Comparing Asset Pricing Models by the Conditional Hansen-Jagannathan Distance

Gagliardini, Patrick; Ronchetti, Diego

Published in:
Journal of Financial Econometrics

DOI:
10.1093/jjfinec/nbz013

Journal of Financial Econometrics, 2020, Vol. 18, No. 2, 333–394
doi: 10.1093/jjfinec/nbz013
Advance Access Publication Date: 19 April 2019
Article

Comparing Asset Pricing Models by the Conditional Hansen-Jagannathan Distance*

Patrick Gagliardini[1] and Diego Ronchetti[2]

[1]Università della Svizzera italiana and [2]University of Groningen

Address correspondence to Patrick Gagliardini, Università della Svizzera italiana, Faculty of Economics, Via Buffi 13, CH-6900 Lugano, Switzerland, or e-mail: patrick.gagliardini@usi.ch

Received December 20, 2018; revised February 15, 2019; editorial decision February 21, 2019; accepted March 5, 2019

Abstract

We compare nonnested parametric specifications of the stochastic discount factor (SDF) using the conditional Hansen–Jagannathan (HJ-) distance. This distance measures the discrepancy between a parametric model-implied SDF and the admissible SDF's satisfying all the conditional (dynamic) no-arbitrage restrictions, instead of just few unconditional no-arbitrage restrictions for managed portfolios chosen through the instrument selection. We estimate the conditional HJ-distance by a generalized method of moments estimator and establish its large sample properties for model selection purposes. We compare empirically several SDF models including multifactor beta pricing specifications and some recently proposed SDF models that are conditionally linear in consumption growth.

Key words: asset pricing model comparison, generalized method of moments, Hansen–Jagannathan distance, nonparametric estimation, stochastic discount factor

JEL classification: C12, C14, G12

Modern asset pricing theories can be formulated in terms of the stochastic discount factor (SDF), which is a positive random process representing time discount and risk adjustment in the pricing of future risky payoffs. Any SDF matching the dynamic no-arbitrage pricing restrictions for a set of test assets is considered admissible for the market of the same assets.

* The authors would like to thank Joel Horowitz, Frank Kleibergen, Belen Nieto, Eric Renault, Olivier Scaillet, Allan Timmermann, Fabio Trojani, Daniel Wilhelm, an Editor, an Associate Editor, two anonymous referees, and 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, the NESG 2015 in Maastricht, the SoFiE 2015 in Aarhus, the Finance Forum 2015 in Madrid, the EFA 2015 in Wien, the EUROFIDAI-AFFI 2015 meeting in Paris, the Econometric Study Group 2017 workshop in Bristol, the 2018 Workshop on Quantitative Finance in Rome, and seminars at the University of Geneva, University of Groningen and Tinbergen Institute, for useful comments.

Asset pricing models introduce parameterizations of the admissible SDF's as functions of the variables generating the risk premia. For example, in a setting characterized by a representative investor with time-separable preferences, the Euler equations for the solution of the investment/consumption problem lead to a parameterized 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 parametric SDF models, and there is even more latitude in this choice when reduced-form SDF specifications are considered.

Not nested and potentially misspecified asset pricing models are often compared in terms of the Hansen–Jagannathan (HJ-) distance (Hansen and Jagannathan, 1997), which measures the distance between the parametric SDF implied by a model and the set of admissible SDF's. This measurement is done in a space where just static no-arbitrage pricing restrictions for a set of managed portfolios of the test assets are ensured to hold. So defined, the HJ-distance is a measure of static (unconditional) pricing errors on a limited set of portfolios of the test assets, and not the dynamic (conditional) pricing errors on the test assets themselves.

In this article, we introduce a novel version of the HJ-distance which measures the dynamic pricing errors on the test assets, and fully exploits the conditioning information when comparing the performance of competing asset pricing models. We describe its properties at the population level, and its relation with the distance considered in Hansen and Jagannathan (1997). The motivation of our analysis is understanding theoretically and empirically to which extent the ranking of misspecified asset pricing models is modified, when the conditioning information is fully taken into account in assessing the pricing performance of competing models. We propose a nonparametric estimator for the new distance based on the information-theoretic generalized method of moments (GMM) approach. We study its large sample properties, which we deploy to construct a model selection procedure. We illustrate the use of our estimation and model selection procedures in an empirical comparison of popular competing asset pricing models for short term T-bills and the six size- and value-based Fama–French (FF-) research portfolios of U.S. publicly traded equities. To distinguish between the two distances, we refer to the distance described in Hansen and Jagannathan (1997) as the *unconditional HJ-distance*, and to the one introduced in this article as the *conditional HJ-distance*.

In order to introduce the framework of our article, let us consider an economy with N risky assets traded in discrete time. Let $R_t := [R_{1,t} \dots R_{N,t}]'$ denote the vector of gross returns between dates t and $t+1$ for the N assets, and let $\mathcal{I}_t$ be the information available to investors at date t. The Absence of Arbitrage Opportunities (AAO) in the market is equivalent to the existence of a scalar stochastic process $\{M_{t,t+1}\}$ such that the SDF $M_{t,t+1}$ between t and $t+1$ is (i) positive, (ii) measurable with respect to (w.r.t.) the information $\mathcal{I}_{t+1}$ and (iii) satisfying the no-arbitrage restriction $\mathrm{E}\left[M_{t,t+1}R_{t+1}|\mathcal{I}_t\right] = 1_N$, where $\mathrm{E}[\cdot|\mathcal{I}_t]$ denotes the conditional expectation operator under the historical probability measure given information $\mathcal{I}_t$ and 1_N is the N-dimensional vector of ones (see, e.g., Harrison and Kreps, 1979, and Hansen and Richard, 1987). 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 $\mathcal{X} \subseteq \mathbb{R}^L$. Then, the property (iii) for an admissible SDF can be rewritten as

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

for any $x \in \mathcal{X}$, where 0_N is the N-dimensional vector of ones.

In a parametric asset pricing model the admissible SDF $M_{t,t+1}$ is replaced by a candidate SDF, which is a known function m of an observable random vector Y_{t+1} parameterized by a vector θ with unknown value. 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 $\theta \in \Theta \subseteq \mathbb{R}^p$ constitutes a parametric SDF family. The parametric asset pricing model is correctly specified if at least one admissible SDF belongs to this parametric SDF family. In this case, the AAO assumption in Equation (1) implies the N-dimensional vector conditional moment restriction

$$\mathrm{E}[m(Y_{t+1};\theta_0)R_{t+1} - 1_N | X_t = x] = 0_N \tag{2}$$

for any $x \in \mathcal{X}$, where θ_0 is the unknown true value 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 GMM framework (see 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 conditional moment restriction in Equation (2). To create this unconditional moment restriction, we select a $(q \times N)$-dimensional instrument matrix $Z(X_t)$ for any date t, that is a deterministic function of the state variables vector X_t with $q \geq p$. Under the hypothesis of correct model specification, using the instrument matrix $Z(X_t)$ and the law of iterated expectations, we derive the following q-dimensional vector unconditional moment restriction:

$$\mathrm{E}\left[Z(X_t)(m(Y_{t+1};\theta_0)R_{t+1} - 1_N)\right] = 0_q, \tag{3}$$

where $\mathrm{E}[\cdot]$ denotes the unconditional expectation operator under the historical probability measure. The vector $Z(X_t)(m(Y_{t+1};\theta_0)R_{t+1} - 1_N)$ is interpreted as a set of pricing errors for a collection of q managed portfolios, realized by taking dynamic positions in the N traded assets. The rows of instrument matrix $Z(X_t)$ are the weights of these managed portfolios (see, e.g., Section 4B in Cochrane, 1996). If the asset pricing model is correctly specified, that is, Equation (2) holds, then also Equation (3) is satisfied, while the converse implication is not true.

The value of the minimized GMM criterion multiplied by the 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 all admissible SDF's satisfying the unconditional moment restrictions for the chosen instrument matrix. This distance is defined as

$$d_Z := \min_{\theta \in \Theta} \min_{M_{t,t+1} \in \mathcal{M}_Z} \mathrm{E}\left[\left(M_{t,t+1} - m(Y_{t+1};\theta)\right)^2\right]^{1/2}, \tag{4}$$

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 $Z(X_t)$, that is

$$\mathcal{M}_Z := \Big\{M_{t,t+1} \in L^2(\mathcal{I}_{t+1}) : \mathrm{E}\big[Z(X_t)(M_{t,t+1}R_{t+1} - 1_N)\big] = 0_q\Big\}, \quad (5)$$

where we indicate by $L^2(\mathcal{I}_{t+1})$ the linear space of 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:

$$d_Z = \min_{\theta\in\Theta} \Big(\mathrm{E}\big[Z(X_t)(m(Y_{t+1};\theta)R_{t+1} - 1_N)\big]'\Omega_Z \mathrm{E}\big[Z(X_t)(m(Y_{t+1};\theta)R_{t+1} - 1_N)\big]\Big)^{1/2}, \quad (6)$$

where $\Omega_Z := \mathrm{E}\Big[Z(X_t)R_{t+1}R'_{t+1}Z(X_t)'\Big]^{-1}$ is the inverse of the $(q \times q)$-dimensional matrix collecting the second unconditional moments of the scaled assets' gross returns. The HJ-distance is suitable for comparing two possibly nonnested 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. The large sample properties of the unconditional HJ-distance 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, 2014). 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). Almeida and Garcia (2012) and Ghosh, Julliard, and Taylor (2016) have considered different discrepancy measures to assess the unconditional distance of the parametric SDF family from the set of admissible SDF's. We stress that a model ranking based on the unconditional HJ-distance relies on just a few out of infinitely many managed portfolios with weights that are functions of the conditioning information vector. In particular, it depends on the selected instrument matrix.

Nagel and Singleton (2011) have proposed to estimate the 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 Equation (2). Specifically, Nagel and Singleton (2011) have implemented the optimal instrument matrix by nonparametric methods (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 version of Nagel and Singleton (2011), 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 distance-maximizing choice of the instrument matrix. However, in Nagel and Singleton (2011) the conditional HJ-distance is neither estimated nor used for model selection among possibly misspecified models.

In this article, we derive the conditional HJ-distance by extending Equations (4) and (5) 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 Equation (1), and not just the unconditional moment restrictions holding for a particular choice of the instrument matrix.

The article has two main theoretical contributions. First, we study in detail the differences between the conditional and unconditional HJ-distances and their implications. We provide upper and lower bounds for the difference $\delta^2 - d_Z^2$ that are valid for general SDF families. 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. We also show that the conditional HJ-distance admits interpretations in terms of maximal mispriced conditional Sharpe ratio and, in linear SDF models, a conditional version of the Gibbons, Ross, and Shanken (1989) statistic, as well as in terms of Sharpe ratio improvement along the lines of Barillas and Shanken (2017). The second theoretical contribution is the definition of a sample analog 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 $m(Y_{t+1};\theta)R_{t+1} - 1_N$ given the conditioning information vector X_t by kernel regression methods. We establish large sample distributional results which allow us to develop a model selection procedure based on the conditional HJ-distance.

Our empirical contribution consists in the comparison of thirteen parametric SDF specifications for the gross returns on the six U.S. equity FF-research portfolios and short term T-Bill in terms of the unconditional and conditional HJ-distances. We include in our analysis preference-based SDF models with the time-separable Constant Relative Risk-Aversion (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), and their linearizations for small values of the logarithmic consumption growth and returns on the market portfolio. Three SDF specifications, considered also in Nagel and Singleton (2011), are conditionally linear in logarithmic consumption growth, and correspond to the Consumption-based Capital Asset Pricing Model (CCAPM) with time-varying risk premia. This variation of the coefficients over time is modeled through either consumption to wealth ratio (Lettau and Ludvigson, 2001), corporate bond spread (Jagannathan and Wang, 1996) or labor income-to-consumption ratio (Santos and Veronesi, 2006). The remaining 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 other beta pricing models. 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. We show that a clear superior empirical performance of the unconditional beta pricing models is obtained only when we base the ranking on the unconditional HJ-distance with particular instrument matrices. The empirical ranking of the models is sensitive to the selection of the instrument matrix. When adopting the conditional HJ-distance for model comparison, we find that some SDF specifications that are conditionally linear in consumption growth have superior performance. In fact, the conditional pricing errors of some multifactor beta specifications are more volatile but less correlated with the variable generating the conditioning information, thus explaining the reversed ranking when conditional or unconditional HJ-distances are considered.

In a recent paper, Antoine, Proulx, and Renault (2018) also consider the conditional HJ-distance for the study of misspecification. Their work is complementary to our analysis since the focus of their paper is on the pseudo-true SDF parameter implied by the minimization of the conditional HJ-distance rather than on the properties of the conditional HJ-distance itself. They provide insightful economic interpretations for the pseudo-true SDF and study the properties of estimators thereof based on fixed bandwidth methods.

1 The Conditional HJ-Distance

We introduce in Section 1.1 the conditional HJ-distance. In Section 1.2 we describe its theoretical properties and discuss the difference with the unconditional HJ-distance. Finally, in Section 1.3 we discuss a modification of the conditional HJ-distance in case the competing asset pricing models are for sure able to correctly price a benchmark asset, such as a risk-free asset.

1.1 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:

$$\mathcal{M} := \left\{ M_{t,t+1} \in L^2(\mathcal{I}_{t+1}) : \mathrm{E}\left[M_{t,t+1}R_{t+1} - 1_N | X_t = x\right] = 0_N \text{ for any } \mathrm{x} \in \mathcal{X} \right\}. \tag{7}$$

Let $\{m(\cdot;\theta) : \theta \in \Theta\}$ be a parametric SDF family. Adapting the definition of unconditional HJ-distance given in Equation (4) to the setting with dynamic pricing restrictions, we define the conditional HJ-distance δ as

$$\delta := \min_{\theta \in \Theta} \min_{M_{t,t+1} \in \mathcal{M}} \mathrm{E}\left[\left(M_{t,t+1} - m(Y_{t+1};\theta)\right)^2\right]^{1/2}. \tag{8}$$

It is the minimal L^2 distance between the SDF's in the parametric family and the SDF's that matches the dynamic pricing restrictions. By solving the inner constrained optimization problem in the right-hand side (RHS) of Equation (8) we get the next proposition.

Proposition 1. The conditional HJ-distance is equal to

$$\delta = \min_{\theta \in \Theta} \mathrm{E}\left[e(X_t;\theta)'\Omega(X_t)e(X_t;\theta)\right]^{1/2}, \tag{9}$$

where

$$e(X_t;\theta) := \mathrm{E}\left[h(Y_{t+1}, R_{t+1};\theta)|X_t\right] \tag{10}$$

is the N-dimensional vector of conditional pricing errors for parameter value $\theta \in \Theta$, defined by means of the vector function $h(Y_{t+1}, R_{t+1};\theta) := m(Y_{t+1};\theta)R_{t+1} - 1_N$, and $\Omega(X_t) := \mathrm{E}\left[R_{t+1}R'_{t+1}|X_t\right]^{-1}$ is the inverse of the $(N \times N)$-dimensional conditional second-moments matrix of assets' gross returns vector R_{t+1} given the conditioning information vector X_t.

Proof. See Appendix A.1. □

The expression in Equation (9) corresponds to the conditional HJ-distance proposed in the working paper version of Nagel and Singleton (2011).[1] The conditional HJ-distance equals the square root of the minimized expectation of a quadratic form of the conditional pricing error vector for the N test assets. It is null if, and only if, the conditional pricing error vector vanishes at all dates for a parameter value, that is, the conditional moment restriction in Equation (2) is satisfied and the parametric SDF family $\{m(\cdot;\theta) : \theta \in \Theta\}$ is correctly specified. In this case, the minimizer of the criterion in Equation (9) coincides with the true parameter value θ_0. Differently, when the asset pricing model is misspecified, the minimization problem in Equation (9) defines the pseudo-true parameter value, which we denote by $\theta_* \in \Theta$. In this case, $\delta > 0$ is the L^2-distance between the candidate SDF $m(Y_{t+1};\theta_*)$ and the closest admissible SDF in set $\mathcal{M}$. We stress that, in contrast to the unconditional HJ-distance d_Z in Equations (4) and (6), the conditional HJ-distance δ does not depend on any instrument matrix $Z(X_t)$. □

1.2 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 the values of these distances can differ. We start by establishing an inequality between the two distances. Let us consider the definition of set $\mathcal{M}$ of admissible SDF's satisfying the conditional moment restrictions in Equation (7) and the definition of set $\mathcal{M}_Z$ of admissible SDF's satisfying the unconditional moment restrictions derived from the chosen instrument matrix in Equation (5). By the law of iterated expectations we deduce that $\mathcal{M} \subseteq \mathcal{M}_Z$. Therefore, the conditional HJ-distance, which is a L^2-discrepancy minimized over the subset $\mathcal{M}$, is not smaller than the unconditional HJ-distance, which is the corresponding measure minimized over the bigger set $\mathcal{M}_Z$:

$$\delta \geq d_Z. \tag{11}$$

Intuitively, for a given asset pricing model the value of the conditional HJ-distance is generally larger than its unconditional counterpart because the former captures any violation of the *dynamic* pricing restrictions for the test assets, and not just nonzero time averages of the pricing errors for a chosen set of managed portfolios.

In the rest of this section we study the discrepancy between the two distances in more depth, and 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. For this purpose, we introduce in Subsection 1.2.1 a representation of the two HJ-distances as vector norms in a suitable Hilbert space. We then provide in Subsection 1.2.2 an upper and lower bound for the difference between these two norms, for any parametric SDF family. We particularize the results in Subsection 1.2.3 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.

1 Differently from our definition, some authors refer to the argument of the unconditional expectation operator in Equation (9) as squared conditional HJ-distance (see, e.g., Balduzzi and Robotti, 2010, and Fang, Ren, and Yuan, 2011).

1.2.1 Representing the HJ-distances as weighted L^2-norms

We introduce the Hilbert space $L^2_\Omega(\mathcal{X})$ of N-dimensional square integrable vector functions endowed with the inner product $\langle \phi, \tilde{\phi} \rangle_{L^2_\Omega(\mathcal{X})} := \mathrm{E}\big[\phi(X_t)'\Omega(X_t)\tilde{\phi}(X_t)\big]$ and associated norm $\|\phi\|_{L^2_\Omega(\mathcal{X})} := \langle \phi, \phi \rangle^{1/2}_{L^2_\Omega(\mathcal{X})}$, for any N-dimensional vector functions $\phi, \tilde{\phi} \in L^2_\Omega(\mathcal{X})$. Compared to the standard L^2 scalar product and norm in function spaces, here the expectation is taken w.r.t. the stationary distribution of the state variables process $\{X_t\}$, and the inverse conditional second-moments matrix of assets' gross returns is used for weighting. From Equation (9) the conditional HJ-distance can be written as the minimized $L^2_\Omega(\mathcal{X})$-norm of the conditional pricing error vector:

$$\delta = \min_{\theta \in \Theta} \; \|e(\cdot;\theta)\|_{L^2_\Omega(\mathcal{X})} = \|e(\cdot;\theta_*)\|_{L^2_\Omega(\mathcal{X})}. \tag{12}$$

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 $Z(X_t)$ by means of a $(N \times q)$-dimensional matrix function A of vector X_t as

$$Z(X_t) = A(X_t)'\Omega(X_t). \tag{13}$$

The columns of the matrix $A(X_t)$ are the portfolio weights in the instrument matrix $Z(X_t)$ scaled by the conditional second moment matrix of the assets' gross returns, that is $A(X_t)' = Z(X_t)\Omega(X_t)^{-1}$. In Appendix A.2 we prove that the unconditional HJ-distance is equal to

$$d_Z = \min_{\theta \in \Theta} \; \|\mathcal{P}_A[e(\cdot;\theta)]\|_{L^2_\Omega(\mathcal{X})} = \|\mathcal{P}_A[e(\cdot;\theta_Z)]\|_{L^2_\Omega(\mathcal{X})}, \tag{14}$$

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 Equation (4), or equivalently in Equation (6). Thus, the unconditional HJ-distance only accounts for a component of the conditional pricing error vector, namely the one which is along the scaled instrument matrix $A(X_t)$. We highlight the consequences of this finding in the next two subsections, which deal with general and linear parametric SDF families, respectively.

1.2.2 Results for general parametric SDF families

In this subsection we refine the relation between the two HJ-distances given by Inequality (11) by means of the representations introduced in Subsection 1.2.1 for general parametric SDF families.

Proposition 2. The conditional HJ-distance δ and the unconditional HJ-distance d_Z are such that

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

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 Appendix A.3. □

The upper and lower bounds for the difference between the two squared HJ-distances given in Proposition 2 depend (among other quantities) on the instrument matrix $Z(X_t)$ through matrix $A(X_t)$. First, the lower bound $\|\mathcal{P}_A^{\perp}[e(\cdot;\theta_*)]\|^2_{L^2_\Omega(\mathcal{X})}$ implies that the difference $\delta^2 - d_Z^2$ is not smaller than 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 $A(X_t)$. The pricing error vector is computed at the minimizer of the conditional HJ-distance θ_*. The intuition behind this finding is as follows. When the instrument matrix $Z(X_t)$ 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 $Z(X_t)$ 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. Second, the upper bound $\|\mathcal{P}_A^{\perp}[e(\cdot;\theta_Z)]\|^2_{L^2_\Omega(\mathcal{X})}$ implies that the difference $\delta^2 - d_Z^2$ is capped by the $L^2_\Omega(\mathcal{X})$-norm of the orthogonal projection of the conditional pricing error vector $e(X_t;\theta_Z)$ onto the space that is unspanned by the columns of matrix $A(X_t)$. Here, the pricing error vector is computed at the minimizer of the unconditional HJ-distance.

The following corollary follows as a consequence of Proposition 2.

Corollary 1. *Given any 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 Corollary 1, let us consider the case of constant instrument matrix, that is $Z(X_t) = \mathbf{I}_N$, and test asset returns with constant second moment, that is $\Omega = \Omega_Z = \mathrm{E}\left[R_{t+1}R'_{t+1}\right]^{-1}$. First, from Equation (6) 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 Proposition 2 and the invariance of the trace operator under cyclical permutations, the lower bound $\|\mathcal{P}_A^{\perp}[e(\cdot;\theta_*)]\|^2_{L^2_\Omega(\mathcal{X})}$ for the difference $\delta^2 - d_Z^2$ can be written as $\mathrm{E}\big[(e(X_t;\theta_*) - \mathrm{E}[e(X_t;\theta_*)])' \ \Omega(e(X_t;\theta_*) - \mathrm{E}[e(X_t;\theta_*)])\big] = \mathrm{Tr}[\Omega\mathrm{V}[e(X_t;\theta_*)]]$. 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.

Nagel and Singleton (2011) have already shown that the small average pricing errors of a misspecified model, that are obtained when estimation is based on unconditional moment restrictions, may hide very large time-variation in conditional pricing errors (see, e.g., Figures 3 and 4 of Nagel and Singleton, 2011). The contribution of Proposition 2 is to show that this finding is generally valid beyond specific asset pricing models and instrument choices, and this effect is related to the component of the conditional pricing error vector that is unspanned by the selected instruments.

In the following corollary of Proposition 2, we give sufficient and necessary conditions for the two HJ-distances to coincide for a given parametric SDF family.

Corollary 2. *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 $A(X_t)$.*

Hence, a sufficient condition for the two HJ-distances to coincide is that the selected instruments fully capture the information in the conditional pricing error vector for parameter value θ_Z. This sufficient condition on the instrument matrix is rather restrictive and is model dependent.

Let us now consider the implications of Proposition 2 when comparing the values of the conditional and unconditional HJ-distances for distinct parametric SDF families.

Corollary 3. *The unconditional and conditional HJ-distances can yield different rankings for the degree of misspecification of competing parametric SDF families.*

Let us illustrate Corollary 3 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 $\tilde{\mathcal{F}}$ is preferred over the SDF family $\mathcal{F}$ in terms of the unconditional HJ-distance computed using the instrument matrix $Z(X_t)$. By Proposition 2 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 $\tilde{\delta} \geq \delta$.[2] Therefore, the parametric SDF family $\mathcal{F}$ is preferred over $\tilde{\mathcal{F}}$ in terms of the conditional HJ-distance (i.e., there is a reverse ranking compared to that implied by the unconditional HJ-distance).

1.2.3 Results for linear SDF families

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

$$m(Y_{t+1};\theta) = Y'_{t+1}\theta, \tag{15}$$

where the elements of the p-dimensional vector Y_{t+1} are functions of the priced factors and possibly some variables included in the conditioning information vector. Examples of such linear SDF specifications are the (scaled) beta pricing models and the conditional CCAPM models considered by Nagel and Singleton (2011), and are among the models included in our empirical analysis in Section 4. The conditional pricing error vector at date t is

$$e(X_t;\theta) = B(X_t)\theta - 1_N, \tag{16}$$

for any $\theta \in \Theta$, where the $(N \times p)$-dimensional matrix function $B(X_t) := \mathrm{E}[R_{t+1}Y'_{t+1}|X_t]$ consists of the conditional cross-moments of assets' gross returns and SDF variables given the vector X_t. Adapting the expression for the conditional HJ-distance δ given in Equation (12) 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 constant vector 1_N onto the space spanned by the columns of matrix function $B(X_t)$.

The following proposition gives explicit expressions for the difference of the squared conditional and unconditional HJ-distances in terms of the conditional pricing error vector

2 From Equation (14) written for family $\tilde{\mathcal{F}}$, the quantity $\delta^2 - \tilde{d}_Z^2$ depends on the SDF family $\tilde{\mathcal{F}}$ only via $\mathcal{P}_A[\tilde{e}(\cdot;\tilde{\theta}_Z)]$, that is 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 $\delta^2 - \tilde{d}_Z^2$.

$e(X_t;\cdot)$ valued at θ_* and θ_Z. It also provides two sufficient and necessary conditions for the conditional and unconditional HJ-distances to coincide.

Proposition 3. For the linear SDF specification given in Equation (15) and the parameterization of the instrument matrix given in Equation (13), we have

$$\delta^2 - d_Z^2 = \left\|\mathcal{P}_A^{\perp}[e(\cdot;\theta_*)]\right\|^2_{L^2_\Omega(\mathcal{X})} + \left\|\mathcal{P}_{\mathcal{P}_A[B]}[e(\cdot;\theta_*)]\right\|^2_{L^2_\Omega(\mathcal{X})} \tag{17}$$

$$= \left\|\mathcal{P}_A^{\perp}[e(\cdot;\theta_Z)]\right\|^2_{L^2_\Omega(\mathcal{X})} - \left\|e(\cdot;\theta_Z) - e(\cdot;\theta_*)\right\|^2_{L^2_\Omega(\mathcal{X})}, \tag{18}$$

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 $\mathcal{P}_A[B]$. Moreover

$$\delta = d_Z \quad \Leftrightarrow \quad \mathcal{P}_A^{\perp}[e(\cdot;\theta_*)](X_t) = 0_N \quad \Leftrightarrow \quad \mathcal{P}_A^{\perp}[e(\cdot;\theta_Z)](X_t) = 0_N. \tag{19}$$

Proof. See Appendix A.4. □

In Equation (17), 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 the scaled instruments matrix function $A(X_t)$, while the second one is the orthogonal projection of the column space of conditional cross-moments matrix function $B(X_t)$ onto the column space of $A(X_t)$. We deduce that the difference between the two HJ-distances is large when the variability of the conditional pricing error vector is either unexplained by the instruments or collinear with the projection of the cross-moments $B(X_t)$ onto the instruments. Equation (18) provides another expression for the difference $\delta^2 - d_Z^2$. Note that the first terms in the RHS of Equations (17) and (18) correspond to the lower and upper bounds, respectively, for the difference $\delta^2 - d_Z^2$ given in Proposition 2 for general SDF families. Thus, Proposition 3 explains how far $\delta^2 - d_Z^2$ is from these lower and upper bounds in the case of linear SDF families.

