) = \theta_1 + \theta_2 MKT_{t+1} + \bar{\theta}' F_{t+1}, \tag{3.8}$$

where F_t is a vector of factors, $\theta = [\theta_1 \; \theta_2 \; \bar{\theta}']'$, with dimension of vector $\bar{\theta}$ which varies across models, and $Y_t = [MKT_t \; F_t']'$. In the CAPM specification the vector F_t is null (Treynor (1962), Sharpe (1964), Lintner (1965) and Mossin (1966)). The FF3 specification corresponds to the Fama-French 3-factors model, so that F_t includes size and value factors in this specification (Fama and French (1993, 1998)). The vector F_t is augmented by short and long term reversal factors in the FFREV specification, by the momentum factor in the FFM specification (Carhart (1997)), and by the liquidity factor in the FFL specification (Pastor and Stambaugh (2003)). In the FF5 specification, vector F_t includes size, profitability and investment factors (Fama and French (2014) and Hou, Xue and Zhang (2014)). We do not include the value factor in this specification because Fama and French (2014) show that for the U.S. equity return in the period 1963-2013 the value factor can be explained by the other factors in their model. Finally, the vector F_t includes the industry-adjusted value factor, the momentum factor and the profitability factor in the NM specification (Novy-Marx (2013)).

B. Data

We describe in this section the data used for the empirical comparison of the asset pricing models illustrated in Sec. III.A. Ideally the estimation of these models and their comparison by means of the HJ-distances should be based on (i) the largest set of variables that are observable to investors taken as components of the conditioning information vector used to set prices and (ii) an infinite number of functions of the conditioning information vector taken as components of the instrument matrix. The estimation would be based in this way on the largest number of moment restrictions, and it would account for the actual information set available to investors. Unfortunately, the estimation of the parameters by optimal GMM and the computation of the sample HJ-distances for the competing SDF specifications based on finite samples are precise and stable only if they are based on a limited set of moment restrictions. Moreover, because of the curse of dimensionality affecting the regression estimator in Eq. (2.19), the computation of the sample conditional HJ-distance is reliable only assuming just few variables generating the important information used by investors to set prices.[19]

Therefore, for the estimation of the asset pricing models by optimal GMM and their comparison by means of the HJ-distances we select small sets of (i) variables which comove with the remaining risk factors and gross returns on the test assets as variables representing the information used for pricing, and (ii) functions of the conditioning information vector as components of the instrument matrix.

The estimation of the asset pricing models and their comparison by means of the HJ-distances can be affected by the chosen conditioning information vector. This choice may indeed artificially favor a particular asset pricing model when the performance is valued on the basis of the conditional pricing errors. This problem would appear directly in the computation of the conditional HJ-distance. Moreover, as explained in Sec. I, the instrument matrix, which is a function of the conditioning information vector, may favor a particular asset pricing model. As explained in the introduction, choosing the instrument matrix Z_t means focusing on a particular set of managed portfolios obtained by taking dynamic positions in the test assets. Some models may explain well the variability of the returns on these portfolios, and not explain the variability on portfolios obtained by taking distinct positions on the same test assets. We describe in this section our choices to avoid favoring a model in the empirical comparison while keeping the computation of the estimators feasible. We illustrate in Subsec. B.1 the proxies for the risk factors and the test assets, and in Subsec. B.2 the variable generating information and the instrument matrix used for the computation of the unconditional and conditional HJ-distances.

B.1 Proxies for the risk factors and test assets

The study spans the period from July 1963 to Dec. 2014 and it relies on proxies for the risk factors and returns on equity portfolios often taken as representative of the U.S. market for publicly traded equities at the monthly frequency.

[19] Considering a low-dimensional conditioning information vector is a a common practice. For example, Nagel and Singleton (2011) consider singularly the corporate bond spread DEF_t, the transitory component of financial wealth CAY_t and the labor income to consumption ratio YC_t as the one-dimensional variable X_t generating the information to estimate the DEF_t, CAY_t and YC_t specifications.

To create the time series of the risk factors in the structural preference-based asset pricing models and their linear approximations and extensions, we consider the aggregate personal consumption expenditures included in the U.S. National Income and Product Accounts (NIPA), provided by the U.S. Bureau of Economic Analysis of the Department of Commerce, and the estimates of the total U.S. population provided by the U.S. Census Bureau. We consider net real growth rates of per capita total consumption and consumption of durable goods from the NIPA chain-type price indexes. These rates are taken as proxies for the personal consumption of nondurables and services C_t and the personal consumption of durables $C_{D,t}$. We consider the monthly Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity as corporate bond spread DEF_t. We derive monthly estimates of the transitory component of financial wealth CAY_t by linear interpolation through levels at following end-of-quarters of the Lettau and Ludvigson (2001) quarterly proxies for the consumption to wealth ratio.[20] Following Santos and Veronesi (2006) we consider the sum of wages and salaries, transfer payments and other labor income minus personal contributions for social insurance and taxes to create the time series of the labor income to consumption ratio YC_t.[21]

In the empirical study of the linear factor models we consider the proxies for the risk factors given by those researchers who first proposed the competing reduced form specifications of the SDF, even in the case of the same risk factor in distinct asset pricing models. We consider the monthly Fama-French (FF) proxy for the factor MKT_t. We consider the monthly FF proxies $SMB3_t$ (Small-Minus-Big) and $HML3_t$ (High-Minus-Low) for size and value factors constructed from 6 value-weight portfolios formed on size and book-to-market in the FF3 specification. Together with the FF proxies $SMB3_t$ and $HML3_t$ we use the FF proxies ST_t and LT_t for the short- and long-term reversal factors in the FFREV specification, the FF proxy UMD_t (Up-Minus-Down) for the momentum factor in the FFM specification, and the Pastor-Stambaugh proxy L_t for the liquidity factor in the FFL specification. In the FF5 specification we use the FF proxies $SMB5_t$, RMW_t (Robust-Minus-Weak) and CMA_t (Conservative-Minus-Aggressive) for the size, value, profitability and investment factors constructed from 18 value-weight port-

[20]The quarterly proxies for the consumption to wealth ratio and the details on its construction are available on Martin Lettau's website http://faculty.haas.berkeley.edu/lettau/. The time series for CAY_t stops in July 2013.

[21]The data, seasonally adjusted at annual rates, are obtained from the Bureau of Economic Analysis. Given the availability of the data, the SDF specification FFL is estimated only until July 2014.

folios formed on (i) size and book-to-market, (ii) size and operating profitability, and (iii) size and investment. We use the Novy-Marx proxies HML_t^*, UMD_t^* and PMU_t^* for the industry-adjusted value, momentum and profitability factor in the NM specification.[22] From the CRSP database we obtain total cum-dividend value weighted returns $R_{m,t}$ and their counterpart returns without dividends $R_{m,t}^{[x]}$ on the CRSP value-weighted portfolio created with stocks traded at NYSE, AMEX, NASDAQ and ARCA. Using these returns we construct the total dividend yield $D_{m,t} := R_{m,t}/R_{m,t}^{[x]} - 1$ following Cochrane (2008).[23]

We consider the six value-weighted FF research portfolios adjusted for inflation and dividends and the 1-month T-bill as the $N = 7$ test assets. The six value-weighted Fama-French (FF) research portfolios explain the largest part of the cross-sectional variation of U.S. equity returns. All the nominal returns on the FF research portfolios are deflated by the monthly Consumer Price Index for All Urban Consumers to obtain ex-post real returns. We use real excess returns for the cum-dividend equity portfolios.[24] Each of the six portfolios SG_t, SN_t, SV_t, BG_t, BN_t and BV_t obtained in this way exemplify the stochastic behavior of inflation- and dividends-adjusted gross excess return on a growth (G), neutral (N) or value (V) stock of a company with small (S) or big (B) market capitalization, as their names indicate. For the 1-month T-bill we consider real gross returns.

Table I reports time series length, minimum, maximum, mean, standard deviation, skewness and excess kurtosis for the proxies for risk factors, conditioning variables and returns on the test assets.

B.2 Conditioning information vector and instrument matrix

In the empirical study, the selection of the components of vector X_t and the instrument matrix Z_t should be driven by the analysis of the stochastic process of the conditional moment function $h(Y_{t+1}, R_{t+1}; \theta_0)$ and economic and financial variables at time t. Indeed, the function is the argu-

[22] All the data and the 1-month T-bill gross rate $R_{f,t}$ are taken from the websites of Kenneth French, Robert Stambaugh and Robert Novy-Marx. Given the availability of the data, the SDF specification FFL is estimated and tested until Dec. 2013., and the SDF specification NM until Dec. 2012. The data have been retrieved on Jan. 25th, 2015 from http://www.bea.gov/, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html, http://finance.wharton.upenn.edu/ stambaug/ and http://rnm.simon.rochester.edu/.

[23] Data have been retrieved from the Wharton Research Data Services database. The returns on the CRSP value portfolio give $D_{m,t} = 19.91$ on May 1969, which is 10 times bigger in magnitude than any other value in the time series.

[24] We repeat the analysis using the cum- and ex-dividend equity portfolios.

ment in the conditional expectation in Eq. (1.2) and the lagged variables are its conditioning set. However, the time series of the conditional moment function $h(Y_{t+1}, R_{t+1}; \theta_0)$ is not observed but estimated, and there is no unique consensus in the scientific community about which variables actually predict stock returns. In addition, it is customary to model probabilities of distinct risk factors on the basis of different information sets. For example, consumption growth and the factors included in beta pricing models are often assumed as predicted by different financial and macroeconomic variables. This in turn translates into having a *most natural* choice for the components of vector X_t and the instrument matrix Z_t depending on the SDF specification. Unfortunately, variables chosen in different SDF specifications are typically not perfectly dependent on each other.

For a fair empirical comparison of competing parametric SDF specifications, the information set that is available to investors should be the same across SDF specifications. Therefore, the same conditioning information set and the same instrument matrix should be taken for distinct SDF specifications, and they should include all the variables considered as most natural for the SDF specifications taken singularly. However, because of empirical feasibility, we must choose a limited set of variables as components of vector X_t and a limited set of functions to create the matrix Z_t. Any choice of this kind represents the risk of artificially favoring a model in a model fit contest with competing models based on the HJ-distances.

Since the test equity portfolios are representative of the market for publicly traded U.S. equities, we use the dividend to price ratio $D_{m,t}/P_{m,t}$ of the CRSP value-weighted portfolio as first component of the vector X_t. The remaining components of vector X_t are chosen on the basis of the risk factors included in the SDF. The conditional moment function $h(Y_{t+1}, R_{t+1}; \theta_0)$ is not directly observable, so that we focus on the comovements between the vectors Y_{t+1} and R_{t+1} and the cross-section of risk factors and gross returns on the test assets at time t. Our strategy is to select vector X_t on the basis of linear regression coefficients and partial correlations among variables at following times. In any SDF specification we first select the risk factors at time t that are correlated at least at the 5% significance level with at least one component of vector Y_{t+1}. We then regress linearly each component of vector Y_{t+1} onto the selected variables, and moreover we compute the partial correlation coefficient of the regressand and each regressor once accounting for the variability in the other reressors. In this selection we choose $SMB3_t$, $HML3_t$, UMD_t and RMW_t as representative of size, value, momentum and profitability factor, respectively.

In Table II we report the pointwise sample contemporaneous correlations of the risk factors included in the asset pricing models. We find several correlations that are statistically different from 0 at the 1% confidence level. In particular, the growth of personal consumption of nondurables and services C_{t+1}/C_t and the ratio of personal consumption of nondurables and services to consumption of durables $C_{D,t}/C_t$ are positively correlated. The three risk factors DEF_t, CAY_t and YC_t are pairwise negatively correlated. The market factor MKT_t, the size factor $SMB3_t$ and the value factor $HML3_t$ are correlated with almost every other risk factor considered in the beta pricing models. The dividend to price ratio $D_{m,t}/P_{m,t}$ is not correlated with any contemporaneous risk factor.

We consider the first sample autocorrelation, partial autocorrelation and cross-correlations of all the risk factors. We report in Table III the autocorrelation coefficients, and in Table IV the first sample cross-correlations just in the case one of these correlations at different times is higher than 0.12 in absolute value. We obtain the following information from these tables. The risk factors DEF_t, CAY_t and YC_t are strongly serially correlated. The correlation among the three variables at the same time and the persistence in each time series can explain the cross-correlations at different times. The risk factor C_t/C_{t-1} is serially correlated. The risk factors $C_{D,t}/C_t$, $HML3_t$, RMW_t, CMA_t and LT_t are mildly serially correlated. The market factor MKT_{t-1} is correlated with several factors at the following time t. The results we obtain are for the choice $X_t = [C_t/C_{t-1}\ D_{m,t}/P_{m,t}]'$ in the CRRA, EZ, DCRRA,LCRRA, LEZ and LDCRRA specifications; $X_t = [C_t/C_{t-1}\ u_t]'$, with either DEF_t, CAY_t or YC_t in place of u_t in the DEF, CAY and YC specifications, respectively; and $X_t = [MKT_t\ D_{m,t}/P_{m,t}]'$ in the remaining specifications.

Finally, we perform several robustness checks on the choice of test assets, proxies for the risk factors, instrument matrix and variable generating information. These checks, that are not reported in this paper because of space limits, support the results obtained with our original choice and provide results that are similar to the ones described in Sec. III.C. We repeat the analysis using different sets of 29 FF research portfolios created on the basis of size, book-to-market, operating profitability, investment, momentum, short reversal, long reversal and liquidity. Among the research portfolios, 16 exemplify the stochastic behavior of assets with a combination of high/low size, book-to-market, operating profitability and investment; other 4 research portfolios derive from the combination of high/low size and momentum; other equity research portfolios derive from the combination of high/low size and long-term reversal; other 4 equity research portfolios derive from the combination of high/low size and short-term reversal.[25] We also consider the Pastor-Stambaugh liquidity portfolio restricting the analysis to the period Dec. 1967 to Dec. 2012.[26]

Since there is typically no distinction between consumption and disposable income in general equilibrium models, we consider disposable income from the FRED database of the St. Louis Fed as an alternative proxy for consumption C_t. We also consider quarterly data on compensation of employees from the U.S. Department of Commerce, Bureau of Economic Analysis, as an alternative proxy for the labor income (see also Santos and Veronesi (2006)). The CAY specification is also estimated and tested using proxies for the monthly transitory component of financial wealth CAY_t obtained by other interpolation methods aside the linear interpolation through levels at following end-of-quarters. We consider the values of the variables between two end-of-quarters as (i) fixed at the previous end-of-quarter levels, (ii) fixed at the following end-of-quarter levels, (iii) fixed at the midpoint between the previous and the following end-of-quarter levels, and (iv) fixed at the closest end-of-quarter levels. In addition, we also consider the FF proxies $SMB5_t$, $HML5_t$, RMW_t and CMA_t constructed using 4 or 16 value-weight portfolios jointly formed on size, book-to-market, operating profitability, and investment.

Following Goyal and Welch (2008) we consider also singularly each of the following variables as first component of vector X_t in place of the total dividend-to-price ratio: earnings to price ratio, stock variance, cross-sectional premium, book-to-market ratio, net equity expansion, percent equity issuing, treasury bills, long term rate of returns, long term yield, term spread, default yield spread, inflation, investment to capital ratio.[27]

Finally, we repeat the comparison of all the SDF specifications using the same vector X_t

[25] The details on the construction of the equity portfolios and the different variants of the FF proxies for the size, book-to-market, operating profitability, investment, momentum, short reversal, long reversal and liquidity factors can be found at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html

[26] The details on the portfolio construction can be found at http://finance.wharton.upenn.edu/ stambaug/.

[27] The data have been retrieved on May 1st, 2015 from http://www.hec.unil.ch/agoyal/. Given the availability of the data, the analysis is carried only up to Dec. 2013.

summarizing the information used for pricing in each asset pricing specification. We use either the vector $[D_{m,t}/P_{m,t}\ C_t/C_{t-1}\ MKT_t\ DEF_t]'$, $[D_{m,t}/P_{m,t}\ C_t/C_{t-1}\ MKT_t\ CAY_t]'$ or $[D_{m,t}/P_{m,t}\ C_t/C_{t-1}\ MKT_t\ YC_t]'$ as vector X_t.

