Comparing Asset Pricing Models by the
Conditional Hansen-Jagannathan Distance

PATRICK GAGLIARDINI and DIEGO RONCHETTI

This version: September 2014
Please do not circulate without authors’ permission†

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

JEL classification: C12, C14, G12.

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

†We thank the participants at the AMES 2014 in Taipei, the CMES 2014 in Xiamen, the EEA-ESEM 2014 in Toulouse and the EEF seminar in Groningen for useful comments.
Modern asset pricing theories can be formulated in terms of the Stochastic Discount Factor (SDF). The SDF accounts for time discount and risk adjustment in the pricing of a future risky payoff. 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 leads to a SDF that is proportional to the marginal rate of substitution between current and future consumption. As several preference specifications are possible in theory, we are confronted with a large set of alternative SDF models, and there is even more latitude in this choice when reduced-form SDF specifications are considered. Therefore, a central question in empirical asset pricing is how to select the most adequate model among a set of competing non-nested parametric SDF families.

In this paper we adopt the point of view that all competing asset pricing models are potentially misspecified and we compare them in terms of their distance from the set of admissible SDF’s. A SDF is considered admissible if it matches the dynamic no-arbitrage pricing restrictions implied by a set of test assets. The goals of our paper are (1) the introduction of a conditional version of the Hansen-Jagannathan (HJ-) distance (Hansen and Jagannathan (1997)), which is the distance currently used to compare asset pricing models, (2) the description of the properties of this new distance and (3) the illustration of its use in an empirical comparison of some popular SDF families proposed in the asset pricing literature. We then refer in this paper to the distance described in Hansen and Jagannathan (1997) as the unconditional HJ-distance, and to the newly introduced distance as the conditional HJ-distance. The latter distance captures the conditional pricing errors implied by the SDF model for any managed portfolio of the test assets, and not just for some particular portfolios like the unconditional HJ-distance. Hence, the conditional HJ-distance fully exploits the conditioning information when comparing the performance of competing asset pricing models.

In order to introduce the framework of our paper, let us consider an economy with $N$ assets traded in discrete time.\footnote{In the empirical application discussed in Sec. III six portfolios representative of the market for publicly traded U.S. equities and the 1-month T-Bill are taken as test assets.} Let $P_t := [P_{1,t} \ldots P_{N,t}]'$ denote the vector of the trading prices at time $t$ for the $N$ assets, and let $\mathcal{I}_t$ be the information available to investors at time $t$. The Absence of Arbitrage Opportunities (AAO) in the market is equivalent to the existence of a scalar
stochastic process \{M_{t,t+1}\} such that the random variable \(M_{t,t+1}\) is (i) positive, (ii) measurable with respect to (w.r.t.) the information \(\mathcal{I}_{t+1}\) and (iii) satisfying the no-arbitrage restriction

\[ P_t = E[M_{t+1} \mid \mathcal{I}_t], \]

where \(E[\cdot \mid \mathcal{I}_t]\) denotes the conditional expectation operator under the historical probability measure given information \(\mathcal{I}_t\) (see, e.g., Harrison and Kreps (1979), and Hansen and Richard (1987)).\(^2\) As in most of the modern empirical asset pricing literature, we refer to any random variable \(M_{t,t+1}\) satisfying properties (ii) and (iii) as an admissible SDF between dates \(t\) and \(t+1\). Let us assume that the information \(\mathcal{I}_t\) is generated by the Markov process of \(d_X\) random state variables collected in vector \(X_t\) admitting values in set \(\mathcal{X} \subseteq \mathbb{R}^{d_X}\). Moreover, let \(R_t := [R_{1,t} \ldots R_{N,t}]'\) be the \(N\)-dimensional vector of the assets’ gross returns \(R_{i,t} = P_{i,t}/P_{i,t-1}\), for any \(i = 1, \ldots, N\). Then, the property (iii) for an admissible SDF can be rewritten as

\[ E[M_{t,t+1}R_{t+1} - 1_N \mid X_t = x] = 0_N, \tag{1.1} \]

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

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

\[ E[h(Y_{t+1}; \theta_0) \mid X_t = x] = 0_N \tag{1.2} \]

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

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

\(^2\)To simplify the exposition we consider assets that do not pay any dividend. If dividends are paid, expected future prices in the AAO are the sum of expected trading prices and dividends.
which represents the conditional moment function of the econometric problem, and the unknown value $\theta_0$ of the SDF parameter vector. To ensure the identification of the unknown parameter value in parameter set $\Theta$, it is customary to assume that the value $\theta_0$ 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 unknown parameter value and the testing of the correct model specification are typically addressed in a Generalized Method of Moments (GMM) framework (see, e.g., Hansen (1982), and Hansen and Singleton (1982)). The method is based on the minimization of the GMM criterion, which is a quadratic form of a sample counterpart of a vector unconditional moment restriction derived from the vector conditional moment restriction in Eq. (1.2). To create this unconditional moment restriction, we select a $(q \times N)$-dimensional instrument matrix $Z_t$ for any date $t$, that is a function of the components of the state variables vector $X_t$. Under the hypothesis of correct model specification, using the instrument matrix $Z_t$ and the law of iterated expectations, we derive the following $q$-dimensional vector unconditional moment restriction:

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

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

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

---

3 Being the instrument matrix a function of vector $X_t$, the dynamic positions are determined by this conditioning variable vector.
all admissible SDF’s satisfying the unconditional moment restrictions for the chosen instrument matrix. This distance is defined as

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

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

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

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

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

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

\[ \Omega_Z := \mathbb{E} \left[ Z_t R_{t+1} R_{t+1}' Z_t' \right]^{-1}. \tag{1.8} \]

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

\(^4\)The optimal weighting matrix used in the Hansen’s J statistic (that is, the matrix that minimizes the asymptotic variance of the estimator of the SDF parameter vector) varies across models. If, for the chosen instrument matrix, a model has very volatile unconditional pricing errors, considering the inverse of the unconditional variance-covariance matrix of the scaled assets’ gross returns may lead artificially to a small unconditional HJ-distance.
Garcia (2012) have considered different discrepancy measures to assess the unconditional distance of the parametric SDF family from the set of admissible SDF’s.

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

In this paper we derive the conditional HJ-distance by extending Eqs. (1.5) and (1.6) in a conditional setting. Specifically, we define the conditional HJ-distance \( \delta \) as the \( L^2 \)-discrepancy between the candidate parametric SDF family and the set of SDF’s satisfying the conditional moment restrictions in Eq (1.2), and not just the unconditional moment restrictions in Eq (1.4) holding for a particular choice of the instrument matrix. The paper has two main theoretical contributions. First, we study in detail the difference between the conditional and unconditional HJ-distances. We provide upper and lower bounds for the difference \( \delta^2 - d^2_Z \) that are valid for general SDF families. In particular, we are able to characterize the difference \( \delta^2 - d^2_Z \) explicitly for families of SDF that are conditionally linear in the priced risk factors. We show how this difference is related to the component of the conditional pricing error vector which is unspanned by the instrument matrix. We demonstrate that the difference between conditional and unconditional HJ-distances can be arbitrarily large, and that the ranking for the degree of model misspecification between two misspecified SDF families can be reversed, depending on which distance is
used for the comparison. The second theoretical contribution is the definition of a sample analogue of the conditional HJ-distance and the description of its large sample properties, for both correctly specified and misspecified models. In constructing the sample conditional HJ-distance we estimate the conditional expectation of the moment vector \( h(Y_{t+1}; \theta) \) given the state variable vector \( X_t \) by kernel regression methods. The large sample results allow us to develop a model selection procedure based on the conditional HJ-distance.

Our empirical contribution consists in the comparison of fourteen parametric SDF specifications for the U.S. equity and short term T-Bill markets in terms of the conditional HJ-distance. Two specifications are preference-based SDF models with the time-separable CRRA utility and time-nonseparable preferences of Epstein and Zin (1989, 1991) (see Stock and Wright (2000)). A specification corresponds to the linearization of the preference-based SDF model with time-separable CRRA utility for small values of the logarithmic consumption growth. Three SDF specifications, considered also in Nagel and Singleton (2011), are conditionally linear in logarithmic consumption growth and correspond to dynamic versions of the linearized Consumption-based Capital Asset Pricing Model (CCAPM) with time-varying risk premia. This time variation is due to 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)). Another SDF specification is the linearization of the preference-based SDF model with time-nonseparable preferences of Epstein and Zin (1989, 1991) including personal consumption of nondurables, services and durables. The last five SDF specifications considered in our empirical analysis correspond to the Capital Asset Pricing Model (CAPM, Treynor (1962), Sharpe (1964), Lintner (1965) and Mossin (1966)) and some of its extension largely used in empirical studies. We consider the three-factor Fama and French (1993, 1998) model and its extensions including either the maturity risk and the default risk factors (Fama and French (1993, 1998)), or a momentum factor (Carhart (1997)), or a liquidity factor (Pastor and Stambaugh (2003)). We also include the four-factor model introduced in Novy-Marx (2013). To highlight the differences between the results obtained by relying on the unconditional and conditional HJ-distances, we report the results of empirical analyses based on the two distances. The unconditional HJ-distance is computed for different choices of the instrument matrix, and the conditional HJ-distance is
computed for different choices of the variable generating information.

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

I. Conditional HJ-distance

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

A. Definition of conditional HJ-distance

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

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

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

$$\delta := \min_{\theta \in \Theta} \min_{M_{t,t+1} \in \mathcal{M}} E \left[ \left( (M_{t,t+1} - m(Y_{t+1}; \theta))^2 \right)^{1/2} \right].$$  (2.2)
Differently than set $M_Z$ and unconditional HJ-distance $d_Z$, the set $M$ and the conditional HJ-distance $\delta$ do not depend on any instrument matrix $Z_t$. Let us consider the minimization problem that defines the squared distance $\delta^2$ in Eq. (2.2). For a given value of parameter $\theta \in \Theta$, the Lagrangian function $L$ for the constrained minimization w.r.t. the admissible SDF $M_{t,t+1} \in M$ is given by

$$L(\theta) = E \left[ (m(Y_{t+1}; \theta) - M_{t,t+1})^2 \right] + 2 \int_X \lambda(x)'E \left[ M_{t,t+1} R_{t+1} - 1_N | X_t = x \right] f_X(x) dx$$

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

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

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

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

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

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

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

where $\Omega(X_t)$ is the inverse of the $(N \times N)$-dimensional conditional second moment matrix of the assets’ gross returns vector $R_t$ given $X_t$, i.e.

$$\Omega(X_t) := E \left[ R_{t+1} R_{t+1}' | X_t \right]^{-1}, \quad (2.5)$$
and $e(X_t; \theta)$ is the $N$-dimensional vector of the conditional pricing errors, i.e.

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

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

$$
\delta = \min_{\theta \in \Theta} E \left[ \lambda(X_t)^t E \left[ R_{t+1} R_{t+1}^t | X_t \right] \lambda(X_t) \right]^{1/2} = \min_{\theta \in \Theta} E \left[ e(X_t; \theta)^t \Omega(X_t) e(X_t; \theta) \right]^{1/2}. \tag{2.7}
$$

The last expression corresponds to the conditional HJ-distance proposed in Nagel and Singleton (2008).\footnote{Differently from our definition, some authors refer to the argument of the unconditional expectation operator in Eq. (2.7) as squared conditional HJ-distance (see, e.g., Balduzzi and Robotti (2010), and Fang, Ren and Yuan (2011)).} This distance is null if, and only if, the conditional moment restriction in Eq. (1.2) is satisfied, that is, the parametric SDF family $\{m(\cdot; \theta) : \theta \in \Theta\}$ is correctly specified. Differently, when the asset pricing model is misspecified, the minimization problem in Eq. (2.2) (or equivalently in Eq. (2.7)) defines the pseudo-true parameter value $\theta_* \in \Theta$. Moreover, the minimized criterion gives the $L^2$-distance between the candidate SDF $m(Y_{t+1}; \theta_*)$ and the closest admissible SDF in set $\mathcal{M}$.

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

**B. Comparison of conditional and unconditional HJ-distances**

If a model is correctly specified, both conditional and unconditional HJ-distances are null. However, if the model is misspecified these distances can differ, and in this section we study the discrepancy between them. However, before describing in detail this discrepancy, let us first get an intuition about it. Let us consider the definition of set $\mathcal{M}$ of admissible SDF’s satisfying the conditional moment restrictions in Eq. (2.1) and the definition of set $\mathcal{M}_Z$ of admissible SDF’s
satisfying the unconditional moment restrictions derived from the chosen instrument matrix in Eq (1.6). By the law of iterated expectations and considering the definitions of the unconditional and conditional HJ-distances in Eqs. (1.5) and (2.2) we deduce that $\mathcal{M} \subseteq \mathcal{M}_Z$. Therefore, the conditional HJ-distance, which is a measure minimized over just a subset of $\mathcal{M}$, is not smaller than the unconditional HJ-distance, which is the corresponding measure minimized over the entire set $\mathcal{M}$:

$$\delta \geq d_Z.$$ (2.8)

In this section we study the difference between the two distances, and we characterize the situations in which this discrepancy is large, and leading to different rankings of competing SDF families, and situations in which it is null. We show in Subsec. B.1 how to retrieve the unconditional HJ-distance from the conditional HJ-distance. We introduce in Subsec. B.2 a representation of the two HJ-distances as vector norms in a suitable Hilbert space. We then provide in Subsec. B.3 an upper and lower bound for the difference between these two norms, for any parametric SDF family. We particularize the results in Subsec. B.4 to the case of SDF families that are conditionally linear w.r.t. the priced risk factors, providing exact expressions for the difference between the two norms.

**B.1 Retrieving the unconditional HJ-distance from the conditional HJ-distance**

Let us consider the expression of the unconditional HJ-distance in Eq. (1.7) and the expression of the conditional HJ-distance in Eq. (2.7). If we want to retrieve the unconditional HJ-distance from the conditional HJ-distance, we have to replace (i) the conditional pricing error vector $e(X_t; \theta)$ with the unconditional cross-moment between it and the instrument matrix $\mathbb{E}\left[Z_t e(X_t; \theta)\right] = \mathbb{E}\left[Z_t h(Y_{t+1}; \theta)\right]$,

and (ii) the matrix $\Omega(X_t)$ of conditional second moments of the test assets’ gross returns with the matrix $\Omega_Z$ of unconditional cross-moments of just the portfolios derived from the chosen instrument matrix $Z_t$. Intuitively, we have to replace conditional moments with unconditional ones, possibly leaving valuable information aside.
B.2 Interpreting the HJ-distances as weighted $L^2$-norms

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

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

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

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

The value $\theta_*$ is the pseudo-true value which minimizes the absolute value of the conditional moment restriction in Eq. (1.2). In order to represent also the unconditional HJ-distance as a $L^2_{\Omega}(\mathcal{X})$-norm, let us parameterize without loss of generality the instrument matrix $Z_t$ by means of a $(N \times q)$-dimensional matrix function $A$ of vector $X_t$ as follows:

$$Z_t = A(X_t)^{\prime} \Omega(X_t).$$  (2.11)

With this parameterization and using the law of iterated expectations the inverse of matrix $\Omega_Z$ and the unconditional pricing error vector become

$$\Omega_Z^{-1} = E \left[ A(X_t)^{\prime} \Omega(X_t) A(X_t) \right],$$

and

$$E \left[ Z_t h(X_{t+1}; \theta) \right] = E \left[ A(X_t)^{\prime} \Omega(X_t) e(X_t; \theta) \right],$$  (2.12)

for any $\theta \in \Theta$. The elements of matrix $\Omega_Z^{-1}$ are the $L^2_{\Omega}(\mathcal{X})$-inner products between the columns of the matrix function $A(X_t)$ in the Hilbert space $L^2_{\Omega}(\mathcal{X})$. The elements of the unconditional

$\begin{align*}
A(X_t)^{\prime} &= Z_t \Omega(X_t)^{-1}.
\end{align*}$

---

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

$$A(X_t)^{\prime} = Z_t \Omega(X_t)^{-1}.$$
pricing error vector are the \( L^2_{\Omega}(\mathcal{X}) \)-inner products between the columns of the matrix function \( A(X_t) \) and the conditional pricing errors vector function \( e(X_t; \theta) \) (see App. A.1). Therefore, from Eq. (1.7) the unconditional HJ-distance under the hypothesis of model misspecification is

\[
d_Z = \min_{\theta \in \Theta} \|P_A[e(\cdot; \theta)]\|_{L^2_{\Omega}(\mathcal{X})} = \|P_A[e(\cdot; \theta^Z)]\|_{L^2_{\Omega}(\mathcal{X})},
\]

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

**B.3 Results for general parametric SDF families**

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

**PROPOSITION 1:** The conditional HJ-distance \( \delta \) and the unconditional HJ-distance \( d_Z \) are such that

\[
\|P_A^+[e(\cdot; \theta^Z)]\|_{L^2_{\Omega}(\mathcal{X})}^2 \geq \delta^2 - d_Z^2 \geq \|P_A^+[e(\cdot; \theta^*)]\|_{L^2_{\Omega}(\mathcal{X})}^2,
\]

where \( P_A^+[\cdot] := I_N - P_A[\cdot] \) 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 \( I_N \) is the \( N \)-dimensional identity matrix.

