Article:

Bravo, Francesco orcid.org/0000-0002-8034-334X (2022) Misspecified semiparametric model selection with weakly dependent observations. Journal of Time Series Analysis. pp. 558-586. ISSN 1467-9892

https://doi.org/10.1111/jtsa.12628

Misspecified semiparametric model selection with weakly dependent observations

Francesco Bravo*
University of York

August 2021

Abstract

This paper proposes a general methodology to estimate and discriminate (select) between two possibly misspecified semiparametric models with weakly dependent observations. The proposed estimator and test statistics are based on exponential tilting and can be considered as useful semiparametric extensions of Vuong's (1989) and Shi's (2015) approaches to model selection. Monte Carlo evidence and an empirical application to Fama & French's (1993) three factor model suggest that the proposed methodology has competitive finite sample properties and is useful in practice.

Keywords: Mixing, Stochastic discount factor, Stochastic equicontinuity, Uniform size control.

*I am grateful to two referees for useful comments and suggestions that improved considerably the original paper. The usual disclaimer applies.
Address correspondence to: Department of Economics, University of York, York YO10 5DD, UK. E-mail: francesco.bravo@york.ac.uk. Web Page: https://sites.google.com/a/york.ac.uk/francescobravo/

1 Introduction

In this paper we propose to extend the model selection approach of Vuong (1989) and Kitamura (2000) to misspecified semiparametric moment conditions (estimating equations) models. Semiparametric moment conditions models are very useful and flexible extensions to the generalized instrumental variables models often used in the economic and financial literature - see for example Hansen & Singleton (1982) - and to the generalized estimating equations and quadratic inference functions models that are very popular in the statistical literature - see for example Liang & Zeger (1986) and Qu, Lindsay & Li (2000). Misspecification often arises in economics and finance: for example, many asset pricing models are likely to be misspecified, see for example Hansen & Jagannathan (1997) and more recently Gospodinov, Kan & Robotti (2013). Some examples of misspecified moment conditions models in the context of generalized method of moments (GMM) estimation are given by Hall & Inoue (2003). Misspecification is also of interest in time series modelling, see for example Kunitomo & Yamamoto (1985), Dahlhaus & Wefelmeyer (1996) and McElroy (2016), and can affect semiparametric moment conditions models, as shown for example by Ai & Chen (2007) and Chen, Liao & Sun (2014). Misspecification is theoretically interesting, see for example Dahlhaus & Wefelmeyer (1996) and Hall & Inoue (2003), and empirically relevant because, as shown for example by Gospodinov, Kan & Robotti (2014), using standard statistical methods in misspecified models can result in very misleading inferences. It seems therefore useful to develop model selection procedures for misspecified semiparametric moment conditions models.

The model selection procedure we consider is based on a two-step semiparametric estimator that has an information theoretic interpretation, which is important because it provides a natural extension to semiparametric moment conditions models of the classical estimation theory of misspecified parametric likelihood models developed by Akaike (1973) and White (1982). To be specific, the estimator we consider is a two-step semiparametric extension of the exponential tilting (ET) estimator suggested by Kitamura (2000) (see also Kitamura & Stutzer (1997)). The estimator is in the same spirit as that considered by Chen & Liao (2015), in the sense that we assume that there is a preliminary consistent (in a suitable norm) estimator of the infinite dimensional parameter. This preliminary estimator can be obtained using an available auxiliary model, or could be the result of profile estimation (such as in the partial linear model discussed in Section 5 below) or the result of the first-step of an iterative estimation procedure, often called backfitting, where in the first-step all the unknown parameters are estimated nonparametrically - see Carroll & van Keilegom (2007) for a discussion on the differences between profiling and backfitting estimation in semiparametric models. As in Kitamura (2000), the resulting two-step estimator is robust against misspecification (hence it can be used for inference) and has an information theoretic interpretation in terms of minimizing the Kullback-Leibler (KL) divergence between the distribution of a possibly misspecified semiparametric model and the true (unknown) one. This interpretation allows us to extend Vuong's (1989) model selection theory, which was developed for i.i.d. misspecified parametric likelihood models, to misspecified semiparametric models.