In Equations (19) we provide two equivalent 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 the scaled instrument matrix function $A(X_t)$. These two conditions are both sufficient and necessary for $\delta = d_Z$. In this respect, Proposition 3 is a stronger result than Corollary 2, albeit limited to the case of linear SDF families.

Let us finally relate the conditions in Equations (19) to the optimal instruments. The choice $A(X_t) = B(X_t)$ in Equation (13) yields the optimal instrument matrix for the estimation of the SDF parameter θ_0 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 $q = p$, that is, 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. Differently, if the SDF family is misspecified, we have $\delta > 0$. Therefore, adopting the optimal instruments does not guarantee equality of conditional and unconditional HJ-distances. However, if the column space of $A(X_t)$ includes that of $B(X_t)$, then $\mathcal{P}_A^{\perp}[e(\cdot;\theta_*)] = \mathcal{P}_A^{\perp}[e(\cdot;\theta_Z)] = -\mathcal{P}_A^{\perp}[1_N]$. Therefore, the sufficient and necessary conditions for obtaining $d_Z = \delta$ in Proposition 3 are met if we add instruments to the optimal ones so that the columns of matrix function $A(X_t)$ span the constant vector 1_N.

1.3 The Inclusion of a Reference Asset and Link to Conditional Sharpe Ratios

In this section, we show that the conditional HJ-distance admits interpretations in terms of conditional Sharpe ratios, which are analogue to the ones holding in the unconditional setting. For this purpose, it is convenient to first particularize the conditional HJ-distance under the hypothesis that all the candidate asset pricing models price correctly a reference asset. We take the (conditionally) risk-free asset as the reference asset. Under this hypothesis, all the candidate asset pricing models perfectly explain the cash-flows time discount of the investors.[3] We then introduce a risk-free $(N+1)$-th asset, and value the pricing performance of the competing models on the basis of the explained returns in excess of the gross return on this risk-free asset. Adapting the definition of unconditional modified HJ-distance given in Kan and Robotti (2008) to the setting with dynamic pricing restrictions, we define the conditional modified HJ-distance $\delta_{[m]}$ in the same way as the conditional HJ-distance in Equation (8) by replacing the set $\mathcal{M}$ by its subset

$$\mathcal{M}_{[m]} := \Big\{M_{t,t+1} \in L^2(\mathcal{I}_{t+1}) : \mathrm{E}\big[M_{t,t+1}\big(R_{t+1} - R_{f,t+1}1_N\big)|X_t = x\big] = 0_N \\ \text{and } R_{f,t+1}\mathrm{E}\big[M_{t,t+1}|\mathcal{X}_t = x\big] = 1 \text{ for any } x \in \mathcal{X}\Big\}, \tag{20}$$

where $R_{f,t+1}$ is the gross return on the risk-free asset from date t to date $t+1$. The conditional modified HJ-distance $\delta_{[m]}$ is the minimal L^2 distance between the SDF's in the parametric family and the SDF's that matches the dynamic pricing restrictions for risk-free asset and excess returns on the test assets. By solving the optimization problem through similar steps as the ones taken to show Proposition 1 we obtain the expression

$$\delta_{[m]} = \min_{\theta\in\Theta} \mathrm{E}\Big[e(X_t;\theta)'\Omega_{[m]}(X_t)e(X_t;\theta)\Big]^{1/2}, \tag{21}$$

where e is the vector of conditional pricing errors defined in Equation (10) and $\Omega_{[m]}(X_t) := \mathrm{V}[R_{t+1}|X_t]^{-1}$ is the inverse of the $(N \times N)$-dimensional conditional variance–covariance matrix of the test assets' gross returns vector R_{t+1} given X_t. See Appendix A.5 for a derivation of Equation (21). If the asset pricing model is misspecified, this criterion defines the pseudo-true parameter value $\theta_{[m]*} \in \Theta$.

Let us now consider a possibly misspecified candidate asset pricing model with SDF $m(Y_{t+1};\theta)$, which prices correctly the risk-free asset. Let us denote its model-implied conditional expected gross returns vector given the conditioning information vector X_t as

$$\mathrm{E}_\theta[R_{t+1}|X_t] := R_{f,t+1}1_N - R_{f,t+1}\mathrm{Cov}[m(Y_{t+1};\theta), R_{t+1}|X_t] \tag{22}$$

for any $\theta \in \Theta$. Let us consider the set Z of real-valued instrument vectors $Z(X_t)$, that is, the set of instrument matrices introduced in Equation (3) for $q=1$. Building on the derivations

3 In our empirical application in Section 4 we identify the SDF scale through the no-arbitrage restriction for short-term T-bills, and we do not a priori assume that the competing asset pricing models price them correctly.

in Hodrick and Zhang (2001) and extending them in the conditional setting, in Appendix A.6 we show that

$$\delta_{[\mathrm{m}]}^2 = \mathrm{E}\left[\max_{Z\in Z} \frac{1}{R_{f,t+1}^2} \frac{\left(Z(X_t)'\left(\mathrm{E}[R_{t+1}|X_t] - \mathrm{E}_{\theta_{[\mathrm{m}]*}}[R_{t+1}|X_t]\right)\right)^2}{\mathrm{V}\left[Z(X_t)'R_{t+1}|X_t\right]}\right]. \quad (23)$$

Thus, the squared conditional modified HJ-distance represents the unconditional expectation of the maximum squared discounted mispriced conditional Sharpe ratio for a managed portfolio constructed from the test assets.

The next interpretation holds for linear SDF families.

1.3.1 Results for linear SDF families

Let us consider the case with the excess returns for the N test assets being conditionally linear-affine in other k additional excess returns given the conditioning information vector X_t. In this case we have $R_{t+1} = \alpha(X_t) + B(X_t)f_{t+1} + \epsilon_{t+1}$, for the N-dimensional vector $R_t := \left[R_{1,t} - R_{f,t} \ldots R_{N,t} - R_{f,t}\right]'$ and the k-dimensional vector $f_t := \left[R_{N+1,t} - R_{f,t} \ldots R_{N+k,t} - R_{f,t}\right]'$. The N-dimensional vector of errors ϵ_t is such that $\mathrm{E}[\epsilon_{t+1}|X_t] = 0_N$ and $\mathrm{Cov}[\epsilon_{t+1}, f_{t+1}|X_t] = 0_{N\times k}$, so that $\alpha(X_t)$ and $B(X_t)$ are the vector of alphas and the matrix of factor betas conditional on the state X_t. Moreover, we consider a linear SDF $m(Y_{t+1};\theta) = R_{f,t+1}^{-1} + (f_{t+1} - \mathrm{E}[f_{t+1}|X_t])'\theta$, with $Y_{t+1} = \left[f_{t+1}' \, X_t'\right]'$. This SDF model prices correctly the risk-free asset.

The square conditional Sharpe ratio for the augmented vector of excess gross returns $R_t^{[\mathrm{Aug}]} := \left[f_t' \, R_t'\right]'$ is $\mathrm{Sh}\left[R_{t+1}^{[\mathrm{Aug}]}|X_t\right]^2 = \mathrm{E}\left[R_{t+1}^{[\mathrm{Aug}]}|X_t\right]'\mathrm{V}\left[R_{t+1}^{[\mathrm{Aug}]}|X_t\right]^{-1}\mathrm{E}\left[R_{t+1}^{[\mathrm{Aug}]}|X_t\right]$. Using the conditional factor structure and simple matrix algebra we have that

$$\mathrm{Sh}\left[R_{t+1}^{[\mathrm{Aug}]}|X_t\right]^2 = \mathrm{Sh}[f_{t+1}|X_t]^2 + \alpha(X_t)'\mathrm{V}[\epsilon_{t+1}|X_t]^{-1}\alpha(X_t). \quad (24)$$

We show in Appendix A.7 that the conditional modified HJ-distance for the vector $R_t^{[\mathrm{Aug}]}$ is such that

$$\begin{aligned}\delta_{[\mathrm{m}]}^{[\mathrm{Aug}]2} &= \mathrm{E}\left[\frac{1}{R_{f,t+1}^2}\alpha(X_t)'\mathrm{V}[\epsilon_{t+1}|X_t]^{-1}\alpha(X_t)\right] \\ &\quad + \mathrm{E}\left[\frac{1}{R_{f,t+1}^2}\mathrm{E}[f_{t+1}|X_t]'\left(\mathrm{V}[f_{t+1}|X_t]^{-1} - \mathrm{E}[\mathrm{V}[f_{t+1}|X_t]]^{-1}\right)\mathrm{E}[f_{t+1}|X_t]\right].\end{aligned} \quad (25)$$

The first term in the RHS yields a link of the conditional modified HJ-distance with a conditional version of the quadratic form of alphas considered in Gibbons, Ross, and Shanken (1989). The second term in the RHS involves the conditional heteroskedasticity of the factors. In fact, the considered SDF has constant risk prices, while the latter should account for the conditional heteroskedasticity of the risk factors. From Equation (24), we have that $\mathrm{E}\left[\frac{1}{R_{f,t+1}^2}\alpha(X_t)'\mathrm{V}[\epsilon_{t+1}|X_t]^{-1}\alpha(X_t)\right] = \mathrm{E}\left[\frac{1}{R_{f,t+1}^2}\left(Sh\left[R_{t+1}^{[\mathrm{Aug}]}|X_t\right]^2 - Sh[f_{t+1}|X_t]^2\right)\right]$. This provides an interpretation as expected improvement in squared conditional Sharpe ratios for the first term in the RHS of Equation (25), which extends results in Barillas and Shanken (2017) to a conditional setting.

2 Nonparametric Estimation of the Conditional HJ-Distance

The conditional HJ-distance δ, and the true value θ_0 of the SDF parameter vector if the model is correctly specified, or the pseudo-true value θ_* in case of model misspecification, are unobservable features of the data generating process (DGP). In this section, we describe an estimation methodology for these quantities based on a sample of T time series observations of the stationary joint process for state variables X_t, risk factors Y_{t+1} and gross returns R_{t+1}. The use of the proposed estimator 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 introduce the nonparametric estimator of the conditional HJ-distance in Section 2.1. We discuss its asymptotic properties in Section 2.2 for the case of correct model specification, and in Section 2.3 for the case of model misspecification.

2.1 The Estimator

The conditional HJ-distance can be estimated by replacing the unconditional expectation in the criterion in Equation (9) by a sample average, and the conditional expectation in the definition of the conditional pricing error vector in Equation (10) by a nonparametric regression. We consider kernel smoothing and denote by K a kernel function on set $\mathbb{R}^L$ and by b_T a positive bandwidth, which depends on the sample size T and converges to zero as T tends to infinity. 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 $[Y'_{t+1}\ R'_{t+1}]'$ and X_t, which is defined as

$$\hat{e}_T(X_t;\theta) := \sum_{i=1}^{T-1} w(X_t, X_i) h(Y_{i+1}, R_{i+1};\theta), \tag{26}$$

where the kernel weighting function w is defined as

$$w(X_t, X_i) := K\big((X_t - X_i)/b_T\big) \Big/ \sum_{j=1}^{T-1} K\big((X_t - X_j)/b_T\big), \tag{27}$$

and we omit to indicate the dependence of w on the bandwidth b_T for expository purpose. Thus, vector $\hat{e}_T(X_t;\theta)$ is a weighted sample average of the vectors $h(Y_{i+1}, R_{i+1};\theta)$, 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 sample conditional HJ-distance $\hat{\delta}_T$ is defined by

$$\hat{\delta}_T^2 := \min_{\theta\in\Theta} \mathcal{Q}_T(\theta) = \mathcal{Q}_T(\hat{\theta}_T), \qquad \mathcal{Q}_T(\theta) := \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(X_t)\hat{e}_T(X_t;\theta)'\hat{\Omega}_T(X_t)\hat{e}_T(X_t;\theta), \tag{28}$$

where

$$\hat{\Omega}_T(X_t) := \left(\sum_{i=1}^{T-1} w(X_t, X_i) R_{i+1} R'_{i+1}\right)^{-1} \tag{29}$$

is the inverse of the kernel regression estimator of matrix $\Omega(X_t)^{-1} = \mathrm{E}[R_{t+1}R'_{t+1}|X_t]$. The indicator variable $\mathbf{1}(X_t)$ is equal to 1 when X_t is in a given compact subset $\mathcal{X}_\star \subset \mathcal{X}$ of the state variables support independent on T, and 0 otherwise.

The indicator variable $\mathbf{1}(X_t)$ excludes the values of the state variable vector outside the set $\mathcal{X}_\star$ from the sample average. Technically, the indicator variable $\mathbf{1}(X_t)$ is a trimming factor to control boundary effects in the kernel regressions.[4] More importantly, as noted by Aït-Sahalia et al. (2001) considering a subset of the state variables support allows to focus on those states which are more relevant for the analysis. For instance, set $\mathcal{X}_\star$ could be chosen to target some specific market conditions, such as booms or recession periods, when comparing two asset pricing models. Since the criterion function $\mathcal{Q}_T$ involves the indicator variable $\mathbf{1}(X_t)$ for set $\mathcal{X}_\star$, the sample conditional HJ-distance $\hat{\delta}_T$ is not a consistent estimator of δ but instead of the quantity

$$\delta_\star := \min_{\theta \in \Theta} \mathrm{E}\left[\mathbf{1}(X_t) e(X_t;\theta)' \Omega(X_t) e(X_t;\theta)\right]^{1/2}. \tag{30}$$

The quantity $\delta_\star$ is 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 have $\delta_\star = \min_{\theta \in \Theta} \min_{M_{t,t+1} \in \mathcal{M}_\star} \mathrm{E}\left[\mathbf{1}(X_t)(M_{t,t+1} - m(Y_{t+1};\theta))^2\right]^{1/2}$ indeed, where the set $\mathcal{M}_\star$ is defined as in Equation (7) with set $\mathcal{X}_\star$ in place of set $\mathcal{X}$. We refer to $\delta_\star$ as the trimmed conditional HJ-distance. Because of the trimming factor we have $\delta_\star \leq \delta$. We denote by $\theta_\star$ the minimizer of the criterion in Equation (30), which we assume to be unique in parameter set Θ to ensure its identification. Parameter value $\theta_\star$ coincides with the true parameter value θ_0 if the SDF family is correctly specified, but generally differs from the pseudo-true parameter θ_* under misspecification. However, if set $\mathcal{X}_\star$ is chosen sufficiently large, we can make $\delta_\star$ and $\theta_\star$ arbitrarily close to δ and θ_* in the latter case.

The sample conditional HJ-distance $\hat{\delta}_T$, and the SDF parameter estimator $\hat{\theta}_T$ defined as the minimizer of the criterion in (28), 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 the finite dimensional parameter θ, and to an Euclidean Empirical Likelihood (EEL) estimator of Antoine, Bonnal, and Renault (2007) with nonoptimal weighting matrix. The EEL estimator is a member of the class of information-theoretic 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-theoretic 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-theoretic 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

4 See, for example, Tripathi and Kitamura (2003) and Su and White (2014) for the introduction of trimming factors in test statistics to account for the inaccuracy of nonparametric estimation in tail regions.

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 consistency of the specification tests under the alternative hypothesis of model misspecification, mostly in a i.i.d. data framework.[5] On the other hand, Hall and Inoue (2003) investigate in depth the asymptotic properties of GMM estimators in misspecified unconditional moment restrictions models. In a conditional moment restrictions setting, Antoine, Proulx, and Renault (2018) advocate the use of estimators of the pseudo-true SDF parameter with fixed bandwidth. Instead, the large sample behavior of the estimator $\hat{\theta}_T$ of the pseudo-true value $\theta_\star$ 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 with bandwidth b_T tending to zero as T grows are unexplored issues. These issues are the topics of the next section.

The sample conditional HJ-distance $\hat{\delta}_T$ and the SDF parameter estimator $\hat{\theta}_T$ require nonparametric estimators for the conditional pricing errors. The latter are affected by the curse of dimensionality and are reliable only if the dimension L of the conditioning information vector X_t is low, say $L \leq 4$. Low dimensionality of the vector defining the conditional moment restriction can be obtained by selecting a subset of the available information and using the law of iterated expectations as in Nagel and Singleton (2011). For sake of clarity, let us stress that this dimensionality problem is a different issue than the problem of instrument selection which has been discussed in Section 1. Even when the state variables vector generating information X_t is low dimensional, to estimate the unconditional HJ-distance the econometrician has nonetheless to choose a finite number q of deterministic functions of vector X_t to create the instrument matrix $Z(X_t)$.

2.2 Large Sample Properties under Correct Model Specification

Let $\mathcal{F} := \{m(\cdot;\theta) : \theta \in \Theta\}$ be a correctly specified parametric SDF family, and let $\hat{\delta}_T$ and $\hat{\theta}_T$ be the sample conditional HJ-distance and related estimator of the true SDF parameter value defined in (28). Under a set of regularity conditions collected in Appendix B, for $T \to \infty$ the estimator $\hat{\theta}_T$ is consistent and asymptotically normal (see Appendix C.1) and the asymptotic distribution of statistic $\hat{\delta}_T^2$ is given in Proposition 4 below. 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 $\{X_t\}$. These conditions are standard in the literature on nonparametric estimation and testing. We use the symbol $\xrightarrow{\mathcal{D}}$ to denote convergence in distribution.

Proposition 4. If the parametric SDF family is correctly specified, the sample conditional HJ-distance $\hat{\delta}_T$ is such that $Tb_T^{L/2}(\hat{\delta}_T^2 - a_T) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_0^2)$ as $T \to \infty$, for the positive centering variable

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

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

and the asymptotic variance $\sigma_0^2 := 2\left(\int_{\mathbb{R}^L} \mathcal{K}(x)^2 dx\right)\int_{\mathcal{X}_\star} \mathrm{Tr}[V_0(x)\Omega(x)V_0(x)\Omega(x)]dx$, with kernel convolution $\mathcal{K}(x) := \int_{\mathbb{R}^L} K(u)K(u-x)du$ and conditional variance–covariance matrix $V_0(x) := \mathrm{V}\left[h(Y_{t+1}, R_{t+1}; \theta_0)|X_t = x\right]$ of the moment function valued at the true parameter value θ_0.

Proof. See Appendix D.1. □

The proof of Proposition 4 follows closely the proof of Theorem 4.1 in Tripathi and Kitamura (2003) by extending their results to a setting with serially dependent data and generic weighting matrix. The asymptotic distribution in Proposition 4 of the squared sample HJ-distance sharply differs from the asymptotic distribution of the squared sample 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 $\hat{\delta}_T^2$ is Gaussian (instead of a linear combination of independent chi-square variables as for its unconditional counterpart), 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 Equation (2) corresponds to an infinity of unconditional moment restrictions.[6] Asymptotic normality is derived by applying a Central Limit Theorem for quadratic forms in Yoshihara (1976, 1992).

The test of the null hypothesis $\delta = 0$ (correct specification) against the alternative hypothesis $\delta > 0$ (model misspecification) is unilateral. The rejection rule for a test of the correct specification of the SDF family with asymptotic significance level α is

$$\hat{\delta}_T^2 \geq a_T + c_{1-\alpha}\frac{\hat{\sigma}_0}{\sqrt{Tb_T^{L/2}}}, \tag{31}$$

where $c_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal distribution, and $\hat{\sigma}_0$ is a consistent estimator of σ_0 obtained by replacing the matrices $V_0(x)$ and $\Omega(x)$ in Proposition 4 by kernel estimators.

2.3 Large Sample Properties 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. Let us assume that SDF 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, that is $\delta_\star > 0$. In this case, $\hat{\theta}_T$ is a consistent and asymptotically normal estimator of $\theta_\star$ (see Appendix C.2) and we give the large sample behavior of the squared sample conditional HJ-distance $\hat{\delta}_T^2$ in the next proposition.

Proposition 5. If the parametric SDF family is misspecified and $\delta_\star > 0$, the sample conditional HJ-distance $\hat{\delta}_T$ is such that $\sqrt{T}\left(\hat{\delta}_T^2 - \delta_\star^2\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_\phi^2\right)$ as $T \to \infty$, for the long-run variance $\sigma_\phi^2 = \sum_{j=-\infty}^{\infty} \mathrm{Cov}\left[\phi(W_t; \theta_\star), \phi\left(W_{t-j}; \theta_\star\right)\right]$ of the stochastic process $\{\phi(W_{t+1}; \theta_\star)\}$

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

defined by means of the scalar function $\phi(W_{t+1};\theta) := 2 \cdot \mathbf{1}(X_t)e(X_t;\theta)' \ \Omega(X_t)h(Y_{t+1}, R_{t+1};\theta) - \mathbf{1}(X_t)\left[e(X_t;\theta)'\Omega(X_t)R_{t+1}\right]^2$, for any $\theta \in \Theta$, and the vector $W_{t+1} := \left[Y'_{t+1}\, R'_{t+1}\, X'_t\right]'$.

Proof. See Appendix D.2. □

Proposition 5 shows that the square sample conditional HJ-distance is a consistent and asymptotically normal estimator of $\delta^2_\star$. The convergence rate is the parametric rate $\sqrt{T}$ and differs from the convergence rate $Tb_T^{L/2}$ under correct model specification given in Proposition 4. The asymptotic variance σ^2_ϕ accounts for uncertainty in estimation of the squared conditional HJ-distance, which is due to (i) kernel smoothing of the conditional moment vector $h(Y_{t+1}, R_{t+1}, \theta_\star)$, (ii) kernel smoothing of the squared returns, and (iii) sample averaging of the quadratic form of the smoothed conditional moment vector. The proof of Proposition 5 shows that estimation of the pseudo-true parameter $\theta_\star$ does not have an impact asymptotically. In fact, the derivative of the criterion at the optimum $\theta_\star$ is equal to zero.

It is interesting to notice that the large sample distribution of $\hat{\delta}_T$ does not depend on the Jacobian matrix of the conditional pricing error $\mathrm{E}\left[\nabla_{\theta'}h(Y_{t+1}, R_{t+1};\theta_\star)|X_t\right] = \nabla_{\theta'}e(X_t;\theta_\star)$. The approximate failure of a full-rank condition for this matrix uniformly in state X_t often observed in consumption-based asset pricing models leads to the problem of weak identification (see, e.g., Stock and Wright, 2000, and Kleibergen and Zhan, 2016). Weak identification may prevent the consistency of estimator $\hat{\theta}_T$ and impact on the asymptotic distribution of $\hat{\delta}_T$. We conjecture that weak identification issues are less acute for estimating the (un)conditional HJ-distance than for estimating the SDF parameter. The analysis of this interesting question is beyond the scope of this article.

A consistent estimator $\hat{\sigma}^2_\phi$ of the asymptotic variance σ^2_ϕ 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 $\phi(W_t;\theta_\star)$ (see, e.g., Andrews and Monahan, 1992, and Newey and West, 1994). Since $\delta^2_\star$ is nonnegative, a confidence interval for it can be obtained by first getting the asymptotic distribution of $\log\left(\hat{\delta}^2_\star\right)$ by the delta method with the logarithmic transformation, and then inverting the logarithmic transformation. The lower and upper bounds of a confidence interval for $\delta^2_\star$ at the asymptotic confidence level α obtained in this way are $\hat{\delta}^2_T\exp\left(c_{\alpha/2}\hat{\sigma}_\phi/\left(\sqrt{T}\hat{\delta}^2_T\right)\right)$ and $\hat{\delta}^2_T\exp\left(c_{1-\alpha/2}\hat{\sigma}_\phi/\left(\sqrt{T}\hat{\delta}^2_T\right)\right)$, where the $(\alpha/2)$- and $(1-\alpha/2)$-quantiles of the standard normal distribution are denoted by $c_{\alpha/2}$ and $c_{1-\alpha/2}$.

3 Model Selection Using the Conditional HJ-Distance

In this section, we describe a model selection procedure for the two competing parametric SDF families $\mathcal{F} = \{m(\cdot;\theta);\theta \in \Theta\}$ and $\tilde{\mathcal{F}} = \{\tilde{m}(\cdot;\tilde{\theta});\tilde{\theta} \in \tilde{\Theta}\}$ both possibly misspecified. As before, every quantity that refers to the second family is denoted as for the first family but with a superscript tilde. The model selection procedure is based on testing the null simple hypothesis $\mathcal{H}_0$ of equal conditional HJ-distances against the alternative composite hypothesis $\mathcal{H}_A$ that the parametric SDF family $\mathcal{F}$ is preferred over $\tilde{\mathcal{F}}$ in terms of the trimmed conditional HJ-distance: $\mathcal{H}_0 : \delta_\star = \tilde{\delta}_\star$ versus $\mathcal{H}_A : \delta_\star < \tilde{\delta}_\star$. 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 asymptotic distribution of $\hat{\delta}_T^2 - \hat{\tilde{\delta}}_T^2$ under the null hypothesis is based on the long-run variance for the process of the difference $\phi(W_t;\theta_\star) - \tilde{\phi}(W_t;\tilde{\theta}_\star)$, where ϕ and $\tilde{\phi}$ are defined as in Proposition 5 for families $\mathcal{F}$ and $\tilde{\mathcal{F}}$. We first consider the case when this process is degenerate and then, under the assumption that this process is not degenerate, we give the asymptotic distribution of $\hat{\delta}_T^2 - \hat{\tilde{\delta}}_T^2$.

The stochastic process $\left\{\phi(W_t;\theta_\star) - \tilde{\phi}(W_t;\tilde{\theta}_\star)\right\}$ degenerates to the zero process in two cases. First, this is the case if both models are correctly specified and they have distinct SDF's, that is under the sub-hypothesis $\mathcal{H}_0^1 := \left\{m(\cdot;\theta_\star) \neq \tilde{m}(\cdot;\tilde{\theta}_\star), \delta_\star = \tilde{\delta}_\star = 0\right\}$. This case can occur for correctly specified models for an incomplete market. Then, the conditional pricing errors at the corresponding true parameter values are zero almost everywhere over the set $\mathcal{X}_\star$, so that $\phi(\cdot;\theta_\star) = \tilde{\phi}(\cdot;\tilde{\theta}_\star) = 0$. Second, the stochastic process $\left\{\phi(W_t;\theta_\star) - \tilde{\phi}(W_t;\tilde{\theta}_\star)\right\}$ degenerates to 0 when the SDF's valued at the pseudo-true SDF parameter vectors coincide, that is, under the sub-hypothesis $\mathcal{H}_0^2 := \left\{m(\cdot;\theta_\star) = \tilde{m}(\cdot;\tilde{\theta}_\star)\right\}$. In this second case we also have that $\phi(\cdot;\theta_\star) = \tilde{\phi}(\cdot;\tilde{\theta}_\star)$.

The stochastic process $\left\{\phi(W_t;\theta_\star) - \tilde{\phi}(W_t;\tilde{\theta}_\star)\right\}$ is not degenerate under the sub-hypothesis $\mathcal{H}_0^3 := \left\{m(\cdot;\theta_\star) \neq \tilde{m}(\cdot;\tilde{\theta}_\star), \delta_\star = \tilde{\delta}_\star > 0\right\}$. Proposition 6 provides the asymptotic distribution of the statistic $\hat{\delta}_T^2 - \hat{\tilde{\delta}}_T^2$ in that case.