**Proof.** See App. A.2. \( \square \)

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

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

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

$$\|P_{A}^\perp[e(\cdot; \theta_*)]\|_{L^2(\Omega)}^2 = E \left[ (e(X_t; \theta_*) - E[e(X_t; \theta_*)])' \Omega (e(X_t; \theta_*) - E[e(X_t; \theta_*)]) \right] = \text{Tr} \left[ \Omega V[e(X_t; \theta_*)] \right].$$

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 $\delta$ can be arbitrarily large.

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

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

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

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

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

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

Let us assume we have $\tilde{d}_Z < d_Z$, that is, the SDF family $\tilde{\mathcal{F}}$ has a lower degree of misspecification than the SDF family $\mathcal{F}$ in terms of the unconditional HJ-distance computed using the instrument matrix $Z_t$. By Prop. 1 applied to the SDF family $\tilde{\mathcal{F}}$ the difference $\tilde{\delta}^2 - \tilde{d}_Z^2$ can be arbitrarily large if the $L^2_{\Omega}(\mathcal{X})$-norm of the orthogonal projection of the conditional pricing error vector $\tilde{e}(\cdot; \tilde{\theta}_\ast)$ onto the space unspanned by the columns of the function matrix $A(X_t)$ is sufficiently large. In particular we can have $\tilde{\delta}^2 - \tilde{d}_Z^2 \geq \delta^2 - d_Z^2$, which implies that $\tilde{\delta} \geq \delta$. Therefore, the parametric SDF family $\tilde{\mathcal{F}}$ has an higher degree of misspecification than the SDF family $\mathcal{F}$ in terms of the conditional HJ-distance. Summarizing the results, using to the two HJ-distances we obtain different rankings for the degree of model misspecification.\footnote{From Eq. (2.13) the quantity $\delta^2 - d_Z^2$ depends on the SDF family $\tilde{\mathcal{F}}$ only via $P_{A}^\perp [e(\cdot; \tilde{\theta}_Z)]$, i.e. the part of the conditional pricing error vector $\tilde{e}(X_t; \tilde{\theta}_Z)$ that is spanned by the instrument matrix. Thus, it is possible to have a large $L^2_{\Omega}(\mathcal{X})$-norm of vector $P_{A}^\perp [e(\cdot; \tilde{\theta}_0)]$ without implications for the value of quantity $\delta^2 - d_Z^2$.}

### B.4 Results for linear SDF families

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

\begin{equation}
\label{eq:linear_sdf}
m(Y_{t+1}; \theta) = \tilde{Y}_{t+1}^\prime \theta,
\end{equation}
where the elements of the \( p \)-dimensional vector \( \tilde{Y}_{t+1} \) are functions of the priced factors and the conditioning information. Examples of such linear SDF specifications are considered by Nagel and Singleton (2011), and are among the models included in our empirical analysis in Sec. III.

The conditional pricing error vector at time \( t \) is

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

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

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

which consists of the conditional cross-moments of assets’ gross returns and SDF factors given \( X_t \). Adapting the expression for the conditional HJ-distance \( \delta \) given in Eq. (2.10) to this case, we see that this distance is the \( L^2_\Omega(X) \)-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 Prop. 2 gives explicit expressions for the difference of the squared conditional and unconditional HJ-distances in terms of the conditional pricing error vector \( e(X_t; \cdot) \) valued at \( \theta^* \) and \( \theta_Z \). It also provides two sufficient and necessary conditions for the conditional and unconditional HJ-distances to coincide.

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

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

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

\[
\delta^2 - d_Z^2 = \| \mathcal{P}_A[e(\cdot; \theta_Z)] \|_{L^2_\Omega(X)}^2 - \| e(\cdot; \theta_Z) - e(\cdot; \theta^*) \|_{L^2_\Omega(X)}^2 \tag{2.17}
\]
Moreover we have

$$\delta = d_Z \iff \mathcal{P}_A^+[e(\cdot; \theta_\ast)](X_t) = 0_N \iff \mathcal{P}_A^+[e(\cdot; \theta_Z)](X_t) = 0_N. \quad (2.18)$$

**Proof.** See App. A.3. \qed

In Eq. (2.16), the difference between the squared conditional and unconditional HJ-distances is written as the sum of the squared $L_2(\mathcal{X})$-norms of the projections of the conditional pricing error vector $e(\cdot; \theta_\ast)$ on two mutually orthogonal spaces. The first one is the linear space that is unspanned by the columns of matrix function $A(X_t)$, while the second one is the linear space that is generated by the columns of matrix function $P^A[B](X_t)$, i.e. the component of the column space of matrix function $B(X_t)$ that is spanned by the columns of matrix function $A(X_t)$. We offer an alternative expression for the difference between the squared conditional and unconditional HJ-distances in Eq. (2.17), where this quantity is written as the difference of the squared $L_2(\mathcal{X})$-norms of two vector functions. The first one is $P^A[e(\cdot; \theta_Z)](X_t)$, that is the residual of the orthogonal projection of the conditional pricing error vector $e(\cdot; \theta_Z)$ onto the columns of matrix function $A(X_t)$. The second one is the difference $e(X_t; \theta_Z) - e(X_t; \theta_\ast)$. Note that the first terms in the right hand sides (r.h.s.) of Eqs. (2.16) and (2.17) correspond to the lower and upper bounds, respectively, for the difference $\delta^2 - d_Z^2$ given in Prop. 1. Thus, Prop. 2 explains how far the difference $\delta^2 - d_Z^2$ is from these lower and upper bounds in case of linear SDF families.

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

C. Sample conditional HJ-distance

The conditional HJ-distance $\delta$, and the value $\theta_0$ of the SDF parameter vector if the model is correctly specified, or the value $\theta_\ast$ in case of model misspecification, are unobservable characteristics of the data generating process. In this section we describe an estimation methodology for these characteristics based on a sample of $T$ time series observations on the stationary state variables process $\{X_t\}$. The conditional HJ-distance can be estimated by replacing the unconditional expectation in the criterion in Eq. (2.7) by a sample average, and the conditional expectation in the definition of the conditional pricing error vector in Eq. (2.6) by a nonparametric regression function. We consider kernel smoothing and denote by $K$ a kernel function on set $\mathbb{R}^{d_X}$ and by $b_T$ a positive scalar bandwidth, which depends on the sample size $T$ and converges to 0 as $T$ tends to infinity.\(^8\) The conditional pricing error vector $e(X_t; \theta)$ is estimated by the Nadaraya-Watson kernel regression estimator computed on the sample of $T - 1$ observations of the pair of vectors $Y_{i+1}$ and $X_i$:

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

(2.19)

where the kernel weighting function $w$ is defined as

$$w(x, \tilde{x}) := K \left( \frac{x - \tilde{x}}{b_T} \right) / \sum_{j=1}^{T} K \left( \frac{x - X_{j}}{b_T} \right),$$

(2.20)

for any $x, \tilde{x} \in \mathcal{X}$. Thus, vector $\hat{e}_T(X_t; \theta)$ is a weighted sample average of the vector $h$, such that the closer is the value $X_i$ of the state variables vector at date $i$ to value $X_t$, the larger is the

---

\(^8\)The different components of vector $X_t$ are rescaled before applying the common bandwidth $b_T$. 
weight for that date. Then, the HJ-distance $\delta$ is estimated by the sample conditional HJ-distance $\hat{\delta}_T$ defined by

$$\hat{\delta}^2_T := \min_{\theta \in \Theta} Q_T(\theta), \quad Q_T(\theta) := \frac{1}{T} \sum_{t=1}^{T} I(X_t) e_T(X_t; \theta)' \hat{\Omega}_T(X_t) e_T(X_t; \theta),$$

(2.21)

where

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

(2.22)

is the inverse of the kernel regression estimator of matrix $E \left[ R_{t+1} R'_{t+1} | X_t \right]$ and it is taken as a consistent estimator of the matrix $\Omega(X_t)$. The indicator variable $I(X_t)$ is equal to 1 when $X_t$ is in a given compact subset $X_\star$ of the state variables support $X$ independent on $T$, and 0 otherwise. This indicator variable is used as a trimming factor to control boundary effects in the kernel regressions. We denote by $\hat{\theta}_T$ the minimizer of the criterion function $Q_T(\theta)$, that is an estimator of value $\theta_0$ in case of correct model specification and value $\theta_\ast$ in case of model misspecification.

Nagel and Singleton (2008) write the condition of null HJ-distance as an unconditional moment restriction for the instrument matrix $Z_t = e(X_t; \theta_0)' \Omega(X_t)$ that depends on the true value $\theta_0$ of the SDF parameter vector. They estimate this instrument matrix by kernel methods and then consider a Hansen test-statistic for model validation.

The sample conditional HJ-distance $\hat{\delta}_T$ and the estimator $\hat{\theta}_T$ are closely related to some test statistics and estimators recently proposed in the econometric literature on conditional moment restrictions models. Specifically, estimator $\hat{\theta}_T$ corresponds to a Minimum Distance estimator of Ai and Chen (2003), and to an Euclidean Empirical Likelihood (EEL) estimator of Antoine, Bonnal and Renault (2007) with a non-optimal weighting matrix. The EEL estimator is a member of the class of information-based GMM estimators. These estimators are defined by searching for the value of the structural parameter $\theta$ and the distribution of the data that minimize the discrepancy between this distribution and the empirical one, subject to the conditional moment

---

9The consistency derives from the convergence in probability of the kernel regression estimator and the continuous mapping theorem applied to the inverse function of the estimator.

10See, e.g., Tripathi and Kitamura (2003) and Su and White (2013) for the introduction of trimming factors in test statistics to account for the poor tail estimation in nonparametric conditional probability distributions. Moreover, as noted by Ait-Sahalia, Bickel and Stoker (2001), considering a subset of the state variables support allows to focus on those states which are more relevant for the analysis.
restrictions implied by the model. The EEL estimator relies on a quadratic distance to measure the discrepancy between distributions. Other choices lead to different information-based GMM estimators, such as the Smoothened Empirical Likelihood estimator based on the Kullback-Leibler divergence in Kitamura, Tripathi and Ahn (2004). For correctly specified models, the information-based GMM estimators are asymptotically equivalent to standard GMM estimators based on the optimal choice of instrument and weighting matrices. Thus, they attain the semi-parametric efficiency bound for estimating the true value of parameter $\theta_0$ from the conditional moment restriction $E[h(Y_{t+1}; \theta_0)|X_t] = 0_N$. The squared sample conditional HJ-distance $\hat{\delta}_T^2$ is asymptotically equivalent to the statistic proposed by Tripathi and Kitamura (2003) to test the conditional moment restriction $E[h(Y_{t+1}; \theta)|X_t] = 0_N$, 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.\footnote{An exception is Gospodinov and Otsu (2012) who allow for serial dependence in the sample.} On the other hand, Hall and Inoue (2003) investigate in depth the asymptotic properties of GMM estimators in misspecified unconditional moment restrictions models. Instead, the large sample behavior of the estimator $\hat{\theta}_T$ of the value $\theta_0$ of the SDF parameter vector for well specified models or value $\theta_*$ for misspecified models, and the large sample behavior of the statistic $\hat{\delta}_T$ as an estimator of the conditional HJ-distance in misspecified SDF models are unexplored issues. These issues are the topics of the next section.

II. Model selection using the conditional HJ-distance

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

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

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

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

\[
Tb_T^{d_x/2} \left( \hat{\delta}_T^2 - a_T \right) \xrightarrow{D} N \left( 0, \sigma_0^2 \right),
\]

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

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

and the asymptotic variance \( \sigma_0^2 \) is defined as

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

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

**Proof.** See App. D.3.
The proof of Prop. 3 follows closely the proof of Th. 4.1 in Tripathi and Kitamura (2003) by extending their results to a setting with serially dependent data. The asymptotic distribution in Prop. 3 of the squared HJ-distance sharply differs from the asymptotic distribution of the sample squared unconditional HJ-distance (see, e.g., Hansen, Heaton and Luttmer (1995), Hansen and Jagannathan (1997), and Kan and Robotti (2009)). Indeed, the asymptotic distribution of the sample squared conditional HJ-distance $\hat{\delta}_T^2$ is Gaussian (instead of a linear combination of independent chi-square variables), it involves a centering term $a_T$, and the normalizing factor is $Tb_T^{d_S/2}$ (instead of $T$). These differences are due to the fact that the conditional moment restriction in Eq. (1.2) corresponds to an infinity of unconditional moment restrictions.\footnote{See also the introduction of Domínguez and Lobato (2004) for a discussion about infinite sets of restrictions implied by conditional moment restrictions.} Asymptotic normality is derived by applying a Central Limit Theorem for quadratic forms derived in Yoshihara (1976, 1989).

The centering term $a_T$ for the asymptotic distribution of $\hat{\delta}_T^2$ is nonnegative, since it is the trace of a sum of products of positive semidefinite matrices. Then, since also $\hat{\delta}_T^2$ is nonnegative, the test of the null hypothesis $\delta = 0$ is unilateral. The rejection rule for a test of the correct specification of the SDF family with $\alpha\%$ asymptotic level is

$$\hat{\delta}_T^2 - a_T \geq c_{\alpha/2} \frac{\hat{\sigma}_0}{\sqrt{Tb_T^{d_S/2}}},$$

where $c_{\alpha/2}$ is the $(1 - \alpha)$-quantile of the standard normal probability distribution, and $\hat{\sigma}_0$ is a consistent estimator of $\sigma_0$ obtained by replacing in its definition the matrices $V_0(x)$ and $\Omega(x)$ by kernel estimators.