Apart from Kitamura (2000), other extensions of Vuong's (1989) model selection theory to misspecified models based on objective functions other than the likelihood function have been proposed in the literature: Christoffersen, Hahn & Inoue (2001) considered a quantile type objective function that is used to compute value at risk (VaR) measures that are widely used in the so-called Riskmetrics (Morgan 1996), while Rivers & Vuong (2002) considered an objective function that can be used for both M and GMM estimation (for the latter, see also Hall & Pelletier (2011)[1]). Chen, Hong & Shum (2007) considered an objective function that can be used to discriminate between a likelihood and a moment condition model[2], while Li (2009) considered a mean squared prediction objective function suitable for simulations type estimators.

The key feature of Vuong's (1989) model selection theory is that it depends on whether the two competing models are *non nested*, *overlapping* or *nested* (see Section 4.1 for a definition), because depending on which case it is, the asymptotic distribution of the selection test statistic is different (respectively a standard normal, a mixture of chi-squared and a standard chi-squared). Because of this different asymptotic behavior, Vuong's (1989) model selection theory is typically based on pretesting to select which distribution to use in the computation of the critical values, and the resulting two-step testing procedure might result in size distortions and/or power loss, especially when the models are non nested but "close" to each others. In a seminal paper, Shi (2015) addressed this problem by proposing a modified Vuong statistic for (parametric) likelihood and moment conditions models that uniformly controls the size; other important modifications of the Vuong statistic include Hsu & Shi (2017), which considered a certain randomization procedure in the context of conditional moment inequalities models, Schennach & Wilhelm (2017), which considered sample splitting and Liao & Shi (2020), which extended Shi's (2015) modified Vuong statistic to semi/nonparametric models.

With the exception of Rivers & Vuong (2002) and Hall & Pelletier (2011), all the results of the above papers are based on the assumption that the data are independent and identically distributed; one important feature of the model selection procedure of this paper is that it allows for weakly dependent observations, which is particularly useful in macroeconomics and finance,

[1] It is important to note that Hall & Pelletier (2011) identified two potential problems with the GMM based model selection procedure, namely that test statistics based on the GMM objective function might provide different conclusions for different choices of the weighting matrix, and that the model comparison itself might not be at all meaningful if different weighting matrices were used.

[2] Interestingly, the model selection procedure proposed by Chen et al. (2007) relies on the empirical likelihood estimator, which as shown by Schennach (2007) might not be robut to misspecification (see also Bravo (2020)), whereas using the ET estimator of this paper would not have this problem.

since macroeconomic and financial data typically exhibit some form of serial dependence. It is important to note that, as opposed to Kitamura & Stutzer (1997), who proposed to use kernel smoothed moment conditions to estimate correctly specified semiparametric moment conditions models to obtain asymptotically efficient estimators, smoothing is not required for the computation of the proposed semiparametric estimator. With misspecification, the notion of efficiency in moment conditions models loses its statistical significance, as the resulting estimators depend crucially on the chosen weight matrix - see for example Hall & Inoue (2003) for this important point. On the other hand, smoothing is still required to obtain consistent estimators of certain long run variances that can be used for inference and/or in the proposed model selection procedures.

In this paper we make the following contributions: first, from a methodological point of view, we propose a general estimation and model selection theory that can be applied to possibly misspecified semiparametric models with weakly dependent observations. To obtain these general results, we assume a number of high level regularity conditions and specify the dependent structure of the observations in terms of either strictly stationary and mixing or nonstationarity and α-mixing. The key assumption is that of stochastic equicontinuity (or uniform asymptotic equicontinuity) with respect to the infinite dimensional parameter. Although very high level, this assumption can be typically verified as long as the parameter space of the infinite dimensional parameter is not "too large" and specific mixing conditions are assumed, see the discussion in Section 3 about possible mixing assumptions and classes of functions whose size is not "too large". Because of the connection between ET and KL divergence, the model selection procedures we propose are in the same spirit as those proposed by Vuong (1989) for parametric likelihood models and, as opposed to those proposed by Hall & Pelletier (2011), do not depend on the choice of the weighting matrix. To be specific, we first consider what can be defined as a naive (but not trivial) extension of Vuong's (1989) two-step model selection procedure, which is a useful generalization of those proposed by Kitamura (2000), Rivers & Vuong (2002) and Hall & Pelletier (2011) for parametric models. We then consider an extension to Shi's (2015) procedure, which, as mentioned before, does not require pretesting and achieves uniform size control. The extension is based on the same local asymptotic theory used by Shi (2015), which is widely used in the study of local power, near unit root and other nonstandard asymptotic problems. As in Shi's (2015), the proposed local asymptotic theory results in a Vuong statistic with a nonstandard asymptotic distribution, which, however, is allowed to smoothly transition to a standard normal (see Section 4.2 for more details). While uniform size control is a very desirable statistical property, there are two main reasons as to why the naive semiparametric extension to Vuong's two-step procedure presented in Section 4.1 below might still be very useful in practice. First, the proposed semiparametric extension of Shi's (2015) procedure requires simulation to obtain the critical values. Second, as pointed out by Shi (2015), the (second order) bias characterizing the Vuong statistic tends to favor more complex models, as a result of which it might show poor finite sample properties when the two competing models are very different in terms of their dimension (parameters to be estimated). On the other hand, when the dimension of the competing models is comparable, the effect of the bias might negligible.