C. Sample unconditional and conditional HJ-distances

We present in this section the estimation and testing results for unconditional and conditional HJ-distances. First, we describe the kernel bandwidth selection for the computation of the conditional HJ-distance in Subsec. C.1. Then, we describe separately the results for the unconditional and conditional HJ-distance in Subsec. C.2 and Subsec. C.3, respectively.

C.1 Optimal GMM estimation

The sample conditional pricing error vector $\hat{e}_T(X_t;\theta,H)$ defined in Eq. (2.19) is a nonparametric kernel regression function, which depends on the kernel bandwidth. Therefore, for the purpose of the bandwidth matrix selection, we first consider an estimate of the SDF process in each asset pricing model by optimal GMM $\hat{\theta}_{GMM,T}$. We then use the time series for $h(Y_t,R_t;\hat{\theta}_{GMM,T})$ as a "pilot" time series for the bandwidth selection.

The columns of instrument matrix used in the GMM computation are a constant, vector X_t and its first lagged value. The estimate of the SDF process is obtained by optimal GMM in its iterative form.[28] The 21 orthogonality conditions for the econometric problem correspond to the no-arbitrage restrictions on the six value-weighted FF research portfolios and the 1-month T-bill combined with the instrument matrix. The weighting matrix for the GMM criterion is the inverse of the matrix of second moments of the instruments at the first step, and the inverse of the positive semi-definite spectral density matrix with 12-months Bartlett window proposed in Newey and West (1987) in the following steps. The numerical optimization is conducted relying on the Broyden–Fletcher–Goldfarb–Shanno algorithm, and the value $1e-6$ is chosen as arbitrary threshold for the convergence of the minimization of the GMM criterion.[29]

C.2 Sample unconditional HJ-distance

[28]See, e.g., Hall (2005) Sec. 2.4 and 3.6 for a discussion on the gain in finite-sample efficiency obtained by the iterated GMM.

[29]We obtain similar estimates relying on steepest descent, Gauss-Newton, Levenberg–Marquardt and Davidon–Fletcher–Powell algorithms.

We compute the sample unconditional HJ-distance $\hat{d}_{Z,T}$, taken as an estimator of the unconditional HJ-distance d_Z, by the usual HJ-statistic. We rely on a constant, vector X_t and its first lagged value to create the instrument matrix. We then have 35 unconditional moment restrictions for the econometric problem (that is, the restrictions in Eq. (1.4)). The sample unconditional HJ-distance $\hat{d}_{Z,T}$ corresponds to the empirical GMM criterion with (non-optimal) weighting matrix Ω_Z. The GMM estimation is performed in its iterative form.

We report in Table VII the value of the squared sample unconditional HJ-distance $\hat{d}_{Z,T}$ and its minimizer $\hat{\theta}_{Z,T}$, for all the competing SDF specifications. We report in the columns with headers L and U the lower and upper bounds at the asymptotic 5%-confidence level under the alternative hypothesis of model misspecification. Overall, beta pricing models show the best performance in terms of the unconditional HJ-distance for distinct choices of the instrument matrix. Similar results are obtained using different instruments and different test assets. [. . .]

C.3 Sample conditional HJ-distance

We consider the time series of vector $h(Y_t, R_t; \hat{\theta}_{GMM,T})$, for vector function $h(\cdot, \cdot; \cdot)$ defined in Eq. (1.3), and where $\hat{\theta}_{GMM,T}$ is the optimal GMM estimate of the SDF parameter vector based on the instrument matrix described in Sec. B.2. We apply to this time series the leave-one-out cross-validation criterion to select the bandwidth matrix. This matrix, that we denote by H_{OPT}, is the minimizer w.r.t. H of the time-averaged squared error, that is

$$H_{OPT} := \arg\min_{H \in \mathbb{R}^{L \times L}} CV(H), \quad CV(H) := \sum_{i=1}^{T-1} \Big(h(Y_{i+1}, R_{i+1}; \hat{\theta}_{GMM,T}) - \hat{e}_{T;-i}(X_i; \hat{\theta}_{GMM,T}, H) \Big)^2, \tag{3.9}$$

where $\hat{e}_{T;-i}$ is the Nadaraya-Watson kernel regression estimator defined similarly as in Eq. (2.19) but computed on the sample missing of the i-th observation. The optimal bandwidth matrix H_{OPT} converges in probability to the matrix H_0 which minimizes the average square error from the nonparametric regression. In addition, the ratio of the average square errors computed at H_{OPT} and H_0 converges to 1 in probability. We report the results for the Gaussian kernel function $K(x) = (2\pi)^{-1} \exp(-0.5x'x)$, for which convolution $\mathcal{K}(x) = (4\pi)^{-1} \exp(-0.25x'x)$ and the integral of the squared convolution $\int_{\mathbb{R}^2} \mathcal{K}(u)^2 du = (8\pi)^{-1}$ are known in closed form. However, the intuition behind the ranking of the competing models does not change using different specifications of the kernel function, such as the rectangular or the Epanechnikov kernel function.

We report in Table IX the squared sample conditional HJ-distance and its minimizer implicitly defined in Eq. (2.10), for all the competing SDF specifications. The estimates are computed relying on the 7 conditional moment restrictions derived from no-arbitrage for the six FF value-weighted research portfolios and the 1-month T-bill. Similarly as in the analysis based on the unconditional HJ-distance in the previous subsection, we report in the columns with headers L and U the lower and upper bounds at the asymptotic 5%-confidence level under the hypothesis of misspecification. We do not reject the hypothesis of correct model specification at the asymptotic 5%-confidence level just for the NM, FF5 and FFM specifications. [. . .]

IV. Conclusion

[. . .]

REFERENCES

Ai, Chunrong, and Xiaohong Chen, 2003, Efficient estimation of models with conditional moment restrictions containing unknown functions, *Econometrica* 71, 1795–1843.

Aït-Sahalia, Yacine, Peter J. Bickel, and Thomas M. Stoker, 2001, Goodness-of-fit tests for kernel regression with an application to option implied volatilities, *Journal of Econometrics* 105, 363–412.

Almeida, Caio, and René Garcia, 2012, Assessing misspecified asset pricing models with empirical likelihood estimators, *Journal of Econometrics* 170, 519–537.

Anatolyev, Stanislav, and Nikolay Gospodinov, 2011, Methods for estimation and inference in modern econometrics. Chapman & Hall/CRC, London.

Andrews, Donald W. K., and Cristopher J. Monahan, 1992, An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica*, 60, 953–966.

Antoine, Bertille, Hélène Bonnal, and Eric 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.

Balduzzi, Pierluigi, and Cesare Robotti, 2010, Asset pricing models and economic risk premia: a decomposition, *Journal of Empirical Finance* 17, 54–80.

Bekaert, Geert, and Jun Liu, 2008, Conditioning information and variance bounds on pricing kernels, *Review of Financial Studies* 17, 339–378.

Bierens, Herman J., 1990, A consistent conditional moment test of functional form, *Econometrica* 58, 1443–1458.

Bierens, Herman J., and Werner Ploberger, 1997, Asymptotic theory of integrated conditional moment test, *Econometrica* 65, 1129–1151.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chabi-Yo, Fousseni, 2008, Conditioning information and variance bounds on pricing kernels with higher-order moments: theory and evidence, *Review of Financial Studies* 21, 181–231.

Chamberlain, Gary, 1987, Asymptotic efficiency in estimation with conditional moment restrictions, *Journal of Econometrics* 34, 305–334.

Cochrane, John H., 2008, The dog that did not bark: a defense of return predictability, *Review of Financial Studies* 21, 1533–1575.

Domínguez, Manuel A., and Ignacio N. Lobato, 2004, Consistent estimation of models defined by conditional moment restrictions, *Econometrica* 72, 1601–1615.

Dunn, Kenneth B. and Kenneth J. Singleton, 1986, Modeling the term structure of interest rates under nonseparable utility and durability of goods, *Journal of Financial Economics*, 17, 27–55.

Eichenbaum, Martin S., and Lars Peter Hansen, 1990, Estimating models with intertemporal substitution using aggregate time series data, *Journal of Business and Economic Statistics*, 8, 53–69.

Epstein, Larry G., and Stanley E. Zin, 1989, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework, *Econometrica*, 937–969.

Epstein, Larry G., and Stanley E. Zin, 1991, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: an empirical analysis, *Journal of Political Economy*, 263–286.

Escanciano, Carlos J., 2006, A consistent diagnostic test for regression models using projections, *Econometric Theory* 22, 1030–1051.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 1998, Value versus growth: the international evidence, *Journal of Finance* 53, 1975–2000.

Fama, Eugene F., and Kenneth R. French, 2014, A five-factor asset pricing model, Forthcoming in *Journal of Financial Economics*.

Fang, Ying, Yu Ren, and Yufei Yuan, 2011, Nonparametric estimation and testing of stochastic discount factor, *Finance Research Letters* 8, 196–205.

Ferson, Wayne, and George Constantinides, 1991, Habit formation and durability in aggregate consumption, empirical tests, *Journal of Financial Economics*, 29, 199–240.

Goyal, Amit, and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Gospodinov, Nikolay, Raymond Kan and Cesare Robotti, 2012, Analytical solution for the constrained Hansen-Jagannathan distance under multivariate ellipticity, *Working paper*.

Gospodinov, Nikolay, Raymond Kan and Cesare Robotti, 2013, Chi-squared tests for evaluation and comparison of asset pricing models, *Journal of Econometrics* 173, 108–125.

Gospodinov, Nikolay, and Taisuke Otsu, 2012, Local GMM estimation of time series models with conditional moment restrictions, *Journal of Econometrics* 170, 476–490.

Hall, Alastair R., 2005, Generalized method of moments. Oxford University Press, Oxford.

Hall, Alastair R., and Atsushi Inoue, 2003, The large sample behaviour of the generalized method of moments estimator in misspecified models, *Journal of Econometrics* 114, 361–394.

Hansen, Bruce E., 2008, Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24, 726–748.

Hansen, Lars Peter, 1982, Large sample properties of generalized method of moments estimators, *Econometrica*, 1029–1054.

Hansen, Lars Peter, John Heaton, and Erzo G. J. Luttmer, 1995, Econometric evaluation of asset pricing models, *Review of Financial Studies* 8, 237–274.

Hansen, Lars Peter, and Ravi Jagannathan, 1997, Assessing specification errors in stochastic discount factor models, *Journal of Finance* 52, 557–590.

Hansen, Lars Peter, and Scott F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica*, 587–613.

Hansen, Lars Peter, and Kenneth J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica*, 1269–1286.

Harrison, Michael J., and David M. Kreps, 1979, Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory* 20, 381–408.

Hodrick, Robert J., and Xiaoyan Zhang, 2001, Evaluating the specification errors of asset pricing models, *Journal of Financial Economics* 62, 327–376.

Hou, Kewei, Chen Xue, and Lu Zhang, 2014, Digesting anomalies: an investment approach, Forthcoming in *Review of Financial Studies*.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

Kan, Raymond, and Cesare Robotti, 2009, Model comparison using the Hansen-Jagannathan distance, *Review of Financial Studies* 22, 3449–3490.

Kitamura, Yuichi, Gautam Tripathi, and Hyungtaik Ahn, 2004, Empirical likelihood-based inference in conditional moment restriction models, *Econometrica* 72, 1667–1714.

Lettau, Martin, and Sydney Ludvigson, 2001, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.

Lewellen, Jonathan, Stefan Nagel, and Jay Shanken, 2010, A skeptical appraisal of asset-pricing tests, *Journal of Financial Economics* 96, 175–194.

Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13–37.

Mossin, Jan, 1966, Equilibrium in a capital asset market, *Econometrica* 34, 768–783.

Nagel, Stefan, and Kenneth J. Singleton, 2008, Estimation and evaluation of conditional asset pricing models, *Working paper*.

Nagel, Stefan, and Kenneth J. Singleton, 2011, Estimation and evaluation of conditional asset pricing models, *Journal of Finance* 66, 873–909.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 55, 703–708.

Newey, Whitney K., 1993, Efficient estimation of models with conditional moment restrictions, in *Handbook of Statistics* 11, 419–454.

Newey, Whitney K., 1994, Kernel estimation of partial means and a general variance estimator, *Econometric Theory* 10, 1–21.

Newey, Whitney K., and Kenneth D. West, 1994, Automatic lag selection in covariance matrix estimation, *Review of Economic Studies*, 61, 631–653.

Novy-Marx, Robert, 2013, The other side of value: the gross profitability premium, *Journal of Financial Economics*, 108, 1–28.

Parker, Jonathan A., and Christian Julliard, 2005, Consumption risk and the cross section of expected returns, *Journal of Political Economy* 113, 185–222.

Pastor, Lubos, and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

Santos, Tano, and Pietro Veronesi, 2006, Labor income and predictable stock returns, *Review of Financial Studies* 19, 1–44.

Sharpe, William F., 1964, Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.

Silverman, Bernard W., 1986, Density estimation for statistics and data analysis. Chapman & Hall/CRC, London.

Stinchcombe, Maxwell B., and Halbert White, 1998, A consistent diagnostic test for regression models using projections, *Econometric Theory* 14, 295–325.

Stock, James H., and Jonathan H. Wright, 2000, GMM with weak identification, *Econometrica* 68, 1055–1096.

Su, Liangjun, and Halbert White, 2014, Testing conditional independence via empirical likelihood, *Journal of Econometrics* 182, 27–44.

Treynor, Jack L., 1962, Toward a theory of market value of risky assets. Unpublished manuscript. A final version was published in 1999, in Asset Pricing and Portfolio Performance: Models, Strategy and Performance Metrics. Risk Books, London.

Tripathi, Gautam, and Yuichi Kitamura, 2003, Testing conditional moment restrictions, *The Annals of Statistics* 31, 2059–2095.

Yogo, Motohiro, 2006, A consumption-based explanation of expected stock returns, *Journal of Finance*, 61, 539–580.

Yoshihara, Ken-ichi, 1976, Limiting behavior of U-statistics for stationary, absolutely regular processes, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 35, 237–252.

Yoshihara, Ken-ichi, 1989, Limiting behavior of generalized quadratic forms generated by absolutely regular processes. In *Proceedings of the Fourth Prague Conference on Asymptotic Statistics*, 539–547.

Table I: Sample unconditional statistics of deflated real excess returns on the test equity portfolios, real gross return for the T-bill and proxies for the priced factors included in the asset pricing models. The following descriptive statistics are reported in the same order: sample size, mean, median, standard deviation, skewness, excess kurtosis, minimum value, maximum value, $25th$ and $75th$ percentiles. The standard error for skewness and excess kurtosis from normal approximation for each risk factor is 0.10 and 0.20, respectively.

	T	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
C_t/C_{t-1}	618	.96	1.03	1.0021	.00746	-.029	2.650
$C_{D,t}/C_t$	618	.85	1.14	1.0047	.02654	-.127	4.776
DEF_t	618	.29	6.01	2.0375	.80815	.934	2.841
CAY_t	601	-.04	.03	-.0009	.01884	.048	-.907
YC_t	613	.61	.87	.7372	.07359	.154	-1.188
MKT_t	618	-23.24	16.10	.5074	4.45640	-.538	1.970
$SMB3_t$	618	-16.40	22.02	.2398	3.10271	.536	5.559
$HML3_t$	618	-12.61	13.88	.3627	2.86053	.010	2.590
RMW_t	618	-17.60	12.24	.2482	2.13958	-.407	11.605
CMA_t	618	-6.76	8.93	.3231	1.99049	.256	1.389
UMD_t	618	-34.72	18.39	.6890	4.22668	-1.425	11.322
ST_t	618	-14.54	16.22	.4957	3.14127	.374	5.726
LT_t	618	-7.79	14.51	.2957	2.51457	.646	2.724
L_t	606	-.38	.29	-.0001	.05654	-1.211	6.483
$D_{m,t}/P_{m,t}$	618	-4.73	19.91	.0440	.92039	16.068	354.473
SG_t	618	.68	1.28	1.0091	.06812	-.340	1.810
SN_t	618	.72	1.27	1.0129	.05401	-.501	2.905
SV_t	618	.72	1.30	1.0143	.05548	-.410	3.492
BG_t	618	.77	1.21	1.0089	.04606	-.319	1.928
BN_t	618	.80	1.17	1.0094	.04300	-.367	2.161
BV_t	618	.77	1.21	1.0109	.04634	-.448	2.725

Table II: Sample contemporaneous correlations of the proxies for the priced factors included in the asset pricing models. Pointwise correlation coefficient estimates and standard errors from normal approximation are reported. The symbols "†" and "∗" indicate significant correlations at the 0.01 and 0.05 level, respectively.