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

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

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

(3.2)

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

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

where the set $\mathcal{M}_*$ is defined as in Eq. (2.1) with set $\mathcal{X}_*$ in place of set $\mathcal{X}$, and $\theta_*$ is the minimizer of the criterion in Eq. (3.2). The distance $\delta_*$ generally differs from $\delta$, and similarly $\theta_*$ generally differs from $\theta_*$. However, if the set $\mathcal{X}_*$ is chosen sufficiently large, the quantities $\delta_*$ and $\theta_*$ can be made arbitrarily close to $\delta$ and $\theta_*$, respectively. Moreover, we assume that the conditional pricing error $e(X_t; \theta_*)$ does not vanish almost surely on the subset $\mathcal{X}_*$ of the state variable support, i.e.

$$
\delta_* > 0.
$$

(3.3)

The large sample behaviour of the squared sample conditional HJ-distance $\hat{\delta}_T^2$ is given in the next Prop. 4.\textsuperscript{13}

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

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

where the asymptotic variance $\sigma_*^2$ is defined as

$$
\sigma_*^2 := \sum_{l=-\infty}^{\infty} \text{Cov} [ \varepsilon_l, \varepsilon_{l-1} ] + \mathbb{E} \left[ (\eta_{t+1} - \mathbb{E}[\eta_{t+1}|X_t])^2 \right]
$$

\textsuperscript{13}The large sample distribution of estimator $\hat{\theta}_T$ for a misspecified SDF family is given in App. E.1.
for the scalar random variables

\[ \varepsilon_t := I(X_t)e(X_t; \theta_\star)\Omega(X_t)e(X_t; \theta_\star) \]

and

\[ \eta_{t+1} := 2I(X_t)e(X_t; \theta_\star)\Omega(X_t)h(Y_{t+1}; \theta_\star) - I(X_t)e(X_t; \theta_\star)\Omega(X_t)R_{t+1}^2. \]

**Proof.** See App. E.2.

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

A consistent estimator \( \hat{\sigma}_\star^2 \) of the asymptotic variance \( \sigma_\star^2 \) 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 \( \varepsilon_t \) (see e.g. Andrews and Monahan (1992) and Newey and West (1994)). The confidence interval for \( \delta_\star^2 \) with \( \alpha \% \) asymptotic level is obtained by the delta method for the stochastic convergence in Prop. 4 with the logarithmic transformation. The lower confidence bounds \( L \) and the upper confidence bound \( U \) are defined as

\[
L := \hat{\delta}_T^2 \exp \left[ -c_{\alpha/2} \frac{\hat{\sigma}_\star}{\sqrt{T} \hat{\delta}_T^2} \right], \quad U := \hat{\delta}_T^2 \exp \left[ c_{\alpha/2} \frac{\hat{\sigma}_\star}{\sqrt{T} \hat{\delta}_T^2} \right].
\]  

(3.4)
C. **Model selection**

We propose a way to rank $M$ competing asset pricing models for whom the hypothesis of correct model specification is rejected in a test with $\alpha\%$ asymptotic level. This ranking procedure is based on the conditional HJ-distance and it accounts for the statistical uncertainty in the estimation procedure. Let us denote by $L_j$ and $U_j$ the confidence bounds in Eq. (3.4) computed for the $j$-th model, for any $j = 1, \ldots, N$.

**PROPOSITION 5:** Given any $i$-th and $j$-th competing asset pricing models for whom the hypothesis of correct model specification is rejected in a test with $\alpha\%$ asymptotic level, the $i$-th model has a lower degree of model misspecification than the $j$-th model on the basis of the conditional HJ-distance if $U_i < L_j$.

III. **Empirical comparison of parametric SDF specifications**

In this section we compare different no arbitrage asset pricing models for publicly traded U.S. equities and short term T-bills during the last 50 years, on the basis of both unconditional and conditional HJ-distances. In Sec. III.A we describe the competing asset pricing models in terms of the associated SDF’s. In Sec. III.B we describe the data used for the analysis. In Sec. III.C we compare the models on the basis of both unconditional and conditional HJ-distances and discuss the differences.

A. **Competing asset pricing models**

The competing asset pricing models are (i) two structural preference-based asset pricing models, (ii) five linear approximations and corrections of the structural models to include additional priced factors and time-varying risk premia, and (iii) six reduced form specifications inspired by linear factor models. In the two structural preference-based asset pricing models the SDF is implied by the behavior of a representative agent who invests, saves and consumes and we refer to these SDF specifications as CRRA and EZ specifications (see, e.g., Stock and Wright (2000) for the GMM estimation of these models). The SDF of the CRRA specification is linearized in the LCRRA specification, and the same linearization is augmented by an additional priced factor.
in the DLCRRA specification. The coefficients of this linearization are time-varying in the DEF, CAY, YC specifications, which allow for time-varying risk premia. The reduced form specifications are “beta pricing models”, which move from the CAPM and account for stylized facts in the cross-section of equity returns. We refer to these specifications as the CAPM, FF, FF5, FFM, FFL and NM specifications. In the remaining part of the section we describe more precisely all the SDF specifications.

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

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

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

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

\[ m(Y_{t+1}; \theta) = \theta_1^{\theta_3} (C_{t+1}/C_t)^{-\theta_2 \theta_3} R_{m,t+1}^{\theta_3 - 1} , \]

where \( R_{m,t} \) is the gross return on the market portfolio and \( Y_{t+1} = [C_{t+1}/C_t \, R_{m,t+1}]' \). The parameter vector \( \theta = [\theta_1 \, \theta_2 \, \theta_3]' \) includes the time discount rate \( \theta_1 \) and the two additional parameters \( \theta_2 \) and \( \theta_3 \). The parameterization is such that the risk-aversion is \( 1 - \theta_3(1 - \theta_2) \) and the Elasticity of Intertemporal Substitution (EIS) is \( 1/\theta_2 \). The EZ specification reduces to the CRRA specification if \( \theta_3 = 1 \).

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

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

where \( Y_{t+1} = C_{t+1}/C_t \) and \( \theta = [\theta_1 \, \theta_2]' \).
**DEF, CAY, YC** The three specifications DEF, CAY, YC of the SDF correspond to conditional (or “dynamic”) versions of the linear CCAPM. Each specification is linear in logarithmic consumption growth with time-varying coefficients. In particular, these coefficients are affine functions of some conditioning state variables. The three specifications admit the common SDF expression

\[ m(Y_{t+1}; \theta) = (\theta_1 + \theta_2 u_t) + (\theta_3 + \theta_4 u_t) \log \left( \frac{C_{t+1}/C_t}{C_{t+1}/C_t} \right), \]

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

**DLCRRA** The DLCRRA specification corresponds to the linearization of the Durable CCAPM (Yogo (2006)). The Durable CCAPM is an extension of the EZ model to account for the consumption of durables. The linearized SDF has the expression

\[ m(Y_{t+1}; \theta) = \theta_1 + \theta_2 \log \left( \frac{C_{t+1}/C_t}{C_{t+1}/C_t} \right) + \theta_3 \log \left( \frac{C^{(d)}_{t+1}/C^{(d)}_t}{C^{(d)}_{t+1}/C^{(d)}_t} \right) + \theta_4 MKT_{t+1}, \]

(3.5)

where \( C^{(d)}_t \) and \( MKT_t := \log (R_{m,t}) - r_{f,t} \) are the personal consumption of durables and the logarithmic return on the market portfolio in excess of the riskfree rate \( r_{f,t} \) at time \( t \), respectively. In this specification \( Y_{t+1} = \begin{bmatrix} C_{t+1}/C_t & C^{(d)}_{t+1}/C^{(d)}_t & MKT_{t+1} \end{bmatrix}' \) and \( \theta = [\theta_1 \ \theta_2 \ \theta_3 \ \theta_4]' \).\(^{15}\)

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

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

(3.6)

where $F_t$ is a vector of factors, $\theta = [\theta_1 \ \theta_2 \ \bar{\theta}]'$ and $Y_t = [MKT_t \ F_t]'$. In the CAPM specification the vector $F_t$ is null (Treynor (1962), Sharpe (1964), Lintner (1965) and Mossin (1966)). In the FF specification the vector $F_t$ includes the Fama and French (1993, 1998) size factor $SBM_t$ and value factor $HML_t$. In the FF5 specification this vector is augmented by the Fama and French (1993, 1998) short term reversal factor $ST_{rev_t}$ and the long term reversal factor $LT_{rev_t}$. In the FFM and FFL specifications, along the FF value and size factors, the FF momentum factor $UMD_t$ (see Carhart (1997)) and the Pastor and Stambaugh (2003) liquidity factor $L_t$ are considered, respectively. Finally, in the NM specification the vector $F_t$ includes the Novy-Marx (2013) industry-adjusted value factor $\tilde{HML}_t$, momentum factor $\tilde{UMD}_t$ and profitability factor $PMU_t$.

There are three groups of overlapping models. The CRRA specification is nested in the EZ specification. The LCRRA specification is nested in the DEF, CAY, YC and DLCRRA specifications. The FF specification is nested in the FF5, FFM and FFL specifications. Moreover, the DLCRRA, FF, FF5, FFM, FFL and NM specifications overlap since they include the $MKT_t$ factor.

#### B. Data

The study spans the period from Jan. 1959 to Sept. 2012 and focuses on data at the monthly frequency. We obtain data on U.S. consumption from the FRED database of the St. Louis Fed and the Census Bureau of the U.S. Department of Commerce. We consider the sum of per capita personal nondurables and services consumption, seasonally adjusted, as a proxy for consumption $C_t$. Similarly, we consider the per capita personal durables consumption, seasonally adjusted, as a proxy for durables consumption $C_t^{(d)}$. From the FRED database of the St. Louis Fed we also obtain the monthly Moody’s seasoned yield spread between BAA- and AAA-rated corporate bonds to create the monthly time series for the variable $DEF_t$. We derive monthly estimates of the transitory component of financial wealth $CAY_t$ by linear interpolation through...
levels at following end-of-quarters of the Lettau and Ludvigson (2001) quarterly proxies for the consumption to wealth ratio. These quarterly proxies are available on Martin Lettau’s website.\textsuperscript{16} Following Santos and Veronesi (2006) we consider the sum of wages and salaries, transfer payments and other labor income minus personal contributions for social insurance and taxes to create the time series of the labor income to consumption ratio $YC_t$. The data, seasonally adjusted at annual rates, are obtained from the Bureau of Economic Analysis. We consider the monthly Fama-French (FF) proxies for the factors $MKT_t$, $SMB_t$, $HML_t$, $TERM_t$, $DEF_t$ and $UMD_t$; the monthly Pastor-Stambaugh proxy for the factor $L_t$; the monthly Novy-Marx (NM) proxies for the factors $HML_t$, $PMU_t$ and $UMD_t$. The monthly gross returns on the market portfolio $R_{m,t}$ considered in the EZ specification of the SDF are proxied by exponentiating the sum of $MKT_t$ and the 1-month T-bill rate $r_{f,t}$. All these data are taken from the websites of Kenneth French\textsuperscript{17}, Robert Stambaugh\textsuperscript{18} and Robert Novy-Marx\textsuperscript{19}, and because of the availability of data, the analysis of the FFL and NM specifications of the SDF are limited to the periods from Aug. 1962 to Sept. 2012 and from July 1963 to Sept. 2012, respectively.

We consider the six value-weighted Fama-French (FF) research portfolios, which are representative of the U.S. publicly traded equities, and the 1-month T-bill as the $N = 7$ test assets. Each of the six research portfolios exemplify the stochastic behavior of assets with small size and low book-to-market ratio, small size and medium book-to-market, small size and high book-to-market, big size and low book-to-market, big size and medium book-to-market, big size and high book-to-market. As stressed in Nagel and Singleton (2011), considering a low number of assets accounting for “size” and “value” effects is enough to capture most of the cross-sectional return variation in the market for publicly traded U.S. equities (see, e.g., Fama and French (1993) and Lewellen, Nagel and Shanken (2010)).

Table I collects sample mean, standard deviation, minimum and maximum of the variables included in the asset pricing models and the gross returns on the six value-weighted FF research portfolios. Table II reports the correlation among the variables included in the models.

Following Nagel and Singleton (2011) we consider a one-dimensional variable $X_t$ generat-
ing the conditioning information at date $t$ to compute the conditional HJ-distance. This choice prevents from the curse of dimensionality in nonparametric estimation techniques. Therefore, we consider consumption growth, consumption to wealth ratio and labor income to consumption ratio as instrument, for the unconditional HJ-distance, and as variables generating information, for the conditional HJ-distance. The three variables differ in terms of their sample monthly volatility, which, as we can see in Table I, is less than 0.005 for the consumption growth, approximately 0.02 for the consumption to wealth ratio and approximately 0.07 for the labor income to consumption ratio. The three variables display low sample correlation, which, as we can see in Table II, is $-0.12$ between consumption growth and consumption to wealth ratio, $0.21$ between consumption growth and labor income to consumption ratio and $-0.18$ between consumption to wealth ratio and labor income to consumption ratio.

The results reported in Sec. III.C are obtained using the data described so far. However, similar results are obtained using another proxy for consumption and other nineteen proxies for the labor income to consumption ratio. Since there is typically no distinction between consumption and disposable income in general equilibrium models, we consider disposable income from the FRED database of the St. Louis Fed as an alternative proxy for consumption $C_t$. We also consider quarterly data on compensation of employees from the U.S. Department of Commerce, Bureau of Economic Analysis, as an alternative proxy for the labor income (see also Santos and Veronesi (2006)). We obtain monthly estimates by different interpolations methods: we consider the values of the variables between two end-of-quarters as (1) fixed at the previous end-of-quarter levels, (2) fixed at the following end-of-quarter levels, (3) fixed at the midpoint between the previous and the following end-of-quarter levels, and (4) fixed at the closest end-of-quarter levels. We display in Fig. 3 the sketch of the five adopted forms of interpolation between end-of-quarters, that correspond to the points on the horizontal axis where all the lines coincide.

C. Sample unconditional and conditional HJ-distances

We present in this section the estimation and testing results for unconditional and conditional HJ-distances. We first describe separately the results for the unconditional HJ-distance in Subsecs. C.1, and then the results for the conditional HJ-distance in Subsecs. C.2. We compare the
results in Subsec. C.3 and highlight the different assessments of the relative degree of model misspecification based on the two distances.

C.1 Sample unconditional HJ-distance

We compute the sample unconditional HJ-distance \( \hat{d}_{Z,T} \), taken as an estimator of the unconditional HJ-distance \( d_Z \), by replacing population expectations with sample averages. We rely on the constant and the first lagged value of either consumption growth, or consumption to wealth ratio or labor income to consumption ratio as instruments, then obtaining in each case \( qN = 2 \times 7 = 14 \) unconditional moment restrictions for the econometric problem (see Eq. (1.4)). The sample unconditional HJ-distance \( \hat{d}_{Z,T} \) corresponds to the empirical GMM criterion with (non-optimal) weighting matrix \( \Omega_Z \). The GMM estimation is performed in its iterative form.