Proposition 6. If the two competing parametric SDF families $\mathcal{F}$ and $\tilde{\mathcal{F}}$ are misspecified with equal conditional HJ-distances but distinct SDF's, that is, hypothesis $\mathcal{H}_0^3$ holds, as $T \to \infty$ we have that $\sqrt{T}\left(\hat{\delta}_T^2 - \hat{\tilde{\delta}}_T^2\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_\Delta^2)$, where σ_Δ^2 is the long-run variance of process $\left\{\phi(W_t;\theta_\star) - \tilde{\phi}(W_t;\tilde{\theta}_\star)\right\}$.

Proof. See Appendix D.3. □

The three sub-hypotheses $\mathcal{H}_0^i$, for $i = 1, 2, 3$, are mutually exclusive and we can write the null hypothesis $\mathcal{H}_0$ as their union: $\mathcal{H}_0 = \mathcal{H}_0^1 \cup \mathcal{H}_0^2 \cup \mathcal{H}_0^3$. Degeneracy of the test statistic on a subset of the null hypothesis is a well-known problem in the model selection literature. In a classical likelihood framework, Vuong (1989) proposes a multiple stage testing approach. In the following subsections, we adapt the multiple stage testing approach of Gospodinov, Kan, and Robotti (2013) to our conditional setting, and consider separately the three cases of nested, strictly nonnested and overlapping competing SDF asset pricing models.

3.1 Nested Models

Let us assume in this section 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 $\mathcal{F}$, that is, $\tilde{\mathcal{F}} \subset \mathcal{F}$. Being nested, the models cannot have the same conditional HJ-distance while having distinct

SDF's, and therefore hypotheses $\mathcal{H}_0^1$ and $\mathcal{H}_0^3$ cannot hold. Therefore, in this case we have $\mathcal{H}_0 = \mathcal{H}_0^2$, and we should test if the SDF's valued at the pseudo-true SDF parameter vectors coincide or not. The SDF's coincide, that is $m(\cdot;\theta_\star) = \tilde{m}(\cdot;\tilde{\theta}_\star)$, if r well-chosen transformations of the parameter $\theta_\star$ of the nesting family are zero, where $r = p - \tilde{p}$ is the difference in the parameter dimensions of the two SDF families. By this consideration, we can express the sub-hypothesis as $\mathcal{H}_0^2 = \{\psi(\theta_\star) = 0_r\}$ for an r-dimensional vector of restrictions ψ. Thus, we can test hypothesis $\mathcal{H}_0$ by means of the Wald test statistic $\xi_T^W := T\psi(\hat{\theta}_T)'\left(\nabla_{\theta'}\psi(\hat{\theta}_T)\hat{\Sigma}_\star\nabla_\theta\psi(\hat{\theta}_T)'\right)^{-1}\psi(\hat{\theta}_T)$, where the $(p \times r)$-dimensional matrix $\nabla_\theta\psi(\theta)'$ is the Jacobian matrix of the restriction vector, and $\hat{\Sigma}_\star$ is a consistent estimator of the asymptotic variance $\Sigma_\star$ of estimator $\hat{\theta}_T$ defined in Lemma 2 in Appendix C.2.

Proposition 7. If the family $\tilde{\mathcal{F}}$ is nested into family $\mathcal{F}$, under hypothesis $\mathcal{H}_0$ as $T \to \infty$ we have $\xi_T^W \xrightarrow{\mathcal{D}} \chi_r^2$.

Proof. See Appendix D.4. □

The null hypothesis is rejected when ξ_T^W is above the critical value of the χ_r^2 distribution for the selected asymptotic significance level. Alternatively, we can test the parametric hypothesis $\psi(\theta_\star) = 0_r$ by means of the Lagrange multiplier test statistic or the quasi-likelihood ratio test statistic (see Newey and McFadden, 1994).

3.2 Nonnested Models

Let us assume in this section that the SDF families are nonnested, that is, $\mathcal{F} \cap \tilde{\mathcal{F}} = \emptyset$. In this case, the SDF's valued at the pseudo-true SDF parameter vectors cannot coincide and $\mathcal{H}_0 = \mathcal{H}_0^1 \cup \mathcal{H}_0^3$. We propose a two-stage testing procedure. First, we test $\mathcal{H}_0^1$ by means of the statistic $\hat{\delta}_T^2 + \hat{\tilde{\delta}}_T^2$, whose large sample distribution in given in Proposition 8 below. Second, if $\mathcal{H}_0^1$ is rejected, we test the sub-hypothesis $\mathcal{H}_0^3$ using statistic $\sqrt{T}(\hat{\delta}_T^2 - \hat{\tilde{\delta}}_T^2)$, whose large sample distribution is given is Proposition 6. We neglect the effect of pretesting on the level of the test.

Proposition 8. If the two competing parametric SDF families $\mathcal{F}$ and $\tilde{\mathcal{F}}$ have both null conditional HJ-distance, the asymptotic behavior of the statistic $\hat{\delta}_T^2 + \hat{\tilde{\delta}}_T^2$ is, as $T \to \infty$, $Tb_T^{L/2}\left(\hat{\delta}_T^2 + \hat{\tilde{\delta}}_T^2 - a_T - \tilde{a}_T\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_{00}^2)$, for the asymptotic variance $\sigma_{00}^2 := \sigma_0^2 + \tilde{\sigma}_0^2 + 4\left(\int_{\mathbb{R}^L}\mathcal{K}(u)^2 du\right)\int_{\mathcal{X}_\star}\mathrm{Tr}\left[\Gamma_0(x)\Omega(x)\Gamma_0(x)'\Omega(x)\right]dx$, where $\Gamma_0(x) := \mathrm{Cov}\left[\tilde{h}(Y_{t+1}, R_{t+1};\tilde{\theta}_0), h(Y_{t+1}, R_{t+1};\theta_0)|X_t = x\right]$ and the variances σ_0^2, $\tilde{\sigma}_0^2$ and the centering terms a_T, $\tilde{a}_T$ are defined as in Proposition 4 for families $\mathcal{F}$, $\tilde{\mathcal{F}}$.

Proof. See Appendix D.5. □

The rejection region for the test of $\mathcal{H}_0^1$ at the asymptotic significance level α is:

$$\hat{\delta}_T^2 + \hat{\tilde{\delta}}_T^2 \geq a_T + \tilde{a}_T + c_{1-\alpha}\frac{\hat{\sigma}_{00}}{\sqrt{Tb_T^{L/2}}}, \tag{32}$$

and the rejection region for the test of $\mathcal{H}_0^3$ in favor of $\mathcal{H}_A$ at the asymptotic significance level α is:

$$\hat{\delta}_T^2 - \hat{\tilde{\delta}}_T^2 \geq c_{1-\alpha} \frac{\hat{\sigma}_\Delta}{\sqrt{T}}, \tag{33}$$

where $\hat{\sigma}_{00}^2$ and $\hat{\sigma}_\Delta^2$ are consistent estimators of variances σ_{00}^2 and σ_Δ^2. While both steps of the testing procedure have asymptotic size α, the overall test does not have asymptotic size α because of the sequential nature. Mimicking the arguments in Vuong (1989) in the likelihood framework, it is easy to show that the test is asymptotically conservative.

Proposition 9. The sequential test defined by rejection regions (32) and (33) is asymptotically conservative, that is, $\limsup_{T\to\infty} \mathbb{P}[\text{reject}\, \mathcal{H}_0 | \mathcal{H}_0] \leq \alpha$.

Proof. See Appendix D.6. □

Proposition 9 shows that we can upper bound asymptotically the size of the test, for any given DGP under the null. Leeb and Poetscher (2005) among others point out that the statistical properties of sequential/pretest procedures might suffer from lack of uniformity, that is, for any given sample size there may exist DGP's under the null with large size distortions. In a likelihood framework, Shi (2015) and Schennach and Wilhelm (2017) introduce modifications of the model selection test of Vuong (1989) that achieve uniformity.[7] We could build on these results to get uniform model selection tests in our moment-based framework. We leave this interesting extension for future research.

3.3 Overlapping Models

Let us assume in this section that the SDF families are overlapping, which means that they contain common SDF's and are not nested. In this case, any of the three sub-hypotheses $\mathcal{H}_{0,1}$, $\mathcal{H}_{0,2}$ and $\mathcal{H}_{0,3}$ can hold, and we cannot simplify the null hypothesis $\mathcal{H}_0$. In this case, we first test individually the sub-hypotheses $\mathcal{H}_{0,1}$ and $\mathcal{H}_{0,2}$, and if we reject them, we test the sub-hypothesis $\mathcal{H}_{0,3}$.

The sub-hypothesis $\mathcal{H}_0^1$ can be tested as in Subsection 3.2 with the rejection region (32). The test of sub-hypothesis $\mathcal{H}_0^2$ can be based on the observation that the SDF's valued at the pseudo-true SDF parameter vectors $\theta_\star$ and $\tilde{\theta}_\star$ coincide if r transformations of $\theta_\star$ and $\tilde{r}$ transformations of $\tilde{\theta}_\star$ are zero, for some dimensions r and $\tilde{r}$. We can indeed consider the $(r+\tilde{r})$-dimensional vector $\overline{\psi}(\overline{\theta}) := \left[\psi(\theta)' \, \tilde{\psi}(\tilde{\theta})'\right]'$ collecting the r- and $\tilde{r}$-dimensional vectors ψ and $\tilde{\psi}$ of restrictions for $\overline{\theta} = \left[\theta' \, \tilde{\theta}'\right]'$, and write $\mathcal{H}_0^2 = \left\{\overline{\psi}\left(\overline{\theta}_\star\right) = 0_{(r+\tilde{r})}\right\}$. We test hypothesis $\mathcal{H}_0^2$ by using the Wald test statistic $\overline{\xi}_T^W := T\overline{\psi}\left(\hat{\overline{\theta}}_T\right)'$ $\left(\nabla_{\overline{\theta}'}\overline{\psi}\left(\hat{\overline{\theta}}_T\right)\hat{\overline{\Sigma}}_\star \nabla_{\overline{\theta}}\overline{\psi}\left(\hat{\overline{\theta}}_T\right)'\right)^{-1} \overline{\psi}\left(\hat{\overline{\theta}}_T\right)$, where $\hat{\overline{\Sigma}}_\star$ is a consistent estimator of the asymptotic variance of estimator $\hat{\overline{\theta}}_T := \left[\hat{\theta}_T' \, \hat{\tilde{\theta}}_T'\right]'$.

Proposition 10. If the SDF families $\mathcal{F}$ and $\tilde{\mathcal{F}}$ are overlapping and $\mathcal{H}_0^2$ holds with $\overline{\psi}\left(\overline{\theta}_\star\right) = 0_{(r+\tilde{r})}$, then as $T \to \infty$ we have $\overline{\xi}_T^W \xrightarrow{\mathcal{D}} \chi^2_{r+\tilde{r}}$.

7 We are grateful to Daniel Wilhelm for pointing these results to us.

Proof. See Appendix D.7. □

If the sub-hypotheses $\mathcal{H}_{0,1}$ and $\mathcal{H}_{0,2}$ are both rejected, we test the sub-hypothesis $\mathcal{H}_{0,3}$ using the rejection rule in Inequality (33). Similarly as in Subsection 3.2, we can show that this sequential testing procedure is conservative.

4 Empirical Comparison of Parametric SDF Specifications for U.S. Equity Portfolios and Short-Term T-Bills

In this section, we show that thirteen parametric models for U.S. equity portfolios and short-term T-bills largely adopted in the financial literature, whose hypothesis of correct model specification is rejected at the 5% significance level on the basis of the Hansen's J statistic, are differently ranked on the basis of the unconditional HJ-distance with different instrument matrices. We then discuss their ranking on the basis of the conditional HJ-distance. We introduce the competing models in Section 4.1, and the data used for analyzing them in Section 4.2. In Section 4.3, we describe the different model rankings based on the unconditional HJ-distance for distinct instrument matrices, and the model ranking based on the conditional HJ-distance. We also discuss several robustness checks.

4.1 Competing Asset Pricing Models

We consider two groups of competing asset pricing models. The first group includes the linear specification of the CAPM and some of its extensions to account for some cross-sectional stylized facts in the U.S. equity market. These beta pricing models admit the following common expression of the SDF:

$$m_i(Y_{t+1};\theta) = \theta_1 + \theta_2 MKT_{t+1} + \overline{\theta}' F_{t+1}^{[i]}, \qquad i = \mathsf{CAPM}, \mathsf{FF3}, \mathsf{FF5}, \mathsf{FFM}, \mathsf{FFL}, \mathsf{NM},$$

for the parameter vector $\theta = \left[\theta_1\ \theta_2\ \overline{\theta}'\right]'$ and vector $Y_t = \left[MKT_t\ F_t^{[i]\prime}\right]'$, where MKT_t is the market return factor and the vector $F_t^{[i]}$ collects additional potentially priced risk factors. The vectors $\overline{\theta}$ and $F_t^{[i]}$ have the same dimension, which depends on the model specification. Vector $F_t^{[\mathsf{CAPM}]}$ is null. Vector $F_t^{[\mathsf{FF3}]}$ is composed by size and value factors (Fama and French, 1993, 1998), and it is augmented by either the momentum factor to create vector $F_t^{[\mathsf{FFM}]}$ (Carhart, 1997) or the liquidity factor to create vector $F_t^{[\mathsf{FFL}]}$ (Pastor and Stambaugh, 2003). Vector $F_t^{[\mathsf{FF5}]}$ includes size, value, profitability and investment factors (Fama and French, 2015).[8] Vector $F_t^{[\mathsf{NM}]}$ includes the industry-adjusted value factor, the momentum factor, and the profitability factor (Novy-Marx, 2013).

The second group of models we consider includes distinct versions of the CCAPM. The CRRA and EZ specifications correspond to structural asset pricing models based on the preferences of a representative economic agent. The SDF is implied by the intertemporal behavior of this agent who invests in the tradable assets, saves and consumes. The CRRA specification coincides with the CCAPM with time-separable preferences and CRRA utility for the representative agent (see, e.g., Hansen and Singleton, 1982), while the EZ specification

8 See also Hou, Xue, and Zhang (2014) for a linear specification of the SDF that includes market, size, profitability, and investment factors.

is implied by the time-nonseparable Epstein and Zin (1989, 1991) preferences. The two SDF specifications are as follows:

$$m_{\mathsf{CRRA}}(Y_{t+1};\theta) = \theta_1 (C_{t+1}/C_t)^{-\theta_2}, \quad m_{\mathsf{EZ}}(Y_{t+1};\theta) = \theta_1^{\theta_3} (C_{t+1}/C_t)^{-\theta_2\theta_3} (1+MKT_{t+1})^{\theta_3 - 1},$$

where C_t is the personal consumption at date t of nondurables and services, $Y_{t+1} = C_{t+1}/C_t$ and $\theta = [\theta_1\, \theta_2]'$ in the CRRA specification, and $Y_{t+1} = [C_{t+1}/C_t\, MKT_{t+1}]'$ and $\theta = [\theta_1\, \theta_2\, \theta_3]'$ in the EZ specification. The EZ specification 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 $\theta_3 = 1$. Stock and Wright (2000) consider the problem of weak instruments for GMM estimators of SDF asset pricing models including these specifications. Often, in empirical studies of structural asset pricing models the SDF is linearized, so that we include in our empirical comparison the linearizations of the CRRA and EZ specifications for small values of logarithmic consumption growth and return on the optimal portfolio. The LCRRA and the LEZ specification correspond to the unconditional (or static) version of the CCAPM and the Market and Consumption-based CAPM, with the following SDF specifications:[9]

$$m_{\mathsf{LCRRA}}(Y_{t+1};\theta) = \theta_1 + \theta_2 \log\,(C_{t+1}/C_t), \quad m_{\mathsf{LEZ}}(Y_{t+1};\theta) = \theta_1 + \theta_2 \log\,(C_{t+1}/C_t) + \theta_3 MKT_{t+1}.$$

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

$$m_i(Y_{t+1};\theta) = \left(\theta_1 + \theta_3 u_t^{[i]}\right) + \left(\theta_2 + \theta_4 u_t^{[i]}\right) \log\,(C_{t+1}/C_t), \qquad i = \mathsf{DEF}, \mathsf{CAY}, \mathsf{YC},$$

where $u_t^{[i]}$ is a conditioning state variable, $Y_{t+1} = \left[C_{t+1}/C_t\, u_t^{[i]}\right]'$ and $\theta = [\theta_1\, \theta_2\, \theta_3\, \theta_4]'$ (see Nagel and Singleton, 2011, for a GMM estimation of these models). The conditioning state variable $u_t^{[i]}$ 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). Each of these SDF's is a linear function of logarithmic consumption growth $\log\,(C_{t+1}/C_t)$, having $\theta_1 + \theta_3 u_t^{[i]}$ and $\theta_2 + \theta_4 u_t^{[i]}$ as time-varying intercept and slope, respectively. As a consequence, these SDF specifications allow for time-varying risk premia. The three specifications reduce to the LCRRA specification for $\theta_3 = \theta_4 = 0$.

Table 1 indicates the relationships between the competing asset pricing models, studied using the data described in following Section 4.2. For example, the specifications CAPM and CRRA overlap because they both include the constant SDF. Also the specifications FF3 and FF5 overlap, because while the latter theoretically nests the former, we use distinct proxies for the size factor.

9 The parameters in the LCRRA specification are renamed after linearizing the SDF in the CRRA specification, and similarly for LEZ.

Table 1 Competing asset pricing models

							$\tilde{\mathcal{F}}$						
	Specification	CAY	CRRA	DEF	EZ	FF3	FF5	FFL	FFM	LCRRA	LEZ	NM	YC
$\mathcal{F}$	CAPM	O	O	O	O	N	N	N	N	O	N	N	O
	CAY		O	O	O	O	O	O	O	I	O	O	O
	CRRA			O	N	O	O	O	O	O	O	O	O
	DEF				O	O	O	O	O	I	O	O	O
	EZ					O	O	O	O	O	O	O	O
	FF3						O	N	N	O	O	O	O
	FF5							O	O	O	O	O	O
	FFL								O	O	O	O	O
	FFM									O	O	O	O
	LCRRA										N	O	N
	LEZ											O	O
	NM												O

Notes: This table reports the relationships between the competing asset pricing models. The letter N means that $\mathcal{F} \subseteq \tilde{\mathcal{F}}$, that is the parametric SDF family $\mathcal{F}$ (indicated along the rows) is nested in the parametric SDF family $\tilde{\mathcal{F}}$ (indicated along the columns). The letter I means that $\tilde{\mathcal{F}} \subseteq \mathcal{F}$. The letter O means that the pair of parametric SDF families $\mathcal{F}$ and $\tilde{\mathcal{F}}$ overlap, that is, $\mathcal{F} \cap \tilde{\mathcal{F}} \neq \emptyset$. Each parametric SDF family contains the constant SDF.

4.2 Data

Our dataset consists of monthly data from July 1963 to July 2016. To construct the time series of monthly observations from the test assets, we consider the 1-month T-bill and the six value-weighted FF cum-dividend research portfolios of publicly traded U.S. equities formed on size and book-to-market. We deflate the gross returns on these assets by the monthly Consumer Price Index for All Urban Consumers rate to obtain ex-post real returns. We denote the real gross returns on the T-bill by $R_{f,t}$ and the inflation- and dividends-adjusted proxy for the gross excess return on a growth (G), neutral (N) or value (V) stock of a company with small (S) or big (B) market capitalization by SG_t, SN_t, SV_t, BG_t, BN_t and BV_t, as each name indicates. As stressed in Section IV of 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). We take the real gross returns on the T-bill and the real excess gross returns on the FF portfolios obtained in this way as returns on the $N = 7$ test assets. Since we analyze excess gross returns on the equity portfolios, instead of gross returns, we adapt Equation (1) in the following way. The no-arbitrage restriction for the T-bill is $R_{f,t}^{-1} = \mathrm{E}[M_{t,t+1}|X_t = x]$ for any $x \in \mathcal{X}$, as this asset serves as the unique test asset of our empirical study with unit price payoff identifying the SDF scale. Differently, the no-arbitrage restriction for the small growth FF portfolio is $\mathrm{E}[M_{t,t+1}SG_{t+1}|X_t = x] = 0$, for any $x \in \mathcal{X}$. The remaining no-arbitrage restrictions are of the same type.

To create the time series of the risk factors in the structural 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 (BEA), and the estimates of the total U.S. population provided by the U.S. Census Bureau.[10] We consider net real growth rates of per capita total consumption from the NIPA chain-type price indexes. These rates are taken as proxies for the personal consumption of nondurables and services.[11] We consider the monthly Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity as corporate bond spread. We derive monthly estimates of the transitory component of financial wealth by linear interpolation through levels at following end-of-quarters of the Lettau and Ludvigson (2001) quarterly proxies for the consumption to wealth ratio.[12] The proxy for monthly total labor income used to construct the labor income-to-consumption ratio is obtained from the interpolation of the sum of wages and salaries, transfer payments and other labor income minus personal contributions for social insurance and taxes (see, e.g., Lettau and Ludvigson, 2001, and Santos and Veronesi, 2006). The data, seasonally adjusted at annual rates, are obtained from the BEA.

We consider the monthly FF proxy for the market factor MKT_t. We consider the monthly FF proxies SMB_t (Small-Minus-Big) and HML_t (High-Minus-Low) for size and value factors constructed from six value-weighted portfolios formed on size and book-to-market in the FF3 specification. Together with the FF proxies SMB_t and HML_t we use 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 proxy HML_t for the value factor and the FF proxies $SMB5_t$, RMW_t (Robust-Minus-Weak) and CMA_t (Conservative-Minus-Aggressive) for the size, profitability and investment factors constructed from eighteen value-weighted portfolios formed on (i) size and book-to-market, (ii) size and operating profitability, and (iii) size and investment. We consider the Novy-Marx proxies HML_t^*, UMD_t^* and PMU_t^* for the industry-adjusted value, momentum and profitability factor in the NM specification.[13] All the factors are expressed in percentage.

Table 2 reports the length of each time series we use, and the sample minimum, maximum, mean, standard deviation, skewness, and excess kurtosis of each variable.

4.3 Sample Unconditional and Conditional HJ-Distances

We present in this section the rankings of the models on the basis of the unconditional and conditional HJ-distances, taking the demeaned version of either the consumption to wealth

10 The full description of the personal consumption expenditures is available at http://www.bea.gov/. All the data are retrieved from the FRED database of the St Louis Fed and the Census Bureau of the U.S. Department of Commerce.

11 Considering the chain-type quantity of the total expenditures for services and nondurable goods allows to accommodate the changing relative importance of the two kind of expenditures. For example, in nominal terms, the ratio of expenditures for services to expenditures for services and nondurable goods varies from 0.53 in 1959 to 0.75 in 2014.

12 The time series for CAY_t ends in July 2015.

13 All the data on the test assets and the market factor have been retrieved from the official websites of Kenneth French, Robert Stambaugh, and Robert Novy-Marx. The FFL specification is estimated and tested from January 1968 until December 2014, and the SDF specification NM until December 2012.

Table 2 Basic time series statistics

	Size	Minimum	Maximum	Average	SD	Skewness	Excess Kurtosis
SG_t	637	−0.33	0.27	0.00	0.07	−0.37	1.77
SN_t	637	−0.29	0.26	0.01	0.05	−0.49	2.86
SV_t	637	−0.28	0.30	0.01	0.06	−0.39	3.39
BG_t	637	−0.24	0.21	0.00	0.05	−0.34	1.89
BN_t	637	−0.21	0.16	0.01	0.04	−0.38	2.22
BV_t	637	−0.23	0.21	0.01	0.05	−0.43	2.49
$R_{f,t}$	637	0.99	1.02	1.00	0.00	0.36	2.91
$MKTt$	637	−0.23	0.16	0.01	0.04	−0.52	1.95
SMB_t	637	−0.17	0.22	0.00	0.03	0.51	5.87
HML_t	637	−0.11	0.13	0.00	0.03	0.04	2.26
UMD	637	−0.35	0.18	0.01	0.04	−1.36	10.70
L_t	576	−0.11	0.22	0.01	0.03	0.40	2.64
$SMB5_t$	637	−0.15	0.19	0.00	0.03	0.38	3.54
$HML5_t$	637	−0.11	0.13	0.00	0.03	0.04	2.26
RMW_t	637	−0.19	0.14	0.00	0.02	−0.36	13.30
CMA_t	637	−0.07	0.10	0.00	0.02	0.28	1.67
HML^*	594	−0.05	0.07	0.00	0.01	0.46	2.30
UMD^*	594	−0.23	0.12	0.01	0.03	−1.95	13.50
PMU^*	594	−0.05	0.07	0.00	0.01	0.45	3.46
C_t/C_{t-1}	637	0.99	1.01	1.00	0.00	0.50	2.74
DEF_t	637	0.29	6.01	2.06	0.81	0.85	2.50
CAY_t	625	−4.73	4.38	−0.24	1.98	−0.18	−0.77
YC_t	592	0.66	0.87	0.75	0.07	0.46	−1.33

Notes: This table reports sample size, time series minimum, maximum, average, standard deviation (SD), skewness, and excess kurtosis for the used monthly variables. Variable $R_{f,t}$ is the real gross returns on the T-bill. Variables SG_t, SN_t, SV_t, BG_t, BN_t, and BV_t are the inflation- and dividends-adjusted proxies for the excess gross return on a growth (G), neutral (N) or value (V) stock of a company with small (S) or big (B) market capitalization, as each name indicates. Variables MKT_t, L_t and CMA_t are proxies for market, liquidity and investment factors, respectively. Variables SMB_t and $SMB5_t$ are proxies for the size factor. Variables HML_t and HML_t^* are proxies for the value factor. Variables RMW_t and PMU_t^* are proxies for the profitability factor. Variables UMD_t and UMD_t^* are proxies for the momentum factor. All the factors are expressed in percentage. Variable C_t/C_{t-1} is the proxy for the growth of personal consumption of nondurables and services. Variable DEF_t is the corporate bond spread, and variables CAY_t and YC_t are the proxies for the consumption to wealth ratio and the labor income-to-consumption ratio, respectively.

ratio or the corporate bond spread as conditioning variable X_t, since these variables are often regarded as possible candidates to the role of return predictors (see, e.g., Goyal and Welch, 2008, and Nagel and Singleton, 2011).[14] We describe in Section 4.3.1 the implementation of the model specification test procedure, and the computation of the sample

14 The computation of the unconditional and conditional HJ-distances requires the numerical inversion of estimates of the matrices Ω_Z^{-1} and $\Omega^{-1}(x)$ for any $x \in \mathcal{X}$. The likelihood of incurring in numerical issues increases with the condition numbers of these matrices. With our data, the condition numbers are lower when demeaned consumption to wealth ratio and demeaned corporate bond spread are taken as conditioning variables instead of their raw counterparts.

unconditional HJ-distance. We illustrate the computation of the conditional HJ-distance in Section 4.3.2. In Section 4.3.3, we highlight the differences between the model rankings based on the two HJ-distances, and discuss the stability of the sample distance minimizers across different choices of the conditioning variable X_t. Finally, in Section 4.3.4 we explain how our results are robust with respect to the selection of the data and the numerical inversion of the second-moments matrices of assets' gross returns.