	$C_{D,t}/C_t$	DEF_t	CAY_t	YC_t	MKT_t	$SMB3_t$	$HML3_t$	RMW_t	CMA_t	UMD_t	ST_t	LT_t	L_t	$D_{m,t}/P_{m,t}$
C_t/C_{t-1}	.214†	-.057	.025	.019	.165†	.119†	-.042	-.029	-.090*	-.011	.039	.043	.034	.065
$C_{D,t}/C_t$		-.002	-.003	-.023	.030	.161†	.009	-.060	.006	-.066	.054	.069	.076	.014
DEF_t			-.166†	-.539†	.063	.040	-.055	.043	.039	-.141†	.000	.018	-.047	-.005
CAY_t				-.176†	.009	-.048	-.018	.100*	.008	.055	-.028	-.032	.071	-.040
YC_t					-.059	.020	.034	-.058	.015	.061	.082*	.048	-.082*	.021
MKT_t						.309†	-.299†	-.205†	-.390†	-.125†	.282†	-.024	.336†	.055
$SMB3_t$							-.226†	-.403†	-.179†	-.003	.161†	.263†	.160†	.019
$HML3_t$								.082*	.702†	-.156†	-.018	.425†	-.087*	-.015
RMW_t									-.100*	.097*	-.059	-.305†	-.002	-.005
CMA_t										-.012	-.124†	.484†	-.137†	-.004
UMD_t											-.283†	-.053	-.034	-.008
ST_t												.079	.071	.019
LT_t													-.027	.009
L_t														.058

Table III: First sample autocorrelation coefficients (A) and sample partial autocorrelation coefficients (PA) of the proxies for the priced factors included in the asset pricing models and the gross returns on the test assets. The standard errors from normal approximation for all the coefficients is 0.04.

Variable	Lag	A	PA	Variable	Lag	A	PA
C_t/C_{t-1}	1	-0.34	-0.34	$HML3_t$	1	0.16	0.16
	2	0.01	-0.13		2	0.04	0.02
	3	0.12	0.1		3	0.04	0.03
	4	-0.05	0.04		4	0.03	0.02
$C_{D,t}/C_t$	1	-0.34	-0.18	RMW_t	1	0.18	0.18
	2	0.01	-0.17		2	0.04	0.01
	3	0.12	-0.12		3	-0.04	-0.05
	4	-0.05	-0.06		4	0.03	0.05
DEF_t	1	0.97	0.97	CMA_t	1	0.14	0.14
	2	0.92	-0.27		2	0.04	0.02
	3	0.88	0.13		3	0.05	0.04
	4	0.84	-0.07		4	-0.06	-0.08
CAY_t	1	0.99	0.99	UMD_t	1	0.06	0.06
	2	0.98	-0.47		2	-0.07	-0.07
	3	0.96	0.08		3	0.01	0.02
	4	0.94	0.26		4	0.04	0.03
YC_t	1	1	1	ST_t	1	-0.02	-0.02
	2	0.99	0.05		2	-0.07	-0.07
	3	0.99	-0.01		3	0.02	0.02
	4	0.98	-0.01		4	-0.05	-0.05
MKT_t	1	0.08	0.08	LT_t	1	0.16	0.16
	2	-0.03	-0.04		2	0.05	0.02
	3	0.02	0.03		3	-0.07	-0.08
	4	0.01	0.01		4	-0.02	0.01
$SMB3_t$	1	0.05	0.05	L_t	1	0	0
	2	0.04	0.04		2	-0.01	-0.01
	3	-0.08	-0.09		3	0.11	0.11
	4	0	0.01		4	-0.04	-0.04

Table IV: Sample pointwise cross-correlations of the proxies for the priced factors included in the asset pricing models at distinct times. For each pair of variables we report the set of cross-correlation only if at least one of the correlations between variables at different time is less than 0.12 in absolute value. For any coefficient the standard error based on the assumption that the series are not cross correlated and that one of the series is white noise is is 0.041.

Lag	$C_t/C_{t-1}, DEF_{t-j}$	$C_{D,t}/C_t, MKT_{t-j}$	DEF_t, CAY_{t-j}	DEF_t, YC_{t-j}	DEF_t, L_{t-j}	CAY_t, YC_{t-j}	$MKT_t, SMB3_{t-j}$	MKT_t, ST_{t-j}	MKT_t, L_{t-j}	$SMB3_t, ST_{t-j}$	$SMB3_t, L_{t-j}$	$HML3_t, CMA_{t-j}$	$HML3_t, LT_{t-j}$	RMW_t, ST_{t-j}	CMA_t, ST_{t-j}	CMA_t, LT_{t-j}
-7	-.022	-.016	-.143	-.471	-.134	-.181	.033	-.067	-.005	.005	-.038	-.013	-.047	-.018	.028	-.017
-6	-.035	.108	-.138	-.481	-.125	-.180	-.001	-.078	.007	.052	-.060	.045	-.004	.023	.086	.019
-5	-.038	.017	-.133	-.490	-.109	-.179	-.063	.017	.089	.010	.043	-.038	.011	-.020	-.029	.009
-4	-.047	.064	-.132	-.500	-.123	-.178	-.028	.018	-.012	-.029	-.014	-.062	-.072	.061	.026	-.048
-3	-.039	.043	-.137	-.511	-.131	-.177	-.092	.042	.010	.047	.022	.033	-.042	-.042	-.009	-.019
-2	-.047	-.009	-.144	-.520	-.108	-.177	-.063	-.004	-.034	-.020	-.062	.036	.006	.112	-.001	.043
-1	-.065	.132	-.155	-.528	-.091	-.177	.063	.020	.057	.086	.040	.122	.132	.026	.093	.147
0	-.062	.028	-.166	-.538	-.050	-.176	.310	.287	.335	.166	.162	.701	.421	-.059	-.128	.484
1	-.097	.078	-.174	-.544	-.025	-.178	.214	.047	.160	.025	.130	.115	.072	-.159	-.075	.108
2	-.116	-.054	-.182	-.548	-.035	-.180	-.001	-.156	.081	-.078	.004	.027	-.037	.054	.203	-.021
3	-.116	-.011	-.187	-.552	-.011	-.181	.017	-.074	-.003	-.190	.080	.048	-.008	.170	.009	-.038
4	-.127	.028	-.188	-.554	.024	-.183	-.058	-.043	.053	.006	-.041	-.018	-.005	-.038	.053	.022
5	-.118	.017	-.184	-.556	.028	-.184	-.094	-.010	.013	-.053	-.066	-.017	.021	-.038	.117	.057
6	-.113	-.010	-.179	-.559	.031	-.186	-.022	-.019	.056	.020	.040	.010	-.045	-.039	.047	.041
7	-.116	-.023	-.173	-.561	.011	-.187	-.045	-.007	.004	.093	-.057	.009	.029	-.033	-.009	.028

Table V: Sample Hansen's J statistic $\hat{J}_Z$ and optimal GMM estimates of the SDF parameter vectors $\hat{\theta}_{GMM} = [\hat{\theta}_{1,GMM} \ \ldots \ \hat{\theta}_{6,GMM}]'$ for the competing SDF specifications. The optimal GMM in implemented in its iterative form. The orthogonality conditions are the no-arbitrage restrictions on the six value-weighted FF research portfolios and the 1-month T-bill combined with the instrument matrix, which includes the constant and the variables reported under the column "Instrument", for a total of 21 orthogonality conditions considered in the estimation of each asset pricing model. The weighting matrix for the GMM criterion the inverse of the matrix of second moments of the instruments at the first step, and the inverse of the positive semi-definite spectral density matrix with 12-months Bartlett window proposed in the following steps. The value $1e-6$ is chosen as arbitrary threshold for the convergence of the minimization of the GMM criterion. Each estimate is given with standard error and p-value in brackets. In any model specification, the probability $\mathbb{P}[\chi^2_{df} > \hat{J}_Z]$, where df is 21 minus the number of scalar parameters in the specification, is null.

Specification	Instruments	$\hat{J}_Z$	$\hat{\theta}_{1,GMM}$	$\hat{\theta}_{2,GMM}$	$\hat{\theta}_{3,GMM}$	$\hat{\theta}_{4,GMM}$	$\hat{\theta}_{5,GMM}$	$\hat{\theta}_{6,GMM}$
FFM	$MKT_t, D_{m,t}/P_{m,t}$	31.61	0.59 (4.66; 0.93)	-5.24 (1.48; 0.14)	-3.56 (1.53; 0.02)	-1.56 (2.64; 0.56)	10.74 (3.33; 0.00)	-
DEF	$C_t/C_{t-1}, DEF_t$	38.84	0.86 (0.11; 0.20)	162.27 (43.37; 0.00)	0.11 (0.06; 0.08)	-114.28 (24.49; 0.00)	-	-
FFREV	$MKT_t, D_{m,t}/P_{m,t}$	61.47	16.23 (4.83; 0.00)	-3.15 (1.03; 0.00)	1.12 (1.21; 0.35)	-2.58 (1.72; 0.13)	0.06 (3.36; 0.98)	-10.64 (4.23; 0.01)
CAY	$C_t/C_{t-1}, CAY_t$	61.50	1.00 (0.01; 0.66)	-0.06 (3.69; 0.99)	-0.04 (1.78; 0.98)	0.00 (941.57; 1.00)	-	-
FFL	$MKT_t, D_{m,t}/P_{m,t}$	63.11	-326.39 (144.37; 0.02)	-5.04 (1.07; 0.00)	-6.87 (1.39; 0.00)	-8.72 (1.47; 0.00)	348.10 (146.12; 0.02)	-
LEZ	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	72.40	2.32 (0.13; 0.00)	0.58 (0.18; 0.00)	-1.32 (0.13; 0.00)	-	-	-
FF3	$MKT_t, D_{m,t}/P_{m,t}$	89.25	2.87 (0.37; 0.00)	-0.62 (0.65; 0.34)	-1.21 (0.15; 0.00)	-0.04 (0.88; 0.96)	-	-
NM	$MKT_t, D_{m,t}/P_{m,t}$	91.96	-41.88 (996.10; 0.97)	0.02 (1.19; 0.99)	-56.06 (475.58; 0.91)	140.82 (47.91; 0.00)	-41.91 (572.05; 0.94)	-
EZ	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	90.11	1.46 (130.95; 0.99)	4.07 (944.90; 0.99)	0.02 (3.71; 0.99)	-	-	-
CRRA	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	90.38	1.00 (0.00; 0.00)	0.07 (0.02; 0.00)	-	-	-	-
LCRRA	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	90.38	1.00 (0.00; 0.00)	-0.07 (0.02; 0.01)	-	-	-	-
CAPM	$MKT_t, D_{m,t}/P_{m,t}$	89.50	3.09 (0.08; 0.00)	-2.09 (0.08; 0.00)	-	-	-	-
YC	$C_t/C_{t-1}, YC_t$	95.65	1.01 (0.35; 0.98)	-0.04 (199.27; 1.00)	0.02 (0.46; 0.97)	-0.03 (259.15; 1.00)	-	-
DCRRA	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	5496.51	1.00 (0.00)	-38.00 (4.83)	-0.90 (0.01)	0.00 (197.17)	0.01 (0.06)	-
LDCRRA	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	2415.10	1.00 (0.00)	22.11 (2.65)	5.90 (2.16)	-118.22 (2.54)	-0.99 (0.00)	-
FF5	$MKT_t, D_{m,t}/P_{m,t}$	1298.69	1.65 (0.00)	-0.65 (0.00)	(2.53 (0.01)) $\times 10^{-3}$	(-1.25 (0.03)) $\times 10^{-3}$	(-0.75 (0.03)) $\times 10^{-3}$	(0.25 (0.04)) $\times 10^{-3}$

Table VI: Sample Hansen's J statistic as in Table V but with a different instrument matrix. The instrument matrix includes the constant and the variables reported under the column "Instrument" along with their first lagged values. Therefore, 35 orthogonality conditions are considered in the estimation of each asset pricing model. Every comment for Table V applies also to this table.

Specification	Instruments	$\hat{J}_Z$	$\hat{\theta}_{1,GMM}$	$\hat{\theta}_{2,GMM}$	$\hat{\theta}_{3,GMM}$	$\hat{\theta}_{4,GMM}$	$\hat{\theta}_{5,GMM}$	$\hat{\theta}_{6,GMM}$
FFM	$MKT_t, D_{m,t}/P_{m,t}$	61.18	-1.71 (2.40; 0.26)	-3.95 (1.04; 0.00)	-0.84 (1.21; 0.49)	-0.38 (2.05; 0.85)	7.86 (1.63; 0.00)	-
FFREV	$MKT_t, D_{m,t}/P_{m,t}$	81.91	15.38 (2.56; 0.00)	-5.75 (0.86; 0.00)	-2.27 (0.81; 0.00)	-8.40 (1.09; 0.00)	2.68 (1.24; 0.03)	-0.59 (1.69; 0.73)
FF3	$MKT_t, D_{m,t}/P_{m,t}$	94.44	3.10 (0.27; 0.00)	-1.54 (0.16; 0.00)	-1.17 (0.13; 0.00)	0.61 (0.39; 0.11)	-	-
CAY	$C_t/C_{t-1}, CAY_t$	95.93	1.00 (0.00; 0.00)	-0.12 (0.05; 0.02)	-0.02 (0.02; 0.19)	-10.84 (7.04; 0.12)	0.00 (941.57; 1.00)	-
FFL	$MKT_t, D_{m,t}/P_{m,t}$	96.42	39.99 (32.39; 0.23)	-4.70 (0.48; 0.00)	3.63 (0.73; 0.00)	-6.50 (0.86; 0.00)	-31.37 (32.60; 0.34)	-
CAPM	$MKT_t, D_{m,t}/P_{m,t}$	98.61	3.67 (0.12; 0.00)	-2.66 (0.12; 0.00)	-	-	-	-
NM	$MKT_t, D_{m,t}/P_{m,t}$	104.24	53.43 (48.54; 0.28)	-1.19 (0.14; 0.00)	-9.32 (39.52; 0.81)	-26.43 (8.35; 0.00)	-15.49 (24.62; 0.53)	-
LEZ	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	106.91	2.11 (0.04; 0.00)	0.49 (0.09; 0.00)	-1.11 (0.04; 0.00)	-	-	-
DEF	$C_t/C_{t-1}, DEF_t$	111.33	0.99 (0.00; 0.00)	-0.57 (1.52; 0.71)	0.00 (0.00; 0.00)	0.66 (0.89; 0.46)	-	-
EZ	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	129.05	1.24 (17.94; 0.98)	-1.46 (95.24; 0.99)	0.03 (1.85; 0.99)	-	-	-
CRRA	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	129.48	1.00 (0.00; 0.00)	-0.04 (0.01; 0.00)	-	-	-	-
LCRRA	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	129.48	1.00 (0.00; 0.00)	0.04 (0.01; 0.00)	-	-	-	-
YC	$C_t/C_{t-1}, YC_t$	130.05	1.01 (0.00; 0.00)	0.74 (1.28; 0.57)	-0.01 (0.00; 0.00)	-1.03 (1.67; 0.54)	-	-
DCRRA	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	5496.51	1.00 (0.00)	-38.00 (4.83)	-0.90 (0.01)	0.00 (197.17)	0.01 (0.06)	-
LDCRRA	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	2415.10	1.00 (0.00)	22.11 (2.65)	5.90 (2.16)	-118.22 (2.54)	-0.99 (0.00)	-
FF5	$MKT_t, D_{m,t}/P_{m,t}$	1298.69	1.65 (0.00)	-0.65 (0.00)	(2.53 (0.01)) $\times 10^{-3}$	(-1.25 (0.03)) $\times 10^{-3}$	(-0.75 (0.03)) $\times 10^{-3}$	(0.25 (0.04)) $\times 10^{-3}$

Table VII: Sample squared unconditional HJ-distance $\hat{d}_T^2$ and its minimizer $\hat{\theta}_Z = [\hat{\theta}_{1,Z} \ \ldots \ \hat{\theta}_{6,Z}]'$ for the competing SDF specifications. The non-optimal GMM in implemented in its iterative form. The orthogonality conditions are the no-arbitrage restrictions on the six value-weighted FF research portfolios and the 1-month T-bill combined with the instrument matrix, which includes the constant and the variables reported under the column "Instrument", for a total of 21 orthogonality conditions considered in the estimation of each asset pricing model. The weighting matrix for the GMM criterion is the inverse of the matrix of second moments of the instruments at the first step, and the inverse of the positive semi-definite spectral density matrix with 12-months Bartlett window proposed in the following steps. The value $1e-6$ is chosen as arbitrary threshold for the convergence of the minimization of the GMM criterion. Each estimate is given with standard error and p-value in brackets. In any model specification, the probability $\mathbb{P}[\chi^2_{df} > \hat{J}_Z]$, where df is 21 minus the number of scalar parameters in the specification, is null. The symbol $*$ close to the specification name indicates that we do not reject the hypothesis of correct model specification at the asymptotic 5%-confidence level. In all the other cases we report in columns L and U the confidence bounds for the squared unconditional HJ-distance under the alternative hypothesis of model misspecification. The pointwise estimate of the SDF parameter vector minimizing the unconditional HJ-distance is also given. In any model specification, the probability $\mathbb{P}[\chi^2_{df} > \hat{J}_Z]$, where df is 21 minus the number of scalar parameters in the specification, is null.