We report in Table III the pointwise estimates of the unconditional HJ-distance \( d_Z \) and its minimizer \( \theta_Z \), for all the different model specifications. We report in the columns with headers \( L \) and \( U \) the lower and upper bounds at the asymptotic 5%-confidence level under the alternative hypothesis of model misspecification. We do not reject the hypothesis of correct model specification just for the NM and FFM specifications at the asymptotic 5%-confidence level when we consider constant and first lagged value of consumption growth as instrumental variables. In all the remaining cases we do reject the hypothesis of correct model specification. Overall, beta pricing models show the best performance in terms of the unconditional HJ-distance for distinct choices of the instrument matrix. First, let us consider the comparison in the case with constant and first lagged value of consumption growth taken as instrumental variables. [ … ]

C.2 Sample conditional HJ-distance

To compute the sample conditional HJ-distance \( \hat{\delta}_T \) defined in Eq (2.21) we need to know the value of the sample conditional pricing error vector \( \hat{e}_T(X_t; \theta) \). Since this vector is a nonparametric kernel regression function, we need to choose the kernel bandwidth for its computation, and we choose to perform the kernel bandwidth by cross-validation. For any model specification, we first obtain an estimate of the SDF process by optimal GMM, and then we consider the time series to select the bandwidth by cross-validation for each component of the conditional pricing
error vector. The estimate of the SDF process is obtained by optimal GMM in its iterative form
with constant and first two lagged values of the consumption growth combined in the instru-
ment matrix and optimal weighting matrix. The $7 \times 3 = 21$ orthogonality conditions are the
no-arbitrage restrictions on the six value-weighted FF research portfolios and the 1-month T-bill
combined with the instrument matrix. We take as weighting matrix for the GMM criterion the
inverse of the matrix of second moments of the instruments at the first step, and the inverse of
the positive semi-definite spectral density matrix with 12-months Bartlett window proposed in
Newey and West (1987) in the following steps. The value $1e-6$ is chosen as arbitrary threshold
for the convergence of the minimization of the GMM criterion.\footnote{See, e.g., Hall (2005) sec.
2.4 and 3.6 for a discussion on the gain in finite-sample efficiency obtained by the
iterated GMM.} We then consider the time se-
ries of vector $h(Y_t; \hat{\theta}_{GMM,T})$, for vector function $h(\cdot; \cdot)$ defined in Eq. (1.3), and where $\hat{\theta}_{GMM,T}$
is the GMM estimate of the SDF parameter vector just described. We then apply to this time
series the leave-one-out cross-validation criterion to select the bandwidth. The criterion selects
the bandwidth as the componentwise minimizer w.r.t. $b_T$ of the quantity
\begin{equation}
\sum_{i=1}^{T-1} \left( h(Y_{i+1}; \hat{\theta}_{GMM,T}) - \hat{e}_{T;-i}(X_t; \hat{\theta}_{GMM,T}) \right)^2,
\end{equation}
where $\hat{e}_{T;-i}$ is the Nadaraya-Watson kernel regression estimator defined similarly as in Eq.
(2.19) but computed on the sample missing of the $i$-th observation. We report the results for
the Gaussian kernel function $K(x) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2}x^2\right)$, for which convolution $K(x) =
\frac{1}{4\sqrt{\pi}} \exp \left(-\frac{1}{4}x^2\right)$ and the integral of the squared convolution $\int_{\mathbb{R}^d} K(u)^2 du = \frac{1}{\sqrt{8\pi}}$ are
known in closed form. The Gaussian kernel function $K$ and its convolution $K$ defined in Prop.
3 are displayed in Fig. 2. However, using different specifications of the kernel function, such as
the rectangular or the Epanechnikov, we obtain similar results. We report in Table V the band-
width selected for the different models, and for different choices of the conditioning variable. In
some cases, indicated by $\ast$ in the table, the variation of the kernel smoothed function around its
average is extremely low. As a consequence, the leave-one-out cross-validation criterion tends
to lead to oversmoothing. Therefore, in these cases we use the set of bandwidths selected by the
rule of thumb for the nonparametric kernel estimation of the probability density of the condi-
tioning variable, that is \( b_T = 0.9 \min \{ s, R/1.34 \} T^{-\frac{1}{2}} \) as suggested in Silverman (1986), where \( s \) is the sample volatility of the conditioning variable and \( R \) is its sample interquartile range. The level of labor to income consumption ratio has a downward trend and its sample pdf is not unimodal. Therefore we consider the first differences of this variable in the bandwidth selection by the Silverman’s rule of thumb. We plot in Fig. 1 an example of the Nadaraya-Watson kernel regression estimator for the conditional pricing error vector. In this example the conditioning variable \( X_t \) is consumption growth, and the vector \( 10^{-3}[4.13 \ 3.79 \ 3.95 \ 3.33 \ 3.70 \ 3.41 \ 7.28]' \) is selected by leave-one-out cross-validation as the vector of bandwidths.

We report in Table IV the estimates of the conditional HJ-distance \( \delta \) and its minimizer \( \theta_* \) implicitly defined in Eq. (2.10), for all the different model specifications. The estimates are computed relying on the 7 conditional moment restrictions derived from no-arbitrage for the six FF value-weighted research portfolios and the 1-month T-bill. Similarly as in the analysis based on the unconditional HJ-distance in the previous subsection, we report in the columns with headers \( L \) and \( U \) the lower and upper bounds at the asymptotic 5%-confidence level under the hypothesis of misspecification. We do not reject the hypothesis of correct model specification at the asymptotic 5%-confidence level just for the NM and FFM specifications when we consider consumption growth as the variable generating conditioning information. \([\ldots]\)

**C.3 Differences between sample unconditional and conditional HJ-distances**

On the basis of the unconditional HJ-distance the reduced-form beta-pricing models show a lower degree of model misspecification than the other specifications, for all the three choices of instrument matrix. We do not find the same distinct degree of model misspecification between reduced form models on the one side and structural models and their approximations and extensions on the other side when we analyze the asset pricing models on the basis of the conditional HJ-distance. We find indeed in this second analysis structural models and their approximations and extensions showing a lower degree of model misspecification than several reduce form beta pricing models. \([\ldots]\)
IV. Conclusion

[...]


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


Table I: Sample unconditional statistics of the variables included in the asset pricing models and the gross returns on the six value-weighted FF research portfolios.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_{f,t} )</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.01</td>
</tr>
<tr>
<td>( C_{t+1}/C_t )</td>
<td>1.00</td>
<td>0.00</td>
<td>0.99</td>
<td>1.02</td>
</tr>
<tr>
<td>( C_{t+1}^{d}/C_t^{d} )</td>
<td>1.00</td>
<td>0.03</td>
<td>0.85</td>
<td>1.14</td>
</tr>
<tr>
<td>( DEF_t )</td>
<td>1.02</td>
<td>0.46</td>
<td>0.32</td>
<td>3.38</td>
</tr>
<tr>
<td>( CAY_t )</td>
<td>0.00</td>
<td>0.02</td>
<td>-0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>( YC_t )</td>
<td>0.75</td>
<td>0.07</td>
<td>0.66</td>
<td>0.87</td>
</tr>
<tr>
<td>( MKT_t )</td>
<td>0.47</td>
<td>4.47</td>
<td>-23.24</td>
<td>16.10</td>
</tr>
<tr>
<td>( SMB_t )</td>
<td>0.22</td>
<td>3.05</td>
<td>-16.39</td>
<td>22.00</td>
</tr>
<tr>
<td>( HML_t )</td>
<td>0.38</td>
<td>2.83</td>
<td>-12.60</td>
<td>13.84</td>
</tr>
<tr>
<td>( UMD_t )</td>
<td>0.74</td>
<td>4.18</td>
<td>-34.74</td>
<td>18.39</td>
</tr>
<tr>
<td>( ST_{rev,t} )</td>
<td>0.53</td>
<td>3.10</td>
<td>-14.51</td>
<td>16.23</td>
</tr>
<tr>
<td>( LT_{rev,t} )</td>
<td>0.29</td>
<td>2.50</td>
<td>-7.78</td>
<td>14.47</td>
</tr>
<tr>
<td>( L_t )</td>
<td>-0.03</td>
<td>0.06</td>
<td>-0.46</td>
<td>0.20</td>
</tr>
<tr>
<td>( HML_t )</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.05</td>
<td>0.07</td>
</tr>
<tr>
<td>( PMU_t )</td>
<td>0.00</td>
<td>0.03</td>
<td>-0.23</td>
<td>0.12</td>
</tr>
<tr>
<td>( UMD_t )</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.05</td>
<td>0.07</td>
</tr>
<tr>
<td>( SMLO_t )</td>
<td>1.05</td>
<td>0.07</td>
<td>0.73</td>
<td>1.33</td>
</tr>
<tr>
<td>( SMME_t )</td>
<td>1.02</td>
<td>0.05</td>
<td>0.76</td>
<td>1.31</td>
</tr>
<tr>
<td>( SMHI_t )</td>
<td>1.02</td>
<td>0.06</td>
<td>0.76</td>
<td>1.36</td>
</tr>
<tr>
<td>( BILO_t )</td>
<td>1.01</td>
<td>0.05</td>
<td>0.80</td>
<td>1.24</td>
</tr>
<tr>
<td>( BIEM_t )</td>
<td>1.01</td>
<td>0.04</td>
<td>0.82</td>
<td>1.19</td>
</tr>
<tr>
<td>( BIHI_t )</td>
<td>1.02</td>
<td>0.05</td>
<td>0.80</td>
<td>1.24</td>
</tr>
</tbody>
</table>
Table II: Sample correlation matrix of the priced risk factors included in the asset pricing models.

<table>
<thead>
<tr>
<th></th>
<th>$C_{t+1}/C_t$</th>
<th>$C_{t+1}^{d}/C_{t}^{d}$</th>
<th>$DEF_t$</th>
<th>$CAY_t$</th>
<th>$YC_t$</th>
<th>$MKT_t$</th>
<th>$SMB_t$</th>
<th>$HML_t$</th>
<th>$ST_{rev_t}$</th>
<th>$LT_{rev_t}$</th>
<th>$L_t$</th>
<th>$HML_t$</th>
<th>$PMU_t$</th>
<th>$UMD_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{t+1}/C_t$</td>
<td>1.00</td>
<td>0.21</td>
<td>-0.03</td>
<td>-0.12</td>
<td>0.21</td>
<td>0.06</td>
<td>0.03</td>
<td>-0.04</td>
<td>-0.01</td>
<td>-0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
<td>-0.12</td>
</tr>
<tr>
<td>$C_{t+1}^{d}/C_{t}^{d}$</td>
<td>0.21</td>
<td>1.00</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.06</td>
<td>0.13</td>
<td>0.01</td>
<td>0.00</td>
<td>-0.04</td>
<td>0.04</td>
<td>0.01</td>
<td>0.04</td>
<td>-0.04</td>
<td>-0.10</td>
</tr>
<tr>
<td>$DEF_t$</td>
<td>-0.03</td>
<td>-0.01</td>
<td>1.00</td>
<td>-0.19</td>
<td>-0.11</td>
<td>-0.01</td>
<td>0.02</td>
<td>-0.05</td>
<td>-0.01</td>
<td>0.03</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.14</td>
<td>-0.09</td>
</tr>
<tr>
<td>$CAY_t$</td>
<td>-0.12</td>
<td>-0.01</td>
<td>-0.19</td>
<td>1.00</td>
<td>-0.18</td>
<td>0.02</td>
<td>-0.05</td>
<td>-0.01</td>
<td>0.03</td>
<td>0.00</td>
<td>0.01</td>
<td>0.10</td>
<td>0.04</td>
<td>0.13</td>
</tr>
<tr>
<td>$YC_t$</td>
<td>0.21</td>
<td>0.06</td>
<td>-0.11</td>
<td>-0.18</td>
<td>1.00</td>
<td>-0.02</td>
<td>0.03</td>
<td>0.04</td>
<td>0.02</td>
<td>0.09</td>
<td>0.04</td>
<td>-0.11</td>
<td>0.00</td>
<td>-0.12</td>
</tr>
<tr>
<td>$MKT_t$</td>
<td>0.06</td>
<td>0.13</td>
<td>-0.01</td>
<td>0.02</td>
<td>-0.02</td>
<td>1.00</td>
<td>0.31</td>
<td>-0.30</td>
<td>-0.13</td>
<td>0.29</td>
<td>-0.04</td>
<td>0.27</td>
<td>-0.15</td>
<td>-0.17</td>
</tr>
<tr>
<td>$SMB_t$</td>
<td>0.03</td>
<td>0.01</td>
<td>0.02</td>
<td>-0.05</td>
<td>0.03</td>
<td>0.31</td>
<td>1.00</td>
<td>-0.23</td>
<td>-0.01</td>
<td>0.17</td>
<td>0.25</td>
<td>0.14</td>
<td>-0.07</td>
<td>-0.29</td>
</tr>
<tr>
<td>$HML_t$</td>
<td>-0.04</td>
<td>0.00</td>
<td>-0.05</td>
<td>-0.01</td>
<td>0.04</td>
<td>-0.30</td>
<td>-0.23</td>
<td>1.00</td>
<td>-0.16</td>
<td>0.43</td>
<td>0.07</td>
<td>0.75</td>
<td>-0.18</td>
<td>-0.11</td>
</tr>
<tr>
<td>$UMD_t$</td>
<td>0.09</td>
<td>-0.04</td>
<td>-0.10</td>
<td>0.03</td>
<td>0.02</td>
<td>-0.13</td>
<td>-0.01</td>
<td>-0.16</td>
<td>1.00</td>
<td>-0.29</td>
<td>-0.06</td>
<td>0.04</td>
<td>-0.11</td>
<td>0.93</td>
</tr>
<tr>
<td>$ST_{rev_t}$</td>
<td>-0.01</td>
<td>0.04</td>
<td>-0.02</td>
<td>0.00</td>
<td>0.09</td>
<td>0.29</td>
<td>0.17</td>
<td>-0.02</td>
<td>-0.29</td>
<td>1.00</td>
<td>0.08</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.28</td>
</tr>
<tr>
<td>$LT_{rev_t}$</td>
<td>-0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.04</td>
<td>-0.04</td>
<td>0.25</td>
<td>0.43</td>
<td>-0.06</td>
<td>0.08</td>
<td>1.00</td>
<td>-0.05</td>
<td>0.45</td>
<td>-0.11</td>
<td>-0.20</td>
</tr>
<tr>
<td>$L_t$</td>
<td>0.00</td>
<td>0.04</td>
<td>-0.14</td>
<td>0.10</td>
<td>-0.11</td>
<td>0.27</td>
<td>0.14</td>
<td>-0.07</td>
<td>0.04</td>
<td>-0.02</td>
<td>1.00</td>
<td>0.00</td>
<td>0.03</td>
<td>-0.06</td>
</tr>
<tr>
<td>$HML_t$</td>
<td>0.00</td>
<td>-0.01</td>
<td>-0.06</td>
<td>0.04</td>
<td>-0.01</td>
<td>-0.15</td>
<td>-0.01</td>
<td>0.75</td>
<td>-0.11</td>
<td>-0.01</td>
<td>0.45</td>
<td>0.00</td>
<td>1.00</td>
<td>-0.17</td>
</tr>
<tr>
<td>$PMU_t$</td>
<td>0.07</td>
<td>-0.04</td>
<td>-0.09</td>
<td>0.04</td>
<td>0.00</td>
<td>-0.17</td>
<td>-0.07</td>
<td>-0.18</td>
<td>0.93</td>
<td>-0.28</td>
<td>-0.11</td>
<td>0.03</td>
<td>-0.17</td>
<td>1.00</td>
</tr>
<tr>
<td>$UMD_t$</td>
<td>-0.12</td>
<td>-0.10</td>
<td>0.09</td>
<td>0.13</td>
<td>-0.12</td>
<td>-0.29</td>
<td>-0.29</td>
<td>-0.11</td>
<td>0.24</td>
<td>-0.13</td>
<td>-0.20</td>
<td>-0.06</td>
<td>-0.23</td>
<td>0.28</td>
</tr>
</tbody>
</table>
Table III: Estimate of the squared unconditional HJ-distance and associated SDF vector under different model specifications. The results in the upper part of the table rely on the orthogonality conditions obtained combining (i) the 7 no-arbitrage restrictions holding for the six FF value-weighted research portfolios and the 1-month T-bill and (ii) constant and first lagged value of nondurables consumption growth taken as instrumental variables. In the middle and lower parts of the table, the consumption to wealth ratio and the labor income to consumption ratio are considered as instrument in place of the consumption growth, respectively. The symbol * close to the specification name indicates that we do not reject the hypothesis of correct model specification at the asymptotic 5%-confidence level. In all the other cases we report in columns $L$ and $U$ the confidence bounds for the squared unconditional HJ-distance under the alternative hypothesis of model misspecification. The SDF vector minimizing the unconditional HJ-distance is also given.