4.3.1 Model specification test and sample unconditional HJ-distance

For every competing asset pricing model, we test the hypothesis of correct model specification by GMM with optimal weighting matrix. We base the test on $7 \times 2 = 14$ orthogonality conditions, taking either $I_7 \otimes [1\, X_t]'$ or $I_7 \otimes \left[1\, X_t^2\right]'$ as instrument matrix. These conditions correspond to Equation (3) and represent null unconditional pricing errors on seven test assets and seven additional managed portfolios created by long and short positions in each of the test assets. In particular, the time-varying weights of these additional portfolios are determined by the value of the function of X_t included in the instrument matrix (i.e., either X_t or X_t^2).

The hypothesis of correct specification of any asset pricing model is rejected at the 5% significance level for at least one instrument matrix. We report in Tables 3 and 4 the sample Hansen's J statistic $\hat{J}_Z$, for the demeaned consumption to wealth ratio and demeaned corporate bond spread as variable X_t, respectively. The adopted instrument matrix is indicated by the text on the left part of each panel. We also report the degrees of freedom of the asymptotic distributional limit of $\hat{J}_Z$, and the probability that the value of the statistic is smaller than this limit. The estimation is performed by optimal GMM implemented in its iterative form, with value 10^{-6} as arbitrary threshold on the improvement of the criterion for the numerical convergence.[15] For the two instrument matrices that are built on the consumption to wealth ratio and its square, the two specifications FF5 and FFM are not rejected at the 5% significance level. However, this hypothesis is rejected when including other nonlinear functions of the demeaned consumption to wealth ratio in the instrument matrix, and not only when the corporate bond spread is included as for example in the results of the upper panel of Figure 4.[16] Therefore, we treat all the competing models as potentially misspecified when either the consumption to wealth ratio or the corporate bond spread is taken as conditioning variable X_t, and proceed to rank them on the basis of the unconditional HJ-distance.

The computation of the unconditional HJ-distance corresponds to a one-step GMM with nonoptimal weighting matrix. We keep the same implementation features, such as the minimization algorithm, the starting points for the minimization and the numerical convergence criterion, as for the model specification test procedure. We report in Tables 5 and 6 the value of the squared sample unconditional HJ-distance and its minimizer $\hat{\theta}_Z$. For many

15 See, for example, Sections 2.4 and 3.6 in Hall (2005) for a discussion on the gain in finite-sample efficiency obtained by the iterated GMM.

16 For example, the hypothesis of correct SDF specification for the FF5 specification is rejected at the 5% significance level including in the instrument matrix the absolute value of the demeaned consumption to wealth ratio, its square and cube. We reject this hypothesis for the FFM specification at the same significance level when we include the demeaned consumption to wealth ratio and its cube in the instrument matrix.

Table 3 Unconditional statistics

	Specification	$\hat{J}_Z$	df	$\mathbb{P}[\chi^2_{df} > \hat{J}_Z]$	$\hat{\theta}^{[GMM]}_{1,Z}$	$\hat{\theta}^{[GMM]}_{2,Z}$	$\hat{\theta}^{[GMM]}_{3,Z}$	$\hat{\theta}^{[GMM]}_{4,Z}$	$\hat{\theta}^{[GMM]}_{5,Z}$	$\hat{\theta}^{[GMM]}_{6,Z}$
$Z(X_t) = I_7 \otimes [1\, X_t]'$	CAPM	61.71	12	0.00	1.03	−3.79				
					(0.01)	(0.89)				
	CAY	11.12	10	0.35	−0.11	581.00	−0.20	136.00		
					(0.30)	(160.00)	(0.10)	(55.00)		
	CRRA	65.90	12	0.00	0.99	−2.89				
					(0.00)	(0.77)				
	DEF	34.31	10	0.00	0.25	576.00	0.04	−138.00		
					(0.64)	(197.00)	(0.28)	(82.90)		
	EZ	55.23	11	0.00	0.99	−1.58	−2.23			
					(0.01)	(0.88)	(4.22)			
	FF3	45.02	10	0.00	1.04	−2.68	−0.74	−6.97		
					(0.02)	(0.99)	(1.19)	(1.56)		
	FF5	5.35	8	0.72	1.10	−3.41	−8.98	−12.70	−26.50	15.20
					(0.04)	(1.77)	(2.92)	(5.68)	(6.01)	(11.90)
	FFL	30.10	9	0.00	1.11	−3.80	−1.50	−7.42	−13.20	
					(0.04)	(1.30)	(1.65)	(1.72)	(5.92)	
	FFM	14.42	9	0.11	1.19	−7.13	−2.05	−12.60	−17.30	
					(0.07)	(1.92)	(1.90)	(3.07)	(5.95)	
	LCRRA	65.91	12	0.00	0.99	2.90				
					(0.00)	(0.77)				
	LEZ	57.01	11	0.00	1.03	−8.95	−2.64			
					(0.01)	(5.23)	(0.95)			
	NM	17.00	9	0.05	1.23	−5.83	−28.50	−14.70	−4.51	
					(0.07)	(1.61)	(5.27)	(8.39)	(11.00)	
	YC	6.28	10	0.79	−2.89	5030.00	3.26	−5640.00		
					(3.87)	(2340.00)	(5.07)	(2850.00)		
$Z(X_t) = I_7 \otimes [1\, X_t^2]'$	CAPM	68.60	12	0.00	1.00	−0.83				
					(0.01)	(1.11)				
	CAY	24.08	10	0.01	0.24	365.00	−0.29	67.40		
					(0.24)	(121.00)	(0.17)	(58.80)		
	CRRA	67.00	12	0.00	1.00	0.15				
					(0.05)	(0.05)				
	DEF	43.10	10	0.00	0.61	453.00	−0.06	−113.00		
					(0.50)	(196.00)	(0.23)	(90.10)		
	EZ	57.41	11	0.00	1.02	−1.18	11.20			
					(0.04)	(0.92)	(15.80)			
	FF3	59.42	10	0.00	1.02	−2.99	2.96	−3.13		
					(0.01)	(1.02)	(0.91)	(1.33)		
	FF5	11.63	8	0.17	1.12	−5.89	−3.58	−9.33	−25.50	0.92
					(0.04)	(1.65)	(2.29)	(5.60)	(5.59)	(10.90)
	FFL	29.90	9	0.00	1.12	−4.13	−1.28	−6.93	−14.20	
					(0.04)	(1.34)	(1.63)	(1.78)	(5.35)	
	FFM	11.60	9	0.24	1.19	−6.82	−2.21	−12.60	−18.80	
					(0.06)	(1.80)	(1.96)	(2.73)	(4.76)	
	LCRRA	69.91	12	0.00	1.00	−0.33				
					(0.11)	(0.07)				
	LEZ	65.82	11	0.00	1.00	1.91	−1.56			
					(0.16)	(3.32)	(0.85)			
	NM	12.50	9	0.19	1.31	−6.96	−33.80	−16.70	−16.20	
					(0.07)	(1.70)	(5.67)	(5.52)	(11.50)	
	YC	27.41	10	0.00	−2.98	2530.00	4.26	−2800.00		
					(3.73)	(2020.00)	(4.88)	(2540.00)		

Notes: This table reports the sample Hansen's J statistic $\hat{J}_Z$, degrees of freedom df of its asymptotic distributional limit, probability $\mathbb{P}[\chi^2_{df} > \hat{J}_Z]$ that $\hat{J}_Z$ is smaller than this limit, and estimate $\hat{\theta}^{[GMM]}_Z = [\hat{\theta}^{[GMM]}_{1,Z} \ldots \hat{\theta}^{[GMM]}_{p,Z}]'$ of the SDF parameter vector obtained by GMM with optimal weighting matrix, for all the competing SDF specifications. Demeaned consumption to wealth ratio is taken as variable X_t, and the instrument matrix is indicated by the text on the left. Standard errors are reported in parentheses below the parameter estimates.

Table 4 Unconditional statistics

	Specification	$\hat{J}_Z$	df	$\mathbb{P}[\chi^2_{df} > \hat{J}_Z]$	$\hat{\theta}_{1,Z}^{[GMM]}$	$\hat{\theta}_{2,Z}^{[GMM]}$	$\hat{\theta}_{3,Z}^{[GMM]}$	$\hat{\theta}_{4,Z}^{[GMM]}$	$\hat{\theta}_{5,Z}^{[GMM]}$	$\hat{\theta}_{6,Z}^{[GMM]}$
$Z(X_t) = I_7 \otimes [1\, X_t]'$	CAPM	89.20	12	0.00	1.01	1.85				
					(0.00)	(0.28)				
	CAY	43.09	10	0.00	0.89	65.90	−0.48	176.00		
					(0.13)	(68.20)	(0.18)	(61.80)		
	CRRA	75.80	12	0.00	1.00	0.32				
					(0.00)	(0.36)				
	DEF	63.11	10	0.00	0.54	166.00	0.11	−24.40		
					(0.26)	(122.00)	(0.07)	(42.40)		
	EZ	66.10	11	0.00	1.00	−1.37	−1.69			
					(0.01)	(0.92)	(2.34)			
	FF3	55.08	10	0.00	1.02	−1.77	−0.58	−6.47		
					(0.01)	(0.83)	(1.31)	(1.46)		
	FF5	37.80	8	0.00	1.06	−3.24	−3.96	−9.39	−8.97	5.28
					(0.02)	(1.24)	(2.01)	(3.88)	(3.76)	(7.53)
	FFL	26.63	9	0.00	1.12	−3.22	−1.15	−6.95	−15.00	
					(0.03)	(1.25)	(1.66)	(1.81)	(3.62)	
	FFM	42.12	9	0.00	1.07	−4.02	−2.52	−8.31	−2.29	
					(0.02)	(1.21)	(1.36)	(1.62)	(0.97)	
	LCRRA	75.81	12	0.00	1.00	−0.32				
					(0.00)	(0.36)				
	LEZ	67.32	11	0.00	1.03	−7.06	−2.49			
					(0.03)	(9.49)	(1.00)			
	NM	30.80	9	0.00	1.18	−4.78	−26.30	−8.20	−2.42	
					(0.06)	(1.47)	(4.86)	(3.08)	(8.52)	
	YC	34.30	10	0.00	2.04	3190.00	−2.32	−3720.00		
					(2.35)	(1750.00)	(3.23)	(2190.00)		
$Z(X_t) = I_7 \otimes [1\, X_t^2]'$	CAPM	63.80	12	0.00	1.00	−1.19				
					(0.00)	(0.44)				
	CAY	9.57	10	0.48	0.39	151.00	−0.40	−210.00		
					(0.23)	(82.40)	(0.18)	(120.00)		
	CRRA	69.14	12	0.00	1.00	0.27				
					(0.00)	(0.10)				
	DEF	29.34	10	0.00	−0.32	618.00	0.32	−156.00		
					(0.45)	(200.00)	(0.12)	(60.40)		
	EZ	58.50	11	0.00	1.00	−1.45	−0.60			
					(0.00)	(0.90)	(0.97)			
	FF3	41.01	10	0.00	1.04	−2.98	−1.32	−7.24		
					(0.02)	(0.96)	(1.29)	(1.48)		
	FF5	6.34	8	0.61	1.13	−4.48	−8.21	−9.96	−26.30	5.45
					(0.04)	(1.85)	(2.66)	(5.27)	(5.63)	(11.30)
	FFL	21.22	9	0.01	1.13	−3.61	−1.58	−7.20	−15.80	
					(0.04)	(1.34)	(1.72)	(1.85)	(2.99)	
	FFM	37.60	9	0.00	1.06	−3.75	−2.17	−7.71	−1.35	
					(0.02)	(1.13)	(1.36)	(1.51)	(0.50)	
	LCRRA	69.00	12	0.00	1.00	0.35				
					(0.00)	(0.27)				
	LEZ	59.90	11	0.00	1.02	−4.36	−2.50			
					(0.01)	(3.15)	(0.97)			
	NM	12.21	9	0.20	1.26	−6.92	−33.20	3.44	−47.20	
					(0.06)	(1.66)	(5.83)	(2.65)	(15.30)	
	YC	25.62	10	0.00	−6.15	−1440.00	9.03	2110.00		
					(2.36)	(711.00)	(3.32)	(1020.00)		

Notes: Sample Hansen's J statistic $\hat{J}_Z$, degrees of freedom df of its asymptotic distributional limit, probability $\mathbb{P}[\chi^2_{df} > \hat{J}_Z]$ that $\hat{J}_Z$ is smaller than this limit, and estimate $\hat{\theta}_Z^{[GMM]} = [\hat{\theta}_{1,Z}^{[GMM]} \dots \hat{\theta}_{p,Z}^{[GMM]}]'$ of the SDF parameter vector obtained by GMM with optimal weighting matrix, for all the competing SDF specifications. Demeaned corporate bond spread is taken as variable X_t, and the instrument matrix is indicated by the text on the left. Standard errors are reported in parentheses below the parameter estimates.

Table 5 Sample squared unconditional HJ-distance

	Specification	$T\hat{d}^2_{Z,T}$	$\hat{\theta}_{1,Z}$	$\hat{\theta}_{2,Z}$	$\hat{\theta}_{3,Z}$	$\hat{\theta}_{4,Z}$	$\hat{\theta}_{5,Z}$	$\hat{\theta}_{6,Z}$
$Z(X_t) = I_7 \otimes [1\, X_t]'$	CAY	16.71	−0.03	515.00	−0.11	91.10		
		[8.50: 30.19]	(0.42)	(214.00)	(0.11)	(61.10)		
	YC	19.61	−0.67	3490.00	0.61	−3800.00		
		[14.72: 29.75]	(4.02)	(2690.00)	(5.21)	(3210.00)		
	FF5	19.80	1.03	−4.62	−7.90	−5.28	−22.20	3.28
		[12.37: 28.88]	(0.06)	(1.84)	(4.39)	(7.69)	(7.66)	(15.50)
	NM	25.37	1.16	−5.92	−28.30	−7.81	−12.80	
		[15.55: 44.68]	(0.08)	(2.01)	(6.63)	(10.80)	(15.40)	
	FFM	27.20	1.07	−4.67	−4.10	−9.53	−11.70	
		[15.57: 48.27]	(0.08)	(2.53)	(2.97)	(3.01)	(6.18)	
	FFL	39.99	0.91	−1.43	−1.02	−4.32	−3.84	
		[16.33: 45.10]	(0.06)	(1.43)	(1.74)	(1.50)	(6.68)	
	LEZ	43.06	0.59	128.98	−1.25			
		[28.11: 58.76]	(0.17)	(80.30)	(1.41)			
	FF3	43.42	0.89	−1.58	−3.27	−4.70		
		[31.22: 63.97]	(0.04)	(1.40)	(2.13)	(1.50)		
	DEF	43.66	0.59	94.70	0.00	22.10		
		[32.11: 62.80]	(0.70)	(210.00)	(0.32)	(94.10)		
	CAPM	48.70	0.85	−1.51				
		[33.00: 61.67]	(0.04)	(1.32)				
	EZ	53.10	1.00	−0.54	−2.54			
		[38.35: 72.07]	(0.12)	(1.56)	(44.20)			
	LCRRA	54.30	1.00	−10.20				
		[42.91: 74.83]	(0.16)	(20.40)				
	CRRA	103.07	1.00	−0.25				
		[80.17: 140.96]	(0.03)	(0.07)				
$Z(X_t) = I_7 \otimes [1\, X_t^2]'$	FF5	3.51	1.16	−5.55	−5.37	−19.40	−39.00	13.40
		[2.04: 6.64]	(0.08)	(3.01)	(4.08)	(14.30)	(13.20)	(28.50)
	FFM	21.20	1.53	−14.10	−17.40	−21.70	−57.50	
		[12.59: 38.23]	(0.40)	(7.85)	(15.20)	(13.30)	(41.00)	
	NM	28.23	1.43	−8.40	−51.70	10.70	−83.10	
		[17.63: 49.50]	(0.19)	(2.96)	(18.20)	(17.70)	(50.90)	
	CAY	37.60	−1.34	1700.00	−0.97	739.00		
		[23.23: 62.79]	(3.70)	(2890.00)	(11.50)	(1190.00)		
	DEF	42.50	−5.83	891.00	3.07	−271.00		
		[28.11: 70.54]	(4.26)	(816.00)	(1.94)	(422.00)		
	FFL	70.79	1.13	−5.42	3.16	−3.91	−20.00	
		[49.25: 104.80]	(0.10)	(2.13)	(2.92)	(3.00)	(18.30)	
	FF3	73.47	1.03	−4.24	1.68	−4.85		
		[54.64: 107.72]	(0.02)	(1.59)	(2.17)	(2.47)		
	EZ	102.01	0.89	2.78	−1.34			
		[77.64: 146.55]	(0.24)	(1.57)	(12.31)			
	YC	102.02	6.67	−4030.00	−5.31	5320.00		
		[75.12: 139.32]	(12.80)	(10100.00)	(16.70)	(12500.00)		
	CRRA	108.57	0.57	−223.26				
		[81.13: 150.91]	(0.18)	(105.44)				
	LEZ	115.81	1.32	−159.08	−3.50			
		[89.48: 160.75]	(0.30)	(148.80)	(1.76)			
	CAPM	157.00	1.01	−2.75				
		[119.78: 210.67]	(0.02)	(1.55)				
	LCRRA	159.01	4.79	−2.72				
		[124.16: 218.39]	(11.40)	(1.54)				

Notes: This table shows the sample squared unconditional HJ-distance multiplied by the sample size $T\hat{d}^2_{Z,T}$, and minimizers $\hat{\theta}_Z = [\hat{\theta}_{1,Z} \ldots \hat{\theta}_{p,Z}]'$ of the corresponding criterion, for demeaned consumption to wealth ratio as conditioning variable X_t. The instrument matrix is indicated by the text on the left. The 95%-confidence intervals for the population squared unconditional HJ-distance multiplied by the sample size and the standard errors for its minimizers are reported in parentheses below the corresponding estimates.

Table 6 Sample squared unconditional HJ-distance

	Specification	$T\hat{d}^2_{Z,T}$	$\hat{\theta}_{1,Z}$	$\hat{\theta}_{2,Z}$	$\hat{\theta}_{3,Z}$	$\hat{\theta}_{4,Z}$	$\hat{\theta}_{5,Z}$	$\hat{\theta}_{6,Z}$
$Z(X_t) = I_7 \otimes [1\, X_t]'$	FF5	41.02	1.12	−4.54	−7.98	1.84	−22.20	−12.30
		[23.39: 75.89]	(0.04)	(1.70)	(2.49)	(5.91)	(5.07)	(11.40)
	FFL	48.19	1.12	−4.91	−1.54	−7.03	−14.90	
		[29.06: 89.78]	(0.04)	(1.56)	(1.72)	(2.27)	(4.00)	
	NM	62.71	1.24	−5.45	−26.50	−10.20	−14.40	
		[38.58: 105.36]	(0.07)	(1.63)	(5.67)	(8.07)	(13.10)	
	FFM	71.50	1.13	−5.41	−2.12	−10.10	−8.77	
		[45.39: 122.52]	(0.05)	(1.64)	(1.51)	(2.49)	(5.50)	
	YC	73.00	5.35	−672.00	−6.80	1300.00		
		[45.08: 117.79]	(2.25)	(1210.00)	(3.25)	(1610.00)		
	CAY	73.60	0.66	118.00	−0.40	−19.40		
		[53.25: 111.81]	(0.17)	(89.30)	(0.26)	(74.10)		
	EZ	77.71	0.75	2.11	−104.00			
		[56.79: 112.47]	(0.13)	(1.60)	(65.40)			
	CRRA	77.80	1.00	0.51				
		[59.75: 112.19]	(0.03)	(0.10)				
	DEF	80.73	0.20	349.00	0.14	−50.00		
		[60.01: 111.82]	(0.36)	(171.00)	(0.09)	(54.50)		
	FF3	85.60	1.04	−3.19	−1.88	−5.78		
		[63.83: 116.19]	(0.02)	(1.46)	(1.38)	(2.35)		
	LEZ	101.02	1.00	18.80	−1.17			
		[79.46: 140.72]	(0.14)	(7.53)	(1.05)			
	CAPM	103.16	1.00	−1.25				
		[79.28: 139.06]	(0.01)	(1.02)				
	LCRRA	104.35	1.00	21.20				
		[82.48: 142.61]	(0.14)	(8.00)				
$Z(X_t) = I_7 \otimes [1\, X_t^2]'$	CAY	16.01	0.18	152.00	−0.26	−371.00		
		[8.01: 29.34]	(0.36)	(151.00)	(0.32)	(207.00)		
	FF5	18.30	1.06	−4.09	−9.23	3.52	−28.60	−10.90
		[9.89: 32.11]	(0.05)	(2.41)	(2.86)	(6.01)	(7.41)	(13.30)
	FFL	19.60	1.14	−7.09	−0.50	−9.65	−16.60	
		[11.45: 33.36]	(0.05)	(2.30)	(2.00)	(3.28)	(3.92)	
	YC	31.63	−4.09	−1870.00	6.25	2640.00		
		[19.02: 52.02]	(3.12)	(1000.00)	(4.30)	(1380.00)		
	DEF	43.70	0.46	372.00	0.02	−77.70		
		[27.01: 71.22]	(0.42)	(187.00)	(0.11)	(47.80)		
	EZ	43.96	0.71	1.98	−116.00			
		[31.33: 65.14]	(0.20)	(2.08)	(95.80)			
	NM	50.61	1.06	−2.71	−6.94	6.86	−44.40	
		[36.22: 72.34]	(0.09)	(2.37)	(10.10)	(4.14)	(15.80)	
	FFM	61.91	0.92	−0.14	−1.36	0.86	−0.23	
		[46.46: 87.87]	(0.03)	(1.57)	(1.39)	(2.44)	(2.03)	
	FF3	61.93	0.92	−0.04	−1.32	0.99		
		[46.77: 85.06]	(0.03)	(1.54)	(1.37)	(2.40)		
	CAPM	64.90	0.94	0.03				
		[50.65: 92.16]	(0.03)	(2.35)				
	LEZ	65.71	1.00	19.20	0.14			
		[50.89: 90.53]	(0.03)	(35.30)	(2.36)			
	LCRRA	66.02	1.00	19.60				
		[50.68: 89.04]	(0.03)	(41.60)				
	CRRA	69.11	1.00	−0.19				
		[54.08: 96.58]	(0.01)	(0.05)				

Notes: This table reports the sample squared unconditional HJ-distance multiplied by the sample size $T\hat{d}^2_{Z,T}$, and minimizers $\hat{\theta}_Z = [\hat{\theta}_{1,Z} \ldots \hat{\theta}_{p,Z}]'$, for demeaned corporate bond spread as conditioning variable X_t. The instrument matrix is indicated by the text on the left. The 95%-confidence intervals for the population squared unconditional HJ-distance multiplied by the sample size and the standard errors for its minimizers are reported in parentheses below the corresponding estimates.

instrument matrices, the sample unconditional HJ-distance does not suggest to prefer any particular SDF specification. This is the case, for example, for the instrument matrix that includes the corporate bond spread as reported in the upper panel of Table 6. In this case the 95%-confidence intervals of all the unconditional HJ-distances, except the one for specification FF5, overlap.[17] The impossibility to identify the superior performance of an asset pricing model in terms of the unconditional HJ-distance among competing models recurs frequently in empirical equity studies.[18] However, in our empirical analysis we do find different model rankings on the basis of the unconditional HJ-distance depending on distinct instrument matrix, in this way illustrating empirically the theoretical arguments of Section 1.2. As reported in Table 5, the consumption-based CAY and YC specifications have smaller sample unconditional HJ-distance than beta pricing models, such as the FF5, FFM and FF3 specifications, when using the demeaned consumption to wealth ratio as instrument, albeit the confidence intervals for the distance overlap in some cases. The ranking between these classes of models is reversed when the square of demeaned consumption to wealth ratio is used as instrument. In the latter case, the FF5 specification has indeed smaller sample unconditional HJ-distance than the CAY and YC specifications, and with nonoverlapping 95%-confidence intervals. Moreover, model ranking inversion can also occur between distinct conditional versions of the CCAPM. For example, the specification YC is preferred to the specification DEF when the instrument matrix includes the demeaned consumption to wealth ratio (top panel of Table 5), while the opposite happens when the instrument matrix includes just its square (lower panel of the same table).

4.3.2 Sample conditional HJ-distance

The computation of the sample conditional HJ-distance requires the nonparametric estimation of the conditional pricing error vector and the conditional second-moments matrix of assets' gross returns. We select the bandwidths for the nonparametric kernel regression by leave-one-out cross-validation. For each model specification, we select the bandwidths for the estimation of the conditional pricing error vector using the time series of the conditional moment vector with SDF parameter estimated by GMM as explained in Section 4.3.1. This time series serves just as a preliminary estimate for the sole purpose of selecting the bandwidths.[19] For each component of the conditional pricing error vector, we choose the

17 The 95%-confidence interval of the unconditional HJ-distance for the specification FF5 does not overlap with those for the specifications LEZ, CAPM, and LCRRA. While a proper model comparison test requires a statistic, here we consider the individual confidence intervals to simplify. We consider test statistics in the conditional setting in Section 4.3.3.

18 See, for example, the empirical analysis of the CAPM, LCRRA, DEF, FF3, FF5 and other asset pricing specifications, using the gross returns on the T-bill and gross returns on FF research value-weighted portfolios sorted by size and book-to-market ratio in excess of the T-bill, at the monthly frequency from January 1952 to December 2006, in Kan and Robotti (2009).

19 The value $\hat{\theta}_Z^{[GMM]}$ used to create the preliminary time series of the conditional moment function is reported in the upper panels of Tables 3 and 4, in case X_t is either the demeaned consumption to wealth ratio or the demeaned corporate bond spread.

bandwidth that minimizes the time-averaged squared error for the Nadaraya–Watson kernel regression estimator computed without an observation:

$$b_{j,T} := \arg\min_{b \in \mathcal{B}} \; CV_j(b), \quad CV_j(b) := \sum_{i=1}^{T-1} \left(h_j\left(Y_{i+1}, R_{i+1}; \hat{\theta}_Z^{[GMM]}\right) - \hat{e}_{j,T;-i}\left(X_i; \hat{\theta}_Z^{[GMM]} | b\right) \right)^2,$$

for any $j = 1, \ldots, 7$, where h_j is the j-th component of vector h defined in Proposition 1, $\hat{e}_{j,T;-i}(\cdot; \cdot | b)$ is the j-th component of the Nadaraya–Watson kernel regression estimator defined similarly as in Equation (26) but computed on the sample missing of the i-th observation and for the value b of the kernel bandwidth. The minimization is over set $\mathcal{B} := [0.1 \times b_S : 10 \times b_S]$, where b_S is the bandwidth chosen by the rule of thumb for the nonparametric kernel estimation of the probability density of the conditioning variable as suggested in Silverman (1986), that is, $b_S := 0.9 \min\{s, R/1.34\} T^{-\frac{1}{5}}$, for the sample volatility s of the conditioning variable and its sample interquartile range R. We proceed similarly for the selection of the $N(N+1)/2 = 28$ kernel bandwidths that are used to compute estimator $\hat{\Omega}_T$ given in Equation (29) and are common to all the model specifications. We report the results obtained for the Gaussian density as kernel function, and inter-quintile range of the support of the conditioning variable as set $\mathcal{X}_\star$ for the indicator variable $\mathbf{1}(X_t)$ in (28). We take the average value of the chosen bandwidths for the estimation of the conditional pricing errors as the bandwidth to compute the rejection region for the test of correct model specification given in Inequality (31). The empirical results are not unduly sensitive on these bandwidths and kernel choices.