Specification	Instruments	$\hat{d}_T^2$	$\hat{\theta}_{1,Z}$	$\hat{\theta}_{2,Z}$	$\hat{\theta}_{3,Z}$	$\hat{\theta}_{4,Z}$	$\hat{\theta}_{5,Z}$	$\hat{\theta}_{6,Z}$
FFM	$MKT_t, D_{m,t}/P_{m,t}$	35.44	39.51 (9.33; 0.00)	-5.69 (1.84; 0.00)	-1.95 (2.12; 0.36)	-9.62 (2.80; 0.00)	-21.04 (6.33; 0.00)	-
DEF	$C_t/C_{t-1}, DEF_t$	51.38	0.84 (0.13; 0.24)	172.97 (48.55; 0.00)	0.10 (0.07; 0.15)	-112.06 (28.19; 0.00)	-	-
FFL	$MKT_t, D_{m,t}/P_{m,t}$	73.42	-156.55 (146.54; 0.28)	-4.49 (1.16; 0.00)	-4.92 (1.32; 0.00)	-8.24 (1.27; 0.00)	175.26 (148.03; 0.24)	-
FFREV	$MKT_t, D_{m,t}/P_{m,t}$	73.91	10.34 (4.12; 0.02)	-4.16 (1.35; 0.00)	-2.16 (1.73; 0.21)	-3.21 (2.75; 0.24)	9.68 (3.87; 0.01)	-9.48 (5.53; 0.09)
EZ	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	76.36	0.99 (0.00; 0.00)	0.00 (0.00; 0.0.3)	-217.47 (14.44; 0.00)	-	-	-
LEZ	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	79.19	2.39 (0.17; 0.12)	0.26 (0.17; 0.12)	-1.39 (0.11; 0.00)	-	-	-
FF3	$MKT_t, D_{m,t}/P_{m,t}$	79.61	12.38 (0.99; 0.00)	-2.93 (0.89; 0.00)	-2.57 (0.68; 0.00)	-5.86 (1.07; 0.96)	-	-
CAY	$C_t/C_{t-1}, CAY_t$	83.38	1.00 (0.01; 0.70)	-0.02 (4.04; 1.00)	-0.04 (1.96; 0.98)	0.00 (1034.53; 1.00)	-	-
YC	$C_t/C_{t-1}, YC_t$	84.14	1.01 (0.45; 0.98)	-0.01 (257.25; 1.00)	-0.02 (0.59; 0.97)	-0.01 (334.55; 1.00)	-	-
CRRA	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	89.73	1.00 (0.00; 0.00)	0.05 (0.03; 0.11)	-	-	-	-
LCRRA	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	89.73	1.00 (0.00; 0.00)	-0.05 (0.03; 0.01)	-	-	-	-
CAPM	$MKT_t, D_{m,t}/P_{m,t}$	115.31	3.25 (0.13; 0.00)	-2.24 (0.13; 0.00)	-	-	-	-
NM	$MKT_t, D_{m,t}/P_{m,t}$	124.24	-38.95 (902.51; 0.97)	1.73 (1.19; 0.15)	-88.86 (428.27; 0.84)	136.99 (67.60; 0.04)	-9.93 (506.42; 0.98)	-
DCRRA	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	5496.51	1.00 (0.00)	-38.00 (4.83)	-0.90 (0.01)	0.00 (197.17)	0.01 (0.06)	-
LDCRRA	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	2415.10	1.00 (0.00)	22.11 (2.65)	5.90 (2.16)	-118.22 (2.54)	-0.99 (0.00)	-
FF5	$MKT_t, D_{m,t}/P_{m,t}$	1298.69	1.65 (0.00)	-0.65 (0.00)	(2.53 (0.01)) $\times 10^{-3}$	(-1.25 (0.03)) $\times 10^{-3}$	(-0.75 (0.03)) $\times 10^{-3}$	(0.25 (0.04)) $\times 10^{-3}$

Table VIII: Sample Hansen's J statistic as in Table VII but with a different instrument matrix. The instrument matrix includes the constant and the variables reported under the column "Instrument" along with their first lagged values. Therefore, 35 orthogonality conditions are considered in the estimation of each asset pricing model. Every comment for Table VII applies also to this table.

Specification	Instruments	$\hat{d}_T^2$	$\hat{\theta}_{1,Z}$	$\hat{\theta}_{2,Z}$	$\hat{\theta}_{3,Z}$	$\hat{\theta}_{4,Z}$	$\hat{\theta}_{5,Z}$	$\hat{\theta}_{6,Z}$
FFREV	$MKT_t, D_{m,t}/P_{m,t}$	82.45	14.13 (2.38; 0.00)	-3.67 (1.07; 0.00)	-2.96 (0.84; 0.00)	-7.07 (1.16; 0.00)	1.48 (1.36; 0.28)	-0.86 (1.77; 0.62)
FFM	$MKT_t, D_{m,t}/P_{m,t}$	84.99	14.83 (0.99; 0.00)	-3.03 (0.89; 0.02)	-3.50 (0.73; 0.00)	-6.74 (1.09; 0.00)	-0.52 (0.41; 0.21)	-
FFL	$MKT_t, D_{m,t}/P_{m,t}$	86.00	20.65 (76.31; 0.80)	-3.38 (0.99; 0.00)	-4.48 (1.09; 0.00)	-7.82 (1.20; 0.00)	-3.90 (77.26; 0.96)	-
EZ	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	89.61	0.99 (0.00; 0.00)	0.00 (0.00; 0.0.3)	-244.63 (7.10; 0.00)	-	-	-
LEZ	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	93.93	2.38 (0.05; 0.12)	0.31 (0.11; 0.00)	-1.38 (0.05; 0.00)	-	-	-
YC	$C_t/C_{t-1}, YC_t$	98.63	1.01 (0.00; 0.00)	-2.41 (2.09; 0.25)	-0.02 (0.00; 0.00)	3.08 (2.70; 0.25)	-	-
DEF	$C_t/C_{t-1}, DEF_t$	106.38	0.99 (0.00; 0.01)	-0.82 (1.45; 0.57)	0.00 (0.00; 0.60)	0.51 (0.86; 0.55)	-	-
CAY	$C_t/C_{t-1}, CAY_t$	108.48	1.00 (0.00; 0.00)	-0.06 (0.05; 0.18)	-0.03 (0.02; 0.13)	-9.82 (9.01; 0.28)	-	-
CRRA	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	111.79	1.00 (0.00; 0.00)	0.02 (0.03; 0.44)	-	-	-	-
LCRRA	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	111.79	1.00 (0.00; 0.00)	-0.02 (0.03; 0.43)	-	-	-	-
CAPM	$MKT_t, D_{m,t}/P_{m,t}$	121.19	3.38 (0.09; 0.00)	-2.37 (0.09; 0.00)	-	-	-	-
NM	$MKT_t, D_{m,t}/P_{m,t}$	134.13	1.00 (27.24; 1.00)	-0.01 (0.11; 0.96)	0.00 (27.68; 1.00)	0.00 (5.48; 1.00)	0.00 (13.16; 1.00)	-
DCRRA	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	5496.51	1.00 (0.00)	-38.00 (4.83)	-0.90 (0.01)	0.00 (197.17)	0.01 (0.06)	-
LDCRRA	$C_t/C_{t-1}, D_{m,t}/P_{m,t}$	2415.10	1.00 (0.00)	22.11 (2.65)	5.90 (2.16)	-118.22 (2.54)	-0.99 (0.00)	-
FF5	$MKT_t, D_{m,t}/P_{m,t}$	1298.69	1.65 (0.00)	-0.65 (0.00)	(2.53 (0.01)) $\times 10^{-3}$	(-1.25 (0.03)) $\times 10^{-3}$	(-0.75 (0.03)) $\times 10^{-3}$	(0.25 (0.04)) $\times 10^{-3}$
FF3	$MKT_t, D_{m,t}/P_{m,t}$	135.93	1.00 (0.06; 1.00)	-0.01 (0.05; 0.94)	0.00 (0.04; 0.99)	0.00 (0.09; 1.00)	-	-

Table IX: Sample squared conditional HJ-distance $\hat{\delta}_T^2$ and its minimizer $\hat{\theta}_* = [\hat{\theta}_{1,*} \ \ldots \ \hat{\theta}_{6,*}]'$ for the competing SDF specifications. The conditional moment restrictions from no-arbitrage for the six FF value-weighted research portfolios and the 1-month T-bill are considered. The results are for the variable $X_t = [C_t/C_{t-1} \ YC_t]'$. The symbol $*$ close to the specification name indicates that we do not reject the hypothesis of correct model specification at the asymptotic 5%-confidence level. In all the other cases we report in columns L and U the confidence bounds for the squared conditional HJ-distance under the alternative hypothesis of misspecification. The pointwise estimate of the SDF parameter vector minimizing the conditional HJ-distance is also given.

SDF specification	$\hat{\delta}_T^2$	$\hat{\theta}_{1,*}$	$\hat{\theta}_{2,*}$	$\hat{\theta}_{3,*}$	$\hat{\theta}_{4,*}$	$\hat{\theta}_{5,*}$	$\hat{\theta}_{6,*}$	L	U
NM*	0.04	1.23	-0.09	-29.21	-0.90	-32.45	-		
FF5	0.05	0.71	-0.06	-0.02	-0.05	-0.02	-0.01		
FFM*	0.07	-0.07	-0.02	-0.02	-0.08	0.04	-		
DEF	0.13	2.01	371.40	-1.25	-321.30	-	-	0.09	0.17
DCRRA	0.13	1.01	-0.67	-12.33	-0.78	-0.11	-	0.11	0.17
EZ	0.15	0.99	-1.84	-4.58	-	-	-	0.11	0.17
CAY	0.16	1.40	-82.41	-69.98	5779.52	-	-	0.14	0.18
LDCRRA	0.17	0.81	19.66	7.95	-6.00	-0.21	-	0.16	0.18
LEZ	0.19	1.26	-3.14	-11.11	-	-	-	0.18	0.22
YC	0.19	2.76	-302.70	-2.50	405.82	-	-	0.17	0.21
FFL	0.19	0.93	-0.04	-0.02	-0.01	-0.48	-	0.16	0.22
FFREV	0.19	1.26	-1.99	-0.76	-0.04	-0.07	-0.01	0.18	0.22
FF3	0.19	1.00	-0.06	-0.02	-0.06	-	-	0.16	0.22
CAPM	0.20	0.96	-0.05	-	-	-	-	0.17	0.23
CRRA	0.20	0.87	-10.31	-	-	-	-	0.17	0.23
LCRRA	0.20	0.87	8.93	-	-	-	-	0.16	0.24

Appendix A. Proof of the results in Section I.B

In this section we show the proofs of the relations between conditional and unconditional HJ-distances given in Sec. I.B. In these proofs, we denote the $(q \times m)$-dimensional $L^2_\Omega(\mathcal{X})$-inner product between any $(N \times q)$-dimensional matrix function $\Phi(X_t) := [\phi_1(X_t) \; \ldots \; \phi_q(X_t)]$ and any $(N \times m)$-dimensional matrix function $\Psi(X_t) := [\psi_1(X_t) \; \ldots \; \psi_m(X_t)]$ with square integrable elements in the following way:

$$\langle \Phi, \Psi \rangle_{L^2_\Omega(\mathcal{X})} := \begin{bmatrix} \langle \phi_1, \psi_1 \rangle_{L^2_\Omega(\mathcal{X})} & \ldots & \langle \phi_1, \psi_m \rangle_{L^2_\Omega(\mathcal{X})} \\ \vdots & \ddots & \vdots \\ \langle \phi_q, \psi_1 \rangle_{L^2_\Omega(\mathcal{X})} & \ldots & \langle \phi_q, \psi_m \rangle_{L^2_\Omega(\mathcal{X})} \end{bmatrix}, \tag{A1}$$

for the definition of the $L^2_\Omega(\mathcal{X})$-inner product given in Eq. (2.9). Consider now the matrix $\Phi(X_t)$ such that its columns are linearly independent vectors in $L^2_\Omega(\mathcal{X})$. We denote by $\mathcal{P}_\Phi$ the orthogonal projection operator onto the linear subspace of $L^2_\Omega(\mathcal{X})$ spanned by the columns of the matrix function Φ. When this orthogonal projection operator is applied to vector function ψ and it is valued at the generic value $x \in \mathcal{X}$ we use the following notation:

$$\mathcal{P}_\Phi[\psi](x) := \Phi(x) \langle \Phi, \Phi \rangle^{-1}_{L^2_\Omega(\mathcal{X})} \langle \Phi, \psi \rangle_{L^2_\Omega(\mathcal{X})}. \tag{A2}$$

Appendix 1. Proof of Equation (2.13)

Let us express the inverse of matrix Ω_Z and the unconditional pricing error vector given in Eq. (2.12) by means of the matrix function A implicitly defined in Eq. (2.11) and the $L^2_\Omega(\mathcal{X})$-inner products defined in Eq. (A1):

$$\mathrm{E}\left[Z_t h(Y_{t+1}, R_{t+1}; \theta)\right] = \langle A, e(\cdot; \theta) \rangle_{L^2_\Omega(\mathcal{X})} \qquad \text{and} \qquad \Omega_Z^{-1} = \langle A, A \rangle_{L^2_\Omega(\mathcal{X})}.$$

We can use these two quantities to express the orthogonal projection operator $\mathcal{P}_A$ onto the linear space spanned by the columns of matrix function $A(X_t)$ applied to the conditional pricing error vector $e(X_t; \theta)$:

$$\mathcal{P}_A[e(\cdot; \theta)](X_t) = A(X_t) \langle A, A \rangle^{-1}_{L^2_\Omega(\mathcal{X})} \langle A, e(\cdot; \theta) \rangle_{L^2_\Omega(\mathcal{X})} = A(X_t) \Omega_Z \mathrm{E}\left[Z_t h(Y_{t+1}, R_{t+1}; \theta)\right]. \tag{A3}$$

Let us consider the $L^2_\Omega(\mathcal{X})$-norm of this vector:

$$\begin{aligned} \|\mathcal{P}_A[e(\cdot; \theta)]\|_{L^2_\Omega(\mathcal{X})} &= \left(\mathrm{E}\left[Z_t h(Y_{t+1}, R_{t+1}; \theta)\right]' \Omega_Z \mathrm{E}\left[A(X_t)' \Omega(X_t) A(X_t)\right] \Omega_Z \mathrm{E}\left[Z_t h(Y_{t+1}, R_{t+1}; \theta)\right]\right)^{1/2} \\ &= \left(\mathrm{E}\left[Z_t h(Y_{t+1}, R_{t+1}; \theta)\right]' \Omega_Z \mathrm{E}\left[Z_t h(Y_{t+1}, R_{t+1}; \theta)\right]\right)^{1/2}. \end{aligned}$$

Therefore, this norm is the criterion minimized by the unconditional HJ-distance d_Z in Eq. (1.7), and Eq. (2.13) follows.