Second, we establish the asymptotic normality of a general two-step semiparametric ET estimator under an asymptotically orthogonality condition (see Assumption A3(ii) below), which is satisfied by many commonly used semiparametric models such as partial linear, (non)linear regression with unknown heteroskedasticity and single index. This result extends and/or complements results obtained by Bravo (2020), who considered a general class of estimators for misspecified semiparametric moment conditions models with identically and independently distributed observations, and by Bravo, Chu & Jacho-Chavez (2017), who considered generalized empirical likelihood estimators for correctly specified semiparametric moment conditions models with weakly dependent observations. As a final methodological contribution, we illustrate the general applicability of the proposed estimator and model selection procedures with an example in which we provide a set of more primitive assumptions that can be used to verify the high level assumptions used to prove the previous results. We also report Monte Carlo evidence about the finite sample properties of the proposed model selection statistics; the results are encouraging as they suggest that both the naive and the uniform extension to Vuong's (1989) model selection approach are useful for semiparametric models.

Finally, from an applied point of view, we apply the proposed model selection procedure to the widely used Fama & French's (1993) three factor version of the stochastic discount factor (SDF) model, which is the basis for many assets pricing theories. The use of ET and KL divergence in the context of SDF models is not new: Stutzer (1995), Stutzer (1996) and Kitamura & Stutzer (2002) used KL to construct bounds for the SDF as an alternative to the so-called HJ distance (Hansen & Jagannathan 1997) that provides minimum variance bounds for the SDF. More recently Ghosh, Julliard & Taylor (2017) used ET explicitly to obtain bounds for the SDF that are tighter than those given by the HJ distance. The semiparametric ET estimator of this paper could also be used to construct similar bounds, but we do not pursue this here. Rather, we focus on model selection and consider two different specifications of the three factor model, one based on a fully parametric specification and the other one based on a novel semiparametric one. As a standard ET based misspecification test indicate that indeed both models are misspecified, we use the proposed tests to check whether either of the two models should be preferred (i.e. discrimination is possible) and find that the semiparametric specification should be chosen.

The rest of the paper is structured as follows: next section introduces the two-step semiparametric ET estimator of this paper and illustrates its connection between ET and the KL divergence. Section 3 develops the asymptotic theory for the estimators, whereas Section 4 presents the tests for model comparison. Sections 5 and 6, respectively, illustrate the results of Sections 3 and 4 with an example and a simulation study, and present the empirical application. Section 7 contains some concluding remarks. All the proofs are contained in an online supplemental Appendix.

The following notation is used throughout the paper: "$'$" indicates transpose, "$\otimes$" denotes the Kronecker product, $\|\cdot\|$" and "$\|\cdot\|_{\mathcal{F}}$" denote, respectively, the standard Euclidean (Frobenius) norm for random vectors (matrices) and a functional norm such as the sup norm for a pseudo-metric space of functions $\mathcal{F}$, "tr" and "vec" are the trace and vec operators; finally, for any vector v, $v^{\otimes 2} = vv'$.