4.3.3 Comparison of sample unconditional and conditional HJ-distances

We report in the two panels of Table 7 the ranking of the models based on the conditional HJ-distance for the demeaned consumption to wealth ratio and demeaned corporate bond spread as information variable X_t, respectively. For each asset pricing model, we reject the hypothesis of correct model specification by means of the rejection rule given in Inequality (31) based on the conditional HJ-distance. The p-values obtained from the asymptotic distribution under correct model specification in Proposition 4 are smaller than 0.05 for all models, for both choices of the information variable. Consequently, we report in Table 7 the 95%-confidence intervals for the squared conditional HJ-distance multiplied by the sample size using the large sample distribution in Proposition 5 (see also the discussion at the end of Section 2.3).

Interestingly, the ranking based on the conditional HJ-distance, which avoids the selection of an instrument matrix that may favor a model specification at the expenses of another, is quite similar for the two discussed distinct choices of the variable X_t. The specifications DEF and YC, which are differently ranked on the basis of the unconditional HJ-distance with the proposed instrument matrices, provide an example. Their relative ranking is in some cases reverted, as for example when the instrument matrix includes demeaned consumption to wealth ratio (see Table 5). Moreover, they are not always included among the top four specifications on the basis of the sample unconditional HJ-distance. Nonetheless, they both appear among the top four specifications for the model ranking based on the sample conditional HJ-distance, in both cases with either demeaned consumption to wealth ratio or demeaned corporate bond spread as conditioning variable. Moreover, the beta pricing specifications, such as the FF5, FFM and FF3 specifications,

Table 7 Sample squared conditional HJ-distance

X_t	Specification	$T\hat{\delta}_T^2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$	$\hat{\theta}_6$
Demeaned consumption to wealth ratio	NM	75.71	1.30	−4.91	−12.10	−19.90	−86.00	
		[43.58: 142.16]	(0.28)	(4.87)	(37.20)	(28.90)	(91.20)	
	DEF	84.40	1.11	35.11	−0.10	2.79		
		[48.96: 150.65]	(6.46)	(93.20)	(3.64)	(51.20)		
	CAY	94.10	1.07	−83.90	0.64	−478.00		
		[58.91: 162.57]	(5.32)	(315.23)	(1.85)	(267.10)		
	YC	95.66	0.97	−359.00	−0.31	643.00		
		[60.77: 163.96]	(0.05)	(312.21)	(22.10)	(287.43)		
	FF5	145.03	1.13	−5.17	−4.73	−6.06	−8.72	−4.36
		[95.81: 243.91]	(0.11)	(4.03)	(6.96)	(19.10)	(15.70)	(43.20)
	FFM	159.09	1.09	−4.14	−1.58	−8.89	−7.05	
		[110.92: 233.31]	(0.23)	(6.20)	(8.90)	(25.60)	(19.10)	
	FFL	161.18	1.09	−4.31	2.81	−7.99	−1.51	
		[116.33: 231.13]	(0.17)	(3.13)	(4.80)	(4.95)	(30.00)	
	LEZ	162.07	0.83	95.40	−2.32			
		[119.99: 228.43]	(0.32)	(226.00)	(2.13)			
	FF3	171.46	1.14	−4.05	−3.71	−12.10		
		[126.45: 233.20]	(0.05)	(1.82)	(2.75)	(2.98)		
	EZ	174.11	1.00	0.49	0.21			
		[132.97: 244.18]	(0.03)	(0.13)	(3.20)			
	CAPM	188.96	0.98	−2.54				
		[143.09: 257.47]	(0.04)	(1.89)				
	CRRA	206.96	0.97	0.14				
		[160.52: 282.33]	(0.12)	(11.60)				
	LCRRA	289.01	1.09	−21.40				
		[225.33: 395.82]	(0.30)	(22.70)				
Demeaned corporate bond spread	YC	51.52	1.03	−399.00	−1.52	1180.00		
		[29.02: 94.85]	(0.46)	(144.67)	(1.56)	(492.31)		
	DEF	84.20	0.97	−52.60	−0.01	58.10		
		[49.97: 153.43]	(0.28)	(33.20)	(0.18)	(6.44)		
	EZ	90.32	0.95	0.12	0.44			
		[55.01: 152.21]	(0.03)	(1.07)	(2.30)			
	FFL	113.11	1.15	−5.05	−5.41	−9.34	−24.80	
		[69.65: 188.45]	(0.06)	(2.78)	(2.89)	(3.97)	(4.78)	
	NM	114.62	1.12	−5.61	−49.20	−2.65	47.00	
		[73.15: 186.76]	(0.09)	(2.63)	(13.20)	(10.40)	(47.80)	
	CRRA	126.08	0.97	1.76				
		[90.37: 189.31]	(0.01)	(0.29)				
	FF5	144.81	1.14	−5.17	−4.73	−6.06	−87.10	−4.37
		[104.41: 207.43]	(0.05)	(2.77)	(2.98)	(6.94)	(8.32)	(17.80)
	FFM	149.54	0.71	−3.11	−5.19	0.71	30.20	
		[109.54: 208.29]	(0.07)	(1.69)	(1.86)	(2.98)	(7.12)	
	CAY	152.19	0.82	139.00	−0.22	310.00		
		[115.10: 212.03]	(0.20)	(191.00)	(0.33)	(256.00)		
	LEZ	176.68	0.45	258.00	−0.52			
		[136.15: 250.76]	(0.17)	(37.70)	(2.77)			
	FF3	235.29	1.03	−2.78	−5.65	−0.11		
		[177.66: 318.27]	(0.03)	(1.91)	(1.42)	(2.88)		
	LCRRA	247.48	0.52	229.00				
		[192.73: 339.30]	(0.29)	(126.03)				
	CAPM	413.67	1.00	−1.85				
		[325.29: 569.42]	(0.03)	(4.02)				

Notes: This table reports the sample squared conditional HJ-distance multiplied by the sample size $T\hat{\delta}_T^2$, and minimizers $\hat{\theta} = [\hat{\theta}_1 \ldots \hat{\theta}_p]'$ of the corresponding criterion. The conditioning variable X_t is indicated by the text on the left. The 95%-confidence intervals for the population squared conditional HJ-distance multiplied by the sample size and the standard errors for its minimizers are reported in parentheses below the corresponding estimates.

constantly feature larger sample conditional HJ-distances compared to the consumption-based DEF and YC specifications. Overall, a message of Table 7 is that SDF models which are conditionally linear in consumption growth perform well compared to multifactor beta pricing specifications, despite the smaller number of parameters, when the conditional pricing errors of the two models are properly accounted for. The 95%-confidence intervals are rather large and often overlap across models. For most of the pairs of models, we obtain p-values that are larger than 0.1 for the tests described in Section 3, preventing us from rejecting the hypothesis of equal conditional HJ-distance at the 10% significance level. However, there are some exceptions. These exceptions are the cases with a well performing model on the basis of the conditional HJ-distance and one of the specifications CAPM, CRRA and LCRRA for the demeaned consumption to wealth ratio as conditioning variable, or one of the specifications FFM, CAY, LEZ, FF3, LCRRA, and CAPM for the demeaned corporate bond spread as conditioning variable. An example is the pair of nested specifications YC and LCRRA. We reject the hypothesis of equal conditional HJ-distances at the 5% significance level, with a p-value for the related test of 0.05 or 0.08, when either the demeaned corporate bond or the demeaned consumption to wealth ratio is taken as conditioning variable, respectively. As explained in Section 3.3, the alternative hypothesis is of a smaller conditional HJ-distance for the YC specification than the LCRRA specification. Therefore, while the conditional HJ-distance indicates that both model are misspecified, it provides considerable support for preferring the YC specification to the LCRRA specification.

Comparing the upper panel of Table 7 with Table 5 (respectively, the lower panel of Table 7 with Table 6), we see that the conditional HJ-distance is larger than its unconditional counterpart for each model, as explained theoretically in Section 1.2.[20] The difference between the values of the two HJ-distances is numerically large for most specifications, illustrating the fact that ad-hoc instruments can fail to capture a substantial part of the variability of conditional pricing errors.

We also analyze the time series of the statistic $\hat{e}_T(X_t;\hat{\theta}_T)'\hat{\Omega}_T(X_t)\hat{e}_T(X_t;\hat{\theta}_T)$, whose average over the sample dates t with $X_t \in \mathcal{X}_\star$ yields the sample conditional HJ-distance in Equations in (28). The variability of this statistic over time is due to the variation of the conditional pricing error vector $\hat{e}_T(X_t;\hat{\theta}_T)$ and to the variation of the conditional second-moments matrix of assets' gross returns $\hat{\Omega}_T(X_t)$. Figure 1 reports the time series of the demeaned corporate bond spread and the demeaned consumption to wealth ratio. Figure 2 displays the value of the statistic $\hat{e}_T(X_t;\hat{\theta}_T)'\hat{\Omega}_T(X_t)\hat{e}_T(X_t;\hat{\theta}_T)$ as a function of time in the form of a smoothed heatmap, with the demeaned corporate bond spread as conditioning variable X_t. The darker the color, the higher is the value of the statistic for a precise model specification and time. Overall, this statistic does not clearly indicate a different relative performance of the asset pricing models during distinct periods. However, for all the competing models its value is typically lower when the volatility of the corporate bond spread is relatively low, and when the correlation among the excess gross returns on the six equity portfolios is relatively mild. This situation typically realizes when the shocks to corporate bond spreads have relatively low intensity, and there are relatively more possibilities to

20 The only exception is the YC specification with constant and squared demeaned consumption wealth ratio as instruments.

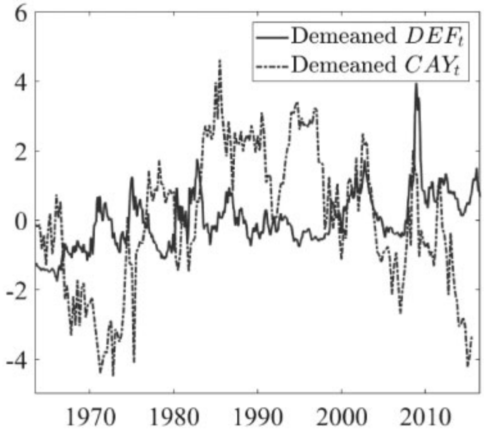

Figure 1 The figure shows the time series of the demeaned corporate bond spread DEF_t and the demeaned consumption to wealth ratio CAY_t as functions of time.

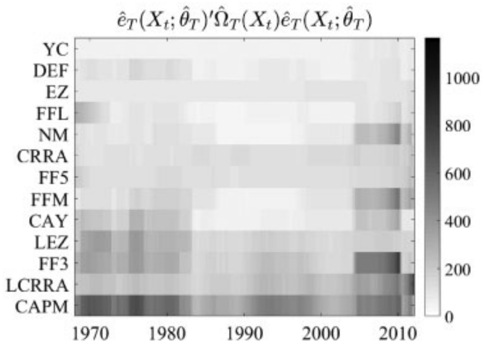

Figure 2 The smoothed heatmap shows the quadratic form of the pricing error vector $\hat{e}_T(X_t;\hat{\theta}_T)$ weighted by the inverse of the conditional second-moments matrix of assets' gross returns $\hat{\Omega}_T(X_t)$, for each model specification, as a function of time. The vector $\hat{\theta}_T$ is the minimizer of the conditional HJ-distance. The demeaned corporate bond spread is the conditioning variable X_t. The model specifications are presented with ascending sample conditional HJ-distance (see the lower panel of Table 7). Before 1985 and after 2005 the unconditional variance of the corporate bond spread is more than three times the value during the period in between. In those periods the statistic has generally higher values, and accordingly the color is darker in the corresponding areas.

diversify an equity portfolio. The value of the statistic $\hat{e}_T(X_t;\hat{\theta}_T)'\hat{\Omega}_T(X_t)\hat{e}_T(X_t;\hat{\theta}_T)$ reaches indeed its largest values before 1985 and after 2005, for all the model specifications. Before 1985 the sample unconditional variance of the corporate bond spread is 0.40, and each pairwise sample correlations between the excess gross returns on the equity portfolios is in the range $[0.75 : 0.98]$. Between 1985 and 2005 the variance is 0.17, and the correlations are in the range $[0.60 : 0.97]$. After 2005 the variance is 0.78, and the correlations are in the range [0.83 : 0.97].[21] The period starting from mid-80 s is characterized by low volatility for many U.S. macroeconomic series and is often referred to as the Great Moderation (see Stock and Watson, 2002). This period of low volatility ended with the financial crisis of 2007–2009.

As an example of different time series properties of the conditional pricing errors leading to distinct model rankings on the basis of the conditional and unconditional HJ-distances, let us consider the errors in the YC and FF5 specifications, again taking the demeaned corporate bond spread as conditioning variable X_t. We plot in Figure 3 the estimates of these errors as functions of time, and report in Table 8 their time series averages, standard deviations, cross-correlations, and correlations with the contemporaneous demeaned corporate bond spread. The absolute value of each average error for an equity portfolio is lower in the YC specification than in the FF5 specification. Moreover, apart from the error for the growth stock with small market capitalization, the errors for the equity portfolios in the YC specification are less volatile than those in the FF5 specification. In both specifications, the errors for the T-bill are larger and more volatile than those for the equity portfolios. As we already mentioned in the first paragraph of Section 4.2, the T-bill identifies the scale of the SDF, and it is more sensitive to a misspecification of the SDF than the equity portfolios (see, e.g., Section V.A in Nagel and Singleton, 2011, for a discussion).[22] The errors for the equity portfolios are positively correlated among each other, and negatively correlated with the contemporaneous corporate bond spread and the errors for the T-bill, with all the correlations being significant at the 1% level. Interestingly, the correlations between errors and corporate bond spread are stronger in the specification YC, with the exception of the error for the neutral stock with small market capitalization.[23] Therefore, albeit the errors in the YC model appear smaller than those in the FF5 model, the latter errors can lead to smaller components of the sample unconditional moment function (i.e., the sample counterpart of the LHS of Equation 3) and in turn to a smaller sample unconditional HJ-distance for the FF5 specification, when the demeaned corporate bond spread in included in the instrument matrix (see the upper panel of Table 6).

The main focus of our discussion is not on the distinct economic interpretation of the values of the minimizers of sample HJ-distances, taken as estimates of the true or pseudo-true model parameters. However, we end this section by commenting on the stability of the parameter estimates obtained by GMM with optimal weighting matrix and on the stability

21 All the correlations are significant at the 1% level. The bootstrapped 95%-confidence intervals for the unconditional variance of the corporate bond spread based on 1000 bootstrap samples are [0.31 : 0.49], [0.10 : 0.17], and [0.49 : 1.10] in the three discussed periods, respectively.

22 The errors for the T-bill are scattered around 0.55 in the YC specification and around 0.15 in the FF5 specification (see Table 8). Nagel and Singleton (2011) estimate conditional pricing errors for gross returns on the T-bill and excess gross returns on four FF U.S. equity portfolios from the second quarter of 1952 to the end of 2006, using optimal weighting matrices in the estimation procedure. Their local polynomial regression estimates of the errors for the T-bill in the DEF specification with DEF_t as conditioning variable are also not scattered around 0 (see Figure 4 of the article).

23 The raw cross-moment $T^{-1}\sum_{t=1}^{T} \hat{e}_i(X_t;\hat{\theta}_T) DEF_t$, with the pricing error $\hat{e}_i$ for the i-th equity portfolio, lies in the range [−0.01 : 0.00] under both specifications. The corresponding moment for the T-bill is 0.09 in the YC specification and 0.02 in the FF5 specification.

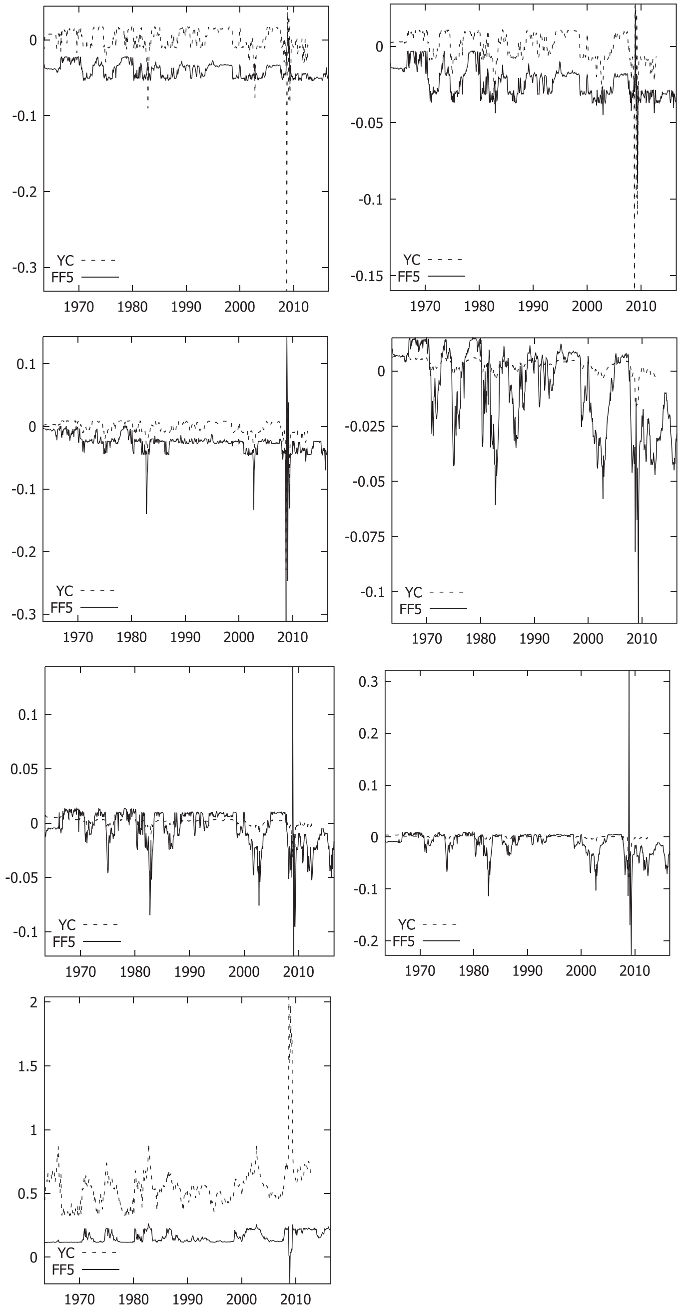

Figure 3 The figure shows the time series of the conditional pricing errors $\hat{e}_i$ valued at the minimizer of the conditional HJ-distance in the two specifications YC and FF5. The errors are displayed from left to right, and then top to bottom, for $i = SG, SN, SV, BG, BN, BV, Tbill$, respectively. See the caption of Table 8 for a brief description of these errors.

Table 8 Sample time series properties of the conditional pricing errors

		YC							FF5						
		$\hat{e}_{SG}$	$\hat{e}_{SN}$	$\hat{e}_{SV}$	$\hat{e}_{BG}$	$\hat{e}_{BN}$	$\hat{e}_{BV}$	$\hat{e}_{Tbill}$	$\hat{e}_{SG}$	$\hat{e}_{SN}$	$\hat{e}_{SV}$	$\hat{e}_{BG}$	$\hat{e}_{BN}$	$\hat{e}_{BV}$	$\hat{e}_{Tbill}$
Average		0.00	0.00	0.00	0.00	0.00	0.00	0.55	−0.04	−0.02	−0.03	−0.01	0.00	−0.01	0.15
SD		0.02	0.01	0.02	0.00	0.00	0.01	0.18	0.01	0.01	0.02	0.02	0.02	0.03	0.04
Correlations	X_t	−0.55	−0.64	−0.58	−0.97	−0.89	−0.73	0.76	−0.38	−0.69	−0.51	−0.87	−0.60	−0.53	0.51
	$\hat{e}_{SG}$		0.92	0.93	0.57	0.58	0.36	−0.60		0.85	0.28	0.59	0.52	0.46	−0.74
	$\hat{e}_{SN}$			0.99	0.66	0.62	0.38	−0.68			0.55	0.83	0.73	0.71	−0.70
	$\hat{e}_{SV}$				0.61	0.59	0.37	−0.65				0.61	0.68	0.74	−0.44
	$\hat{e}_{BG}$					0.93	0.83	−0.87					0.86	0.79	−0.72
	$\hat{e}_{BN}$						0.89	−0.81						0.96	−0.72
	$\hat{e}_{BV}$							−0.79							−0.67

Notes: This table shows the sample time series average, standard deviation (SD), correlations for the conditional pricing errors $\hat{e}_i$ valued at the minimizer of the conditional HJ-distance, and their correlations with the demeaned corporate bond spread X_t, in the two specifications YC and FF5. The demeaned corporate bond spread is the conditioning variable. The seven conditional pricing errors are for the real inflation- and dividends-adjusted proxy for the gross excess return on a growth (G), neutral (N) or value (V) stock of a company with small (S) or big (B) market capitalization, and for the gross returns on the T-bill, as each subscript indicates.

of the minimizers of the two HJ-distances across different choices of the conditioning variable X_t. We report in Tables 3 and 4 the GMM estimate $\hat{\theta}_Z^{[GMM]}$ of the SDF parameter vector, whose dimension depends on the model. For example, it is $\hat{\theta}_Z^{[GMM]} = \left[\hat{\theta}_{1,Z}^{[GMM]} \ \ldots \ \hat{\theta}_{4,Z}^{[GMM]}\right]'$ in the CAY and YC specifications, and $\hat{\theta}_Z^{[GMM]} = \left[\hat{\theta}_{1,Z}^{[GMM]} \ \ldots \ \hat{\theta}_{5,Z}^{[GMM]}\right]'$ in the NM specification. In several cases, we obtain a different estimate of a pseudo-true value of the same parameter when using distinct instrument matrices. Again, this is due to the fact that each estimation procedure corresponds to an empirical analysis of a different set of managed portfolios. For example, our pointwise estimate of the weakly identified coefficient of relative risk aversion in the CRRA specification (see, e.g., Hall, 2005) is positive (negative, respectively) when the demeaned corporate bond spread (demeaned consumption to wealth ratio) is taken as variable X_t (see Tables 3 and 4). Only when this coefficient is positive the corresponding power utility function of the form $\left(C_t^{1-\theta_2}\right)/(1-\theta_2)$ for the representative agent is concave. Similar issues arise in the estimation of the other model specifications. To exemplify the impact of different instrument matrices on the sign of the parameter estimates, we report in Table 9 the results of the estimation of the parameters of the CAPM specification by optimal GMM for several distinct instrument matrices. When the consumption to wealth ratio is the unique conditioning variable X_t, we obtain a negative estimate for θ_2. In this case, the proxy $-\overline{R}_f \widehat{\mathrm{Cov}}\left[m_{\mathsf{CAPM}}\left(Y_{t+1}; \hat{\theta}_Z^{[GMM]}\right), SG_t\right]$ of the monthly unconditional risk premium for the SG_t portfolio, where $\overline{R}_f$ is the mean gross T-bill return and $\widehat{\mathrm{Cov}}[\cdot,\cdot]$ denotes the sample covariance, is positive. The same happens for all the equity portfolios. For example, with an estimate of 1.01 for θ_1 and –2.51 for θ_2, the proxies for the equity portfolio monthly risk premia are in the range $[4.4 : 6.6] \times 10^{-3}$. Differently, the proxies for the risk premia created from the same estimate for θ_1 and an estimate of 1.85 for θ_2 are in the range $[-3.3 : -4.9] \times 10^{-3}$. A similar discussion can be

Table 9 Unconditional statistics

X_t	$Z(X_t)$	$\hat{J}_Z$	df	$\mathbb{P}[\chi^2_{df} > \hat{J}_Z]$	$\hat{\theta}^{[GMM]}_{1,Z}$	$\hat{\theta}^{[GMM]}_{2,Z}$
0	I_7	51.40	5	0.00	1.01	−2.51
					(0.01)	(0.96)
$X_{1,t}$	$I_7 \otimes [1\, X_{1,t}]'$	61.70	12	0.00	1.03	−3.79
					(0.01)	(0.89)
	$I_7 \otimes [1\, X_{1,t}\, X^2_{1,t}]'$	75.53	19	0.00	1.01	−2.83
					(0.01)	(0.60)
	$I_7 \otimes [1\, X_{1,t}\, X^2_{1,t}\, X^3_{1,t}]'$	84.80	26	0.00	1.01	−2.43
					(0.00)	(0.50)
	$I_7 \otimes [1\, X_{1,t}\, X^2_{1,t}\, X^3_{1,t}\, X^4_{1,t}]'$	92.11	33	0.00	1.01	−2.64
					(0.01)	(0.53)
$[X_{1,t}\, X_{2,t}]'$	$I_7 \otimes [1\, X_{1,t}\, X^2_{1,t}\, X^3_{1,t}\, X^4_{1,t}\, X_{2,t}]'$	120.58	40	0.00	1.00	1.05
					(0.00)	(0.07)
	$I_7 \otimes [1\, X_{1,t}\, X^2_{1,t}\, X^3_{1,t}\, X_{2,t}\, X^2_{2,t}]'$	114.19	40	0.00	1.00	1.25
					(0.00)	(0.07)
	$I_7 \otimes [1\, X_{1,t}\, X^2_{1,t}\, X_{2,t}\, X^2_{2,t}\, X^3_{2,t}]'$	113.03	40	0.00	1.00	1.68
					(0.00)	(0.05)
	$I_7 \otimes [1\, X_{1,t}\, X_{2,t}\, X^2_{2,t}\, X^3_{2,t}\, X^4_{2,t}]'$	115.01	40	0.00	1.00	1.74
					(0.00)	(0.03)
$X_{2,t}$	$I_7 \otimes [1\, X_{2,t}\, X^2_{2,t}\, X^3_{2,t}\, X^4_{2,t}]'$	104.17	33	0.00	1.01	1.93
					(0.00)	(0.04)
	$I_7 \otimes [1\, X_{2,t}\, X^2_{2,t}\, X^3_{2,t}]'$	98.20	26	0.00	1.01	1.87
					(0.00)	(0.07)
	$I_7 \otimes [1\, X_{2,t}\, X^2_{2,t}]'$	91.23	19	0.00	1.00	1.64
					(0.00)	(0.13)
	$I_7 \otimes [1\, X_{2,t}]'$	89.21	12	0.00	1.01	1.85
					(0.00)	(0.28)

Notes: Sample Hansen's J statistic $\hat{J}_Z$ and GMM estimate $\hat{\theta}^{[GMM]}_Z = [\hat{\theta}^{[GMM]}_{1,Z}\ \hat{\theta}^{[GMM]}_{2,Z}]'$ of the SDF parameter vector for the specification CAPM, under different choices of the conditioning variable X_t. The conditioning variable X_t and the instrument matrix $Z(X_t)$ are reported in the first two columns, where the demeaned consumption to wealth ratio and the demeaned corporate bond spread are denoted by $X_{1,t}$ and $X_{2,t}$, respectively. Each GMM estimate is computed with its own optimal weighting matrix. The table includes also the degrees of freedom *df* of the asymptotic distributional limit of $\hat{J}_Z$, and the probability $\mathbb{P}[\chi^2_{df} > \hat{J}_Z]$ that $\hat{J}_Z$ is smaller than this limit. Standard errors are reported in parentheses below the parameter estimates.

done regarding the stability of the minimizers of the unconditional HJ-distance. Differently, the minimizers of the conditional HJ-distance are more consistent across different choices of the conditioning information variable X_t than the GMM estimates of the model parameters and the minimizers of the unconditional HJ-distance (see Table 7). This characteristic of the estimates is in line with the findings in Nagel and Singleton (2011): an estimation procedure that accounts for the dynamic pricing restrictions instead of just few static restrictions provides estimates usually affected by a lower degree of statistical uncertainty.