Appendix 2. Proof of Proposition 1

The criterion $\|\mathcal{P}_A[e(\cdot;\theta)]\|_{L^2_\Omega(\mathcal{X})}$ valued at its global minimum θ_Z, which is implicitly defined in Eq. (2.13), cannot be greater than the same criterion at any other value in set Θ. In particular we have $d_Z^2 \leqslant \|\mathcal{P}_A[e(\cdot;\theta_*)]\|^2_{L^2_\Omega(\mathcal{X})}$. From the Pythagorean theorem for inner product spaces we have that

$$\|e(\cdot;\theta_*)\|^2_{L^2_\Omega(\mathcal{X})} = \|\mathcal{P}_A[e(\cdot;\theta_*)]\|^2_{L^2_\Omega(\mathcal{X})} + \|\mathcal{P}_A^\perp[e(\cdot;\theta_*)]\|^2_{L^2_\Omega(\mathcal{X})}. \tag{A4}$$

Using the last equation and inequality and the definition of conditional HJ-distance δ given in Eq. (2.10) we get

$$d_Z^2 \leqslant \delta^2 - \|\mathcal{P}_A^\perp[e(\cdot;\theta_*)]\|^2_{L^2_\Omega(\mathcal{X})},$$

which gives the lower bound for the difference of the squared HJ-distances in Prop. 1. The Pythagorean theorem for inner product spaces implies also that

$$\|e(\cdot;\theta_Z)\|^2_{L^2_\Omega(\mathcal{X})} = \|\mathcal{P}_A[e(\cdot;\theta_Z)]\|^2_{L^2_\Omega(\mathcal{X})} + \|\mathcal{P}_A^\perp[e(\cdot;\theta_Z)]\|^2_{L^2_\Omega(\mathcal{X})} = d_Z^2 + \|\mathcal{P}_A^\perp[e(\cdot;\theta_Z)]\|^2_{L^2_\Omega(\mathcal{X})}, \tag{A5}$$

where we use the expression for d_Z given in Eq. (2.13). By a similar reasoning as for criterion $\|\mathcal{P}_A[e(\cdot;\theta)]\|_{L^2_\Omega}$, the criterion $\|e(\cdot;\theta)\|^2_{L^2_\Omega(\mathcal{X})}$ valued at its global minimum θ_*, which is implicitly defined in Eq. (2.10), cannot be greater than the same criterion at any other value in set Θ. In particular we have that

$$\delta^2 \leqslant \|e(\cdot;\theta_Z)\|^2_{L^2_\Omega(\mathcal{X})}. \tag{A6}$$

From the last inequality and Eq. (A5) we get

$$\delta^2 \leqslant d_Z^2 + \|\mathcal{P}_A^\perp[e(\cdot;\theta_Z)]\|^2_{L^2_\Omega(\mathcal{X})}, \tag{A7}$$

which gives the upper bound for the difference of the squared HJ-distances in Prop. 1.

Appendix 3. Proof of Proposition 2

i) Conditional and unconditional HJ-distances for linear SDF families

For the proof of Prop. 2 we particularize the formulas of the conditional and unconditional HJ-distances to the case of an SDF that is linear in the risk factors. From Eqs. (2.10) and (2.15) the squared conditional HJ-distance is

$$\delta^2 = \min_{\theta\,\in\,\Theta} \|B\theta - 1_N\|^2_{L^2_\Omega(\mathcal{X})} = \|B\theta_* - 1_N\|^2_{L^2_\Omega(\mathcal{X})}. \tag{A8}$$

The minimizer of the quadratic optimization problem in Eq. (A8) is given by

$$\theta_* = \langle B, B\rangle^{-1}_{L^2_\Omega(\mathcal{X})}\langle B, 1_N\rangle_{L^2_\Omega(\mathcal{X})}. \tag{A9}$$

Similarly, from Eq. (2.13) the squared unconditional HJ-distance is

$$d_Z^2 = \min_{\theta\,\in\,\Theta} \|\mathcal{P}_A[B\theta - 1_N]\|^2_{L^2_\Omega(\mathcal{X})} = \|\mathcal{P}_A[B\theta_Z - 1_N]\|^2_{L^2_\Omega(\mathcal{X})}, \tag{A10}$$

where the minimizer θ_Z is given by

$$\theta_Z = \langle \mathcal{P}_A[B], \mathcal{P}_A[B] \rangle_{L^2_\Omega(\mathcal{X})}^{-1} \langle \mathcal{P}_A[B], \mathcal{P}_A[1_N] \rangle_{L^2_\Omega(\mathcal{X})}. \tag{A11}$$

From Eqs. (A9) and (A11) we have

$$B\theta_* = \mathcal{P}_B\left[1_N\right] \qquad \text{and} \qquad \mathcal{P}_A\left[B\theta_Z\right] = \mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_A\left[1_N\right], \tag{A12}$$

where the symbol $\circ$ denotes operator composition, and we use the notation for projection operators given in Eq. (A2). Using the expression for $B\theta_*$ in Eq. (A12) into Eq. (2.15), we obtain the following expression for the conditional pricing error vector at time t for the parameter value θ_*:

$$e\left(X_t; \theta_*\right) = B(X_t)\theta_* - 1_N = \mathcal{P}_B[1_N](X_t) - 1_N = -\mathcal{P}_B^{\perp}[1_N](X_t). \tag{A13}$$

From Eq. (A8) the squared conditional HJ-distance is

$$\delta^2 = \|\mathcal{P}_B^{\perp}[1_N]\|^2_{L^2_\Omega(\mathcal{X})}. \tag{A14}$$

Using the expression for $\mathcal{P}_A\left[B\theta_Z\right]$ in Eq. (A12) we obtain

$$\mathcal{P}_A[B\theta_Z - 1_N] = \mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_A\left[1_N\right] - \mathcal{P}_A\left[1_N\right] = -\mathcal{P}^{\perp}_{\mathcal{P}_A[B]} \circ \mathcal{P}_A[1_N]. \tag{A15}$$

From Eq. (A10) the unconditional HJ-distance is

$$d_Z^2 = \|\mathcal{P}^{\perp}_{\mathcal{P}_A[B]} \circ \mathcal{P}_A[1_N]\|^2_{L^2_\Omega(\mathcal{X})}. \tag{A16}$$

The last expression for the squared unconditional HJ-distance corresponds to the expression for the squared conditional HJ-distance in Eq. (A14) with $\mathcal{P}_A[B]$ instead of B, and $\mathcal{P}_A[1_N]$ instead of 1_N.

ii) Proof of Equation (2.16)
From the Pythagorean theorem for inner product spaces we have

$$\|B\theta - 1_N\|^2_{L^2_\Omega(\mathcal{X})} = \|\mathcal{P}_A[B\theta - 1_N]\|^2_{L^2_\Omega(\mathcal{X})} + \|\mathcal{P}_A^{\perp}[B\theta - 1_N]\|^2_{L^2_\Omega(\mathcal{X})}, \tag{A17}$$

for any $\theta \in \Theta$. The first term in the r.h.s. of the last equation is a quadratic function of vector θ, minimized at θ_Z, where it assumes value d_Z^2 (see Eq. (A10)), so that

$$\|\mathcal{P}_A[B\theta - 1_N]\|^2_{L^2_\Omega(\mathcal{X})} = d_Z^2 + \|\mathcal{P}_A\left[B\left(\theta - \theta_Z\right)\right]\|^2_{L^2_\Omega(\mathcal{X})}, \tag{A18}$$

for any $\theta \in \Theta$. By using Eq. (A18) into Eq. (A17) we get

$$\|B\theta - 1_N\|^2_{L^2_\Omega(\mathcal{X})} = d_Z^2 + \|\mathcal{P}_A\left[B\left(\theta - \theta_Z\right)\right]\|^2_{L^2_\Omega(\mathcal{X})} + \|\mathcal{P}_A^\perp[B\theta - 1_N]\|^2_{L^2_\Omega(\mathcal{X})}, \tag{A19}$$

for any $\theta \in \Theta$. By evaluating this equation at $\theta = \theta_*$, and using Eq. (A8), we get

$$\delta^2 = d_Z^2 + \|\mathcal{P}_A[B\left(\theta_* - \theta_Z\right)]\|^2_{L^2_\Omega(\mathcal{X})} + \|\mathcal{P}_A^\perp\left[B\theta_* - 1_N\right]\|^2_{L^2_\Omega(\mathcal{X})}. \tag{A20}$$

Let us now rewrite the two terms in the r.h.s. of Eq. (A20). From Eq. (A13) we get

$$\mathcal{P}_A^\perp\left[B\theta_* - 1_N\right] = -\mathcal{P}_A^\perp \circ \mathcal{P}_B^\perp[1_N]. \tag{A21}$$

From Eq. (A12) and the fact that $\mathcal{P}^\perp_{\mathcal{P}_A[B]} \circ \mathcal{P}_A \circ \mathcal{P}_B = 0$ and $\mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_A = \mathcal{P}_{\mathcal{P}_A[B]}$ we have

$$\begin{aligned}
\mathcal{P}_A[B\left(\theta_* - \theta_Z\right)] &= \mathcal{P}_A \circ \mathcal{P}_B[1_N] - \mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_A[1_N] \\
&= \mathcal{P}_A \circ \mathcal{P}_B[1_N] - \mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_A \circ \left(\mathcal{P}_B[1_N] + \mathcal{P}_B^\perp[1_N]\right) \\
&= \mathcal{P}^\perp_{\mathcal{P}_A[B]} \circ \mathcal{P}_A \circ \mathcal{P}_B[1_N] - \mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_A \circ \mathcal{P}_B^\perp[1_N] \\
&= -\mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_B^\perp[1_N].
\end{aligned} \tag{A22}$$

Then, using Eqs. (A21) and (A22) into Eq. (A20) we get

$$\delta^2 - d_Z^2 = \|\mathcal{P}_A^\perp \circ \mathcal{P}_B^\perp[1_N]\|^2_{L^2_\Omega(\mathcal{X})} + \|\mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_B^\perp[1_N]\|^2_{L^2_\Omega(\mathcal{X})}. \tag{A23}$$

Considering that $\mathcal{P}_B^\perp[1_N]$ is the opposite of the conditional pricing error vector $e(\cdot;\theta_*)$ from Eq. (A13), we get Eq. (2.16).

iii) Proof of Equation (2.17)
Let us focus on the quantity $\|B\theta - 1\|^2_{L^2_\Omega(\mathcal{X})}$ for any $\theta \in \Theta$. It is a quadratic function in θ, minimized at θ_* where it assumes value δ^2. Let us equate the expression that describes this property to Eq. (A17):

$$\delta^2 + \|B\left(\theta_* - \theta\right)\|^2_{L^2_\Omega(\mathcal{X})} = \|\mathcal{P}_A\left[B\theta - 1_N\right]\|^2_{L^2_\Omega(\mathcal{X})} + \|\mathcal{P}_A^\perp\left[B\theta - 1_N\right]\|^2_{L^2_\Omega(\mathcal{X})},$$

for any $\theta \in \Theta$. Evaluating this equation at $\theta = \theta_Z$ and using Eq. (A10) we get

$$\delta^2 = d_Z^2 + \|\mathcal{P}_A^\perp\left[B\theta_Z - 1_N\right]\|^2_{L^2_\Omega(\mathcal{X})} - \|B\left(\theta_* - \theta_Z\right)\|^2_{L^2_\Omega(\mathcal{X})}. \tag{A24}$$

Considering that $B\left(\theta_* - \theta_Z\right) = e(\cdot;\theta_*) - e(\cdot;\theta_Z)$ and that the sign of the argument of a $L^2_\Omega(\mathcal{X})$-norm does not matter, we get Eq. (2.17).

iv) Proof of the first equivalence in Equations (2.18)
We have $\delta = d_Z$ if, and only if, the two norms in the r.h.s. of Eq. (A23) are null. This condition realizes when the vectors of which the norms are taken are null, that is

$$\begin{cases} \mathcal{P}_A^{\perp} \circ \mathcal{P}_B^{\perp}[1_N] = 0_N, \\ \\ \mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_B^{\perp}[1_N] = 0_N. \end{cases} \tag{A25}$$

From Eq. (A13), the first condition corresponds to

$$\mathcal{P}_A^{\perp}[e\,(\cdot;\theta_*)] = 0_N. \tag{A26}$$

Moreover, since we have $\left\langle B, \mathcal{P}_B^{\perp}[1_N]\right\rangle_{L^2_\Omega(\mathcal{X})} = 0$ and $\mathcal{P}_A^{\perp}$ is a projection operator, we can write

$$\begin{aligned} \mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_B^{\perp}[1_N] &= \mathcal{P}_A[B] \left\langle \mathcal{P}_A[B], \mathcal{P}_A[B]\right\rangle^{-1}_{L^2_\Omega(\mathcal{X})} \left\langle \mathcal{P}_A[B], \mathcal{P}_B^{\perp}[1_N]\right\rangle_{L^2_\Omega(\mathcal{X})} \\ &= -\mathcal{P}_A[B] \left\langle \mathcal{P}_A[B], \mathcal{P}_A[B]\right\rangle^{-1}_{L^2_\Omega(\mathcal{X})} \left\langle \mathcal{P}_A^{\perp}[B], \mathcal{P}_B^{\perp}[1_N]\right\rangle_{L^2_\Omega(\mathcal{X})} \\ &= -\mathcal{P}_A[B] \left\langle \mathcal{P}_A[B], \mathcal{P}_A[B]\right\rangle^{-1}_{L^2_\Omega(\mathcal{X})} \left\langle B, \mathcal{P}_A^{\perp} \circ \mathcal{P}_B^{\perp}[1_N]\right\rangle_{L^2_\Omega(\mathcal{X})}. \end{aligned}$$

Therefore, the first equation in System (A25) implies the second one (but not necessarily the other way around). Hence, the System (A25) is equivalent to Eq. (A26).

v) Proof of the second equivalence in Equations (2.18)
From Prop. 1, the condition $\mathcal{P}_A^{\perp}[e(\cdot;\theta_Z)] = 0_N$ implies Eq. (A26). Let us now show that the reverse implication holds. If Eq. (A26) holds, we have that

$$\begin{aligned} \langle B, \mathcal{P}_A \circ \mathcal{P}_B^{\perp}[1_N]\rangle_{L^2_\Omega(\mathcal{X})} &= \langle B, \mathcal{P}_B^{\perp}[1_N]\rangle_{L^2_\Omega(\mathcal{X})} - \langle B, \mathcal{P}_A^{\perp} \circ \mathcal{P}_B^{\perp}[1_N]\rangle_{L^2_\Omega(\mathcal{X})} \\ &= \langle B, \mathcal{P}_A^{\perp}[e\,(\cdot;\theta_*)]\rangle = 0_N. \end{aligned}$$

By using $\langle B, \mathcal{P}_A \circ \mathcal{P}_B^{\perp}[1_N]\rangle_{L^2_\Omega(\mathcal{X})} = 0_N$ and the fact that $\mathcal{P}_A$ is a projection operator, from Eq. (A11) we get

$$\begin{aligned} \theta_Z &= \langle B, \mathcal{P}_A[B]\rangle^{-1}_{L^2_\Omega(\mathcal{X})} \langle B, \mathcal{P}_A[1_N]\rangle_{L^2_\Omega(\mathcal{X})} \\ &= \langle B, \mathcal{P}_A[B]\rangle^{-1}_{L^2_\Omega(\mathcal{X})} \langle B, \mathcal{P}_A \circ \mathcal{P}_B[1_N]\rangle_{L^2_\Omega(\mathcal{X})} \\ &= \langle B, B\rangle^{-1}_{L^2_\Omega(\mathcal{X})} \langle B, 1_N\rangle_{L^2_\Omega(\mathcal{X})}. \end{aligned} \tag{A27}$$

Then, by using Eqs. (A9) and (A27) we get $\theta_Z = \theta_*$, and under Eq. (A26) we get $\mathcal{P}_A^{\perp}[e\,(\cdot;\theta_Z)] = 0_N$.