2 Misspecified models and exponential tilting

Let $\{Z_t, t = 0, \pm 1, \pm 2, ...\}$ denote a sequence of strictly stationary random vectors taking values in $\mathcal{Z} \subset \mathbb{R}^{dz}$ with unknown marginal distribution P_0. The semiparametric moment conditions model we consider is defined through a set of moment functions (estimating equations) $g : \mathcal{Z} \times \Theta \times \mathcal{H} \to \mathbb{R}^l$, where $\Theta \subset \mathbb{R}^k$ and $\mathcal{H} = \mathcal{H}_1 \times .... \times \mathcal{H}_m$ is a pseudo-metric space of functions, such that

$$P_{\theta,h} = \{P | E_P (g (Z_t, \theta, h)) = 0\}, \tag{2.1}$$

where $\theta \in \Theta$ and $h \in \mathcal{H}$[3]. Let $\mathcal{P}_g = \cup_{\theta \in \Theta, h \in \mathcal{H}} P_{\theta,h}$, that is, the semiparametric (moment conditions) model P_g consists of a set of distributions indexed by the unknown parameters θ and h that are compatible with the moment restriction (2.1).

Definition 1 *A semiparametric model is said to be correctly specified if* $P_0 \in \mathcal{P}_g$, *for some* $\theta_0 \in \Theta$ *and* $h_0 \in \mathcal{H}$.

Definition 2 *A semiparametric model is said to be misspecified if* $P_0 \notin \mathcal{P}_g$, *for all* $\theta \in \Theta$ *and* $h \in \mathcal{H}$.

The same arguments used by Kitamura & Stutzer (1997) (for *parametric* moment conditions models) can be used here to show that ET estimation asymptotically identifies the element in the set $\mathcal{P}_g$ that is closest in the sense of the KL divergence to P_0, that is the exponential tilting estimator solves the problem

$$\inf_{P \in \mathcal{P}_g} D_{KL} (P, P_0), \tag{2.2}$$

where $D_{KL}(\cdot)$ represents the KL divergence. As in Kitamura & Stutzer (1997), the solution to (2.2) can be characterized as a saddlepoint problem, in which in the first step one solves

[3]Note that, as in Andrews (1994), h is allowed to depend on θ, in which case the assumptions in the next section involving h should be interpreted as uniform in $\theta \in \Theta$.

the unconstrained convex problem $\min_P D_{KL}(P, P_0)$ for a fixed θ and h. Let $P(\theta, h)$ denote the solution of (2.2), which can be expressed as $\min_\lambda \mathcal{M}(\theta, h, \lambda)$, where $\mathcal{M}(\theta, h, \lambda) = E_P(\exp(\lambda' g(Z_t, \theta, h)))$ is the exponential tilting of $g(Z_t, \theta, h)$ and $\lambda = \lambda(\theta, h)$. Then for

$$\mathcal{M}(\theta, h, \lambda(\theta, h)) = \exp(-D_{KL}(P(\theta, h), P_0)) \tag{2.3}$$

θ_* and h_* are the maximizers of (2.3), which, following White's (1982) terminology, we call pseudo true values. Similarly, $\lambda_* = \lambda(\theta_*, h_*)$, the solution to $\min_\lambda \mathcal{M}(\theta, h, \lambda)$, is the pseudo true value for λ.

3 Two-step semiparametric exponential tilting estimation

In this section we investigate the asymptotic properties of a two-step version of the ET estimator when $P_0 \notin \mathcal{P}_g$, that is when the semiparametric model (2.1) is misspecified. In what follows, we drop the dependence on P of the expectation operator E.

To introduce the estimator, we assume that there exists an estimator $\widehat{h}$ (consistent in a suitable norm) for h_*. Then, the two-step semiparametric ET estimator $\widehat{\theta}$ is defined as

$$\left[\widehat{\theta}', \widehat{\lambda}'\right]' = \arg\max_{\theta \in \Theta} \arg\min_{\lambda \in \Lambda(\Theta)} \mathcal{M}_T\left(\theta, \widehat{h}, \lambda\right), \tag{3.1}$$

where $\Lambda(\Theta) \subset \mathbb{R}^l$, $\mathcal{M}_T\left(\theta, \widehat{h}, \lambda\right) = \sum_{t=1}^T \exp\left(\lambda' g_t\left(\theta, \widehat{h}\right)\right)/T$, $\lambda = \lambda(\theta)$ and $g_t(\theta, h) = g(Z_t, \theta, h)$.