4.3.4 Robustness with respect to the construction of the data and the implementation of the statistics

We summarize in this section the checks we made about our choices in the implementation of the HJ-distances, and their impact on the reversing of the model ranking on the basis of the unconditional HJ-distance when changing the instrument matrix. Our robustness checks focus on the construction of the data, and on the numerical implementation of the two distances. The results obtained with the considered alternative procedures do not qualitatively alter our discussion reported in previous sections.

Some robustness checks refer to the construction of the time series for vector Y_t in the different model specifications. There is typically no distinction between consumption and disposable income in general equilibrium models, and we consider disposable income from the FRED database of the St. Louis Fed in place of the personal consumption of nondurables and services. We obtain alternative proxies for the monthly transitory component of financial wealth using 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. Following the discussion in Section 7 of Fama and French (2015) about the redundancy of the value factor HML_t in the FF5 specification, we replace it by the sum of the intercept and residual from its regression onto MKT_t, SMB_t, RMW_t, and CMA_t.[24]

Other robustness checks refer to the choice of the vector X_t in the different model specifications. In the financial literature, several variables have been considered as candidate for the stock return prediction. Following Goyal and Welch (2008) we consider total dividend-to-price ratio, total earnings to price ratio, stock variance, cross-sectional premium, book-to-market ratio, net equity expansion, treasury bills, long-term rate of returns and long-term yield as alternative choices for the variable X_t. To construct the dividend-to-price ratio, we focus on the CRSP value-weighted portfolio created with stocks traded at NYSE, NASDAQ, Amex, and Arca. Following Cochrane (2008), we consider the total value weighted cum-dividend returns $R_{m,t}$ and their counterpart returns without dividends $R^{[x]}_{m,t}$, and construct the total dividend-price ratio $R_{m,t}/R^{[x]}_{m,t} - 1$.[25]

The remaining robustness checks refer to the numerical implementation of the unconditional and conditional HJ-distances. We search for the optimum of the quadratic criteria defining the HJ-distances implementing the steepest descent, Gauss–Newton, Levenberg–Marquardt, Davidon–Fletcher–Powell, and Broyden–Fletcher–Goldfarb–Shanno algorithms, always keeping 10^{-6} as the arbitrary threshold for the convergence. For the nonparametric kernel regressions that are necessary to compute the conditional HJ-distance we consider also the uniform, triangular and Epanechnikov specification of the kernel function. We also repeat the study relying on the leave-one-out cross-validation bandwidth selection on a preliminary SDF estimated by optimal GMM with instrument matrix including

24 The variance inflation factor for the variable HML_t in the FF5 specification is about 1.97. Its strongest correlation is about 0.69 with variable CMA_t.

25 The CRSP portfolio data have been retrieved through the Wharton Research Data Services database, and the proxies for the conditioning variables have been retrieved from the official website of Amit Goyal.

just the square of X_t.[26] Finally, we check for possible impacts of numerical instabilities arising in the inversion of the second-moments matrices of assets' gross returns.[27] The order of magnitude of the condition number of the estimate of the matrix Ω_Z^{-1}, which is defined as $\sqrt{\lambda_{\max}/\lambda_{\min}}$ for its largest and smallest matrix eigenvalues $\lambda_{\max}$ and $\lambda_{\min}$, is 10^2 for every instrument matrix considered for the discussion in previous sections. The estimate of matrix $\Omega^{-1}(x)$ has the same property for any value x of the trimmed support. This order of magnitude of the condition number is typical for second moment matrices of gross returns on equity portfolios.[28] However, to check for the impact of numerical inversion issues to the model ranking, we repeat the entire analysis relying upon statistics that are similarly defined as the unconditional and conditional HJ-distance but with regularized versions of the second-moments matrix of assets' gross returns. The regularization of each matrix consists in inflating each element on its main diagonal by the minimal amount to obtain a condition number of 15. We indeed consider a matrix with this condition number as likely far removed from numerical inversion issues. In practice, our regularization corresponds to estimating higher marginal second moments of the test assets combined with the instruments, while maintaining the same correlations among them. The values of these newly introduced statistics are generally lower than the values of the corresponding unconditional and conditional HJ-distance. This result is to be expected, because the elements on the main diagonal of the inverse of each weighting matrix used for these statistics are higher than the corresponding values used in the HJ-distances. However, we still rank competing asset pricing specifications differently also when we base the study on the regularized counterpart of the unconditional HJ-distance for different instrument matrices. Therefore, the reversion in the model specification ranking based on the unconditional HJ-distance described in Section 4.3.3 is not (at least entirely) due to computational issues regarding the numerical inversion of the second-moments matrix of assets' gross returns.

5 Conclusion

We develop a novel approach for comparing misspecified asset pricing models in terms of the conditional HJ-distance. Unlike the unconditional counterpart, this distance measures the model (in)ability to match the dynamic pricing restrictions for the test assets by fully accounting for the conditional character of such restrictions. This is in stark contrast with the unconditional HJ-distance, which captures solely the component of conditional pricing errors spanned by the chosen instrument matrix, or equivalently, the unconditional pricing errors of a specific set of managed portfolios built from the original test assets. We develop theoretical arguments to show that the difference between the conditional and

26 The estimate $\hat{\theta}_Z^{[GMM]}$ is reported in the two lower panels of Tables 3 and 4, in case X_t is either the demeaned consumption to wealth ratio or the demeaned corporate bond spread.

27 See also Cochrane (2008) for a discussion on the near-singularity of the second-moments matrix of assets' gross returns.

28 This is the case, for example, of the matrix Ω_Z^{-1} in Hodrick and Zhang (2001) constructed using gross returns on the T-bill and gross returns on FF research value-weighted portfolios sorted by size and book-to-market ratio in excess of the T-bill, at the monthly frequency from January 1952 to December 1997.

unconditional HJ-distances of a given parametric SDF family can be large, and that the two distances can yield different rankings when comparing two asset pricing models.

In our empirical analysis, we compare thirteen parametric SDF specifications by computing the conditional and unconditional HJ-distances for a sample of monthly returns on six U.S. equity portfolios and short-term T-bills. For some choices of the instrument matrix in the unconditional approach, we find superior performance for linear SDF specifications arising from the Fama–French model and its extensions, for example with momentum and profitability factors. However, the empirical model ranking is sensitive to the selection of instruments. Moreover, we find that more structural SDF specifications, which are conditionally linear in consumption growth (see Nagel and Singleton, 2011), have superior performance when the conditional HJ-distance is used for model comparison with consumption to wealth ratio or corporate bond spread as conditioning variables.

The conditional HJ-distance of a parametric SDF family is computed for a given choice of the set of test assets and vector of conditioning variables. We checked that the results of our empirical analysis are not unduly sensitive to these choices. However, it is an important question to investigate how the HJ-distance and the implied model ranking depend on the selected test assets and conditioning variables. Conceptually, the number N of test assets and the number L of conditioning variables are not subject to restrictions. Hence, we could consider large cross-sections of test assets returns, such as individual stocks, and conditioning information sets spanned by several state variables. In practice, such extensions are challenging. Indeed, dimension L cannot exceed 3 of 4, say, with realistic sample sizes and standard nonparametric regression methods, due to the curse of dimensionality. Moreover, a large number of assets N implies numerical instabilities when inverting the estimated second-moments matrix of assets' gross returns to obtain the weighting matrix. To address these issues, we could deploy dimensionality-reduction and regularization techniques. We leave such developments for future research.

Supplementary Data

Supplementary data are available at *Journal of Financial Econometrics* online.

Appendix A: Proofs of the Results on the Population HJ-Distance

In this Appendix, we prove the theoretical properties of the conditional HJ-distance given in Section 1.

A.1 Proof of Proposition 1

Let us consider the inner minimization problem that defines the squared distance δ^2 in Equation (8). 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

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

where $\lambda : \mathcal{X} \to \mathbb{R}^N$ 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 $M_{t,t+1}$. Let us denote by $M_{t,t+1}^{[\mathrm{opt}]}$ and $\lambda^{[\mathrm{opt}]}$ the random variable and the vector function that solve the optimization problem. The First-Order Condition (FOC) for optimizing the Lagrangian w.r.t. $M_{t,t+1}$ yields the condition

$$M_{t,t+1}^{[\mathrm{opt}]} := m(Y_{t+1};\theta) - \lambda^{[\mathrm{opt}]}(X_t)'R_{t+1}. \tag{A1}$$

The SDF in Equation (A1) has to satisfy the conditional no-arbitrage restriction in Equation (1) and thus

$$\lambda^{[\mathrm{opt}]}(X_t) := \mathrm{E}\left[R_{t+1}R_{t+1}'|X_t\right]^{-1}\mathrm{E}[m(Y_{t+1};\theta)R_{t+1} - 1_N|X_t] = \Omega(X_t)e(X_t;\theta). \tag{A2}$$

By replacing Equations (A1) and (A2) into the objective function in Equation (8) we get

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

The conclusion follows.

A.2 Proof of Equation (14)

We first derive an explicit expression for the projection $\mathcal{P}_A[e(\cdot;\theta)]$. For any $(N\times q)$-dimensional matrix function $\Phi(X_t) := \left[\phi_1(X_t)\ \ldots\ \phi_q(X_t)\right]$ and $(N\times m)$-dimensional matrix function $\tilde{\Phi}(X_t) := \left[\tilde{\phi}_1(X_t)\ \ldots\ \tilde{\phi}_m(X_t)\right]$ we define the $(q\times m)$-dimensional matrix of $L^2_\Omega(\mathcal{X})$-inner products as

$$\langle\Phi,\tilde{\Phi}\rangle_{L^2_\Omega(\mathcal{X})} := \begin{bmatrix} \langle\phi_1,\tilde{\phi}_1\rangle_{L^2_\Omega(\mathcal{X})} & \cdots & \langle\phi_1,\tilde{\phi}_m\rangle_{L^2_\Omega(\mathcal{X})} \\ \vdots & \ddots & \vdots \\ \langle\phi_q,\tilde{\phi}_1\rangle_{L^2_\Omega(\mathcal{X})} & \cdots & \langle\phi_q,\tilde{\phi}_m\rangle_{L^2_\Omega(\mathcal{X})} \end{bmatrix},$$

where $\langle\phi_i,\tilde{\phi}_j\rangle_{L^2_\Omega(\mathcal{X})} = \mathrm{E}\left[\phi_i(X_t)'\Omega(X_t)\tilde{\phi}_j(X_t)\right]$. Then, the orthogonal projection operator $\mathcal{P}_\Phi$ onto the linear subspace of $L^2_\Omega(\mathcal{X})$ spanned by the columns of matrix function $\Phi(X_t)$ is defined as

$$\mathcal{P}_\Phi[\phi](X_t) = \Phi(X_t)\left(\langle\Phi,\Phi\rangle_{L^2_\Omega(\mathcal{X})}\right)^{-1}\langle\Phi,\phi\rangle_{L^2_\Omega(\mathcal{X})}. \tag{A3}$$

By particularizing this expression to the projection of the conditional pricing error vector onto the scaled instrument matrix, we get $\mathcal{P}_A[e(\cdot;\theta)](X_t) = A(X_t)\left(\langle A,A\rangle_{L^2_\Omega(\mathcal{X})}\right)^{-1}$ $\langle A,e(\cdot;\theta)\rangle_{L^2_\Omega(\mathcal{X})}$. Moreover, by using the parameterization in Equation (13) and the law of iterated expectations, we get $\langle A,A\rangle_{L^2_\Omega(\mathcal{X})} = \mathrm{E}\left[A(X_t)'\Omega(X_t)A(X_t)\right] = \Omega_Z^{-1}$ and $\langle A,e(\cdot;\theta)\rangle_{L^2_\Omega(\mathcal{X})} = \mathrm{E}\left[A(X_t)'\Omega(X_t)e(X_t;\theta)\right] = \mathrm{E}\left[Z(X_t)h(X_{t+1},R_{t+1};\theta)\right]$. Thus, we get

$\mathcal{P}_A[e(\cdot;\theta)]\ \ (X_t) = A(X_t)\Omega_Z\mathrm{E}\big[Z(X_t)h(Y_{t+1},R_{t+1};\theta)\big]$. The $L^2_\Omega(\mathcal{X})$-norm of the vector $\mathcal{P}_A[e(\cdot;\theta)](X_t)$ is

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

Therefore, this norm is the criterion minimized by the unconditional HJ-distance d_Z in Equation (6), and Equation (14) follows.

A.3 Proof of Proposition 2

The criterion $\|\mathcal{P}_A[e(\cdot;\theta)]\|_{L^2_\Omega(\mathcal{X})}$ valued at its global minimum θ_Z is equal to d_Z^2 (see Equation (14)), and cannot be greater than the same criterion at any other argument $\theta\in\Theta$. In particular we have

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

From Equation (12) and the Pythagorean theorem for inner product spaces we have that

$$\delta^2 = \|e(\cdot;\theta_*)\|^2_{L^2_\Omega(\mathcal{X})} = \|\mathcal{P}_A[e(\cdot;\theta_*)]\|^2_{L^2_\Omega(\mathcal{X})} + \|\mathcal{P}_A^\perp[e(\cdot;\theta_*)]\|^2_{L^2_\Omega(\mathcal{X})}. \tag{A5}$$

Using (A4) and (A5) we get the inequality $d_Z^2 \leqslant \delta^2 - \|\mathcal{P}_A^\perp[e(\cdot;\theta_*)]\|^2_{L^2_\Omega(\mathcal{X})}$, which yields the lower bound for the difference of the squared HJ-distances in Proposition 2.

The Pythagorean theorem for inner product spaces implies also that

$$\|e(\cdot;\theta_Z)\|^2_{L^2_\Omega(\mathcal{X})} = \|\mathcal{P}_A[e(\cdot;\theta_Z)]\|^2_{L^2_\Omega(\mathcal{X})} + \|\mathcal{P}_A^\perp[e(\cdot;\theta_Z)]\|^2_{L^2_\Omega(\mathcal{X})} = d_Z^2 + \|\mathcal{P}_A^\perp[e(\cdot;\theta_Z)]\|^2_{L^2_\Omega(\mathcal{X})}, \tag{A6}$$

where we use the expression for d_Z given in Equation (14). The criterion $\|e(\cdot;\theta)\|^2_{L^2_\Omega(\mathcal{X})}$ valued at its global minimum θ_* is equal to δ^2, and cannot be greater than the same criterion at any other argument $\theta\in\Theta$. In particular we have $\delta^2 \leqslant \|e(\cdot;\theta_Z)\|^2_{L^2_\Omega(\mathcal{X})}$. From this inequality and Equation (A6) we get $\delta^2 \leqslant d_Z^2 + \|\mathcal{P}_A^\perp[e(\cdot;\theta_Z)]\|^2_{L^2_\Omega(\mathcal{X})}$, which yields the upper bound for the difference of the squared HJ-distances in Proposition 2.

A.4 Proof of Proposition 3

The proof of Proposition 3 is organized in five steps.

1. *Conditional and unconditional HJ-distances for linear SDF families.*

We start by particularizing the formulas of the conditional and unconditional HJ-distances to the case of an SDF that is linear in the risk factors. From Equations (12) and (16) the squared conditional HJ-distance is

$$\delta^2 = \min_{\theta\in\Theta} \|B\theta - 1_N\|^2_{L^2_\Omega(\mathcal{X})} = \|B\theta_* - 1_N\|^2_{L^2_\Omega(\mathcal{X})}. \tag{A7}$$

The minimizer of the quadratic optimization problem in Equation (A7) is given by

$$\theta_* = \langle B, B \rangle^{-1}_{L^2_\Omega(\mathcal{X})} \langle B, 1_N \rangle_{L^2_\Omega(\mathcal{X})}. \tag{A8}$$

Similarly, from Equation (14) the squared unconditional HJ-distance is

$$d_Z^2 = \min_{\theta \in \Theta} \|\mathcal{P}_A[B\theta - 1_N]\|^2_{L^2_\Omega(\mathcal{X})} = \|\mathcal{P}_A[B\theta_Z - 1_N]\|^2_{L^2_\Omega(\mathcal{X})}, \tag{A9}$$

where the minimizer θ_Z is given by

$$\theta_Z = \langle \mathcal{P}_A[B], \mathcal{P}_A[B] \rangle^{-1}_{L^2_\Omega(\mathcal{X})} \langle \mathcal{P}_A[B], \mathcal{P}_A[1_N] \rangle_{L^2_\Omega(\mathcal{X})}. \tag{A10}$$

From Equations (A8) and (A10) we have

$$B\theta_* = \mathcal{P}_B[1_N] \qquad \text{and} \qquad \mathcal{P}_A[B\theta_Z] = \mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_A[1_N], \tag{A11}$$

where the symbol $\circ$ denotes operator composition, and we use the notation for projection operators given in Equation (A3). Using into Equation (16) the expression for $B\theta_*$ given in Equations in (A11), we obtain the following expression for the conditional pricing error vector at date t for the parameter value θ_*:

$$e(X_t; \theta_*) = B(X_t)\theta_* - 1_N = \mathcal{P}_B[1_N](X_t) - 1_N = -\mathcal{P}_B^\perp[1_N](X_t). \tag{A12}$$

From Equation (A7) the squared conditional HJ-distance is

$$\delta^2 = \|\mathcal{P}_B^\perp[1_N]\|^2_{L^2_\Omega(\mathcal{X})}. \tag{A13}$$

From the second of Equations (A11), $\mathcal{P}_A[B\theta_Z - 1_N] = \mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_A[1_N] - \mathcal{P}_A[1_N]$, so that $\mathcal{P}_A[B\theta_Z - 1_N] = -\mathcal{P}^\perp_{\mathcal{P}_A[B]} \circ \mathcal{P}_A[1_N]$. The squared unconditional HJ-distance is $d_Z^2 = \|\mathcal{P}^\perp_{\mathcal{P}_A[B]} \circ \mathcal{P}_A[1_N]\|^2_{L^2_\Omega(\mathcal{X})}$ from Equation (A9). This expression corresponds to the one for the squared conditional HJ-distance in Equation (A13), with $\mathcal{P}_A[B]$ instead of B, and $\mathcal{P}_A[1_N]$ instead of 1_N.

2. *Proof of Equation (17).*

From the Pythagorean theorem for inner product spaces we have

$$\|B\theta - 1_N\|^2_{L^2_\Omega(\mathcal{X})} = \|\mathcal{P}_A[B\theta - 1_N]\|^2_{L^2_\Omega(\mathcal{X})} + \|\mathcal{P}_A^\perp[B\theta - 1_N]\|^2_{L^2_\Omega(\mathcal{X})}, \tag{A14}$$

for any $\theta \in \Theta$. The first term in the RHS of the last equation is a quadratic function of vector θ, minimized at θ_Z, where it assumes value d_Z^2 (see Equation (A9)), so that

$$\|\mathcal{P}_A[B\theta - 1_N]\|^2_{L^2_\Omega(\mathcal{X})} = d_Z^2 + \|\mathcal{P}_A[B(\theta - \theta_Z)]\|^2_{L^2_\Omega(\mathcal{X})}, \tag{A15}$$

for any $\theta \in \Theta$. By using Equation (A15) into Equation (A14) we get

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

for any $\theta \in \Theta$. By evaluating this equation at $\theta = \theta_*$, and using Equation (A7), we get

$$\delta^2 = d_Z^2 + \|\mathcal{P}_A[B(\theta_* - \theta_Z)]\|^2_{L^2_\Omega(\mathcal{X})} + \|\mathcal{P}_A^\perp[B\theta_* - 1_N]\|^2_{L^2_\Omega(\mathcal{X})}. \tag{A16}$$

Let us now rewrite the two terms in the RHS of Equation (A16). From Equation (A12) we get

$$\mathcal{P}_A^{\perp}[B\theta_* - 1_N] = -\mathcal{P}_A^{\perp} \circ \mathcal{P}_B^{\perp}[1_N]. \tag{A17}$$

From Equations in (A11) 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

$$\begin{aligned}\mathcal{P}_A[B(\theta_* - \theta_Z)] &= \mathcal{P}_A \circ \mathcal{P}_B[1_N] - \mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_A[1_N] \\ &= \mathcal{P}_A \circ \mathcal{P}_B[1_N] - \mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_A \circ \left(\mathcal{P}_B[1_N] + \mathcal{P}_B^{\perp}[1_N]\right) \\ &= \mathcal{P}^{\perp}_{\mathcal{P}_A[B]} \circ \mathcal{P}_A \circ \mathcal{P}_B[1_N] - \mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_A \circ \mathcal{P}_B^{\perp}[1_N] = -\mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_B^{\perp}[1_N].\end{aligned} \tag{A18}$$

Then, using Equations (A17) and (A18) into Equation (A16) we get

$$\delta^2 - d_Z^2 = \|\mathcal{P}_A^{\perp} \circ \mathcal{P}_B^{\perp}[1_N]\|^2_{L^2_\Omega(\mathcal{X})} + \|\mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_B^{\perp}[1_N]\|^2_{L^2_\Omega(\mathcal{X})}. \tag{A19}$$

Considering that $\mathcal{P}_B^{\perp}[1_N]$ is the opposite of the conditional pricing error vector $e(\cdot;\theta_*)$ from Equation (A12), we get Equation (17).

3. *Proof of Equation (18).*
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. Thus, $\|B\theta - 1\|^2_{L^2_\Omega(\mathcal{X})} = \delta^2 + \|B(\theta^* - \theta)\|^2_{L^2_\Omega(\mathcal{X})}$. From Equation (A14) we get $\delta^2 + \|B(\theta_* - \theta)\|^2_{L^2_\Omega(\mathcal{X})} = \|\mathcal{P}_A[B\theta - 1_N]\|^2_{L^2_\Omega(\mathcal{X})} + \|\mathcal{P}_A^{\perp}[B\theta - 1_N]\|^2_{L^2_\Omega(\mathcal{X})}$ for any $\theta \in \Theta$. Evaluating this equation at $\theta = \theta_Z$ and using Equation (A9), we then obtain that $\delta^2 = d_Z^2 + \|\mathcal{P}_A^{\perp}[B\theta_Z - 1_N]\|^2_{L^2_\Omega(\mathcal{X})} - \|B(\theta_* - \theta_Z)\|^2_{L^2_\Omega(\mathcal{X})}$. Considering that $B\theta^* - 1_N = e(\cdot;\theta^*)$ and $B(\theta_* - \theta_Z) = e(\cdot;\theta_*) - e(\cdot;\theta_Z)$, we get Equation (18).
4. *Proof of the first equivalence in Equations (19).*
We have $\delta = d_Z$ if, and only if, the two norms in the RHS of Equation (A19) are null, which is equivalent to:

$$\begin{cases} \mathcal{P}_A^{\perp} \circ \mathcal{P}_B^{\perp}[1_N] = 0_N, \\ \mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_B^{\perp}[1_N] = 0_N. \end{cases} \tag{A20}$$

From Equation (A12), the first condition corresponds to

$$\mathcal{P}_A^{\perp}[e(\cdot;\theta_*)] = 0_N. \tag{A21}$$

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

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

Therefore, the first equation in System (A20) implies the second one (but not necessarily the other way around). Hence, the System (A20) is equivalent to Equation (A21).

5. *Proof of the second equivalence in Equations (19).*

From Proposition 2, the condition $\mathcal{P}_A^{\perp}[e(\cdot;\theta_Z)] = 0_N$ implies Equation (A21). Let us now show that the reverse implication holds. If Equation (A21) holds, we have that

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

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 Equation (A10) we get

$$\begin{aligned}\theta_Z &= \langle B, \mathcal{P}_A[B]\rangle^{-1}_{L^2_\Omega(\mathcal{X})}\langle B, \mathcal{P}_A[1_N]\rangle_{L^2_\Omega(\mathcal{X})} \\ &= \langle B, \mathcal{P}_A[B]\rangle^{-1}_{L^2_\Omega(\mathcal{X})}\langle B, \mathcal{P}_A \circ \mathcal{P}_B[1_N]\rangle_{L^2_\Omega(\mathcal{X})} = \langle B, B\rangle^{-1}_{L^2_\Omega(\mathcal{X})}\langle B, 1_N\rangle_{L^2_\Omega(\mathcal{X})},\end{aligned} \tag{A22}$$

where the last equality is proved by making explicit the projections $\mathcal{P}_A[B]$ and $\mathcal{P}_A \circ \mathcal{P}_B[1_N]$. Then, by using Equations (A8) and (A22) we get $\theta_Z = \theta_*$, and under Equation (A21) we get $\mathcal{P}_A^{\perp}[e(\cdot;\theta_Z)] = 0_N$.