Appendix B. Notation

In this appendix we describe the notation used in App. C-E to derive the large sample properties of the estimators. We denote both the Euclidean norm of a vector and the Frobenius norm of a matrix by $\|\cdot\|$. We denote by $\varphi_{[i]}$ the i-th element of the generic vector φ. If matrix Φ admits an inverse, we denote the element on the i-th row and j-th column of this inverse matrix by $\Phi^{[i,j]}$. We define the $(N \times N)$-dimensional matrix function V and the $(N \times p)$-dimensional matrix function J as

$$J(X_t;\theta) := \mathrm{E}\left[\nabla_{\theta'} h(Y_{t+1}, R_{t+1};\theta)\right| X_t\right], \tag{B1}$$

and

$$V(X_t;\theta) := \mathrm{E}\left[h(Y_{t+1}, R_{t+1};\theta)h(Y_{t+1}, R_{t+1};\theta)'\right| X_t\right], \tag{B2}$$

for any $\theta \in \Theta$. We use the following notation for matrices J and V and vectors e and h valued at the unknown values θ_0 and $\theta_\star$ of the SDF parameter vector:

$$\begin{aligned}
&J_0(X_t) = J(X_t;\theta_0), \quad V_0(X_t) = V(X_t;\theta_0), \quad h_0(Y_t) = h(Y_t, R_t;\theta_0),\\
&J_\star(X_t) = J(X_t;\theta_\star), \quad V_\star(X_t) = V(X_t;\theta_\star), \quad h_\star(Y_t) = h(Y_t, R_t;\theta_\star), \quad e_\star(X_t) = e(X_t;\theta_\star).
\end{aligned}$$

We define the vector $W_t := [Y_t\ X_t]'$, which collects priced factors and conditioning variables and takes value in set $\mathcal{Y} \times \mathcal{X} \subset \mathbb{R}^{d_Y} \times \mathbb{R}^L$, and we denote its realization by $w_t := [y_{t+1}'\ x_t']'$. We define the scalar function h_s as

$$h_s(Y_t) := \sup_{\theta\in\Theta} \|h(Y_t, R_t;\theta)\|,$$

and we consider the kernel estimator $\hat{f}_X$ of the stationary pdf f_X of process $\{X_t\}$ defined as

$$\hat{f}_X(x) := \frac{1}{Tb_T^L} \sum_{j=1}^{T} K\left(\frac{x - X_j}{b_T}\right). \tag{B3}$$

We define a scaled version of the conditional second moment matrix of the assets' gross returns given X_t:

$$H(X_t) := \Omega(X_t)^{-1} f_X(X_t)^2.$$

We also consider the $(N \times N)$-dimensional matrix functions $\hat{H}$ and $\tilde{H}_T$ as

$$\hat{H}(X_t) := \hat{\Omega}_T(X_t)^{-1}\hat{f}_X(X_t)^2, \quad \tilde{H}_T(X_t) := \mathrm{E}\left[\hat{\Omega}_T(X_t)^{-1}\hat{f}_X(X_t)\right| X_t\Big]\ \mathrm{E}\left[\hat{f}_X(X_t)\right| X_t\Big].$$

Finally, we define the p-dimensional stochastic vectors u_t and v_t, the $(p \times p)$-dimensional matrices Q_0, S_0, Q_* and S_* in this way:

$$
\begin{aligned}
&U_t := \mathbf{I}(X_t) J_\star(X_t)' \Omega(X_t) e_\star(X_t), \\
&V_{t+1} := \mathbf{I}(X_t) J_\star(X_t)' \Omega(X_t) \left(h_\star(Y_{t+1}) - \mathrm{E}\left[h_\star(Y_{t+1}) | X_t\right]\right) \\
&-\mathbf{I}(X_t) J_\star(X_t)' \Omega(X_t) \left(R_{t+1} R_{t+1}' - \mathrm{E}\left[R_{t+1} R_{t+1}' | X_t\right]\right) \Omega(X_t) e_\star(X_t) \\
&\qquad +\mathbf{I}(X_t) \left(\nabla_\theta m(Y_{t+1}; \theta_\star) - \mathrm{E}\left[\nabla_\theta m(Y_{t+1}; \theta_\star) | \, X_t\right]\right) R_{t+1}' \Omega(X_t) e_\star(X_t), \\
&Q_0 := \mathrm{E}\left[\mathbf{I}(X_t) J_0(X_t)' \Omega(X_t) J_0(X_t)\right], \quad S_0 := \mathrm{E}\left[\mathbf{I}(X_t) J_0(X_t)' \Omega(X_t) V_0(X_t) \Omega(X_t) J_0(X_t)\right] \\
&Q_\star := \mathrm{E}\left[\mathbf{I}(X_t) J_\star(X_t)' \Omega(X_t) J_\star(X_t)\right] + \mathrm{E}\left[\mathbf{I}(X_t) \nabla_{\theta\theta'} m(Y_{t+1}; \theta_\star) R_{t+1}' \Omega(X_t) e_\star(X_t)\right] \\
&S_\star := \sum_{l=-\infty}^{\infty} \mathrm{Cov}\left[u_t, u_{t-l}\right] + \mathrm{E}\left[\left(\eta_{t+1} - \mathrm{E}\left[v_{t+1} | \, X_t\right]\right)^2\right],
\end{aligned}
$$

for the random variables ε_t and η_t defined in Prop. 4.

Appendix C. Regularity assumptions

In this appendix we list the regularity assumptions used in Apps. D and E to derive the large sample properties of the estimators.

ASSUMPTION 1: *The stochastic process for the variable W_t defined in App. B is strictly stationary and strong mixing with mixing coefficients $\alpha(h) = O(\rho^h)$ for $h \in \mathbb{N}$ and $\rho \in (0,1)$.*

ASSUMPTION 2: *The pdf $f_X(\cdot)$ of process $\{X_t\}$ and the matrix function $V(\cdot;\theta)$ defined in Eq.* (B2)*, for any $\theta \in \Theta$, are of differentiability class $C^1(\mathcal{X})$.*

ASSUMPTION 3: *The compact set $\mathcal{X}_\star$ considered in the indicator function in Eq.* (2.21) *is contained in the interior of set $\mathcal{X}$ and it is such that $\inf_{x \in \mathcal{X}_\star} f_X(x) > 0$.*

ASSUMPTION 4: *The quantity $\mathrm{E}\left[\|h(Y_{t+1}, R_{t+1}; \theta)\|^n | X_t\right]$ is bounded on $\mathcal{X}$, uniformly in $\theta \in \Theta$, for a positive constant n.*

ASSUMPTION 5: *The bandwidth $b_T = o(1)$ is such that*

$$
\sqrt{T} b_T^3 = o(1), \qquad \frac{\log(T)}{T b_T^L} = o(1).
$$

ASSUMPTION 6: *The kernel function K is finite and such that ...*

Ass. 1-6 are standard in nonparametric analysis and yield the uniform convergence of kernel estimators over the set $\mathcal{X}_\star$. By adopting Ass. 3 we avoid the boundary problems that are typical in nonparametric kernel estimation. The compactness of set $\mathcal{X}_\star$ is useful to handle with the remainder terms in the asymptotic expansions of the kernel estimators of conditional expectations. Ass. 5 allows to simplify the expression of the asymptotic distribution of the estimators under model misspecification.

ASSUMPTION 7: *The parameter values θ_0 and $\theta_\star$ belong to the interior of set $\Theta \subset \mathbb{R}^p$.*

ASSUMPTION 8: *The vector function $h(Y_t, R_t; \theta)$ is of differentiability class $C^1(\Theta)$ and such that*

$$\mathrm{E}\left[\sup_{\theta \in \Theta} \|h(Y_t, R_t; \theta)\|^2\right] < \infty.$$

ASSUMPTION 9: *The largest eigenvalue of matrix $\Omega(x)$ is bounded from above, and the smallest eigenvalue is bounded from below away from zero, uniformly in $x \in \mathcal{X}$.*

Ass. 9 implies that matrix $\Omega(x)$ is positive definite for any $x \in \mathcal{X}$, and $\|\Omega(x)\|$ is bounded on $\mathcal{X}$. Moreover, Ass. 3 and 9 imply that

$$\sup_{x \in \mathcal{X}_\star} \|H(x)^{-1}\| < \infty. \tag{C1}$$

ASSUMPTION 10: *Consider positive integers i, j, k such that $1 \leqslant i \leqslant N$ and $1 \leqslant j,\ k \leqslant p$. There exists an open ball $\mathcal{N}_0$ around θ_0 such that $\theta \mapsto h(Y_t, R_t; \theta)$ is twice continuously differentiable and such that*

$$\sup_{\theta \in \mathcal{N}_0} \left|\nabla_{\theta_{[j]}} h_{[i]}(Y_t; \theta)\right| \leqslant l_1(Y_t), \qquad \textit{and} \qquad \sup_{\theta \in \mathcal{N}_0} \left|\nabla_{\theta_{[j]}\theta_{[k]}} h_{[i]}(Y_t; \theta)\right| \leqslant l_2(Y_t),$$

$\mathbb{P}$-a.s., for some real-valued functions l_1 and l_2 of vector Y_t such that $\mathrm{E}\left[|l_1(Y_t)|^\eta\right] < \infty$, for $\eta \geqslant 6$, and $\mathrm{E}\left[|l_2(Y_t)|^2\right] < \infty$.

Ass. 10 is used to expand function $h(y, R; \theta)$ in a second order Taylor series w.r.t. parameter θ at around $\theta = \theta_0$, for any $R \in \mathbb{R}^N$.

ASSUMPTION 11: *For any pair of SDF families $\mathcal{F}$ and $\tilde{\mathcal{F}}$, the processes $\{\phi_t\}$ and $\{\tilde{\phi}_t\}$ defined as in Prop. 4 are such that*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{bmatrix} \phi_t \\ \tilde{\phi}_t \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, S),$$

as $T \to \infty$, for the asymptotic variance

$$S := \sum_{l=-\infty}^{\infty} \mathrm{Cov}\left[\begin{bmatrix} \phi_t \\ \tilde{\phi}_t \end{bmatrix}, \begin{bmatrix} \phi_{t-l} \\ \tilde{\phi}_{t-l} \end{bmatrix}\right].$$

As $T \to \infty$, the variables ε_t and η_t defined in Prop. 4 are such that

We use Ass. 11 to derive the asymptotic normality of the sample conditional HJ-distance for a misspecified model.

ASSUMPTION 12: *As $T \to \infty$, the joint bivariate process for the variables U_t and V_{t+1} defined* *in App. B is such that*

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}\begin{bmatrix}U_t\\ V_{t+1}\end{bmatrix}\xrightarrow{\mathcal{D}}\mathcal{N}\left(0_2,\begin{bmatrix}S_u & 0\\ 0 & \mathrm{E}\left[V_{t+1}U_t'\right]\end{bmatrix}\right),$$

for the (2×2)*-dimensional matrix*

$$S_u:=\sum_{j=-\infty}^{\infty}\mathrm{E}\left[v_{t+1}u_t'\right].$$

Appendix D. Large sample properties of the estimators under correct model specification

In this appendix we derive the large sample properties of the estimator $\hat{\theta}_T$ of the SDF parameter vector and the estimator $\hat{\delta}_T$ of the conditional HJ-distance for a correctly specified SDF family.

Appendix 1. Large sample properties of estimator $\hat{\theta}_T$

The asymptotic distribution of estimator $\hat{\theta}_T$ is given in the next lemma.

LEMMA 1: *Under correct model specification, the estimator $\hat{\theta}_T$ is consistent and asymptotically normal with $\sqrt{T}$-rate of convergence. In particular, as $T\to\infty$,*

$$\sqrt{T}\left(\hat{\theta}_T-\theta_0\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0_p,Q_0^{-1}S_0Q_0^{-1}\right).$$

Proof. See App. F of the supplementary material. □

Appendix 2. Useful results on kernel regression estimators

In this section we report some results on kernel regression estimators that are useful to prove the asymptotic normality of the squared sample conditional HJ-distance. Under Ass. 5-8, if the constant n is such that as $T\to\infty$

$$\log(T)/\left(T^{1-2/n}b_T^L\right)\to 0,$$

then from results similar to Lemmas C1, C3.1 and C3.2 in Tripathi and Kitamura (2003) we have that

$$\sup_{x\in\mathcal{X}_\star}\left\|\frac{1}{Tb_T^L}\sum_{j=1}^{T-1}K\left(\frac{x-X_j}{b_T}\right)h_0(Y_{j+1})\right\| = O_p\left(\sqrt{\frac{\log(T)}{Tb_T^L}}\right), \tag{D1}$$

$$\sup_{x\in\mathcal{X}_\star}\left\|\hat{H}(x)^{-1}-H(x)^{-1}\right\| = O_p\left(\sqrt{\frac{\log(T)}{Tb_T^L}}+b_T^2\right)+o_p\left(T^{-1/2+1/n+1/\eta}\right), \tag{D2}$$

$$\sup_{x\in\mathcal{X}_\star}\left\|\hat{H}(x)^{-1}-\tilde{H}_T(x)^{-1}\right\| = O_p\left(\sqrt{\frac{\log(T)}{Tb_T^L}}\right)+o_p\left(T^{-1/2+1/n+1/\eta}\right). \tag{D3}$$

Appendix 3. Proof of Proposition 3

In this section we use the results of Secs. D.1 and D.2 to prove the asymptotic normality of the squared sample conditional HJ-distance stated in Prop. 3. Given the form of the Nadaraya-Watson estimator in Eq. (2.19), we can write the criterion $\mathcal{Q}_T$ in Eq. (2.21) as a sum of quadratic forms of the conditional moment vector h:

$$\mathcal{Q}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{T-1}\sum_{j=1}^{T-1}\mathbf{I}(X_t)w(X_t,X_i)w(X_t,X_j)q(Y_{i+1},X_t,Y_{j+1};\theta),$$

where the scalar statistic q is defined as

$$q(Y_{i+1},X_t,Y_{j+1};\theta) := h(Y_{i+1},R_{i+1};\theta)'\hat{\Omega}_T(X_t)h(Y_{j+1},R_{j+1};\theta),$$

for any $i,j=1,\dots,T-1$ and any and $t=1,\dots,T$. Let us now decompose the criterion $\mathcal{Q}_T$ in a similar way as in Tripathi and Kitamura (2003):

$$\mathcal{Q}_T(\theta) = \sum_{i=1}^{5}\mathcal{Q}_{i,T}(\theta), \tag{D4}$$

where

$$\begin{aligned}
\mathcal{Q}_{1,T}(\theta) &:= \frac{1}{T}\sum_{t=1}^{T-1}\mathbf{I}(X_t)w(X_t,X_t)^2 q(Y_{t+1},X_t,Y_{t+1};\theta),\\
\mathcal{Q}_{2,T}(\theta) &:= \frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{i=1\\ i\neq t}}^{T-1}\mathbf{I}(X_t)w(X_t,X_i)^2 q(Y_{i+1},X_t,Y_{i+1};\theta),\\
\mathcal{Q}_{3,T}(\theta) &:= \frac{1}{T}\sum_{t=1}^{T-1}\sum_{\substack{j=1\\ j\neq t}}^{T-1}\mathbf{I}(X_t)w(X_t,X_t)w(X_t,X_j)q(Y_{t+1},X_t,Y_{j+1};\theta), \quad \mathcal{Q}_{4,T}(\theta)=\mathcal{Q}_{3,T}(\theta),\\
\mathcal{Q}_{5,T}(\theta) &:= \frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{i=1\\ i\neq t}}^{T-1}\sum_{\substack{j=1\\ j\neq i\\ j\neq t}}^{T-1}\mathbf{I}(X_t)w(X_t,X_i)w(X_t,X_j)q(Y_{i+1},X_t,Y_{j+1};\theta),
\end{aligned}$$

for any $\theta \in \Theta$. We first consider the asymptotic behavior of the different terms $\mathcal{Q}_{1,T}(\theta),\ldots,\mathcal{Q}_{5,T}(\theta)$ taken singularly, and in doing it we make use of the equation

$$w(X_t,X_i)=\frac{1}{Tb_T^L\hat{f}_X(X_t)}K\left(\frac{X_t-X_i}{b_T}\right), \tag{D5}$$

that is obtained from Eq. (B3). We then derive the asymptotic distribution of the criterion $\mathcal{Q}_T(\theta)$.

i) Stochastic boundedness of the terms $\mathcal{Q}_{i,T}$ for $i=1,\ldots,4$

Let us consider the first term in the sum of Eq. (D4). From Eq. (D5) and the triangular inequality we get

$$\begin{aligned}
\mathcal{Q}_{1,T}(\theta) &= \frac{1}{T}\sum_{t=1}^{T-1}\mathbf{I}(X_t)\frac{K(0)^2}{T^2b_T^{2L}\hat{f}_X(X_t)^2}h(Y_{t+1},R_{t+1};\theta)'\hat{\Omega}_T(X_t)h(Y_{t+1},R_{t+1};\theta)\\
&= \frac{K(0)^2}{T^2b_T^{2L}}\frac{1}{T}\sum_{t=1}^{T-1}\mathbf{I}(X_t)h(Y_{t+1},R_{t+1};\theta)'\hat{H}(X_t)^{-1}h(Y_{t+1},R_{t+1};\theta)\\
&\leqslant \frac{K(0)^2}{T^2b_T^{2L}}\sup_{1\leqslant t\leqslant T}\|\hat{H}(X_t)^{-1}\|\frac{1}{T}\sum_{t=1}^{T-1}h_s(Y_{t+1})^2,\\
&\leqslant \frac{K(0)^2}{T^2b_T^{2L}}\sup_{1\leqslant t\leqslant T}\left(\|\hat{H}(X_t)^{-1}-H(X_t)^{-1}\|+\|H(X_t)^{-1}\|\right)\frac{1}{T}\sum_{t=1}^{T-1}h_s(Y_{t+1})^2,
\end{aligned}$$

uniformly in $\theta\in\Theta$. Therefore, from Ineq. (C1), Eq. (D2) and Ass. 8, and since function K is finite from Ass. 6, we get

$$\mathcal{Q}_{1,T}(\theta)=O_p\left(\frac{1}{T^2b_T^{2L}}\right) \tag{D6}$$

uniformly in $\theta \in \Theta$. By the property of the trace operator, the second term in the sum of Eq. (D4) evaluated at $\theta = \hat{\theta}_T$ is

$$\mathcal{Q}_{2,T}(\hat{\theta}_T) = a_T. \tag{D7}$$

The third term in the sum of Eq. (D4) evaluated at $\hat{\theta}_T$ is bounded as described in the next Lemma.