Note that the solution of (3.1) corresponds to

$$\mathcal{M}_T\left(\widehat{\theta}, \widehat{h}, \widehat{\lambda}\right) = \exp\left(-\min_{P \in \mathcal{P}_{g_{\widehat{h}}}} D_{KL}(P, P_T)\right),$$

that is the two-step semiparametric ET estimator is the element in the set $\mathcal{P}_{g_{\widehat{h}}} = \cup_{\theta \in \Theta} P_{\theta, \widehat{h}}$ that is the closest in the KL divergence sense to P_T, the empirical distribution of the observations.

To obtain results that are rather general, we make the following assumptions, some of which are rather high level. Let $\phi = [\theta', \lambda']'$, and assume that:

A1 (i) the sequence of random vectors $\{Z_t, t = 0, \pm 1, \pm 2, ...\}$ is strictly stationary and mixing, (ii) $D_{KL}(P(\theta, h_*), P_0)$ is uniquely minimized at θ_*, (iii) the parameter spaces Θ and $\Lambda(\Theta)$ are compact subsets of $\mathbb{R}^k$ and $\mathbb{R}^l$, respectively, (iv) $g_t(\theta, h)$ is twice continuously differentiable with respect to θ *a.s.* in a neighborhood Θ_* of θ_*, (v) $\theta_* \in int(\Theta)$, $\lambda_* \in int(\Lambda(\Theta))$,

A2 (i) $\left\|\widehat{h} - h_\bullet\right\|_{\mathcal{H}} = o_p(1)$ and $\left\|\widehat{\phi} - \phi_*\right\| = o_p(1)$, (ii)

$$\sup_{\lambda \in \Lambda_*(\Theta_*), \theta \in \Theta_*, h \in \mathcal{H}_\epsilon} \left\| \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 \exp(\lambda' g_t(\theta, h))}{(\partial \phi)^{\otimes 2}} - H_g(\phi, h) \right\| = o_p(1),$$

where $\Lambda_*(\Theta_*)$ and Θ_* are neighborhoods of λ_* and θ_* and $\mathcal{H}_\epsilon = \{h | \|h - h_\bullet\|_{\mathcal{H}} \leq \epsilon\}$ for some $\epsilon > 0$, and $H_g(\phi, h) = E\left[\partial^2 \exp(\lambda' g_t(\theta, h)) / (\partial\phi)^{\otimes 2}\right]$, (iii) $H_g(\phi_*, h_*)$ is nonsingular,

A3 (i) the empirical process $v_T(h)$ satisfies the stochastic equicontinuity condition

$$\sup_{h \in \mathcal{H}_\epsilon} \|v_T(h) - v_T(h_*)\| = o_p\left(T^{-1/2}\right),$$

where $v_T(h) = \sum_{t=1}^{T} \left[\partial \exp(\lambda_*' g_t(\theta_*, h)) / \partial\phi - E\left[\partial \exp(\lambda_*' g_t(\theta_*, h)) / \partial\phi\right]\right] / T$,
(ii) $E\left[\partial \exp\left(\lambda_*' g_t\left(\theta_*, \widehat{h}\right)\right) / \partial\phi\right] = o_p\left(T^{-1/2}\right)$,

A4 $T^{1/2} v_T(h_*) \xrightarrow{d} N(0, \Omega_g(\phi_*, h_*))$, where

$$\Omega_g(\phi_*, h_*) = \lim_{T \to \infty} Var\left[\frac{1}{T^{1/2}} \sum_{t=1}^{T} \frac{\partial \exp(\lambda_*' g_t(\theta_*, h_*))}{\partial\phi}\right],$$

assumed to be positive definite.