<table>
<thead>
<tr>
<th></th>
<th>$d^2_J$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
<th>$\theta_4$</th>
<th>$\theta_5$</th>
<th>$\theta_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM</td>
<td>0.02</td>
<td>1.48</td>
<td>-0.12</td>
<td>-47.69</td>
<td>-27.64</td>
<td>-14.09</td>
<td>-</td>
</tr>
<tr>
<td>FFM</td>
<td>0.03</td>
<td>1.31</td>
<td>-0.11</td>
<td>-0.02</td>
<td>-0.16</td>
<td>-0.26</td>
<td>-</td>
</tr>
<tr>
<td>FFS</td>
<td>0.07</td>
<td>1.08</td>
<td>-0.05</td>
<td>0.07</td>
<td>-0.02</td>
<td>-0.31</td>
<td>0.06</td>
</tr>
<tr>
<td>FFL</td>
<td>0.08</td>
<td>0.67</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.07</td>
<td>-11.28</td>
<td>-</td>
</tr>
<tr>
<td>FF</td>
<td>0.08</td>
<td>1.06</td>
<td>-0.06</td>
<td>-0.02</td>
<td>-0.07</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>DLCRRA</td>
<td>0.09</td>
<td>0.70</td>
<td>1.06</td>
<td>7.69</td>
<td>-0.02</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>EZ</td>
<td>0.11</td>
<td>1.30</td>
<td>39.47</td>
<td>-6.03</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>DEF</td>
<td>0.11</td>
<td>1.30</td>
<td>46.80</td>
<td>-1.32</td>
<td>-247.46</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>CAPM</td>
<td>0.12</td>
<td>1.02</td>
<td>-0.01</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>YC</td>
<td>0.12</td>
<td>8.12</td>
<td>-3149.26</td>
<td>-10.85</td>
<td>4394.94</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>CAY</td>
<td>0.14</td>
<td>0.94</td>
<td>18.55</td>
<td>-56.25</td>
<td>-20.43</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>CRRA</td>
<td>0.17</td>
<td>1.00</td>
<td>0.33</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LCRRA</td>
<td>0.17</td>
<td>0.98</td>
<td>1.61</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFS</td>
<td>0.06</td>
<td>1.20</td>
<td>0.00</td>
<td>-0.01</td>
<td>-0.19</td>
<td>-0.45</td>
<td></td>
</tr>
<tr>
<td>NM</td>
<td>0.06</td>
<td>1.21</td>
<td>39.04</td>
<td>-31.44</td>
<td>-7.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFM</td>
<td>0.07</td>
<td>1.20</td>
<td>-0.09</td>
<td>-0.02</td>
<td>-0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFL</td>
<td>0.09</td>
<td>0.89</td>
<td>-0.04</td>
<td>-0.02</td>
<td>-0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FF</td>
<td>0.09</td>
<td>1.06</td>
<td>-0.06</td>
<td>-0.02</td>
<td>-0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EZ</td>
<td>0.11</td>
<td>1.30</td>
<td>39.12</td>
<td>-6.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLCRRA</td>
<td>0.11</td>
<td>0.87</td>
<td>67.07</td>
<td>-43.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEF</td>
<td>0.13</td>
<td>1.00</td>
<td>399.17</td>
<td>-6.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAPM</td>
<td>0.13</td>
<td>1.02</td>
<td>-5.04</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAY</td>
<td>0.15</td>
<td>1.79</td>
<td>-195.41</td>
<td>63.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YC</td>
<td>0.15</td>
<td>-2.30</td>
<td>-715.05</td>
<td>5.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LCRRA</td>
<td>0.17</td>
<td>1.30</td>
<td>-65.21</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM</td>
<td>0.03</td>
<td>1.34</td>
<td>-0.10</td>
<td>-37.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFM</td>
<td>0.04</td>
<td>1.24</td>
<td>-0.10</td>
<td>-0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFS</td>
<td>0.05</td>
<td>1.15</td>
<td>-0.01</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFL</td>
<td>0.08</td>
<td>1.08</td>
<td>-0.06</td>
<td>-0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FF</td>
<td>0.08</td>
<td>1.06</td>
<td>-0.06</td>
<td>-0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEF</td>
<td>0.11</td>
<td>1.30</td>
<td>280.29</td>
<td>-281.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EZ</td>
<td>0.11</td>
<td>0.98</td>
<td>-1.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLCRRA</td>
<td>0.11</td>
<td>0.92</td>
<td>42.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAY</td>
<td>0.12</td>
<td>1.62</td>
<td>-116.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAPM</td>
<td>0.12</td>
<td>1.02</td>
<td>-5.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YC</td>
<td>0.13</td>
<td>12.08</td>
<td>-2556.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRRA</td>
<td>0.16</td>
<td>1.30</td>
<td>69.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LCRRA</td>
<td>0.16</td>
<td>1.19</td>
<td>-41.76</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

42
Table IV: Estimate of the squared conditional HJ-distance and associated SDF vector under different model specifications. The conditional moment restrictions from no-arbitrage for the six FF value-weighted research portfolios and the 1-month T-bill are considered. In the upper, middle and lower part of the table the consumption growth, the consumption to wealth ratio and the labor income to consumption ratio are respectively taken as the variable generating the conditioning information. The symbol ∗ close to the specification name indicates that we do not reject the hypothesis of correct model specification at the asymptotic 5%-confidence level. In all the other cases we report in columns $L$ and $U$ the confidence bounds for the squared conditional HJ-distance under the alternative hypothesis of misspecification. In the last column we group models as explained in Sec. III.C. The SDF vector minimizing the conditional HJ-distance is also given.

<table>
<thead>
<tr>
<th></th>
<th>$\delta^2$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
<th>$\theta_4$</th>
<th>$\theta_5$</th>
<th>$\theta_6$</th>
<th>$L$</th>
<th>$U$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM*</td>
<td>0.03</td>
<td>1.23</td>
<td>-0.09</td>
<td>-29.21</td>
<td>-0.90</td>
<td>-32.45</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFM*</td>
<td>0.05</td>
<td>-0.07</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.08</td>
<td>0.04</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEF</td>
<td>0.11</td>
<td>371.40</td>
<td>-1.25</td>
<td>-321.30</td>
<td>-</td>
<td>-</td>
<td>0.09</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>EZ</td>
<td>0.12</td>
<td>-1.84</td>
<td>-4.58</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.11</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>CAPM</td>
<td>0.13</td>
<td>0.96</td>
<td>-0.05</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.12</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>ZAY</td>
<td>0.16</td>
<td>-1.40</td>
<td>-62.98</td>
<td>5779.52</td>
<td>-</td>
<td>-</td>
<td>0.14</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>DLCRRRA</td>
<td>0.17</td>
<td>0.81</td>
<td>-19.66</td>
<td>7.95</td>
<td>-6.00</td>
<td>-</td>
<td>0.16</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>FF5</td>
<td>0.19</td>
<td>0.91</td>
<td>-0.04</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.04</td>
<td>-0.03</td>
<td>0.15</td>
<td>0.24</td>
</tr>
<tr>
<td>YC</td>
<td>0.19</td>
<td>2.76</td>
<td>-302.70</td>
<td>-2.50</td>
<td>405.82</td>
<td>-</td>
<td>0.17</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>FFL</td>
<td>0.19</td>
<td>0.93</td>
<td>-0.04</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.48</td>
<td>-</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>FF</td>
<td>0.19</td>
<td>1.00</td>
<td>-0.06</td>
<td>-0.02</td>
<td>-0.06</td>
<td>-</td>
<td>-</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>CRRA</td>
<td>0.20</td>
<td>0.87</td>
<td>-10.31</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.17</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>LCRRA</td>
<td>0.20</td>
<td>0.87</td>
<td>8.93</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.16</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>NM</td>
<td>0.07</td>
<td>1.28</td>
<td>-0.08</td>
<td>-37.68</td>
<td>-21.29</td>
<td>-4.58</td>
<td>-</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td>FFL</td>
<td>0.09</td>
<td>0.85</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.07</td>
<td>-2.75</td>
<td>-</td>
<td>0.08</td>
<td>0.09</td>
</tr>
<tr>
<td>FFM</td>
<td>0.10</td>
<td>1.02</td>
<td>-0.06</td>
<td>-0.03</td>
<td>-0.10</td>
<td>-0.07</td>
<td>-</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td>FF5</td>
<td>0.10</td>
<td>0.94</td>
<td>-0.05</td>
<td>-0.06</td>
<td>-0.10</td>
<td>0.01</td>
<td>0.07</td>
<td>0.09</td>
<td>0.11</td>
</tr>
<tr>
<td>FF</td>
<td>0.12</td>
<td>0.95</td>
<td>-0.05</td>
<td>-0.02</td>
<td>-0.07</td>
<td>-</td>
<td>-</td>
<td>0.09</td>
<td>0.15</td>
</tr>
<tr>
<td>DLCRRRA</td>
<td>0.16</td>
<td>1.05</td>
<td>7.74</td>
<td>-37.92</td>
<td>-0.02</td>
<td>-</td>
<td>-</td>
<td>0.15</td>
<td>0.17</td>
</tr>
<tr>
<td>DEF</td>
<td>0.24</td>
<td>-0.33</td>
<td>376.92</td>
<td>0.59</td>
<td>-277.75</td>
<td>-</td>
<td>-</td>
<td>0.23</td>
<td>0.25</td>
</tr>
<tr>
<td>YC</td>
<td>0.26</td>
<td>1.58</td>
<td>-502.95</td>
<td>-0.90</td>
<td>623.87</td>
<td>-</td>
<td>-</td>
<td>0.22</td>
<td>0.31</td>
</tr>
<tr>
<td>CAY</td>
<td>0.27</td>
<td>0.87</td>
<td>-12.74</td>
<td>0.39</td>
<td>159.51</td>
<td>-</td>
<td>-</td>
<td>0.20</td>
<td>0.34</td>
</tr>
<tr>
<td>LCRRA</td>
<td>0.27</td>
<td>0.75</td>
<td>9.60</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.16</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>CAPM</td>
<td>0.31</td>
<td>0.85</td>
<td>-0.02</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.14</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>EZ</td>
<td>0.36</td>
<td>0.70</td>
<td>-4.10</td>
<td>0.62</td>
<td>-</td>
<td>-</td>
<td>0.33</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>CRRA</td>
<td>0.47</td>
<td>0.75</td>
<td>-10.90</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.40</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>FFL</td>
<td>0.10</td>
<td>1.08</td>
<td>-0.07</td>
<td>-0.03</td>
<td>-0.10</td>
<td>-0.12</td>
<td>-</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td>FF5</td>
<td>0.10</td>
<td>0.99</td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.02</td>
<td>0.15</td>
<td>-0.01</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td>DLCRRRA</td>
<td>0.11</td>
<td>0.88</td>
<td>20.25</td>
<td>-16.40</td>
<td>-0.01</td>
<td>-</td>
<td>0.07</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>CAY</td>
<td>0.15</td>
<td>0.85</td>
<td>-10.52</td>
<td>6.74</td>
<td>-708.99</td>
<td>-</td>
<td>-</td>
<td>0.14</td>
<td>0.15</td>
</tr>
<tr>
<td>NM</td>
<td>0.16</td>
<td>0.93</td>
<td>-0.02</td>
<td>-4.31</td>
<td>-5.13</td>
<td>2.06</td>
<td>-</td>
<td>0.15</td>
<td>0.17</td>
</tr>
<tr>
<td>FFL</td>
<td>0.18</td>
<td>0.94</td>
<td>-0.04</td>
<td>-0.04</td>
<td>-0.06</td>
<td>-1.27</td>
<td>-</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td>DEF</td>
<td>0.19</td>
<td>1.13</td>
<td>-47.64</td>
<td>-0.29</td>
<td>2.07</td>
<td>-</td>
<td>0.16</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>FF</td>
<td>0.19</td>
<td>0.85</td>
<td>-0.04</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-</td>
<td>0.15</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>CAPM</td>
<td>0.19</td>
<td>0.80</td>
<td>-0.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.16</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>EZ</td>
<td>0.23</td>
<td>0.70</td>
<td>17.80</td>
<td>0.78</td>
<td>-</td>
<td>-</td>
<td>0.15</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>CRRA</td>
<td>0.16</td>
<td>0.88</td>
<td>-4.58</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.12</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>YC</td>
<td>0.17</td>
<td>1.22</td>
<td>10.18</td>
<td>-0.48</td>
<td>-25.35</td>
<td>-</td>
<td>0.14</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>LCRRA</td>
<td>0.17</td>
<td>0.82</td>
<td>-3.54</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.14</td>
<td>0.20</td>
<td></td>
</tr>
</tbody>
</table>
Table V: Kernel bandwidth used for the estimation of the conditional HJ-distance based on the orthogonality conditions for the six FF value-weighted research portfolios and the 1-month T-bill. A factor $10^{-3}$ multiplies each number. In each column the header indicates the variable taken in first lag as the generator of the conditioning information. For each model the test assets in an orthogonality conditions are ordered as SMLO, SMME, SMHI, BILO, BIME, BIHI, Rf. The bandwidth is chosen by leave-one-out cross-validation, that is by minimizing the criterion in Eq. (3.7). In the cases indicated by * the variation of the kernel smoothed function around its average is extremely low. The leave-one-out cross-validation criterion tends to oversmooth and we use the optimal bandwidth selected by the rule of thumb for the probability density of the conditioning variable, that is $b_T = 0.9 \min \{s, R/1.34\} T^{-1/2}$ as suggested in Silverman (1986), where $s$ and $R$ denote the sample volatility and interquartile range of the observations, respectively. In these case, the maximum variation of the smoothed function valued for the bandwidth selected by leave-one-out cross-validation and Silverman’s rule of thumb is less than 0.01. The level of labor income consumption ratio has a downward trend and its sample pdf is not unimodal. Therefore we consider the first differences of this variable in the bandwidth selection.

| M | $\frac{C_{t+1}}{C_t}$ | CAY | YC_t | M | $\frac{C_{t+1}}{C_t}$ | CAY | YC_t | M | $\frac{C_{t+1}}{C_t}$ | CAY | YC_t | M | $\frac{C_{t+1}}{C_t}$ | CAY | YC_t | M | $\frac{C_{t+1}}{C_t}$ | CAY | YC_t | M | $\frac{C_{t+1}}{C_t}$ | CAY | YC_t |
| NM | 6.87 | 15.82 | 42.68 | 5.91 | * | * | 4.05 | 2.16 | 3.38 | 1.15 | * | * | 4.02 | 7.72 | 21.44 | 3.60 | 6.20 | 22.05 | 3.79 | 8.64 | 24.54 | 3.72 | 11.51 | 44.85 | 3.73 | 8.56 | 23.93 | 3.38 | 24.53 | 23.02 | 4.49 | 29.58 | 10.23 |
| FLM | 2.78 | 31.94 | 19.06 | 3.56 | 10.73 | 16.85 | 1.91 | * | 75.74 | 4.28 | 1.21 | 5.23 | 4.64 | 1.14 | 5.13 | 4.51 | 1.23 | 5.15 | 4.52 | 1.12 | 5.86 | 4.73 | 1.09 | 5.32 | 5.04 | 1.39 | 5.24 | 4.49 | 1.45 | 5.78 |
Figure 1: Example of Nadaraya-Watson kernel regression estimator \( \hat{e}_T(X_t; \hat{\theta}_{GMM,T}) \) for the conditional pricing error vector. The conditioning variable \( X_t \) is consumption growth. The SDF specification is the CRRA and the GMM estimator of the SDF parameter vector is \( \hat{\theta}_{GMM,T} = [1.00 1.69]' \). This vector is the estimate of the SDF parameter vector obtained by GMM with optimal weighting matrix. The test assets are the six value-weighted FF research portfolios (SMLO, SMME, SMHI, BILO, BIME, BIHI) and the 1-month T-bill (Rf). The estimator is computed using the vector \( 10^{-3}[4.13 3.79 3.95 3.33 3.70 3.41 7.28]' \) of bandwidths, which are the optimal ones on the basis of the leave-one-out cross-validation criterion.