LEMMA 2: *Under Assumptions ... we have*

$$\mathcal{Q}_{3,T}(\hat{\theta}_T) = O_p\left(\frac{1}{T^2 b_T^{3L/2}}\right). \tag{D8}$$

Proof. See App. G of the supplementary material. □

The fourth and the third terms in the sum of Eq. (D4) coincide. Since $\hat{\delta}_T^2 = \mathcal{Q}_T(\hat{\theta}_T)$, from Eq. (D4)-(D8) and Ass. 5 we get

$$T b_T^{L/2}\left(\hat{\delta}_T^2 - a_T\right) = T b_T^{L/2} \mathcal{Q}_{5,T}(\hat{\theta}_T) + o_p(1). \tag{D9}$$

ii) Asymptotic expansion of term $\mathcal{Q}_{5,T}(\theta)$

Let us consider the fifth term in the sum of Eq. (D4). From Eq. (D5) and the definition of matrix $\hat{H}(X_t)$, we get

$$\begin{aligned}\mathcal{Q}_{5,T}(\theta) &= \frac{1}{T^3 b_T^{2L}} \sum_{t=1}^{T} \sum_{\substack{i=1 \\ i\neq t}}^{T-1} \sum_{\substack{j=1 \\ j\neq i \\ j\neq t}}^{T-1} \mathbf{I}(x_t) K\left(\frac{X_t - X_i}{b_T}\right) h(Y_{i+1}, R_{i+1};\theta)' \\ &\quad \cdot \hat{H}(X_t)^{-1} h(Y_{j+1}, R_{j+1};\theta) K\left(\frac{X_t - X_j}{b_T}\right),\end{aligned}$$

for any $\theta \in \Theta$. By adding and subtracting the matrix $\tilde{H}_T(X_t)^{-1}$ to the matrix $\hat{H}(X_t)^{-1}$ within the quadratic forms, we can decompose $\mathcal{Q}_{5,T}(\theta)$ as the sum of

$$\begin{aligned}\tilde{\mathcal{Q}}_{5,T}(\theta) &:= \frac{1}{T^3 b_T^{2L}} \sum_{t=1}^{T} \sum_{\substack{i=1 \\ i\neq t}}^{T-1} \sum_{\substack{j=1 \\ j\neq i \\ j\neq t}}^{T-1} \mathbf{I}(X_t) K\left(\frac{X_t - X_i}{b_T}\right) h(Y_{i+1}, R_{i+1};\theta)' \\ &\quad \cdot \tilde{H}_T(X_t)^{-1} h(Y_{j+1}, R_{j+1};\theta) K\left(\frac{X_t - X_j}{b_T}\right)\end{aligned}$$

and

$$\mathcal{Q}_{5,T}(\theta) - \tilde{\mathcal{Q}}_{5,T}(\theta) := \frac{1}{T^3 b_T^{2L}} \sum_{t=1}^{T} \sum_{\substack{i=1 \\ i \neq t}}^{T-1} \sum_{\substack{j=1 \\ j \neq i \\ j \neq t}}^{T-1} \mathbf{I}(X_t) K\left(\frac{X_t - X_i}{b_T}\right) h(Y_{i+1}, R_{i+1}; \theta)' \cdot \left(\hat{H}(X_t)^{-1} - \tilde{H}_T(X_t)^{-1}\right) h(Y_{j+1}, R_{j+1}; \theta) K\left(\frac{X_t - X_j}{b_T}\right),$$

for any $\theta \in \Theta$. The absolute value of the last term is such that

$$\begin{aligned}
|\mathcal{Q}_{5,T}(\theta) - \tilde{\mathcal{Q}}_{5,T}(\theta)| &= \frac{1}{T^3 b_T^{2L}} \left| \mathrm{Tr}\left[\sum_{t=1}^{T} \mathbf{I}(X_t) \left(\hat{H}(X_t)^{-1} - \tilde{H}_T(X_t)^{-1}\right) \right.\right. \\
&\quad \left.\left. \cdot \sum_{\substack{i=1 \\ i \neq t}}^{T-1} K\left(\frac{X_t - X_i}{b_T}\right) h(Y_{i+1}, R_{i+1}; \theta) \sum_{\substack{j=1 \\ j \neq i \\ j \neq t}}^{T-1} K\left(\frac{X_t - X_j}{b_T}\right) h(Y_{j+1}, R_{j+1}; \theta)' \right] \right| \\
&\leqslant \sup_{x \in \mathcal{X}_\star} \|\hat{H}(x)^{-1} - \tilde{H}_T(x)^{-1}\| \sup_{x \in \mathcal{X}_\star} \sup_{t \in [1,T]} \left\| \frac{1}{T b_T^L} \sum_{\substack{i=1 \\ i \neq t}}^{T-1} K\left(\frac{x - X_i}{b_T}\right) h(Y_{i+1}, R_{i+1}; \theta) \right\| \\
&\quad \cdot \sup_{x \in \mathcal{X}_\star} \sup_{\substack{t=1,\dots,T \\ i=1,\dots,T-1 \\ i \neq t}} \left\| \frac{1}{T b_T^L} \sum_{\substack{j=1 \\ j \neq i \\ j \neq t}}^{T-1} K\left(\frac{x - X_j}{b_T}\right) h(Y_{j+1}, R_{j+1}; \theta) \right\|.
\end{aligned}$$

for any $\theta \in \Theta$. From Eqs. (D1) and (D3) we get

$$|\mathcal{Q}_{5,T}(\theta) - \tilde{\mathcal{Q}}_{5,T}(\theta)| = \left(O_p\left(\sqrt{\frac{\log(T)}{T b_T^L}} \right) + o_p\left(T^{-1/2+1/n+1/\eta}\right) \right) O_p\left(\frac{\log(T)}{T b_T^L}\right)$$

uniformly in $\theta \in \Theta$. Thus, from Ass. 5 we get

$$T b_T^{L/2} \mathcal{Q}_{5,T}(\hat{\theta}_T) = T b_T^{L/2} \tilde{\mathcal{Q}}_{5,T}(\hat{\theta}_T) + o_p(1). \tag{D10}$$

We control $\tilde{\mathcal{Q}}_{5,T}(\hat{\theta}_T)$ using Ass. 10. This assumption allows a second-order Taylor expansion of $h(Y_{i+1}, R_{i+1}; \theta)$ at around $\theta = \theta_0$, which valued for $\theta = \hat{\theta}_T$ is

$$h(Y_{i+1}, R_{i+1}; \hat{\theta}_T) = h_0(Y_{i+1}) + \nabla_{\theta'} h(Y_{i+1}, R_{i+1}; \theta_0) \left(\hat{\theta}_T - \theta_0\right) + \mathrm{Rem}\left[Y_{i+1}; \hat{\theta}_T - \theta_0\right], \tag{D11}$$

for a remainder term $\mathrm{Rem}\,[y;\theta]$ that is not greater than $l_2(y)\|\theta\|^2$ for any $y \in \mathcal{Y}$ and function l_2 introduced in Ass. 10.

LEMMA 3: *Under Ass. …we have*

$$\tilde{\mathcal{Q}}_{5,T}(\hat{\theta}_T) = \tilde{\mathcal{Q}}_{5,T}(\theta_0) + O_p(\ldots).$$

Proof. See App. H. □

From Eqs. (D9) and (D10), Lemma 3 and Ass. 5 we get

$$Tb_T^{L/2}\left(\hat{\delta}_T^2 - a_T\right) = Tb_T^{L/2}\tilde{\mathcal{Q}}_{5,T}(\theta_0) + o_p(1). \tag{D12}$$

iii) Asymptotic normality of the squared sample conditional HJ-distance

We now show the asymptotic normality of term $Tb_T^{L/2}\tilde{\mathcal{Q}}_{5,T}(\theta_0)$ in the r.h.s. of Eq. (D12). Then, the asymptotic normality of $Tb_T^{L/2}\left(\hat{\delta}_T^2 - a_T\right)$ follows from the Slutsky's theorem, since these two terms are asymptotically equivalent as pointed out in Eq. (D12). Let us consider the $(N \times N)$-dimensional matrix functions A_T and $\hat{A}_T$ defined as

$$A_T(X_i, X_j) := \int_{\mathcal{X}} \mathbf{I}(x) K\left(\frac{x - X_i}{b_T}\right) \tilde{H}_T(x)^{-1} K\left(\frac{x - X_j}{b_T}\right) f_X(x)dx,$$
$$\hat{A}_T(X_i, X_j) := \frac{1}{T}\sum_{\substack{t=1\\t\neq i\\t\neq j}}^{T} \mathbf{I}(X_t) K\left(\frac{X_t - X_i}{b_T}\right) \tilde{H}_T(X_t)^{-1} K\left(\frac{X_t - X_j}{b_T}\right),$$

for any $i = 1, \ldots, T-1$ and any $j = 1, \ldots, i-1$. The term $\tilde{\mathcal{Q}}_{5,T}(\theta_0)$ can be written as

$$\tilde{\mathcal{Q}}_{5,T}(\theta_0) = \frac{1}{T^2 b_T^{2L}} \sum_{i=1}^{T-1}\sum_{\substack{j=1\\j\neq i}}^{T-1} h_0(Y_{i+1})' \hat{A}_T(X_i, X_j) h_0(Y_{j+1}). \tag{D13}$$

From Ass. 8 the quadratic forms in the r.h.s. of the last equation are L_1-bounded. Then, from the weak law of large numbers we have that $\hat{A}_T(x_i, x_j) - A_T(x_i, x_j) = o_p(1)$, for any $x_i, x_j \in \mathcal{X}$. Moreover matrix A_T is such that $A_T(x, \tilde{x}) = A_T(x, \tilde{x})' = A_T(\tilde{x}, x)$, for any $x, \tilde{x} \in \mathcal{X}$. Therefore, we can write the sum in Eq. (D13) as

$$Tb_T^{L/2}\tilde{\mathcal{Q}}_{5,T}(\theta_0) = \frac{1}{T}\sum_{i=1}^{T-1}\sum_{j=1}^{i-1} g_T(W_i, W_j) + o_p(1), \tag{D14}$$

where the scalar function g_T is defined as

$$g_T(W_i, W_j) := \frac{2}{b_T^{3L/2}} h_0(Y_{i+1})' A_T(X_i, X_j) h_0(Y_{j+1}),$$

for any $i = 1, \ldots, T-1$ and any $j = 1, \ldots, i-1$, and vector W_t defined in App. B. We derive the asymptotic normality of $\frac{1}{T}\sum_{i=1}^{T-1}\sum_{j=1}^{i-1} g_T(W_i, W_j)$ by showing that the regularity conditions for function g_T required by Lemma A.3 in Su and White (2013) (see also Yoshihara (1976, 1989)) hold:

(Symmetry) From the symmetry properties of matrix A_T we have

$$g_T(W_i, W_j) = g_T(W_j, W_i). \tag{D15}$$

(Null expectation) From the law of iterated expectations and since θ_0 satisfies Eq. (1.2) we have

$$\begin{aligned} \mathrm{E}\left[g_T(W_i, w)\right] &= \frac{2}{b_T^{3L/2}} \mathrm{E}\left[h_0(Y_{i+1})' A_T(X_i, x) h_0(y)\right] \\ &= \frac{2}{b_T^{3L/2}} \mathrm{E}\left[\mathrm{E}\left[h_0(Y_{i+1})|X_i\right]' A_T(X_i, x) h_0(y)\right] = 0, \end{aligned} \tag{D16}$$

for any $w = [y' \; x']'$.

(Asymptotic covariance) Let $\bar{W}_1$ be an independent copy of W_1. We have

$$\mathrm{E}\left[g_T(W_1, \bar{W}_1)^2\right] = 2\sigma_0^2 + o(1), \tag{D17}$$

for the asymptotic variance σ_0^2 of statistic $\hat{\delta}_T^2$ in Prop. 3. Indeed we have

$$\begin{aligned} \mathrm{E}\left[g_T(W_1, \bar{W}_1)^2\right] &= \frac{4}{b_T^{3L}} \mathrm{E}\left[\left(h_0(Y_2)' A_T(X_1, \bar{X}_1) h_0(\bar{Y}_2)\right)^2\right] \\ &= \frac{4}{b_T^{3L}} \mathrm{E}\left[h_0(Y_2)' A_T(X_1, \bar{X}_1) h_0(\bar{Y}_2) h_0(\bar{Y}_2)' A_T(X_1, \bar{X}_1) h_0(Y_2)\right], \end{aligned}$$

where we use again the symmetry property of matrix A_T w.r.t. its two arguments and transposition. From the property of invariance under cyclical permutations of the trace operator, we have

$$\mathrm{E}\left[g_T(W_1, \bar{W}_1)^2\right] = \frac{4}{b_T^{3L}} \mathrm{E}\Big[\mathrm{Tr}\big[h_0(Y_2) h_0(Y_2)' A_T(X_1, \bar{X}_1) h_0(\bar{Y}_2) h_0(\bar{Y}_2)' A_T(X_1, \bar{X}_1)\big]\Big].$$

The trace and expectation operators commute. From this property, the law of iterated expectations and the independence between W_1 and $\bar{W}_1$, we get

$$\begin{aligned} \mathrm{E}\left[g_T(W_1, \bar{W}_1)^2\right] &= \frac{4}{b_T^{3L}} \mathrm{Tr}\Big[\mathrm{E}\big[\mathrm{E}\left[h_0(Y_2) h_0(Y_2)' | X_1\right] A_T(X_1, \bar{X}_1) \\ &\qquad \cdot \mathrm{E}\left[h_0(\bar{Y}_2) h_0(\bar{Y}_2)' | \bar{X}_1\right] A_T(X_1, \bar{X}_1)\big]\Big]. \end{aligned}$$

Using the definition of matrix $V_0(x)$ in Prop. 3, the definition of matrix A_T and the independence between X_1 and $\bar{X}_1$ we get

$$
\begin{aligned}
&\mathrm{E}\left[g_T(W_1,\bar{W}_1)^2\right] = \frac{4}{b_T^{3L}}\mathrm{Tr}\left[\mathrm{E}\left[V_0(X_1)A_T(X_1,\bar{X}_1)V_0(\bar{X}_1)A_T(X_1,\bar{X}_1)\right]\right] \\
&= \frac{4}{b_T^{3L}}\int_{\mathcal{X}}\int_{\mathcal{X}}\mathbf{I}(x)\mathbf{I}(\tilde{x})\mathrm{Tr}\Bigg[\mathrm{E}\left[V_0(X_1)K\left(\frac{x-X_1}{b_T}\right)K\left(\frac{\tilde{x}-X_1}{b_T}\right)\right] \\
&\cdot\tilde{H}_T(x)^{-1}\mathrm{E}\left[V_0(\bar{X}_1)K\left(\frac{x-\bar{X}_1}{b_T}\right)K\left(\frac{\tilde{x}-\bar{X}_1}{b_T}\right)\right]\tilde{H}_T(\tilde{x})^{-1}\Bigg]f_X(x)f_X(\tilde{x})dxd\tilde{x}.
\end{aligned}
\tag{D18}
$$