A.5 Proof of Equation (21)

We follow similar steps as in Appendix A.1. Let us consider the inner minimization problem that defines the squared distance $\delta^2_{[\mathrm{m}]}$. For any given value of parameter $\theta \in \Theta$, the Lagrangian function $\mathcal{L}_m$ for the constrained minimization w.r.t. the admissible SDF $M_{t,t+1} \in \mathcal{M}_{[\mathrm{m}]}$ is given by

$$\begin{aligned}\mathcal{L}_m = {} & \mathrm{E}\left[\left(m(Y_{t+1};\theta) - M_{t,t+1}\right)^2\right] + 2\mathrm{E}\left[\lambda_m(X_t)'(M_{t,t+1}R_{t+1} - 1_N)\right] \\ & + 2\mathrm{E}\left[\lambda_f(X_t)\left(M_{t,t+1} - R^{-1}_{f,t+1}\right)\right],\end{aligned}$$

where $R_{f,t+1}$ is the gross return on the risk-free asset from date t to date $t+1$, $\lambda_m : \mathcal{X} \to \mathbb{R}^N$ is a N-dimensional functional Lagrange multipliers vector for the restrictions on the excess returns, $\lambda_f : \mathcal{X} \to \mathbb{R}$ is a functional Lagrange multiplier scalar for the restriction on the reference risk-free asset, and all the other elements are as in Appendix A.1. The Lagrangian function $\mathcal{L}_m$ can then be written as the sum of

$$\mathrm{E}\left[\left(M_{t,t+1} - m(Y_{t+1};\theta) + \lambda_m(X_t)'R_{t+1} + \lambda_f(X_t)\right)^2\right] \tag{A23}$$

and a term that is independent of the admissible SDF $M_{t,t+1}$. Then, the random variable $M^{[\mathrm{opt}]}_{t,t+1}$ solving the FOC for optimizing the Lagrangian w.r.t. $M_{t,t+1}$ is defined as

$$M^{[\mathrm{opt}]}_{t,t+1} := m(Y_{t+1};\theta) - \lambda^{[\mathrm{opt}]}_m(X_t)'R_{t+1} - \lambda^{[\mathrm{opt}]}_f(X_t). \tag{A24}$$

The SDF $M^{[\mathrm{opt}]}_{t,t+1}$ satisfies the conditional restrictions that define $\mathcal{M}_{[\mathrm{m}]}$ in Equation (20). The condition $\mathrm{E}\left[M_{t,t+1}|X_t\right] = R^{-1}_{f,t+1}$ together with the assumption $\mathrm{E}[m(Y_{t+1};\theta)|X_t] = R^{-1}_{f,t+1}$ for any $\theta \in \Theta$, imply that $\lambda^{[\mathrm{opt}]}_f(X_t) = -\lambda^{[\mathrm{opt}]}_m(X_t)'\mathrm{E}[R_{t+1}|X_t]$. Plugging this equation into Equation (A24) we get

$$M^{[\mathrm{opt}]}_{t,t+1} := m(Y_{t+1};\theta) - \lambda^{[\mathrm{opt}]}_m(X_t)'(R_{t+1} - \mathrm{E}[R_{t+1}|X_t]). \tag{A25}$$

The condition $\mathrm{E}\left[M^{[\mathrm{opt}]}_{t,t+1}R_{t,t+1}|X_t\right] = 1_N$ yields

$$\lambda_m^{[\text{opt}]}(X_t) = \mathrm{V}[R_{t+1}|X_t]^{-1}\left(\mathrm{E}\left[m(Y_{t+1};\theta)R_{t,t+1}|X_t\right] - 1_N\right)$$
$$= \Omega_{[\text{m}]}(X_t)\mathrm{E}\left[m(Y_{t+1};\theta)\left(R_{t,t+1} - R_{f,t+1}1_N\right)|X_t\right] = \Omega_{[\text{m}]}(X_t)e(X_t;\theta). \quad \text{(A26)}$$

By replacing Equation (A25) and the last of Equations (A26) into the objective function $\mathrm{E}\left[\left(m(Y_{t+1};\theta) - M_{t,t+1}\right)^2\right]$, and then minimizing w.r.t. $\theta \in \Theta$, we get

$$\delta_{[\text{m}]}^2 = \min_{\theta\in\Theta} \mathrm{E}\left[\lambda_m^{[\text{opt}]}(X_t)'\Omega_{[\text{m}]}(X_t)^{-1}\lambda_m^{[\text{opt}]}(X_t)\right] = \min_{\theta\in\Theta} \mathrm{E}\left[e(X_t;\theta)'\Omega_{[\text{m}]}(X_t)e(X_t;\theta)\right].$$

A.6 Proof of Equation (23)

The conditional pricing restrictions that define the set $\mathcal{M}_{[\text{m}]}$ in Equation (20) imply the following expression for the conditional expected gross returns vector given the conditioning information vector X_t:

$$\mathrm{E}[R_{t+1}|X_t] = R_{f,t+1}1_N - R_{f,t+1}\mathrm{Cov}\left[R_{t+1}, M_{t,t+1}^{[\text{opt}]}|X_t\right]. \quad \text{(A27)}$$

From Equations (22), (A24), and (A27), the conditional prediction error vector given the conditioning information vector X_t is $\mathrm{E}[R_{t+1}|X_t] - \mathrm{E}_\theta[R_{t+1}|X_t] = R_{f,t+1}\mathrm{Cov}[R_{t+1}, \lambda_m^{[\text{opt}]}(X_t)'R_{t+1}|X_t]$, where $\lambda_m^{[\text{opt}]}$ is the optimal Lagrange multipliers vector defined in Equation (A26). Let us now consider the set Z of managed portfolios constructed by means of an instrument vector $Z(X_t)$. That is, let us consider the case $q = 1$ for the generic matrix Z introduced in the introduction of the article. From the Cauchy–Schwarz inequality applied to the counterpart of the last equation written for a single portfolio we have that

$$|Z(X_t)'(\mathrm{E}[R_{t+1}|X_t] - \mathrm{E}_\theta[R_{t+1}|X_t])| = R_{f,t+1}|\mathrm{Cov}\left[Z(X_t)'R_{t+1}, \lambda_m^{[\text{opt}]}(X_t)'R_{t+1}|X_t\right]|$$
$$\leqslant R_{f,t+1}\sqrt{\mathrm{V}\left[\lambda_m^{[\text{opt}]}(X_t)'R_{t+1}|X_t\right]}\sqrt{\mathrm{V}\left[Z(X_t)'R_{t+1}|X_t\right]}.$$

In case $Z \equiv \lambda_m^{[\text{opt}]}$, the last inequality holds as an equation, and, from the expression of the optimal Lagrange multiplier $\lambda_m^{[\text{opt}]}$ in Equation (A26), we have

$$|\lambda_m^{[\text{opt}]}(X_t)'(\mathrm{E}[R_{t+1}|X_t] - \mathrm{E}_\theta[R_{t+1}|X_t])| = R_{f,t+1}\mathrm{V}\left[\lambda_m^{[\text{opt}]}(X_t)'R_{t+1}|X_t\right]$$
$$= R_{f,t+1}e(X_t;\theta)'\Omega_{[\text{m}]}(X_t)e(X_t;\theta).$$

Thus, we have shown that

$$\max_{Z\in Z} \frac{|Z(X_t)'(\mathrm{E}[R_{t+1}|X_t] - \mathrm{E}_\theta[R_{t+1}|X_t])|}{\sqrt{\mathrm{V}\left[Z(X_t)'R_{t+1}|X_t\right]}} = R_{f,t+1}\sqrt{e(X_t;\theta)'\Omega_{[\text{m}]}(X_t)e(X_t;\theta)}.$$

By multiplying both sides of this equation by $R_{f,t+1}^{-1}$, squaring them, taking their unconditional expectation, and finally valuing them at the pseudo-true SDF parameter $\theta = \theta_{[\text{m}]*}$, we obtain Equation (23).

A.7 Proof of Equation (25)

The conditional pricing error vector for the augmented vector of gross returns $R_t^{[\mathrm{Aug}]} := [f_t' \, R_t']'$ is

$$e^{[\mathrm{Aug}]}(X_t;\theta) = \mathrm{E}\left[m(Y_{t+1};\theta)R_{t+1}^{[\mathrm{Aug}]}|X_t\right] = \mathrm{E}[m(Y_{t+1};\theta)|X_t]\mathrm{E}\left[R_{t+1}^{[\mathrm{Aug}]}|X_t\right] + \mathrm{Cov}\left[R_{t+1}^{[\mathrm{Aug}]}, m(Y_{t+1},\theta)|X_t\right] = \frac{1}{R_{f,t+1}}(a(X_t) - \Lambda(X_t)\theta), \tag{A28}$$

where the $(N+k)$-dimensional vector function a and the $((N+k)\times k)$-dimensional matrix function Λ are defined as

$$a(X_t) := \begin{bmatrix} \mathbf{0}_{(k\times 1)} \\ \alpha(X_t) \end{bmatrix} + \begin{bmatrix} \mathbf{I}_k \\ B(X_t) \end{bmatrix} \mathrm{E}[f_{t+1}|X_t], \qquad \Lambda(X_t) := -R_{f,t+1}\begin{bmatrix} \mathbf{I}_k \\ B(X_t) \end{bmatrix} \mathrm{V}[f_{t+1}|X_t]. \tag{A29}$$

We denote the inverse of the $((N+k)\times(N+k))$-dimensional conditional variance–covariance matrix of the augmented vector of excess returns $R_{t+1}^{[\mathrm{Aug}]}$ given the conditioning information vector X_t by $\Omega_{[\mathrm{m}]}^{[\mathrm{Aug}]}(X_t) := \mathrm{V}\left[R_{t+1}^{[\mathrm{Aug}]}|X_t\right]^{-1}$. From the conditional factor structure and the block matrix inverse formula, we have

$$\Omega_{[\mathrm{m}]}^{[\mathrm{Aug}]}(X_t) = \begin{bmatrix} \mathrm{V}[f_{t+1}|X_t]^{-1} + B(X_t)'\mathrm{V}[\epsilon_{t+1}|X_t]^{-1}B(X_t) & -B(X_t)'\mathrm{V}[\epsilon_{t+1}|X_t]^{-1} \\ -\mathrm{V}[\epsilon_{t+1}|X_t]^{-1}B(X_t) & \mathrm{V}[\epsilon_{t+1}|X_t]^{-1} \end{bmatrix}. \tag{A30}$$

Let us plug the last of Equations (A28) in Equation (21). In this way, the squared conditional modified HJ-distance for the model for test assets and tradable factors can be written as $\delta_{[\mathrm{m}]}^{[\mathrm{Aug}]\ 2} = \min_{\theta\in\Theta} \mathrm{E}\left[\frac{1}{R_{f,t+1}^2}(a(X_t) - \Lambda(X_t)\theta)'\Omega_{[\mathrm{m}]}^{[\mathrm{Aug}]}(X_t)(a(X_t)-\Lambda(X_t)\theta)\right]$. The minimizer of the criterion is $\theta_{[\mathrm{m}]*} = \mathrm{E}\left[\frac{1}{R_{f,t+1}^2}\Lambda(X_t)'\Omega_{[\mathrm{m}]}^{[\mathrm{Aug}]}(X_t)\Lambda(X_t)\right]^{-1}\mathrm{E}\left[\frac{1}{R_{f,t+1}^2}\Lambda(X_t)'\Omega_{[\mathrm{m}]}^{[\mathrm{Aug}]}(X_t)a(X_t)\right]$, in analogy to the minimization of a Generalized Least Squares (GLS) criterion. The minimized value of the criterion becomes $\delta_{[\mathrm{m}]}^{[\mathrm{Aug}]2} = \mathrm{E}\left[\frac{1}{R_{f,t+1}^2}a(X_t)'M(X_t)a(X_t)\right]$, for the $((N+k)\times(N+k))$-dimensional matrix

$$M(X_t) := \Omega_{[\mathrm{m}]}^{[\mathrm{Aug}]}(X_t) - \frac{1}{R_{f,t+1}^2}\Omega_{[\mathrm{m}]}^{[\mathrm{Aug}]}(X_t)\Lambda(X_t)\mathrm{E}\left[\frac{1}{R_{f,t+1}^2}\Lambda(X_t)'\Omega_{[\mathrm{m}]}^{[\mathrm{Aug}]}(X_t)\Lambda(X_t)\right]^{-1}\Lambda(X_t)'\Omega_{[\mathrm{m}]}^{[\mathrm{Aug}]}(X_t).$$

From Equations (A29) and (A30) we have that $\Omega_{[\mathrm{m}]}^{[\mathrm{Aug}]}(X_t)\Lambda(X_t) = -R_{f,t+1}\left[\mathbf{I}_k \ 0_{N\times k}'\right]'$, and in turn that $\mathrm{E}\left[\frac{1}{R_{f,t+1}^2}\Lambda(X_t)'\Omega_{[\mathrm{m}]}^{[\mathrm{Aug}]}(X_t)\Lambda(X_t)\right] = \mathrm{E}[\mathrm{V}[f_{t+1}|X_t]]$. From the last two equations, the matrix M can be expressed as

$$\begin{bmatrix} \mathrm{V}[f_{t+1}|X_t]^{-1} - \mathrm{E}[\mathrm{V}[f_{t+1}|X_t]]^{-1} + B(X_t)'\mathrm{V}[\epsilon_{t+1}|X_t]^{-1}B(X_t) & -B(X_t)'\mathrm{V}[\epsilon_{t+1}|X_t]^{-1} \\ -\mathrm{V}[\epsilon_{t+1}|X_t]^{-1}B(X_t) & \mathrm{V}[\epsilon_{t+1}|X_t]^{-1} \end{bmatrix}.$$

From the definition of vector a in Equation (A29) and the fact that

$$M(X_t)\begin{bmatrix} \mathbf{I}_k \\ B(X_t) \end{bmatrix} = \begin{bmatrix} \mathrm{V}[f_{t+1}|X_t]^{-1} - \mathrm{E}[\mathrm{V}[f_{t+1}|X_t]]^{-1} \\ 0_{N\times k} \end{bmatrix},$$

we obtain

$$\begin{aligned} a(X_t)'M(X_t)a(X_t) &= \alpha(X_t)'\mathrm{V}[\epsilon_{t+1}|X_t]^{-1}\alpha(X_t) \\ &\quad + \mathrm{E}[f_{t+1}|X_t]'\left(\mathrm{V}[f_{t+1}|X_t]^{-1} - \mathrm{E}[\mathrm{V}[f_{t+1}|X_t]]^{-1}\right)\mathrm{E}[f_{t+1}|X_t]. \end{aligned}$$

Therefore, by dividing both sides of this equation by $R^2_{f,t+1}$ and taking the unconditional expectation we get Equation (25).

Appendix B. Regularity Assumptions for Large Sample Results

In this Appendix, we list the regularity assumptions used in Appendices C and D to derive the large sample properties of $\hat{\theta}_T$, which is an estimator of either the true SDF parameter value θ_0 or the pseudo-true value θ_*, and of the squared sample conditional HJ-distance $\hat{\delta}^2_T$. For a matrix A we denote by $||A|| = \sqrt{\mathrm{Tr}[AA']}$ its Frobenius norm, which coincides with the standard Euclidean norm for a vector.

Assumption 1. *The stochastic process for the variable* $W_t = \left[Y_t' \, R_t' \, X_{t-1}'\right]'$ *with support* $\mathcal{W} = \mathcal{Y} \times \mathcal{R} \times \mathcal{X}$ *is strictly stationary and strong mixing with mixing coefficients* $\alpha(h) = O\left(\rho^h\right)$ *for* $h \in \mathbb{N}$ *and* $\rho \in (0,1)$.

Assumption 2. *The stationary pdf* $f_X(\cdot)$ *of process* $\{X_t\}$ *and the* $(N\times N)$*-dimensional matrix function* $\mathrm{E}\left[h(Y_{t+1}, R_{t+1};\theta)h(Y_{t+1}, R_{t+1};\theta)'|X_t\right]$ *are of differentiability class* $C^1(\mathcal{X})$*, for any* $\theta \in \Theta$*, where* $h(Y_{t+1}, R_{t+1};\theta) := m(Y_{t+1};\theta)R_{t+1} - 1_N$.

Assumption 3. *The compact set* $\mathcal{X}_\star$ *belongs to 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* $\mathrm{E}\left[\|h(Y_{t+1}, R_{t+1};\theta)\|^n|X_t = x\right]$ *is bounded on* $\mathcal{X}_\star$*, uniformly in* $\theta\in\Theta$*, for a positive constant* n*, that is* $\sup_{\theta\in\Theta}\sup_{x\in\mathcal{X}_\star}\mathrm{E}\left[||h(Y_{t+1}, R_{t+1};\theta)||^n|X_t = x\right] < \infty$.

Assumption 5. *The bandwidth* $b_T > 0$ *converges to 0 as* $T\to\infty$ *such that* $\sqrt{T}b_T^2 = o(1)$ *and* $\frac{\log\,(T)}{Tb_T^L} = o(1)$.

Assumption 6. *The kernel function* K *is such that* $K(x) = K(-x) \geqslant 0$, $\int_{\mathcal{X}} K(u)du = 1$, $\int_{\mathcal{X}} K(u)^2 du < \infty$*, and* $\int_{\mathcal{X}} ||u||^2 K(u)du < \infty$*, for any* $x \in \mathcal{X}$.

Assumption 7. *The true and pseudo-true parameter values θ_0 and $\theta_\star$ belong to the interior of compact set $\Theta \subseteq \mathbb{R}^p$.*

Assumption 8. *The vector function $h(Y_t, R_t; \theta)$ is of differentiability class $C^1(\Theta)$, and it is such that $\mathrm{E}\left[\overline{h}_2(Y_t, R_t)\right] < \infty$, for the function $\overline{h}_2(Y_t, R_t) := \sup_{\theta\in\Theta} \|h(Y_t, R_t; \theta)\|^2$.*

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 $x \in \mathcal{X}$.*

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 the mapping $\theta \mapsto h(Y_t, R_t; \theta)$ is twice continuously differentiable, and such that $\sup_{\theta\in\mathcal{N}_0} |\nabla_{\theta_{[j]}} h_{[i]}(Y_t, R_t; \theta)| \leqslant l_1(Y_t, R_t)$ and $\sup_{\theta\in\mathcal{N}_0} |\nabla_{\theta_{[j]}\theta_{[k]}} h_{[i]}(Y_t, R_t; \theta)| \leqslant l_2(Y_t, R_t)$, $\mathbb{P}$-a.s., for some real-valued function l_1 of vector $\left[Y_t'\, R_t'\right]'$ such that $\mathrm{E}\left[|l_1(Y_t, R_t)|^\eta\right] < \infty$, for $\eta \geqslant 6$, and $\mathrm{E}\left[l_2(Y_t, R_t)^2\right] < \infty$.*

Assumption 11. *The matrix function $J(x; \theta) = \mathrm{E}\left[\nabla_{\theta'} h(Y_{t+1}, R_{t+1}; \theta) | X_t = x\right]$ exists and it is full rank when valued at $\theta = \theta_0$ and $\theta = \theta_\star$, so that* $\mathrm{rank}[J(x; \theta_0)] = \mathrm{rank}[J(x; \theta_\star)] = p$, *for any $x \in \mathcal{X}$.*

Assumption 12. *For any pair of SDF families $\mathcal{F}$ and $\tilde{\mathcal{F}}$, the joint bivariate process for the variables $\phi(W_t; \theta_\star)$ and $\tilde{\phi}\left(W_{t+1}; \tilde{\theta}_\star\right)$ defined as in Proposition 5 is such that*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{bmatrix} \phi(W_{t+1}; \theta_\star) - \mathrm{E}[\phi(W_{t+1}; \theta_\star)] \\ \tilde{\phi}\left(W_{t+1}; \tilde{\theta}_\star\right) - \mathrm{E}\left[\tilde{\phi}\left(W_{t+1}; \tilde{\theta}_\star\right)\right] \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0_2, \begin{bmatrix} \sigma_\phi^2 & \sigma_{\phi,\tilde{\phi}} \\ \sigma_{\phi,\tilde{\phi}} & \sigma_{\tilde{\phi}}^2 \end{bmatrix}\right).$$

Assumption 13. *The r-dimensional restriction vector function ψ used to define the Wald test statistic considered in Proposition 7 is of differentiability class $C^1(\Theta)$, and the $(p \times r)$-dimensional Jacobian matrix $\nabla_\theta \psi(\theta)'$ of the restriction vector has full rank $r \leqslant p$.*

Assumptions 1–6 are standard in nonparametric analysis and yield the uniform convergence of kernel estimators over the set $\mathcal{X}_\star$. By adopting Assumption 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. Assumption 5 on the bandwidth allows to simplify the expression of the asymptotic distribution of the estimators under model misspecification. The condition $\sqrt{T} b_T^2 = o(1)$ is used to remove a bias term and can be relaxed using higher-order kernels. Assumption 9 implies that matrix $\Omega(x)$ is positive definite for any $x \in \mathcal{X}$, and $\|\Omega(x)\|$ is bounded on $\mathcal{X}$. Assumption 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 is a conditional local identification assumption. The matrix function J represents the local sensitivity of the pricing error vector to a change in the SDF parameter vector θ. We use Assumptions 12 and 13 to derive the asymptotic behavior of the sample conditional HJ-distance for a misspecified model and the Wald test statistic used to compare nested model specifications.

Appendix C. Large Sample Results for Estimator $\hat{\theta}_T$

In this section, we prove the large sample properties of the estimator $\hat{\theta}_T$ of the minimizer of the criterion defining the conditional HJ-distance. In this section and in the following Appendix D we use the symbol $\stackrel{\mathbb{P}}{\to}$ to denote convergence in probability.

C.1 Large sample properties of $\hat{\theta}_T$ under correct specification

We first provide the large sample properties for the estimator $\hat{\theta}_T$ of the SDF parameter θ in a correctly specified family. These properties are derived for example in Gospodinov and Otsu (2012).

Lemma 1. *Under Assumptions 1–10, as $T \to \infty$ estimator $\hat{\theta}_T$ is consistent for the true parameter θ_0, that is, $\hat{\theta} \stackrel{\mathbb{P}}{\to} \theta_0$. Moreover, it is asymptotically normal with $\sqrt{T}$-rate of convergence: $\sqrt{T}(\hat{\theta}_T - \theta_0) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0_p, \Sigma_0)$, for the matrix $\Sigma_0 := D_0^{-1} S_0 D_0^{-1}$, defined by $S_0 := \mathrm{E}[\mathbf{1}(X_t) J_0(X_t)' \Omega(X_t) V_0(X_t) \Omega(X_t) J_0(X_t)]$ and $D_0 := \mathrm{E}[\mathbf{1}(X_t) J_0(X_t)' \Omega(X_t) J_0(X_t)]$, for $J_0(x) := J(x;\theta_0)$.*

Proof. See Section E.1 of the Online Appendix. □

C.2 Large sample properties of $\hat{\theta}_T$ under model misspecification

We give in this section the asymptotic properties of the estimator $\hat{\theta}_T$ of the pseudo-true SDF parameter vector $\theta_\star$ in a misspecified family.

Lemma 2. *Under Assumptions 1–10, as $T \to \infty$ estimator $\hat{\theta}_T$ is consistent for $\theta_\star$, that is, $\hat{\theta} \stackrel{\mathbb{P}}{\to} \theta_\star$. Moreover, it is asymptotically normal with $\sqrt{T}$-rate of convergence: $\sqrt{T}(\hat{\theta}_T - \theta_\star) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0_p, \Sigma_\star)$, for the matrix $\Sigma_\star := D_\star^{-1} S_\star D_\star^{-1}$ defined by $D_\star := \mathrm{E}[\mathbf{1}(X_t) J_\star(X_t)' \Omega(X_t) J_\star(X_t)] + \mathrm{E}\left[\mathbf{1}(X_t) \frac{\partial m(Y_{t+1};\theta_\star)}{\partial\theta\partial\theta'} R'_{t+1} \Omega(X_t) e(X_t;\theta_\star)\right]$ and $S_\star := \sum_{l=-\infty}^{\infty} \mathrm{Cov}[u_t, u_{t-l}]$, for $J_\star(x) := J(x;\theta_\star)$ and*

$$u_t := \mathbf{1}(X_t) J_\star(X_t)' \Omega(X_t) h(Y_{t+1}, R_{t+1};\theta_\star)$$
$$-\mathbf{1}(X_t) J_\star(X_t)' \Omega(X_t) R_{t+1} R'_{t+1} \Omega(X_t) e(X_t;\theta_\star) + \mathbf{1}(X_t) \frac{\partial m(Y_{t+1};\theta_\star)}{\partial\theta} R'_{t+1} \Omega(X_t) e(X_t;\theta_\star).$$

Proof. See Section E.2 of the Online Appendix. □

Lemma 2 is the counterpart of the results on the large sample distribution of the GMM estimator of a parameter in misspecified models derived by Hall and Inoue (2003) (see also Anatolyev and Gospodinov, 2011), when this parameter is identified by a set of conditional moment restrictions. We can write u_t as

$$\begin{aligned} u_t = {} & \mathbf{1}(X_t) J_\star(X_t)' \Omega(X_t) e(X_t;\theta_\star) + \mathbf{1}(X_t) J_\star(X_t)' \Omega(X_t) [h(Y_{t+1}, R_{t+1};\theta_\star) - e(X_t;\theta_\star)] \\ & - \mathbf{1}(X_t) J_\star(X_t)' \Omega(X_t) \left[R_{t+1} R_{t+1}{}' - \Omega(X_t)^{-1}\right] \Omega(X_t) e(X_t;\theta_\star) \\ & + \mathbf{1}(X_t) \left[\frac{\partial m(Y_{t+1};\theta_\star)}{\partial\theta} R'_{t+1} - J_\star(X_t)'\right] \Omega(X_t) e(X_t;\theta_\star). \end{aligned}$$

The first term in the RHS has zero expectation (from the FOC of the population estimation problem), while the other terms have null conditional expectation given X_t. Due to the first term, process $\{u_t\}$ is serially correlated, which explains the autocovariances in the long-run variance matrix $S_\star$. Compared to the asymptotic distribution under correct model specification, both the Hessian matrix $D_\star$ and the variance matrix $S_\star$ contain additional terms that are induced by the nonvanishing conditional pricing error $e(X_t;\theta_\star)$.

Appendix D. Proofs of Large Sample Results for the Conditional HJ-Distance

In this Appendix, we prove the large sample properties of the conditional HJ-distance.

D.1 Proof of Proposition 4

In this section, we prove the asymptotic normality of the squared sample conditional HJ-distance $\hat{\delta}_T^2$ under correct model specification stated in Proposition 4. Given the form of the Nadaraya–Watson estimator in Equation (26), we can write the criterion $\mathcal{Q}_T$ in (28) as a weighted sum of quadratic forms of the conditional moment vector h:

$$\mathcal{Q}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{T-1}\sum_{j=1}^{T-1}\mathbf{1}(X_t)w(X_t,X_i)w\left(X_t,X_j\right)h(Y_{i+1},R_{i+1};\theta)'\hat{\Omega}_T(X_t)h\left(Y_{j+1},R_{j+1};\theta\right),$$

for $\theta\in\Theta$. Let us now decompose the criterion $\mathcal{Q}_T$ similar to Tripathi and Kitamura (2003):

$$\mathcal{Q}_T(\theta) = \sum_{i=1}^{5}\mathcal{Q}_{i,T}(\theta), \tag{A31}$$

where

$$\mathcal{Q}_{1,T}(\theta) := \frac{1}{T}\sum_{t=1}^{T-1}\mathbf{1}(X_t)w(X_t,X_t)^2 h(Y_{t+1},R_{t+1}\theta)\hat{\Omega}_T(X_t)h(Y_{t+1},R_{t+1};\theta),$$

$$\mathcal{Q}_{2,T}(\theta) := \frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{i=1\\ i\neq t}}^{T-1}\mathbf{1}(X_t)w(X_t,X_i)^2 h(Y_{i+1},R_{i+1};\theta)\hat{\Omega}_T(X_t)h(Y_{i+1},R_{i+1};\theta),$$

$$\mathcal{Q}_{3,T}(\theta) := \frac{1}{T}\sum_{t=1}^{T-1}\sum_{\substack{j=1\\ j\neq t}}^{T-1}\mathbf{1}(X_t)w(X_t,X_t)w\left(X_t,X_j\right)h(Y_{t+1},R_{t+1};\theta)\hat{\Omega}_T(X_t)h\left(Y_{j+1},R_{j+1};\theta\right) =: \mathcal{Q}_{4,T}(\theta),$$

$$\mathcal{Q}_{5,T}(\theta) := \frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{i=1\\ i\neq t}}^{T-1}\sum_{\substack{j\neq i\\ j\neq t\\ j=1}}^{T-1}\mathbf{1}(X_t)w(X_t,X_i)w\left(X_t,X_j\right)h(Y_{i+1},R_{i+1};\theta)\hat{\Omega}_T(X_t)h\left(Y_{j+1},R_{j+1};\theta\right),$$

for any $\theta\in\Theta$. We bound the first four terms in the next lemma.

Lemma 3. *Under Assumptions 1–11 we have (i)* $\mathcal{Q}_{1,T}(\theta) = O_p\left(\frac{1}{T^2 b_T^{2L}}\right)$ *uniformly in* $\theta\in\Theta$, *(ii)* $\mathcal{Q}_{2,T}\left(\hat{\theta}_T\right) = a_T$, *where* a_T *is defined in Proposition 4, and (iii)* $\mathcal{Q}_{3,T}\left(\hat{\theta}_T\right) = \mathcal{Q}_{4,T}\left(\hat{\theta}_T\right) = O_p\left(\frac{1}{T^2 b_T^{3L/2}}\right)$.