Figure 2: The Gaussian kernel function \( K \) and the associated kernel convolution \( K \) defined in Prop. 3.
Figure 3: Sketch of the five adopted forms of interpolation between end-of-quarters, that correspond to the points on the horizontal axis where all the lines coincide. Summarizing, we consider
Appendix A. Proof of the results in Section I.B

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

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

(A1)

for the definition of the \(L^2_{\Omega}(\mathcal{X})\)-inner product given in Eq. (2.9). Consider now the matrix \(\Phi(X_t)\) such that its columns are linearly independent vectors in \(L^2_{\Omega}(\mathcal{X})\). We denote by \(P_{\Phi}[\psi]\) the orthogonal projection operator onto the linear subspace of \(L^2_{\Omega}(\mathcal{X})\) spanned by the columns of the matrix function \(\Phi\) applied to vector function \(\psi\):

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

(A2)

for any \(x \in \mathcal{X}\).

Appendix 1. Proof of Equation (2.13)

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

\[
E[Z_{t+1}(Y_{t+1}; \theta) | h(Y_t; \cdot; \theta) \rangle = \langle A, e(\cdot; \theta) \rangle_{L^2_{\Omega}(\mathcal{X})} \quad \text{and} \quad \Omega_Z^{-1} = \langle A, A \rangle_{L^2_{\Omega}(\mathcal{X})}.
\]

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

$$\mathcal{P}_A[e(\cdot; \theta)](X_t) = A(X_t)\langle A, A \rangle_{L^2_\Omega(X_t)}^{-1} \langle A, e(\cdot; \theta) \rangle_{L^2_\Omega(X_t)} = A(X_t)\Omega Z E[Z_t h(Y_{t+1}; \theta)]. \quad (A3)$$

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

$$\|\mathcal{P}_A[e(\cdot; \theta)]\|_{L^2_\Omega(X)} = (E[Z_t h(Y_{t+1}; \theta)]')^t \Omega Z E[A(X_t)'\Omega(X_t)A(X_t)] \Omega Z E[Z_t h(Y_{t+1}; \theta)]^{1/2} = (E[Z_t h(Y_{t+1}; \theta)]')^t \Omega Z E[Z_t h(Y_{t+1}; \theta)]^{1/2}.$$ 

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

**Appendix 2. Proof of Proposition 1**

From Eq. (2.13) the criterion $\|\mathcal{P}_A[e(\cdot; \theta)]\|_{L^2_\Omega(X)}$ valued at its minimum $\theta_Z$ cannot be greater than the same criterion at any other value, which implies $d^2_Z \leq \|\mathcal{P}_A[e(\cdot; \theta_*)]\|_{L^2_\Omega(X)}^2$. From the Pythagorean theorem for inner product spaces we have that

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

From the last inequality, Eq. (A4) and the definition of conditional HJ-distance $\delta$ given in Eq. (2.10) we get

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

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

$$\|e(\cdot; \theta_Z)\|_{L^2_\Omega(X)}^2 = \|\mathcal{P}_A[e(\cdot; \theta_Z)]\|_{L^2_\Omega(X)}^2 + \|\mathcal{P}_A^\perp[e(\cdot; \theta_Z)]\|_{L^2_\Omega(X)}^2 = d^2_Z + \|\mathcal{P}_A^\perp[e(\cdot; \theta_Z)]\|_{L^2_\Omega(X)}^2. \quad (A5)$$

where we use the expression for $d_Z$ given in Eq. (2.13). The criterion $\|e(\cdot; \theta)\|_{L^2_\Omega(X)}^2$ in Eq. (2.10) valued at its minimum $\theta_*$ cannot be greater than the same criterion at any other value, and
in particular at $\theta_Z$, which implies

$$\delta^2 \leq \|e(\cdot; \theta_Z)\|_{L_0^2(X)}^2. \quad (A6)$$

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

$$\delta^2 \leq d_Z^2 + \|P_A^{-1}[e(\cdot; \theta_Z)]\|_{L_0^2(X)}^2, \quad (A7)$$

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

**Appendix 3. Proof of Proposition 2**

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

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

$$\delta^2 = \min_{\theta \in \Theta} \|B\theta - 1_N\|_{L_0^2(X)}^2 = \|B\theta^* - 1_N\|_{L_0^2(X)}^2. \quad (A8)$$

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

$$\theta^* = \langle B, B \rangle_{L_0^2(X)}^{-1} \langle B, 1_N \rangle_{L_0^2(X)}. \quad (A9)$$

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

$$d_Z^2 = \min_{\theta \in \Theta} \|P_A[B\theta - 1_N]\|_{L_0^2(X)}^2 = \|P_A[B\theta_Z - 1_N]\|_{L_0^2(X)}^2. \quad (A10)$$

where the minimizer $\theta_Z$ is given by

$$\theta_Z = \langle P_A[B], P_A[B] \rangle_{L_0^2(X)}^{-1} \langle P_A[B], P_A[1_N] \rangle_{L_0^2(X)}. \quad (A11)$$
From Eqs. (A9) and (A11) we have

\[ B\theta_* = \mathcal{P}_B[1_N] \quad \text{and} \quad \mathcal{P}_A[B\theta_Z] = \mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_A[1_N], \quad (A12) \]

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

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

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

\[ \delta^2 = \|\mathcal{P}_B^\perp[1_N]\|_{\mathcal{L}_2^2(X)}. \quad (A14) \]

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

\[ \mathcal{P}_A[B\theta_Z - 1_N] = \mathcal{P}_{\mathcal{P}_A[B]} \circ \mathcal{P}_A[1_N] - \mathcal{P}_A[1_N] = -\mathcal{P}_A^\perp \circ \mathcal{P}_A[1_N]. \quad (A15) \]

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

\[ d_Z^2 = \|\mathcal{P}_A^\perp \circ \mathcal{P}_A[1_N]\|_{\mathcal{L}_2^2(X)}. \quad (A16) \]

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

**ii) Proof of Equation (2.16)**

From the Pythagorean theorem for inner product spaces we have

\[ \|B\theta - 1_N\|_{\mathcal{L}_2^2(X)}^2 = \|\mathcal{P}_A[B\theta - 1_N]\|_{\mathcal{L}_2^2(X)}^2 + \|\mathcal{P}_A^\perp[B\theta - 1_N]\|_{\mathcal{L}_2^2(X)}^2. \quad (A17) \]
for any $\theta \in \Theta$. The first term in the r.h.s. of the last equation is a quadratic function of vector $\theta$, minimized at $\theta_Z$, where it assumes value $d_Z^2$ (see Eq. (A10)), so that

$$\|P_A[B\theta - 1_N]\|_{L_2^2(\mathcal{X})}^2 = d_Z^2 + \|P_A[B(\theta - \theta_Z)]\|_{L_2^2(\mathcal{X})}^2,$$

(A18)

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

$$\|B\theta - 1_N\|_{L_2^2(\mathcal{X})}^2 = d_Z^2 + \|P_A[B(\theta - \theta_Z)]\|_{L_2^2(\mathcal{X})}^2 + \|P_A^+[B\theta - 1_N]\|_{L_2^2(\mathcal{X})}^2,$$

(A19)

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

$$\delta^2 = d_Z^2 + \|P_A[B(\theta_* - \theta_Z)]\|_{L_2^2(\mathcal{X})}^2 + \|P_A^+[B\theta_* - 1_N]\|_{L_2^2(\mathcal{X})}^2.$$

(A20)

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

$$P_A^+[B\theta_* - 1_N] = -P_A^+ \circ P_B^+[1_N].$$

(A21)

From Eq. (A12) and the fact that $P_{\mathcal{A}[B]} \circ P_A \circ P_B = 0$ and $P_{\mathcal{A}[B]} \circ P_A = P_{\mathcal{A}[B]}$ we have

$$P_A[B(\theta_* - \theta_Z)] = P_A \circ P_B[1_N] - P_{\mathcal{A}[B]} \circ P_A[1_N]$$

$$= P_A \circ P_B[1_N] - P_{\mathcal{A}[B]} \circ P_A \circ (P_B[1_N] + P_B^+[1_N])$$

$$= P_{\mathcal{A}[B]} \circ P_A \circ P_B[1_N] - P_{\mathcal{A}[B]} \circ P_A \circ P_B^+[1_N]$$

$$= -P_{\mathcal{A}[B]} \circ P_B^+[1_N].$$

(A22)

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

$$\delta^2 - d_Z^2 = \|P_A^+ \circ P_B^+[1_N]\|_{L_2^2(\mathcal{X})}^2 + \|P_{\mathcal{A}[B]} \circ P_B^+[1_N]\|_{L_2^2(\mathcal{X})}^2.$$

(A23)

Considering that $P_B^+[1_N]$ is the opposite of the conditional pricing error vector $e(\cdot; \theta_*)$ from Eq. (A13), we get Eq. (2.16).
iii) Proof of Equation (2.17)
Let us focus on the quantity $\| B\theta - 1 \|^2_{L^2_\Omega(\mathcal{X})}$ for any $\theta \in \Theta$. It is a quadratic function in $\theta$, minimized at $\theta^*_\omega$ where it assumes value $\delta^2$. Let us equate the expression that describes this property to Eq. (A17):

$$
\delta^2 + \| B (\theta^*_\omega - \theta) \|^2_{L^2_\Omega(\mathcal{X})} = \| P_A [B\theta - 1_N] \|^2_{L^2_\Omega(\mathcal{X})} + \| P_{A}^\bot [B\theta - 1_N] \|^2_{L^2_\Omega(\mathcal{X})},
$$

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

$$
\delta^2 = d^2_Z + \| P_A^\bot [B\theta^*_Z - 1_N] \|^2_{L^2_\Omega(\mathcal{X})} - \| B (\theta^*_\omega - \theta^*_Z) \|^2_{L^2_\Omega(\mathcal{X})}. \quad (A24)
$$

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

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

$$
\begin{align*}
\mathcal{P}_A^\bot \circ \mathcal{P}^\bot_B [1_N] &= 0_N, \\
\mathcal{P}_{P_A[B]} \circ \mathcal{P}^\bot_B [1_N] &= 0_N.
\end{align*}
\quad (A25)
$$

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

$$
\mathcal{P}_A^\bot [e(\cdot; \theta^*_\omega)] = 0_N. \quad (A26)
$$

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

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

52
Therefore, the first equation in System (A25) implies the second one (but not necessarily the other way around). Hence, the System (A25) is equivalent to Eq. (A26).

v) Proof of the second equivalence in Equations (2.18)

From Prop. 1, the condition \( P_A \perp [e(\cdot; \theta_Z)] = 0_N \) implies Eq. (A26). Let us now show that the reverse implication holds. If Eq. (A26) holds, we have that

\[
\langle B, P_A \circ P_B[A \circ P_B[1_N]] \rangle_{\mathbb{L}^2(X)} = \langle B, P_A[A \circ P_B[1_N]] \rangle_{\mathbb{L}^2(X)} - \langle B, P_A[e(\cdot; \theta_*)] \rangle = 0_N.
\]

By using \( \langle B, P_A \circ P_B[A \circ P_B[1_N]] \rangle_{\mathbb{L}^2(X)} = 0_N \) and the fact that \( P_A \) is a projection operator, from Eq. (A11) we get

\[
\theta_Z = \langle B, P_A[B] \rangle_{\mathbb{L}^2(X)}^{-1} \langle B, P_A[1_N] \rangle_{\mathbb{L}^2(X)} - \langle B, P_A[e(\cdot; \theta_*)] \rangle = 0_N.
\]

Then, by using Eqs. (A9) and (A27) we get \( \theta_Z = \theta_* \), and under Eq. (A26) we get \( P_A[e(\cdot; \theta_Z)] = 0_N \).

**Appendix B. Notation**

In this appendix we describe the notation used in App. C-E to derive the large sample properties of the estimators. We denote both the Euclidean norm of a vector and the Frobenius norm of a matrix by \( ||\cdot|| \). We denote by \( \varphi[i] \) the \( i \)-th element of any vector \( \varphi \). If matrix \( \Phi \) admits an inverse, we denote the element on the \( i \)-th row and \( j \)-th column of this inverse matrix by \( \Phi[i,j] \).