Let us write the expectations to whom the trace operator is applied in the r.h.s. of the last expression in integral form:

$$
\begin{aligned}
&\mathrm{E}\left[V_0(X_1)K\left(\frac{x-X_1}{b_T}\right)K\left(\frac{\tilde{x}-X_1}{b_T}\right)\right] \\
&\quad= \int_{\mathcal{X}} V_0(x_1)K\left(\frac{x-x_1}{b_T}\right)K\left(\frac{\tilde{x}-x_1}{b_T}\right)f_X(x_1)dx_1,
\end{aligned}
$$

for any $x,\tilde{x}\in\mathcal{X}$. By changing the variable x_1 with the variable $u=(x-x_1)/b_T$, and using the sets $\mathcal{U}(b,x) := \{u\in\mathbb{R}^L : x-ub\in\mathcal{X}\}$, for any $b\geqslant 0$ and $x\in\mathcal{X}$, we get

$$
\begin{aligned}
&\mathrm{E}\left[V_0(X_1)K\left(\frac{x-X_1}{b_T}\right)K\left(\frac{\tilde{x}-X_1}{b_T}\right)\right] \\
&\quad= b_T^L\int_{\mathcal{U}(b_T,x)} V_0(x-b_Tu)K(u)K\left(u-\frac{x-\tilde{x}}{b_T}\right)f_X(x-b_Tu)du \\
&\quad= b_T^L V_0(x)f_X(x)\int_{\mathcal{U}(b_T,x)} K(u)K\left(u-\frac{x-\tilde{x}}{b_T}\right)du + o(b_T^L),
\end{aligned}
$$

for any $x,\tilde{x}\in\mathcal{X}$. Note that $\mathcal{U}(0,x)=\mathbb{R}^L$ and consider the first-order Taylor expansion of function $\int_{\mathcal{U}_T(b,x)} K(u)K(u-z)\,du$ w.r.t. b at around $b=0$:

$$
\int_{\mathcal{U}_T(b,x)} K(u)K(u-z)\,du = \mathcal{K}(z) + o(1),
\tag{D19}
$$

for the convolution $\mathcal{K}$ of the kernel with itself defined in Prop. 3. Using this expression in the last equation we get

$$
\frac{1}{b_T^L}\mathrm{E}\left[V_0(X_1)K\left(\frac{x-X_1}{b_T}\right)K\left(\frac{\tilde{x}-X_1}{b_T}\right)\right] = V_0(x)f_X(x)\mathcal{K}\left(\frac{x-\tilde{x}}{b_T}\right) + o(1),
$$

for any $x, \tilde{x} \in \mathcal{X}$. Moreover, since $\bar{X}_1$ is an independent copy of X_1, the same expression holds also with $\bar{X}_1$ in place of X_1. Plugging the two expressions into Eq. (D18), and considering that

$$\tilde{H}_T(x) = \Omega(x)^{-1} f_X(x)^2 + o(1),$$

for any $x \in \mathcal{X}$, we get

$$\begin{aligned}
&\mathrm{E}\left[g_T(W_1, \bar{W}_1)^2\right] \\
&= \frac{4}{b_T^L} \int_{\mathcal{X}} \int_{\mathcal{X}} \mathbf{I}(x)\mathbf{I}(\tilde{x}) \mathrm{Tr}\left[V_0(x)\Omega(x)V_0(\tilde{x})\Omega(\tilde{x})\right] \mathcal{K}\left(\frac{x-\tilde{x}}{b_T}\right)^2 d\tilde{x}dx + o(1).
\end{aligned}$$

By changing the variable $\tilde{x}$ with the variable $u = (x - \tilde{x})/b_T$, we get

$$\begin{aligned}
&\mathrm{E}\left[g_T(W_1, \bar{W}_1)^2\right] \\
&= 4 \int_{\mathcal{X}} \int_{\mathcal{U}(b_T,x)} \mathbf{I}(x)\mathbf{I}(x - b_T u) \mathrm{Tr}\left[V_0(x)\Omega(x)V_0(x - b_T u)\Omega(x - b_T u)\right] \\
&\quad \cdot \mathcal{K}(u)^2 \, du dx + o(1) \\
&= 4 \int_{\mathcal{X}} \mathbf{I}(x) \mathrm{Tr}\left[V_0(x)\Omega(x)V_0(x)\Omega(x)\right] \int_{\mathcal{U}(b_T,x)} \mathcal{K}(u)^2 \, du dx + o(1).
\end{aligned}$$

Considering Eq. (D19) for $z = 0$ and the definition of σ_0^2 in Prop. 3, Eq. (D17) follows.

(Null cross-terms) We have

$$\begin{aligned}
\mathrm{E}\left[g_T(W_{j+1}, w) g_T(W_1, \tilde{w})\right] &= \mathrm{E}\left[\mathrm{E}\left[g_T(W_{j+1}, w) \middle| X_{j+1}\right] g_T(W_1, \tilde{w})\right] \\
&= \frac{4}{b_T^{3L}} \mathrm{E}\left[\mathrm{E}\left[h_0(Y_{j+2})' \middle| X_{j+1}\right] A(X_{j+1}, x) h_0(y) g_T(W_1, \tilde{w})\right] = 0,
\end{aligned}$$

for any integer $j \leqslant T - 1$, and any $w, \tilde{w} \in \mathcal{Y} \times \mathcal{X}$.

The above results show that conditions $(i) - (iv)$ in Lemma A.3 of Su and White (2013) are satisfied. Condition (vi) of the same lemma is satisfied as well under Ass. 1. The next lemma shows that the remaining regularity conditions are also satisfied.

LEMMA 4: *There exist constants $\beta > 0$ and $\gamma < 1$ such that we have*

$$\max\left[\max_{i \in [1, T-1]} \|g_T(W_{i+1}, W_1)\|_{4+\beta}, \|g_T(W_1, \bar{W}_1)\|_{4+\beta}\right] = O(T^\gamma),$$

$$\|G_T(W_1, W_1)\|_{2+\beta/2} = o(T^{1/2})$$

and

$$\max\left[\max_{i \in [1, T-1]} \|G_T(W_{i+1}, W_1)\|_2, \|G_T(W_1, \bar{W}_1)\|_2\right] = o(1),$$

where function G_T is defined as

$$G_T(w,z) := \mathrm{E}\left[g_T(W_1,w)g_T(W_1,z)\right]$$

and $\|\cdot\|_p := (\mathrm{E}[|\cdot|^p])^{1/p}$ *is the standard L^p-norm, for any positive integer p.*

Proof. See App. I. □

From Lemma A.3 in Su and White (2013) (see also Yoshihara (1976, 1989)) the statistic

$$\frac{1}{T}\sum_{i=1}^{T-1}\sum_{\substack{j=1\\ j\neq i}}^{T-1} g_T(W_i,W_j) \tag{D20}$$

is asymptotically normal with null mean and asymptotic variance σ_0^2. Then, from Eq. (D14) we have

$$Tb_T^{L/2}\tilde{\mathcal{Q}}_{5,T}(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\sigma_0^2\right).$$

Appendix E. Large sample properties of the estimators under model misspecification

In this appendix we derive the large sample properties of the estimator $\hat{\theta}_T$ of the SDF parameter vector and the estimator $\hat{\delta}_T$ of the conditional HJ-distance for a misspecified SDF family.

Appendix 1. Asymptotic distribution of estimator $\hat{\theta}_T$

The next lemma shows that $\hat{\theta}_T$ is a consistent and asymptotically normal estimator of the pseudo-true parameter vector $\theta_\star$, which is the solution of the minimization problem in Eq. (3.2).

LEMMA 5: *Under regularity conditions, the estimator $\hat{\theta}_T$ converges in probability to $\theta_\star$ and it is asymptotically normal with $\sqrt{T}$-rate of convergence. In particular, as $T\to\infty$,*

$$\sqrt{T}\left(\hat{\theta}_T-\theta_\star\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0_p,\Sigma_\star\right).$$

where $\Sigma_\star := Q_\star^{-1}S_\star Q_\star^{-1}$.

Proof. See App. J. □

When the SDF family is correctly specified, we have $\theta_\star=\theta_0$ and $e(X_t;\theta_\star)=0$, and the asymptotic distribution in Lemma 5 reduces to that in Lemma 1. Lemma 5 is the counterpart of the results on the large sample distribution of the GMM estimator in misspecified models derived by Hall and Inoue (2003) (see also Anatolyev and Gospodinov (2011)), when the structural parameter is identified by a set of conditional moment restrictions.

Appendix 2. Proof of Proposition 4

From Eq. (2.21) we have

$$\hat{\delta}_T^2 = \frac{1}{T}\sum_{t=1}^{T} \mathbf{I}(X_t)\hat{e}_T(X_t;\hat{\theta}_T)'\hat{\Omega}_T(X_t)\hat{e}_T(X_t;\hat{\theta}_T). \tag{E1}$$

We have the following first-order Taylor expansion of $\hat{e}_T(X_t;\hat{\theta}_T)$ around $\hat{\theta}_T = \theta_\star$:

$$\begin{aligned}\hat{e}_T(X_t;\hat{\theta}_T) &= \hat{e}_T(X_t;\theta_\star) + \nabla_{\theta'}\hat{e}_T(X_t;\theta_\star)(\hat{\theta}_T - \theta_\star) + \mathrm{Rem}\left[X_t;\hat{\theta}_T - \theta_\star\right] \\ &= e_\star(X_t) + (\hat{e}_T(X_t;\theta_\star) - e_\star(X_t)) + J_\star(X_t)(\hat{\theta}_T - \theta_\star) + o_p(1),\end{aligned} \tag{E2}$$

because matrix $\nabla_{\theta'}\hat{e}_T(X_t;\theta_\star)$ converges in probability to matrix $J_\star(X_t)$. We can write

$$\begin{aligned}\hat{\Omega}_T(X_t) &= \left(\Omega(X_t)^{-1} + \left(\hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1}\right)\right)^{-1} \\ &= \left(I_N + \left(\hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1}\right)\Omega(X_t)\right)^{-1}\Omega(X_t).\end{aligned}$$

Each of the eigenvalues of the symmetric matrix $\left(\hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1}\right)\Omega(X_t)$ is less than 1 in absolute value. From the weak law of large numbers, since the vector R_t of assets' gross returns is L_2-bounded, we have $\Omega = \hat{\Omega} + o_p(1)$ so that can write

$$\hat{\Omega}_T(X_t) \simeq \Omega(X_t) - \Omega(X_t)\left(\hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1}\right)\Omega(X_t). \tag{E3}$$

By using Approxs. (E2) and (E3) into Eq. (E1), and keeping only the leading terms, we get

$$\begin{aligned}\hat{\delta}_T^2 \simeq\ & \frac{1}{T}\sum_{t=1}^{T}\mathbf{I}(X_t)e_\star(X_t)'\Omega(X_t)e_\star(X_t) + \frac{2}{T}\sum_{t=1}^{T}\mathbf{I}(X_t)e_\star(X_t)'\Omega(X_t)\left(\hat{e}_t(\theta_\star) - e_\star(X_t)\right) \\ & + \frac{2}{\sqrt{T}}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{I}(X_t)e_\star(X_t)'\Omega(X_t)J_\star(X_t)\right)\sqrt{T}(\hat{\theta}_T - \theta_\star) \\ & - \frac{1}{T}\sum_{t=1}^{T}\mathbf{I}(X_t)e_\star(X_t)'\Omega(X_t)\left(\hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1}\right)\Omega(X_t)e_\star(X_t).\end{aligned} \tag{E4}$$

From the weak law of large numbers we have that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{I}(X_t)e_\star(X_t)'\Omega(X_t)J_\star(X_t) = \mathrm{E}\left[\mathbf{I}(X_t)e_\star(X_t)'\Omega(X_t)J_\star(X_t)\right] + o_p(1),$$

and since $\theta_\star$ satisfies the first-order condition associated to the minimization problem in Eq. (3.2) we have that

$$\mathrm{E}\left[\mathbf{I}(X_t)e_\star(X_t)'\Omega(X_t)J_\star(X_t)\right] = 0_p'.$$

Thus, from Lemma 5, the third term in the r.h.s. of Approx. (E4) is $o_p(1/\sqrt{T})$. By subtracting $\delta_\star^2$ on both sides of Approx. (E4) and scaling both sides by $\sqrt{T}$ we get

$$\sqrt{T}\left(\hat{\delta}_T^2 - \delta_\star^2\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\varepsilon_t - \delta_\star^2\right) + \frac{2}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{I}(X_t)e_\star(X_t)'\Omega(X_t)\left(\hat{e}_t(\theta_\star) - e_\star(X_t)\right)$$
$$-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{I}(X_t)e_\star(X_t)'\Omega(X_t)\left(\hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1}\right)\Omega(X_t)e_\star(X_t) + o_p(1), \quad \text{(E5)}$$

where we use the variable $\varepsilon_t = \mathbf{I}(X_t)e_\star(X_t)'\Omega(X_t)e_\star(X_t)$ and the fact that $\delta_\star^2 = \mathrm{E}[\varepsilon_t]$. Let us now define the scalar function

$$a(Y_i, X_t) := 2e_\star(X_t)'\Omega(X_t)h_\star(Y_i) - \left(e_\star(X_t)'\Omega(X_t)R_i\right)^2,$$

such that

$$\begin{aligned}
&\sum_{i=1}^{T-1} w(X_t, X_i)a(Y_{i+1}, X_t) - \mathrm{E}\left[a(Y_{t+1}, X_t)|\, X_t\right] \\
&= 2e_\star(X_t)'\Omega(X_t)\left(\hat{e}_T(X_t;\theta_\star) - e_\star(X_t)\right) \\
&\qquad - e_\star(X_t)'\Omega(X_t)\left(\hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1}\right)\Omega(X_t)e_\star(X_t).
\end{aligned}$$

We can write Eq (E5) using function a as

$$\begin{aligned}
\sqrt{T}\left(\hat{\delta}_T^2 - \delta_\star^2\right) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{I}(X_t)\left(\sum_{i=1}^{T-1} w(X_t, X_i)a(Y_{i+1}, X_t) - \mathrm{E}\left[a(Y_{t+1}, X_t)|\, X_t\right]\right) \\
&\quad \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\varepsilon_t - \mathrm{E}[\varepsilon_t]\right) + o_p(1). \qquad \text{(E6)}
\end{aligned}$$

We control the first term in the r.h.s. of Eq. (E6) by using the next lemma.

LEMMA 6: *Under regularity conditions on function a and under Ass. 5 we have*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{I}(X_t)\sum_{i=1}^{T-1} w(X_t, X_i)a(Y_{i+1}, X_t) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta_{t+1} + o_p(1).$$

Proof. See App. K of the supplementary material. □

From Eq. (E6) and Lemma 6 we get

$$\sqrt{T}\left(\hat{\delta}_T^2 - \delta_\star^2\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{\varepsilon}_t - \mathrm{E}[\tilde{\varepsilon}_t]) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\eta_{t+1} - \mathrm{E}[\eta_{t+1}|X_t]) + o_p(1). \tag{E7}$$

Prop. 4 follows from the fact that from Ass. 11 the term

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\varepsilon_t - \mathrm{E}[\varepsilon_t] + \eta_{t+1} - \mathrm{E}[\eta_{t+1}|X_t])$$

is asymptotically normal, with null mean and asymptotic variance $\sigma_\star^2$ equal to the upper left element of matrix S.

Appendix 3. Proof of Proposition 5

From the asymptotic expansion in Eq. (E7) written for families $\mathcal{F}$ and $\tilde{\mathcal{F}}$ we have

$$\sqrt{T}\left(\hat{\delta}_T^2 - \delta_\star^2\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_t + o_p(1),$$

and

$$\sqrt{T}\left(\hat{\tilde{\delta}}_T^2 - \tilde{\delta}_\star^2\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_t + o_p(1).$$

Under the null hypothesis $\mathcal{H}_0$, by subtracting side by side the two previous equations we get

$$\sqrt{T}\left(\hat{\delta}_T^2 - \hat{\tilde{\delta}}_T^2\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta\phi_t + o_p(1),$$

From Ass. 11, $\sqrt{T}\left(\hat{\delta}_T^2 - \hat{\tilde{\delta}}_T^2\right)$ is asymptotically normal, with mean 0 and variance σ_Δ^2 equal to $S_{1,1} + S_{2,2} - 2S_{1,2}$, where $S_{i,j}$ is the element of matrix S.