Proof. See Section E.3 of the Online Appendix. □

Since $\hat{\delta}_T^2 = \mathcal{Q}_T(\hat{\theta}_T)$, from Equation (A31) and Assumption 5 on the bandwidth b_T we get

$$Tb_T^{L/2}\left(\hat{\delta}_T^2 - a_T\right) = Tb_T^{L/2}\mathcal{Q}_{5,T}(\hat{\theta}_T) + o_p(1). \tag{A32}$$

Let us consider the term $\mathcal{Q}_{5,T}(\hat{\theta}_T)$. From the definition of stochastic matrix $\hat{H}(x) := \hat{\Omega}_T(x)^{-1}\hat{f}_X(x)^2$ and the kernel density estimator $\hat{f}_X(x) := \frac{1}{Tb_T^L}\sum_{j=1}^T K\left(\frac{x-X_j}{b_T}\right)$, we get

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

for any $\theta \in \Theta$. Define the matrix $\tilde{H}_T(X_t) := \mathrm{E}\left[\hat{\Omega}_T(X_t)^{-1}\hat{f}_X(X_t)|X_t\right]\mathrm{E}\left[\hat{f}_X(X_t)|X_t\right]$ and

$$\mathcal{Q}_{6,T}(\theta) := \frac{1}{T^3 b_T^{2L}} \sum_{t=1}^{T} \sum_{\substack{i=1 \\ i\neq t}}^{T-1} \sum_{\substack{j\neq i \\ j\neq t \\ j=1}}^{T-1} \mathbf{1}(X_t) K\left(\frac{X_t - X_i}{b_T}\right) \cdot h(Y_{i+1}, R_{i+1};\theta)'\tilde{H}_T(X_t)^{-1} h(Y_{j+1}, R_{j+1};\theta) K\left(\frac{X_t - X_j}{b_T}\right). \tag{A33}$$

Lemma 4. *Under Assumptions 1–11 we have* $Tb_T^{L/2}\mathcal{Q}_{5,T}(\hat{\theta}_T) = Tb_T^{L/2}\mathcal{Q}_{6,T}(\hat{\theta}_T) + o_p(1) = Tb_T^{L/2}\mathcal{Q}_{6,T}(\theta_0) + o_p(1)$.

Proof. See Section E.4 of the Online Appendix. □

From Equation (A32) and Lemma 4 we get

$$Tb_T^{L/2}\left(\hat{\delta}_T^2 - a_T\right) = Tb_T^{L/2}\mathcal{Q}_{6,T}(\theta_0) + o_p(1). \tag{A34}$$

We now show the asymptotic normality of term $Tb_T^{L/2}\mathcal{Q}_{6,T}(\theta_0)$ in the RHS of Equation (A34). Then, the asymptotic normality of $Tb_T^{L/2}\left(\hat{\delta}_T^2 - a_T\right)$ follows. Let us consider the $(N \times N)$-dimensional matrix functions A_T and $\hat{A}_T$ defined as $A_T(\tilde{x}, \tilde{\tilde{x}}) := \int_{\mathcal{X}} \mathbf{1}(x) K\left(\frac{x-\tilde{x}}{b_T}\right)\tilde{H}_T(x)^{-1}K\left(\frac{x-\tilde{\tilde{x}}}{b_T}\right)f_X(x)dx$ and $\hat{A}_T(\tilde{x}, \tilde{\tilde{x}}) := \frac{1}{T}\sum_{\substack{j\neq i \\ j\neq t \\ j=1}}^{T} \mathbf{1}(X_t)K\left(\frac{X_t-\tilde{x}}{b_T}\right)\tilde{H}_T(X_t)^{-1}K\left(\frac{X_t-\tilde{\tilde{x}}}{b_T}\right)$, for any $\tilde{x}, \tilde{\tilde{x}} \in \mathcal{X}$. The term $\mathcal{Q}_{6,T}(\theta_0)$ can be written as

$$\mathcal{Q}_{6,T}(\theta_0) = \frac{1}{T^2 b_T^{2L}} \sum_{i=1}^{T-1} \sum_{\substack{j \neq i \\ j=1}}^{T-1} h_0(Y_{i+1}, R_{i+1})' \hat{A}_T(X_i, X_j) h_0(Y_{j+1}, R_{j+1}), \tag{A35}$$

where $h_0(Y_t, R_t) := h(Y_t, R_t; \theta_0)$. From Assumption 8 the quadratic forms in the RHS of the last equation are L_1-bounded. 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 Equation (A35) as

$$Tb_T^{L/2} \mathcal{Q}_{6,T}(\theta_0) = \frac{1}{T} \sum_{i=1}^{T-1} \sum_{j=1}^{i-1} g_T(W_i, W_j) + o_p(1), \tag{A36}$$

where the scalar function g_T is defined as

$$g_T(w, \tilde{w}) := \frac{2}{b_T^{3L/2}} h_0(y, R)' A_T(x, \tilde{x}) h_0(\tilde{y}, \tilde{R}), \tag{A37}$$

for any $w = [y'\, R'\, x']', \tilde{w} = \left[\tilde{y}'\, \tilde{R}'\, \tilde{x}'\right]' \in \mathrm{W}$. 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 (2014) (see also Yoshihara, 1976, and Yoshihara, 1992) hold. Specifically, from Assumptions 1 and 2 the process $\{W_t\}$ is strictly stationary and strong mixing, and the preliminary and last conditions for the lemma are satisfied. Lemma 5 below shows that the remaining conditions in Lemma A.3 in Su and White (2014) are satisfied as well.

Lemma 5. *The function g_T is such that for any $w = [y'\, R'\, x']', \tilde{w} = \left[\tilde{y}'\, \tilde{R}'\, \tilde{x}'\right]' \in \mathrm{W}$ and any integer $0 < j \leqslant T-1$ we have (i) $g_T(w, \tilde{w}) = g_T(\tilde{w}, w)$; (ii) $\mathrm{E}[g_T(W_i, w)] = 0$; (iii) $\mathrm{E}\left[g_T\left(W_1, \overline{W}_1\right)^2\right] = 2\sigma_0^2 + o(1)$, where $\overline{W}_1 = \left[\overline{Y}_1'\, \overline{R}_1'\, \overline{X}_1'\right]'$ is an independent copy of $W_1 = [Y_1'\, R_1'\, X_1']'$ and σ_0^2 is defined in Proposition 4; (iv) $\mathrm{E}\left[g_T(W_{j+1}, w) g_T(W_1, \tilde{w})\right] = 0$; and (v) there exist constants $\beta > 0, \gamma < 1$ such that $\|G_T(W_1, W_1)\|_{2+\beta/2} = o(T^{1/2})$, $\max\left[\max_{i \in [1,T-1]} \|G_T(W_{i+1}, W_1)\|_2, \|G_T(W_1, \overline{W}_1)\|_2\right] = o(1)$, and $\max\left[\max_{i \in [1,T-1]} \|g_T(W_{i+1}, W_1)\|_{4+\beta}, \|g_T(W_1, \overline{W}_1)\|_{4+\beta}\right] = O(T^\gamma)$, for the standard L^p-norm, which is defined as $\|\cdot\|_p := \left(\mathrm{E}\left[|\cdot|^p\right]\right)^{1/p}$ for any positive real p, and the function $G_T(w, z) := \mathrm{E}[g_T(W_1, w) g_T(W_1, z)]$.*

Proof. See Section E.5 of the Online Appendix. □

From Lemma A.3 in Su and White (2014) (see also Yoshihara, 1976, and Yoshihara, 1992) the statistic $\frac{1}{T}\sum_{i=1}^{T-1}\sum_{j=1}^{i-1} g_T(W_i, W_j)$ is asymptotically normal with null mean and asymptotic variance σ_0^2. Then, from Equation (A36) we have $Tb_T^{L/2}\mathcal{Q}_{6,T}(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_0^2)$, and the conclusion follows.

D.2 Proof of Proposition 5

From the Equations in (28) we have

$$\hat{\delta}_T^2 = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(X_t)\hat{e}_T\big(X_t;\hat{\theta}_T\big)'\hat{\Omega}_T(X_t)\hat{e}_T\big(X_t;\hat{\theta}_T\big). \tag{A38}$$

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

$$\begin{aligned}\hat{e}_T\big(X_t;\hat{\theta}_T\big) &= \hat{e}_T(X_t;\theta_\star) + \nabla_{\theta'}\hat{e}_T(X_t;\theta_\star)\big(\hat{\theta}_T - \theta_\star\big) + \mathrm{Rem}\big[X_t;\hat{\theta}_T - \theta_\star\big] \\ &= e_\star(X_t) + \big(\hat{e}_T(X_t;\theta_\star) - e_\star(X_t)\big) + J_\star(X_t)\big(\hat{\theta}_T - \theta_\star\big) + o_p(1),\end{aligned} \tag{A39}$$

where $e_\star(X_t) := e(X_t;\theta_\star)$, because matrix $\nabla_{\theta'}\hat{e}_T(X_t;\theta_\star)\xrightarrow{\mathbb{P}}J_\star(X_t) := \nabla_{\theta'}e(X_t;\theta_\star)$, and $\mathrm{Rem}\big[x;\tilde{\theta}\big]$ denotes a remainder term for fixed $x \in \mathcal{X}$. We can write $\hat{\Omega}_T(X_t)$ as

$$\Big(\Omega(X_t)^{-1} + \Big(\hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1}\Big)\Big)^{-1} = \Omega(X_t)\Big(\mathbf{I}_N + \Big(\hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1}\Big)\Omega(X_t)\Big)^{-1}.$$

The norm of matrix $\Big(\hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1}\Big)\Omega(X_t)$ converges to 0 in probability. By the expansion of the matrix inverse function, we get

$$\hat{\Omega}_T(X_t) = \Omega(X_t) - \Omega(X_t)\Big(\hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1}\Big)\Omega(X_t) + o_p(1). \tag{A40}$$

Let us plug Equations (A39) and (A40) into Equation (A38). We get

$$\begin{aligned}\hat{\delta}_T^2 = &\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}(X_t)e_\star(X_t)'\Omega(X_t)e_\star(X_t) \\ &-\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}(X_t)e_\star(X_t)'\Omega(X_t)\Big(\hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1}\Big)\Omega(X_t)e_\star(X_t) \\ &+\frac{2}{T}\sum_{t=1}^{T}\mathbf{1}(X_t)e_\star(X_t)'\Omega(X_t)J_\star(X_t)\big(\hat{\theta}_T - \theta_\star\big) + \frac{2}{T}\sum_{t=1}^{T}\mathbf{1}(X_t)e_\star(X_t)'\Omega(X_t)\big(\hat{e}_T(X_t;\theta_\star) - e_\star(X_t)\big),\end{aligned} \tag{A41}$$

up to negligible terms. From the weak law of large numbers we have that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}(X_t)e_\star(X_t)'\Omega(X_t)J_\star(X_t) = \mathrm{E}\big[\mathbf{1}(X_t)e_\star(X_t)'\Omega(X_t)J_\star(X_t)\big] + o_p(1). \tag{A42}$$

The FOC for the minimization of the criterion in Equation (30) and the Leibniz's rule for differentiation under the integral sign imply $\mathrm{E}\big[\mathbf{1}(X_t)e_\star(X_t)'\Omega(X_t)J_\star(X_t)\big] = 0$. Then, from Equation (A42) and $\hat{\theta}_T - \theta_\star = O_p\big(1/\sqrt{T}\big)$ (see Lemma 2), the third term in the RHS of Equation (A41) is $o_p\big(1/\sqrt{T}\big)$. From this remark and scaling both sides of Equation (A41) by $\sqrt{T}$ we get

$$\sqrt{T}\hat{\delta}_T^2 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{1}(X_t)e(X_t;\theta_\star)'\Omega(X_t)e(X_t;\theta_\star)$$

$$+\frac{2}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{1}(X_t)e_\star(X_t)'\Omega(X_t)\left(\hat{e}_T(X_t;\theta_\star)-e_\star(X_t)\right) \tag{A43}$$

$$-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{1}(X_t)e_\star(X_t)'\Omega(X_t)\left(\hat{\Omega}_T(X_t)^{-1}-\Omega(X_t)^{-1}\right)\Omega(X_t)e_\star(X_t)+o_p(1).$$

Now, let ϕ be the function defined in Proposition 5, namely

$$\phi(w;\theta):=2\cdot\mathbf{1}(x)e(x;\theta)'\Omega(x)h(y,R;\theta)-\mathbf{1}(x)\left[e(x;\theta)'\Omega(x)R\right]^2=\mathbf{1}(x)e(x;\theta)'\Omega(x)e(x;\theta)$$

$$+2\cdot\mathbf{1}(x)e(x;\theta)'\Omega(x)\left[h(y,R;\theta)-e(x;\theta)\right]-\mathbf{1}(x)e(x;\theta)'\Omega(x)\left(RR'-\Omega(x)^{-1}\right)\Omega(x)e(x;\theta),$$

for $w=[y'\,R'\,x']'$. The Nadaraya–Watson kernel regression estimator of $\phi(W_{t+1};\theta)$ conditional on $X_t=x$ is

$$\sum_{i=1}^{T-1}w(x,X_i)\phi(Y_{i+1},R_{i+1},x;\theta)=2\cdot\mathbf{1}(x)e(x;\theta)'\Omega(x)\left(\sum_{i=1}^{T-1}w(x,X_i)h(Y_{i+1},R_{i+1};\theta)-e(x;\theta)\right)$$

$$+\mathbf{1}(x)e(x;\theta)'\Omega(x)e(x;\theta)-\mathbf{1}(x)e(x;\theta)'\Omega(x)\left(\sum_{i=1}^{T-1}w(x,X_i)R_{i+1}R_{i+1}'-\Omega(x)^{-1}\right)\Omega(x)e(x;\theta)$$

for any $x\in\mathcal{X}$ and any $\theta\in\Theta$. Therefore, we can write Equation (A43) using this estimator as

$$\sqrt{T}\hat{\delta}_T^2=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{T-1}w(X_t,X_i)\phi(Y_{i+1},R_{i+1},X_t;\theta_\star)+o_p(1). \tag{A44}$$

We have $\mathrm{E}\left[\phi(W_{t+1};\theta)-\mathbf{1}(X_t)e(X_t;\theta)'\Omega(X_t)e(X_t;\theta)|X_t=x\right]=0$, for any $x\in\mathcal{X}$ and any $\theta\in\Theta$. To study the asymptotic behavior of the RHS of Equation (A44) we use the next Lemma 6. This result is similar to the arguments used in Kitamura, Tripathi, and Ahn (2004, pp. 1696–1698), and Gospodinov and Otsu (2012, p. 487).

Lemma 6. *Consider the square-integrable scalar function $\zeta(w)$, for any $w\in\mathrm{W}$, such that for any $x\in\mathcal{X}$ it is $\mathrm{E}\left[\zeta(W_{t+1})|X_t=x\right]=0$. Under regularity conditions on function ζ and under Assumption 5 we have $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{T-1}w(X_t,X_i)\zeta(Y_{i+1},R_{i+1},X_t)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\zeta(W_{t+1})+o_p(1)$.*

Proof. See Online Appendix E.6. □

Let us consider $\zeta(w)=\phi(w)-\mathbf{1}(x)e(x;\theta)'\Omega(x)e(x;\theta)$ in Lemma 6 and apply it to the first term in the RHS of Equation (A44). We get $\sqrt{T}\hat{\delta}_T^2=\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\phi(W_{t+1};\theta_\star)+o_p(1)$. By subtracting the quantity $\delta_\star^2$ to both sides of the last equation we get

$$\sqrt{T}\left(\hat{\delta}_T^2 - \delta_\star^2\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\left(\phi(W_{t+1};\theta_\star) - \delta_\star^2\right) + o_p(1). \quad \text{(A45)}$$

From Assumption 12 the normalized sum on the RHS of the last equation is asymptotically normal, with null asymptotic mean and σ_ϕ^2 as asymptotic variance. Then Proposition 5 follows.

D.3 Proof of Proposition 6

Subtracting side by side the asymptotic expansion in Equation (A45) written for the SDF family $\mathcal{F}$ and the corresponding expansion written for the SDF family $\tilde{\mathcal{F}}$ we have $\sqrt{T}\left(\hat{\delta}_T^2 - \delta_\star^2 - \hat{\tilde{\delta}}_T^2 + \tilde{\delta}_\star^2\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\left(\phi(W_{t+1};\theta_\star) - \delta_\star^2 - \tilde{\phi}\left(W_{t+1};\tilde{\theta}_\star\right) + \tilde{\delta}_\star^2\right) + o_p(1)$.

Under $\mathcal{H}_0^3$, we have $\delta_\star = \tilde{\delta}_\star$ and the LHS of this equation is $\sqrt{T}\left(\hat{\delta}_T^2 - \hat{\tilde{\delta}}_T^2\right)$. Moreover, under $\mathcal{H}_0^3$ the long-run variance of process $\left\{\phi(W_{t+1};\theta_\star) - \tilde{\phi}\left(W_{t+1};\tilde{\theta}_\star\right)\right\}$ is strictly larger than zero. Then, Proposition 6 follows from Assumption 12.

D.4 Proof of Proposition 7

Under Assumption 13, by the mean value theorem applied to $\psi\left(\hat{\theta}_T\right)$, there exists $\tilde{\theta}$ between $\hat{\theta}_T$ and $\theta_\star$ such that $\sqrt{T}\psi\left(\hat{\theta}_T\right) = \sqrt{T}\psi(\theta_\star) + \nabla_{\theta'}\psi(\tilde{\theta})\sqrt{T}\left(\hat{\theta}_T - \theta_\star\right)$. We then have $\sqrt{T}\psi\left(\hat{\theta}_T\right) = \sqrt{T}\psi(\theta_\star) + \nabla_{\theta'}\psi(\theta_\star)\sqrt{T}\left(\hat{\theta}_T - \theta_\star\right) + o_p(1)$. Under the hypothesis $\mathcal{H}_0^2$ we have $\psi(\theta_\star) = 0_r$, and we get $\sqrt{T}\psi\left(\hat{\theta}_T\right) = \nabla_{\theta'}\psi(\theta_\star)\sqrt{T}\left(\hat{\theta}_T - \theta_\star\right) + o_p(1)$. From Lemma 2 we deduce $\sqrt{T}\psi\left(\hat{\theta}_T\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0_p, \nabla_{\theta'}\psi(\theta_\star)\Sigma_\star\nabla_\theta\psi(\theta_\star)'\right)$. The Continuous Mapping Theorem completes the proof.

D.5 Proof of Proposition 8

From Equations (A34) and (A36) written for families $\mathcal{F}$ and $\tilde{\mathcal{F}}$, we get:

$$Tb_T^{L/2}\left(\hat{\delta}_T^2 + \hat{\tilde{\delta}}_T^2 - a_T - \tilde{a}_T\right) = \frac{1}{T}\sum_{i=1}^{T-1}\sum_{j=1}^{i-1}\bar{g}_T\left(W_i, W_j\right), \quad \text{(A46)}$$

where $\bar{g}_T\left(W_i, W_j\right) = g_T\left(W_i, W_j\right) + \tilde{g}_T\left(W_i, W_j\right)$, function g_T is defined in Equation (A37) for family $\mathcal{F}$, and function $\tilde{g}_T$ is defined analogously for family $\tilde{\mathcal{F}}$. We prove the asymptotic normality of the RHS of Equation (A46) by checking the regularity conditions of Lemma A.3 in Su and White (2014) (see also Yoshihara, 1976, and Yoshihara, 1992). In particular, we have $\mathrm{E}\left[\bar{g}_T\left(W_1, \overline{W}_1\right)^2\right] = 2\sigma_{00}^2 + o(1)$, where $\overline{W}_1$ is an independent copy of W_1. The conclusion follows.

D.6 Proof of Proposition 9

The proof adapts the argument in Vuong (1989, p. 321). The probability of a type-1 error $\mathbb{P}[A \cap B|\mathcal{H}_0]$, for the events $B := \left\{\hat{\delta}_T^2 - \hat{\tilde{\delta}}_T^2 \geqslant c_{1-\alpha}\hat{\sigma}_\Delta T^{-1/2}\right\}$ and $A := \{\hat{\delta}_T^2 + \hat{\tilde{\delta}}_T^2 \geqslant a_T + \tilde{a}_T + c_{1-\alpha}\hat{\sigma}_{00}T^{-1/2}b_T^{-L/4}\}$, is upper bounded as $\mathbb{P}[A \cap B|\mathcal{H}_0] \leqslant \max\left\{\mathbb{P}\left[A \cap B|\mathcal{H}_0^1\right],\right.$

$\mathbb{P}[A \cap B|\mathcal{H}_0^3]\} \leqslant \max\{\mathbb{P}[A|\mathcal{H}_0^1], \mathbb{P}[B|\mathcal{H}_0^3]\}$. Now, $\mathbb{P}[A|\mathcal{H}_0^1] \rightarrow \alpha$ from Proposition 8, and $\mathbb{P}[B|\mathcal{H}_0^3] \rightarrow \alpha$ from Proposition 6. The conclusion follows.

D.7 Proof of Proposition 10

The proof is similar to the proof of Proposition 7 given in Appendix D.4 and it is omitted.

References

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

Aït-Sahalia, Y., P. J. Bickel, and T. M. Stoker. 2001. Goodness-of-Fit Tests for Kernel Regression with an Application to Option Implied Volatilities. *Journal of Econometrics* 105: 363–412.

Almeida, C., and R. Garcia. 2012. Assessing Misspecified Asset Pricing Models with Empirical Likelihood Estimators. *Journal of Econometrics* 170: 519–537.

Anatolyev, S., and N. Gospodinov. 2011. *Methods for Estimation and Inference in Modern Econometrics*. London: Chapman & Hall/CRC.

Andrews, D. W. K., and C. J. Monahan. 1992. An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. *Econometrica* 60: 953–966.

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

Antoine, B., K. Proulx, and E. Renault. 2018. Pseudo-True SDFs in Conditional Asset Pricing Models. *Journal of Financial Econometrics* (forthcoming).

Balduzzi, P., and C. Robotti. 2010. Asset Pricing Models and Economic Risk Premia: A Decomposition. *Journal of Empirical Finance* 17: 54–80.

Barillas, F., and J. Shanken. 2017. Which Alpha? *The Review of Financial Studies* 30: 1316–1338.

Bekaert, G., and J. Liu. 2008. Conditioning Information and Variance Bounds on Pricing Kernels. *Review of Financial Studies* 17: 339–378.

Carhart, M. M. 1997. On Persistence in Mutual Fund Performance. *Journal of Finance* 52: 57–82.

Chabi-Yo, F. 2008. Conditioning Information and Variance Bounds on Pricing Kernels with Higher-Order Moments: Theory and Evidence. *Review of Financial Studies* 21: 181–231.

Chamberlain, G. 1987. Asymptotic Efficiency in Estimation with Conditional Moment Restrictions. *Journal of Econometrics* 34: 305–334.

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

Cochrane, J. H. 2008. The Dog That Did Not Bark: A Defense of Return Predictability. *Review of Financial Studies* 21: 1533–1575.

Domínguez, M. A., and I. N. Lobato. 2004. Consistent Estimation of Models Defined by Conditional Moment Restrictions. *Econometrica* 72: 1601–1615.

Epstein, L. G., and S. E. Zin. 1989. Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. *Econometrica* 57: 937–969.

Epstein, L. G., and S. E. Zin. 1991. Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis. *Journal of Political Economy* 99: 263–286.

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

Fama, E. F., and K. R. French. 1998. Value Versus Growth: The International Evidence. *Journal of Finance* 53: 1975–2000.

Fama, E. F., and K. R. French. 2015. A Five-Factor Asset Pricing Model. *Journal of Financial Economics* 116: 1–22.

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

Ghosh, A., C. Julliard, and A. P. Taylor. 2016. What Is the Consumption-CAPM Missing? An Information-Theoretic Framework for the Analysis of Asset Pricing Models. *Review of Financial Studies* 30: 442–504.

Gibbons, M. R., S. A. Ross, and J. Shanken. 1989. A Test of the Efficiency of a Given Portfolio. *Econometrica* 57: 1121–1152.

Goyal, A., and I. Welch. 2008. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies* 21: 1455–1508.

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

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

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

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

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

Hansen, B. E. 2008. Uniform Convergence Rates for Kernel Estimation with Dependent Data. *Econometric Theory* 24: 726–748.

Hansen, L. P. 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50: 1029–1054.

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

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

Hansen, L. P., and S. R. Richard. 1987. The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models. *Econometrica* 55: 587–613.

Hansen, L. P., and K. J. Singleton. 1982. Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models. *Econometrica* 50: 1269–1286.

Harrison, M. J., and D. M. Kreps. 1979. Martingales and Arbitrage in Multiperiod Securities Markets. *Journal of Economic Theory* 20: 381–408.

Hodrick, R. J., and X. Zhang. 2001. Evaluating the Specification Errors of Asset Pricing Models. *Journal of Financial Economics* 62: 327–376.

Hou, K., C. Xue, and L. Zhang. 2014. Digesting Anomalies: An Investment Approach. *Review of Financial Studies* 28: 650–705.

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

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

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

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

Kleibergen, F., and Z. Zhan. 2016. "Identification Robust Inference on Risk Premia of Mimicking Portfolios of Macroeconomic Factors." Unpublished Working paper, Brown University.

Leeb, H., and B. Poetscher. 2005. Model Selection and Inference: Facts and Fiction. *Econometric Theory* 21: 21–59.

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

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

Lintner, J. 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, J. 1966. Equilibrium in a Capital Asset Market. *Econometrica* 34: 768–783.

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

Newey, W. K. 1993. "Efficient Estimation of Models with Conditional Moment Restrictions." In G. S. Maddala, C. R. Rao, and H. D. Vinod (eds.), *Handbook of Statistics*. Amsterdam: North-Holland.

Newey, W. K., and D. McFadden. 1994. "Large Sample Estimation and Hypothesis Testing." In R. Engle and D. McFadden (eds.), *Handbook of Econometrics*. Amsterdam: North-Holland.

Newey, W. K., and K. D. West. 1994. Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies* 61: 631–653.

Novy-Marx, R. 2013. The Other Side of Value: The Gross Profitability Premium. *Journal of Financial Economics* 108: 1–28.

Parker, J. A., and C. Julliard. 2005. Consumption Risk and the Cross Section of Expected Returns. *Journal of Political Economy* 113: 185–222.

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

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

Schennach, S. M., and D. Wilhelm. 2017. A Simple Parametric Model Selection Test. *Journal of the American Statistical Association* 112: 1663–1674.

Sharpe, W. F. 1964. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance* 19: 425–442.

Shi, X. 2015. A Nondegenerate Vuong Test. *Quantitative Economics* 6: 85–121.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall/CRC.

Stock, J. H., and M. W. Watson. 2002. Has the Business Cycle Changed and Why? *NBER Macroeconomics Annual* 17: 159–218.

Stock, J. H., and J. H. Wright. 2000. GMM with Weak Identification. *Econometrica* 68: 1055–1096.

Su, L., and H. White. 2014. Testing Conditional Independence via Empirical Likelihood. *Journal of Econometrics* 182: 27–44.

Treynor, J. L. 1962. "Toward a Theory of Market Value of Risky Assets." Unpublished Working Paper. A version was published in 1999 in R. A. Korayczyk (ed.), *Asset Pricing and Portfolio Performance: Models, Strategy and Performance Metrics*. London: Risk Books.

Tripathi, G., and Y. Kitamura. 2003. Testing Conditional Moment Restrictions. *The Annals of Statistics* 31: 2059–2095.

Vuong, Q. 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica* 57: 307–333.

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

Yoshihara, K. 1992. Limiting Behavior of Generalized Quadratic Forms Generated by Absolutely Regular Sequences III. *Yokohama Mathematical Journal* 40: 1–9.