We define the \((N \times N)\)-dimensional matrix function \( V \) and the \((N \times p)\)-dimensional matrix function \( J \) as

\[
J(X_t; \theta) := E[\nabla_{\theta} h(Y_{t+1}; \theta) | X_t],
\]

53
\[ V(X_t; \theta) := E \left[ h(Y_{t+1}; \theta) h(Y_{t+1}; \theta)' \right| X_t, \]  

for any \( \theta \in \Theta \). We use the following notation for matrices \( J \) and \( V \) and vectors \( e \) and \( h \) valued at the unknown values \( \theta_0 \) and \( \theta_* \) of the SDF parameter vector:

\[ J_0(X_t) = J(X_t; \theta_0), \quad V_0(X_t) = V(X_t; \theta_0), \quad h_0(Y_t) = h(Y_t; \theta_0), \]
\[ J_*(X_t) = J(X_t; \theta_*), \quad V_*(X_t) = V(X_t; \theta_*), \quad h_*(Y_t) = h(Y_t; \theta_*), \quad e_*(X_t) = e(X_t; \theta_*). \]

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

\[ h_*(Y_t) := \sup_{\theta \in \Theta} \| h(Y_t; \theta) \|, \]

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

\[ \hat{f}_X(x) := \frac{1}{Tb_T^{d_X}} \sum_{j=1}^{T} K \left( \frac{x - X_j}{b_T} \right). \]

We define a scaled version of the conditional second moment matrix of the assets’ gross returns given \( X_t \):

\[ H(X_t) := \Omega(X_t)^{-1} f_X(X_t)^2, \quad h_*(Y_t) := \sup_{\theta \in \Theta} \| h(Y_t; \theta) \|. \]

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

\[ \hat{H}(X_t) := \hat{\Omega}_T(X_t)^{-1} \hat{f}_X(X_t)^2, \quad \hat{H}_T(X_t) := E \left[ \hat{\Omega}_T(X_t)^{-1} \hat{f}_X(X_t) \right| X_t] E \left[ \hat{f}_X(X_t) \right| X_t] \].
Finally, we define the $p$-dimensional stochastic vectors $u_t$ and $v_t$, the $(p \times p)$-dimensional matrices $Q_0$, $S_0$, $Q_*$ and $S_*$ and the $(2 \times 2)$-dimensional matrix $S$ in this way:

\[
    u_t := \mathbf{I}(X_t) J_\star(X_t)' \Omega(X_t) e_\star(X_t),
\]
\[
    v_{t+1} := \mathbf{I}(X_t) J_\star(X_t)' \Omega(X_t) (h_\star(Y_{t+1}) - \mathbb{E}[h_\star(Y_{t+1})|X_t])
\]
\[
- \mathbf{I}(X_t) J_\star(X_t)' \Omega(X_t) \left( R_{t+1} R_{t+1}' - \mathbb{E}[R_{t+1} R_{t+1}' | X_t] \right) \Omega(X_t) e_\star(X_t)
\]
\[
+ \mathbf{I}(X_t) \left( \nabla_\theta m(Y_{t+1}; \theta)|_{\theta = \theta_\star} - \mathbb{E}\left[ \nabla_\theta m(Y_{t+1}; \theta)|_{\theta = \theta_\star} | X_t \right] \right) R_{t+1}' \Omega(X_t) e_\star(X_t),
\]
\[
Q_0 := \mathbb{E}[\mathbf{I}(X_t) J_0(X_t)' \Omega(X_t) J_0(X_t)], \quad S_0 := \mathbb{E}\left[ \mathbf{I}(X_t) J_0(X_t)' \Omega(X_t) V_0(X_t) \Omega(X_t) J_0(X_t) \right] \quad
\]
\[
Q_* := \mathbb{E}[\mathbf{I}(X_t) J_\star(X_t)' \Omega(X_t) J_\star(X_t)] + \mathbb{E}\left[ \mathbf{I}(X_t) \nabla_\theta m(Y_{t+1}; \theta)|_{\theta = \theta_\star} R_{t+1}' \Omega(X_t) e_\star(X_t) \right] \quad
\]
\[
S_* := \sum_{l=-\infty}^{\infty} \text{Cov} [u_t, u_{t-l}] + \mathbb{E} \left[ (\eta_{t+1} - \mathbb{E}[v_{t+1} | X_t])^2 \right] ,
\]

for the random variables $\varepsilon_t$ and $\eta_t$ defined in Prop. 4.

\section*{Appendix C. Regularity assumptions}

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

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

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

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

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

\textbf{ASSUMPTION 5:} The bandwidth $b_T = o(1)$ is such that

\[
    \sqrt{Tb_T^3} = o(1), \quad \frac{\log(T)}{Tb_T^{d_X}} = o(1).
\]
ASSUMPTION 6: *The kernel* $K$ *is such that* ...

Ass. 1-6 are standard in nonparametric analysis and yield the uniform convergence of kernel estimators over the set $X_*$. By adopting Ass. 3 we avoid the boundary problems that are typical in nonparametric kernel estimation. The compactness of set $X_*$ is useful to handle with the remainder terms in the asymptotic expansions of the kernel estimators of conditional expectations.

Ass. 5 allows to simplify the expression of the asymptotic distribution of the estimators under model misspecification.

ASSUMPTION 7: *The parameter values* $\theta_0$ *and* $\theta_*$ *belong to the interior of set* $\Theta \subset \mathbb{R}^p$.

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

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

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

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

Moreover, Ass. 3 and 9 imply that

$$
\sup_{x \in X_*} \| H(x)^{-1} \| < \infty. \quad (C1)
$$

ASSUMPTION 10: *Consider positive integers* $i, j, k$ *such that* $1 \leq i \leq N$ *and* $1 \leq j, k \leq p$.

*There exists an open ball* $N_0$ *around* $\theta_0$ *such that* $\theta \rightarrow h(Y_t; \theta)$ *is twice continuously differentiable and such that*

$$
\sup_{\theta \in N_0} \left| \nabla_{\theta_i} h(Y_t; \theta) \right| \leq l_1(Y_t), \quad \text{and} \quad \sup_{\theta \in N_0} \left| \nabla_{\theta_i \theta_j} h(Y_t; \theta) \right| \leq l_2(Y_t),
$$

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

Ass. 10 is used to expand function $h(y; \theta)$ in a Taylor series w.r.t. parameter $\theta$. 
ASSUMPTION 11: As $T \to \infty$, the variables $\varepsilon_t$ and $\eta_t$ defined in Prop. 4 are such that

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{bmatrix} \varepsilon_t - \mathbb{E}[\varepsilon_t] \\ \eta_{t+1} - \mathbb{E}[\eta_{t+1}|X_t] \end{bmatrix} \xrightarrow{D} \mathcal{N}(0, S),
$$

for the asymptotic variance

$$
S := \begin{bmatrix} \sum_{t=-\infty}^{\infty} \text{Cov}[\varepsilon_t, \varepsilon_{t-l}] & 0 \\ 0 & \mathbb{E}[(\eta_{t+1} - \mathbb{E}[\eta_{t+1}|X_t])^2] \end{bmatrix}.
$$

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

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

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

for the $(2 \times 2)$-dimensional matrix

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

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

In this appendix we derive the large sample properties of the estimator $\hat{\theta}_T$ of the SDF parameter vector and the estimator $\hat{\delta}_T$ of the conditional HJ-distance for a correctly specified SDF family.
Appendix 1. Large sample properties of estimator $\hat{\theta}_T$

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

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

$$\sqrt{T} \left( \hat{\theta}_T - \theta_0 \right) \xrightarrow{D} \mathcal{N} \left( 0, Q_0^{-1} S_0 Q_0^{-1} \right).$$

**Proof.** See App. F of the supplementary material. \hfill $\Box$

Appendix 2. Useful results on kernel regression estimators

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

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

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

$$\sup_{x \in \mathcal{X}} \left\| \frac{1}{T b_T^d X} \sum_{j=1}^{T-1} K \left( \frac{x - X_j}{b_T} \right) h_0(Y_{j+1}) \right\| = O_p \left( \sqrt{\log \left[ T \right] / T b_T^d X} \right), \quad (D1)$$

$$\sup_{x \in \mathcal{X}} \left\| \hat{H}(x)^{-1} - H(x)^{-1} \right\| = O_p \left( \sqrt{\log \left[ T \right] / T b_T^d X} + b_T^d \right) + o_p \left( T^{-1/2+1/n+1/\eta} \right), \quad (D2)$$

$$\sup_{x \in \mathcal{X}} \left\| \hat{H}(x)^{-1} - \hat{H}_T(x)^{-1} \right\| = O_p \left( \sqrt{\log \left[ T \right] / T b_T^d X} \right) + o_p \left( T^{-1/2+1/n+1/\eta} \right). \quad (D3)$$

Appendix 3. Proof of Proposition 3

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

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

where the scalar statistic $q$ is defined as

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

for any $i, j = 1, \ldots, T - 1$ and any $t = 1, \ldots, T$. Let us now decompose the criterion $Q_T$ in a similar way as in Tripathi and Kitamura (2003):

$$Q_T(\theta) = \sum_{i=1}^{5} Q_{i,T}(\theta), \quad (D4)$$

where

$$Q_{1,T}(\theta) := \frac{1}{T} \sum_{t=1}^{T-1} I(X_t)w(X_t, X_t)^2q(Y_{t+1}, X_t, Y_{i+1}; \theta),$$

$$Q_{2,T}(\theta) := \frac{1}{T} \sum_{t=1}^{T-1} \sum_{i=1}^{T-1} I(X_t)w(X_t, X_i)^2q(Y_{i+1}, X_t, Y_{i+1}; \theta),$$

$$Q_{3,T}(\theta) := \frac{1}{T} \sum_{t=1}^{T-1} \sum_{j=1}^{T-1} I(X_t)w(X_t, X_j)q(Y_{i+1}, X_t, Y_{j+1}; \theta), \quad Q_{4,T}(\theta) = Q_{3,T}(\theta),$$

$$Q_{5,T}(\theta) := \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} I(X_t)w(X_t, X_i)w(X_t, X_j)q(Y_{i+1}, X_t, Y_{j+1}; \theta),$$

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

$$w(X_t, X_i) = \frac{1}{Tb_T^{dx} f_X(X_i)} K \left( \frac{X_t - X_i}{b_T} \right), \quad (D5)$$

We then derive the asymptotic distribution of the criterion $Q_T(\theta)$. 

59
i) **Stochastic boundedness of the terms** $Q_{i,T}$ **for** $i = 1, \ldots, 4$

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

$$Q_{1,T}(\theta) = \frac{1}{T} \sum_{t=1}^{T-1} I(X_t) \frac{K(0)^2}{T^{2b_{T}^{2dX}}} \hat{\Omega}_T(X_t) h(Y_{t+1}; \theta)$$

$$= \frac{K(0)^2}{T^{2b_{T}^{2dX}}} \sum_{t=1}^{T-1} I(X_t) h(Y_{t+1}; \theta) \hat{H}(X_t)^{-1} h(Y_{t+1}; \theta)$$

$$\leq \frac{K(0)^2}{T^{2b_{T}^{2dX}}} \sup_{1 \leq t \leq T} \| \hat{H}(X_t)^{-1} \| \frac{1}{T} \sum_{t=1}^{T-1} h_s(Y_{t+1})^2,$$

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

$$Q_{1,T}(\theta) = O_p \left( \frac{1}{T^{2b_{T}^{2dX}}} \right) \quad (D6)$$

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

$$Q_{2,T}(\hat{\theta}_T) = a_T. \quad (D7)$$

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

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

$$Q_{3,T}(\hat{\theta}_T) = O_p \left( \frac{1}{T^{2b_{T}^{3dX}/2}} \right). \quad (D8)$$

**Proof.** See App. G of the supplementary material. \hfill \Box

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

$$Tb_{T}^{dX/2} \left( \hat{\delta}_T^2 - a_T \right) = Tb_{T}^{dX/2} Q_{5,T}(\hat{\theta}_T) + o_p(1). \quad (D9)$$
ii) Asymptotic expansion of term $Q_{5,T}(\theta)$

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

$$Q_{5,T}(\theta) = \frac{1}{T^3b_T^2d_X} \sum_{t=1}^{T} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} I(x_t) K\left(\frac{X_i - X_j}{b_T}\right) h(Y_{i+1};\theta)',$$

$$\cdot \hat{H}(X_t)^{-1} h(Y_{j+1};\theta) K\left(\frac{X_i - X_j}{b_T}\right),$$

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

$$\tilde{Q}_{5,T}(\theta) := \frac{1}{T^3b_T^2d_X} \sum_{t=1}^{T} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} I(X_t) K\left(\frac{X_i - X_j}{b_T}\right) h(Y_{i+1};\theta)',$$

$$\cdot \hat{H}_T(X_t)^{-1} h(Y_{j+1};\theta) K\left(\frac{X_i - X_j}{b_T}\right),$$

and

$$Q_{5,T}(\theta) - \tilde{Q}_{5,T}(\theta) := \frac{1}{T^3b_T^2d_X} \sum_{t=1}^{T} \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} I(X_t) K\left(\frac{X_i - X_j}{b_T}\right) h(Y_{i+1};\theta)',$$

$$\cdot \left(\hat{H}(X_t)^{-1} - \hat{H}_T(X_t)^{-1}\right) h(Y_{j+1};\theta) K\left(\frac{X_i - X_j}{b_T}\right),$$
for any $\theta \in \Theta$. The absolute value of the last term is such that

$$\left| Q_{5,T}(\theta) - \tilde{Q}_{5,T}(\theta) \right| = \frac{1}{T^{3b_T^2d_X}} \left| \operatorname{Tr} \left[ \sum_{t=1}^{T} I(X_t) \left( \hat{H}(X_t)^{-1} - \tilde{H}_T(X_t)^{-1} \right) \cdot \sum_{i=1}^{T-1} K \left( \frac{X_t - X_i}{b_T} \right) h(Y_{i+1}; \theta) \cdot \sum_{j=1}^{T-1} K \left( \frac{X_t - X_j}{b_T} \right) h(Y_{j+1}; \theta)' \right] \right|$$

$$\leq \sup_{x \in X} \left\| \hat{H}(x)^{-1} - \tilde{H}_T(x)^{-1} \right\| \sup_{x \in X} \sup_{t=1,...,T} \left\| \frac{1}{Tb_T^{d_X}} \sum_{i=1}^{T-1} K \left( \frac{x - X_i}{b_T} \right) h(Y_{i+1}; \theta) \right\|$$

$$\cdot \sup_{x \in X} \sup_{t=1,...,T} \sup_{i \neq t} \left\| \frac{1}{Tb_T^{d_X}} \sum_{j=1}^{T-1} K \left( \frac{x - X_j}{b_T} \right) h(Y_{j+1}; \theta) \right\| \cdot \left( O_p \left( \sqrt{\log T} \right) + o_p \left( T^{-1/2 + 1/n + 1/\eta} \right) \right) O_p \left( \frac{\log T}{Tb_T^{d_X}} \right)$$

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

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

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

$$Tb_T^{d_X/2} Q_{5,T}(\hat{\theta}_T) = Tb_T^{d_X/2} \tilde{Q}_{5,T}(\hat{\theta}_T) + o_p(1). \quad (D10)$$

We control $\tilde{Q}_{5,T}(\hat{\theta}_T)$ using Ass. 10. To do so, let us first perform a Taylor expansion of $h(Y_{i+1}; \hat{\theta}_T)$ for $\hat{\theta}_T$ around $\theta_0$:

$$h(Y_{i+1}; \hat{\theta}_T) = h_0(Y_{i+1}) + \nabla_{\theta} h(Y_{i+1}; \theta)|_{\theta=\theta_0} \left( \hat{\theta}_T - \theta_0 \right) + \text{Rem} \left[ Y_{i+1}; \hat{\theta}_T - \theta_0 \right], \quad (D11)$$

where the remainder term $\text{Rem} \left[ y; \theta \right]$ is not greater than $l_2(y)(\|\theta\|)^2$ for any $y \in \mathcal{Y}$ and function $l_2$ introduced in Ass. 10.

**Lemma 3:** Under Ass. . . . we have

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

62
Proof. See App. H.

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

$$ Tb_T^{d_x/2} \left( \frac{\delta^2_T}{T} - a_T \right) = T b_T^{d_x/2} \bar{Q}_{5,T}(\theta_0) + o_P(1). $$  \hfill (D12)

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

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

$$ A_T(X_i, X_j) := \int_\mathcal{X} I(x)K \left( \frac{x - X_i}{b_T} \right) \tilde{H}_T(x)^{-1} K \left( \frac{x - X_j}{b_T} \right) f_X(x)dx, $$

$$ \hat{A}_T(X_i, X_j) := \frac{1}{T} \sum_{\substack{t=1 \atop t \neq i}}^{T} I(X_{t})K \left( \frac{X_t - X_i}{b_T} \right) \tilde{H}_T(X_t)^{-1} K \left( \frac{X_t - X_j}{b_T} \right), $$

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

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

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 $x_i, x_j \in \mathcal{X}$, and that matrix $A_T$ is such that $A_T(x, \tilde{x}) = A_T(x, \tilde{x})' = A_T(\tilde{x}, x)$, for any $x, \tilde{x} \in \mathcal{X}$. Therefore, we can write the sum in Eq. (D13) as

$$ T b_T^{d_x/2} \bar{Q}_{5,T}(\theta_0) = \frac{1}{T} \sum_{i=1}^{T-1} \sum_{j=1}^{i-1} g_T(W_i, W_j) + o_P(1), $$  \hfill (D14)

where the scalar function $g_T$ is defined as

$$ g_T(W_i, W_j) := \frac{2}{b_T^{d_x/2}} h_0(Y_{i+1})' A_T(X_i, X_j) h_0(Y_{j+1}). $$
for any $i = 1, \ldots, T - 1$ and any $j = 1, \ldots, i - 1$, and vector $W_i$ defined in App. B. We derive the asymptotic normality of $\sum_{i=1}^{T-1} \sum_{j=1}^{i-1} g_T(W_i, W_j)$ by showing that the regularity conditions for function $g_T$ required by Lemma A.3 in Su and White (2013) (see also Yoshihara (1976, 1989)) hold:

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

$$g_T(W_i, W_j) = g_T(W_j, W_i). \quad (D15)$$

**(Null expectation)** From the law of iterated expectations and since $\theta_0$ satisfies Eq. (1.2) we have

$$E[g_T(W_i, w)] = \frac{2}{b_T^{3d/2}} E[h_0(Y_{i+1})' A_T(X_i, x) h_0(y)]$$

$$= \frac{2}{b_T^{3d/2}} E \left[ E[h_0(Y_{i+1})|X_i]' A_T(X_i, x) h_0(y) \right] = 0, \quad (D16)$$

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

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

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

for the asymptotic variance $\sigma_0^2$ of statistic $\delta_T^2$ in Prop. 3. Indeed we have

$$E[g_T(W_1, \bar{W}_1)^2] = \frac{4}{b_T^{3d/2}} E \left[ (h_0(Y_2)' A_T(X_1, \bar{X}_1) h_0(Y_2))^2 \right]$$

$$= \frac{4}{b_T^{3d/2}} E \left[ h_0(Y_2)' A_T(X_1, \bar{X}_1) h_0(Y_2) h_0(Y_2)' A_T(X_1, \bar{X}_1) h_0(Y_2) \right],$$

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

$$E[g_T(W_1, \bar{W}_1)^2] = \frac{4}{b_T^{3d/2}} E \left[ \text{Tr} \left[ h_0(Y_2) h_0(Y_2)' A_T(X_1, \bar{X}_1) h_0(Y_2) h_0(Y_2)' A_T(X_1, \bar{X}_1) \right] \right].$$

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

$$
E[g_T(W_1, \bar{W}_1)^2] = \frac{4}{b_T^d} \text{Tr} \left[ E \left[ E \left[ h_0(Y_2) h_0(Y_2)' | X_1 \right] A_T(X_1, \bar{X}_1) \right] \cdot E \left[ h_0(Y_2) h_0(Y_2)' | X_1 \right] A_T(X_1, \bar{X}_1) \right].
$$

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

$$
E[g_T(W_1, \bar{W}_1)^2] = \frac{4}{b_T^d} \text{Tr} \left[ E \left[ V_0(X_1) A_T(X_1, \bar{X}_1)V_0(\bar{X}_1) A_T(X_1, \bar{X}_1) \right] \right]
\begin{align*}
&= 4 \frac{b_T^d}{b_T^d} \int_{\mathcal{X}} \int_{\mathcal{X}} \mathbf{I}(x) \mathbf{I}(\bar{x}) \text{Tr} \left[ E \left[ V_0(X_1) K \left( \frac{x - X_1}{b_T} \right) K \left( \frac{\bar{x} - \bar{X}_1}{b_T} \right) \right] \right] \\
&\quad \cdot \tilde{H}_T(x)^{-1} E \left[ V_0(\bar{X}_1) K \left( \frac{x - \bar{X}_1}{b_T} \right) K \left( \frac{\bar{x} - \bar{X}_1}{b_T} \right) \right] \tilde{H}_T(\bar{x})^{-1} f_X(x) f_X(\bar{x}) dx d\bar{x}.
\end{align*}
$$

(D18)

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

$$
E \left[ V_0(X_1) K \left( \frac{x - X_1}{b_T} \right) K \left( \frac{\bar{x} - \bar{X}_1}{b_T} \right) \right] = \int_{\mathcal{X}} V_0(x_1) K \left( \frac{x - x_1}{b_T} \right) K \left( \frac{\bar{x} - \bar{x}_1}{b_T} \right) f_X(x_1) dx_1,
$$

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

$$
E \left[ V_0(X_1) K \left( \frac{x - X_1}{b_T} \right) K \left( \frac{\bar{x} - \bar{X}_1}{b_T} \right) \right] = b_T^d \int_{\mathcal{U}(b_T, x)} V_0(x - b_T u) K(u) K \left( u - \frac{x - \bar{x}}{b_T} \right) f_X(x - b_T u) du
$$
$$
= b_T^d V_0(x) f_X(x) \int_{\mathcal{U}(b_T, x)} K(u) K \left( u - \frac{x - \bar{x}}{b_T} \right) du + o(b_T^d),
$$

for any $x, \bar{x} \in \mathcal{X}$. Note that $\mathcal{U}(0, x) = \mathbb{R}^d$ and consider the first-order Taylor expansion

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

for the convolution $\mathcal{K}$ of the kernel with itself
defined in Prop. 3. We have that
\[
\frac{1}{b_T^d} E \left[ V_0(X_1) K \left( \frac{x - X_1}{b_T} \right) K \left( \frac{x - X_1}{b_T} \right) \right] = V_0(x) f_X(x) K \left( \frac{x - \tilde{x}}{b_T} \right) + o(1),
\]
for any \(x, \tilde{x} \in \mathcal{X}\). Moreover, since \(\tilde{X}_1\) is an independent copy of \(X_1\), the same expression holds also for
\[
\frac{1}{b_T^d} E \left[ V_0(\tilde{X}_1) K \left( \frac{x - \tilde{X}_1}{b_T} \right) K \left( \frac{x - \tilde{X}_1}{b_T} \right) \right].
\]
Plugging these two expressions into Eq. (D18), and considering that
\[
\tilde{H}_T(x) = \Omega(x)^{-1} f_X(x)^2 + o(1),
\]
for any \(x \in \mathcal{X}\), we get
\[
E \left[ g_T(W_1, \tilde{W}_1)^2 \right] = 4 \int_{\mathcal{X}} \int_{\mathcal{X}} I(x) I(\tilde{x}) \text{Tr} \left[ V_0(x) \Omega(x) V_0(\tilde{x}) \Omega(\tilde{x}) \right] K \left( \frac{x - \tilde{x}}{b_T} \right)^2 d\tilde{x} dx + o(1).
\]
By changing the variable \(\tilde{x}\) with the variable \(u = (x - \tilde{x})/b_T\), we get
\[
E \left[ g_T(W_1, \tilde{W}_1)^2 \right] = 4 \int_{\mathcal{X}} \int_{U(b_T, x)} I(x) \text{Tr} \left[ V_0(x) \Omega(x) V_0(x) \Omega(x) \right] \cdot K(u)^2 du dx + o(1)
\]
\[
= 4 \int_{\mathcal{X}} I(x) \text{Tr} \left[ V_0(x) \Omega(x) V_0(x) \Omega(x) \right] \int_{U(b_T, x)} K(u)^2 du dx + o(1).
\]
We have the following first-order Taylor expansion:
\[
\int_{U(b_T, x)} K(u)^2 du = \int_{\mathbb{R}^d} K(u)^2 du + o(1).
\]
Thus, from the definition of \(\sigma_0^2\) in Prop. 3, Eq. (D17) follows.
(Null cross-terms) We have

\[
E [ g_T(W_{j+1}, w) g_T(W_1, \tilde{w}) ] = E [ E [ g_T(W_{j+1}, w) | X_{j+1} ] g_T(W_1, \tilde{w}) ] \\
= \frac{4}{b_T^3 \sigma_X} E [ E [ h_0(Y_{j+2}) | X_{j+1} ] A(X_{j+1}, x) h_0(y) g_T(W_1, \tilde{w}) ] = 0,
\]

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

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

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

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

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

and

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

where function \( G_T \) is defined as

\[
G_T(w, z) := E [ g_T(W_1, w) g_T(W_1, z) ]
\]

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

**Proof.** See App. I. \( \square \)

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

\[
\frac{1}{T} \sum_{i=1}^{T-1} \sum_{j=1 \atop j \neq i}^{T-1} g_T(W_i, W_j)
\]

(D19)
is asymptotically normal with null mean and asymptotic variance $\sigma_0^2$. Then, from Eq. (D14) we have

$$Tb_T^{d_5/2} \tilde{Q}_T(\theta_0) \xrightarrow{D} \mathcal{N}(0, \sigma_0^2).$$

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

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

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

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

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

$$\sqrt{T} \left( \hat{\theta}_T - \theta_\star \right) \xrightarrow{D} \mathcal{N}(0, Q_\star^{-1} S_\star Q_\star^{-1}).$$

**Proof.** See App. J. \qed

When the SDF family is correctly specified, we have $\theta_\star = \theta_0$ and $e(X_t; \theta_\star) = 0$, and the asymptotic distribution in Lemma 5 reduces to that in Lemma 1. Lemma 5 is the counterpart of the results on the large sample distribution of the GMM estimator in misspecified models derived by Hall and Inoue (2003) (see also Anatolyev and Gospodinov (2011)), when the structural parameter is identified by a conditional moment restrictions.
Appendix 2. Proof of Proposition 4

From Eq. (2.21) we have

\[
\hat{\delta}_T^2 = \frac{1}{T} \sum_{t=1}^{T} I(X_t) \hat{\epsilon}_T(X_t; \hat{\theta}_T) \hat{\Omega}_T(X_t) \hat{\epsilon}_T(X_t; \hat{\theta}_T) .
\]  

(E1)

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

\[
\hat{\epsilon}_T(X_t; \hat{\theta}_T) = \hat{\epsilon}_T(X_t; \theta_*) + \nabla_{\theta} \hat{\epsilon}_T(X_t; \theta)|_{\theta=\theta_*} (\hat{\theta}_T - \theta_*) + \text{Rem} \left[ X_t; \hat{\theta}_T - \theta_* \right]
\]

\[
\simeq e_*(X_t) + (\hat{\epsilon}_T(X_t; \theta_*) - e_*(X_t)) + J_*(X_t) (\hat{\theta}_T - \theta_*),
\]  

(E2)

since matrix \( \nabla_{\theta} \hat{\epsilon}_T(X_t; \theta)|_{\theta=\theta_*} \) converges in probability to matrix \( J_*(X_t) \). We can write

\[
\hat{\Omega}_T(X_t) = \left( \Omega(X_t)^{-1} + \left( \hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1} \right) \right)^{-1}
\]

\[
= \left( I_N + \Omega(X_t) \left( \hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1} \right) \right)^{-1} \Omega(X_t) .
\]

Each of the eigenvalues of matrix \( \Omega(X_t) \left( \hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1} \right) \) is less than 1 in absolute value

and from its first-order approximation we get

\[
\hat{\Omega}_T(X_t) \simeq \Omega(X_t) - \Omega(X_t) \left( \hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1} \right) \Omega(X_t) .
\]  

(E3)

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

\[
\begin{align*}
\hat{\delta}_T^2 & \simeq \frac{1}{T} \sum_{t=1}^{T} I(X_t) e_*(X_t)' \Omega(X_t) e_*(X_t) + \frac{2}{T} \sum_{t=1}^{T} I(X_t) e_*(X_t)' \Omega(X_t) ( \hat{\epsilon}_T(\theta_*) - e_*(X_t) ) \\
& \quad + \frac{2}{\sqrt{T}} \left( \frac{1}{T} \sum_{t=1}^{T} I(X_t) e_*(X_t)' \Omega(X_t) J_*(X_t) \right) \sqrt{T} (\hat{\theta}_T - \theta_*) \\
& \quad - \frac{1}{T} \sum_{t=1}^{T} I(X_t) e_*(X_t)' \Omega(X_t) \left( \hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1} \right) \Omega(X_t) e_*(X_t) .
\end{align*}
\]  

(E4)
From the Law of Large Numbers we have that
\[
\frac{1}{T} \sum_{t=1}^{T} I(X_t)e_*(X_t)'\Omega(X_t)J_*(X_t) = E \left[ I(X_t)e_*(X_t)'\Omega(X_t)J_*(X_t) \right] + o_p(1),
\]
and since \( \theta_* \) satisfies the first-order condition associated to the minimization problem in Eq. (3.2) we have that
\[
E \left[ I(X_t)e_*(X_t)'\Omega(X_t)J_*(X_t) \right] = 0.
\]
Thus, from Lemma 5, the third term in the r.h.s. of Approximation (E4) is \( o_p(1/\sqrt{T}) \). By subtracting \( \delta^2_\star \) on both sides of Approximation (E4) and scaling both sides by \( \sqrt{T} \) we get
\[
\sqrt{T} \left( \hat{\delta}^2_T - \delta^2_\star \right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\varepsilon_t - \delta^2_\star) + \frac{2}{\sqrt{T}} \sum_{t=1}^{T} I(X_t)e_*(X_t)'\Omega(X_t) (\hat{e}_T(\theta_\star) - e_*(X_t))
\]
\[
- \frac{1}{\sqrt{T}} \sum_{t=1}^{T} I(X_t)e_*(X_t)'\Omega(X_t) \left( \hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1} \right) \Omega(X_t)e_*(X_t) + o_p(1), \tag{E5}
\]
where we use the variable \( \varepsilon_t \) defined in Prop. 4 and the fact that \( \delta^2_\star = E[\varepsilon_t] \). Let us now define the scalar function
\[
a(Y_i, X_t) := 2e_*(X_t)'\Omega(X_t)h_*(Y_i) - (e_*(X_t)'\Omega(X_t)R_i)^2,
\]
such that
\[
\sum_{i=1}^{T-1} w(X_t, X_i)a(Y_{i+1}, X_t) - E \left[ a(Y_{t+1}, X_t) | X_t \right]
\]
\[
= 2e_*(X_t)'\Omega(X_t) (\hat{e}_T(X_t; \theta_\star) - e_*(X_t))
\]
\[
- e_*(X_t)'\Omega(X_t) \left( \hat{\Omega}_T(X_t)^{-1} - \Omega(X_t)^{-1} \right) \Omega(X_t)e_*(X_t).
\]
We can write Eq (E5) using function \( a \) as
\[
\sqrt{T} \left( \hat{\delta}^2_T - \delta^2_\star \right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} I(X_t) \left( \sum_{i=1}^{T-1} w(X_t, X_i)a(Y_{i+1}, X_t) - E \left[ a(Y_{t+1}, X_t) | X_t \right] \right)
\]
\[
- \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\varepsilon_t - E[\varepsilon_t]) + o_p(1). \tag{E6}
\]
We control the first term in the r.h.s. of Eq. (E6) by using the next lemma.

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

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

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

From Eq. (E6) and Lemma 6 we get

$$\sqrt{T} \left( \hat{\sigma}^2 - \delta^2 \right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\varepsilon_t - E[\varepsilon_t]) + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\eta_{t+1} - E[\eta_{t+1}|X_t]) + o_p(1).$$

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

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

is asymptotically normal, with null mean and asymptotic variance

$$\sum_{l=-\infty}^{\infty} \text{Cov} [\varepsilon_t, \varepsilon_{t-l}] + E \left( (\eta_{t+1} - E[\eta_{t+1}|X_t])^2 \right).$$