Assumption A1(i) excludes deterministic and stochastic trends, and specifies the dependence structure of the sequence $\{Z_t\}$ in terms of mixing; for semiparametric models the mixing condition is typically specified as β-mixing (Volkonskii & Rozanov 1959). Many time series models satisfy the β-mixing condition, including nonlinear autoregressive models with possibly additional variables, nonlinear ARCH models and some diffusion models, see for example Carrasco & Chen (2002). Alternatively, the weaker notion of α-mixing can be assumed - see Bradley (2005) for a comparison between the notions of β and α-mixing; see also Doukhan (1994) for a review and examples of α-mixing processes. Assumptions A1(ii)-(iv) are standard in the literature on misspecified nonlinear models, see for example Kitamura (2000) and Hall & Inoue (2003). Assumption A2(i) assumes the consistency of the estimators of the unknown parameters. The consistency of $\widehat{h}$ in A2(i) holds for example in the case of the sup norm for kernel and nonparametric series estimators under standard regularity conditions on h, see for example Masry (1996) and Newey (1997). Given the consistency of $\widehat{h}$, the consistency of $\widehat{\phi} = \left[\widehat{\theta}', \widehat{\lambda}'\right]'$ follows, for example, by the identification condition A1(ii) combined with a suitable uniform law of large numbers and mild smoothness conditions with respect to ϕ, see Proposition 4 in the online Appendix for an example of such conditions. Assumption A2(ii) requires a uniform law of large numbers for the Hessian matrix of the objective function; with semiparametric models, uniform laws of large numbers are typically specified in terms of bracketing numbers and/or entropy conditions (see Van der Vaart & Wellner (1996) for a definition) to control for the complexity of the pseudo metric space $\mathcal{H}_\epsilon$, an envelope condition on $H_g(\theta, \lambda, h)$ and a specific rate for the mixing parameter, see for example Yu (1994) for β-mixing processes and Bravo et al. (2017) for α-mixing processes. Given the compactness assumption A1(iii), the bracketing numbers and entropy conditions are typically satisfied when $\mathcal{H}$ (hence $\mathcal{H}_\epsilon$) is a space of sufficiently smooth functions, such as a Holder or a Sobolev space, or a space of bounded variations functions with compact support (see for example Van der Vaart (1998) and the proof of Proposition 2). Assumption A3(i) assumes the stochastic equicontinuity of the empirical process $v_T(h)$, a key property to obtain the asymptotic distributions of the estimator and test statistics of this paper under Assumption A3(ii), which can be interpreted as an asymptotic orthogonality condition implying that estimation of h_* does not affect the covariance matrix of the asymptotic distribution of the estimator of θ_*. A sufficient condition for A3(ii) to hold is that $E\left[\partial\left(\partial \exp\left(\lambda_*' g_t\left(\theta_*, h_*\right)\right)/\partial\phi\right)/\partial h_j\right] = 0$ $(j = 1, ..., m)$. As in Assumption A2(ii), stochastic equicontinuity typically involves a bracketing number or bracketing entropy condition (see for example Van der Vaart & Wellner (1996) for a definition) to control the complexity of the infinite dimensional parameter space $\mathcal{H}$, see for example Arcones & Yu (1994) and Doukhan, Massart & Rio (1995) for β-mixing processes, and Andrews & Pollard (1994), Bravo et al. (2017) and Mohr (2020) for α-mixing processes among others. Finally A4 requires a central limit theorem, which typically holds under a suitable summability condition on the α or β mixing coefficients, see for example assumptions E1(i) and/or E5(i) in Section 5 for the α mixing case.

Theorem 1 *Under Assumptions A1-A4*

$$T^{1/2}\left[\left(\widehat{\theta}-\theta_*\right)', \left(\widehat{\lambda}-\lambda_*\right)'\right]' \xrightarrow{d} N\left(0, H_g\left(\phi_*, h_*\right)^{-1}\Omega_g\left(\phi_*, h_*\right)H_g\left(\phi_*, h_*\right)^{-1}\right), \tag{3.2}$$

where

$$\begin{aligned}
H_g(\phi, h) &= \begin{bmatrix} H_{g\theta\theta}(\theta,\lambda,h) & H_{g\theta\lambda}(\theta,\lambda,h) \\ H_{g\theta\lambda}(\theta,\lambda,h)' & H_{g\lambda\lambda}(\theta,\lambda,h)\end{bmatrix},\\
H_{g\theta\theta}(\phi,h) &= E\Bigg[\exp\left(\lambda' g_t(\theta,h)\right)\Bigg(\lambda'\otimes I_k \frac{\partial vec\left(\partial g_t(\theta,h)/\partial\theta'\right)}{\partial\theta'} + \left(\left(\frac{\partial g_t(\theta,h)}{\partial\theta'}\right)'\lambda\right)^{\otimes 2} + \\
&\quad \left(\frac{\partial g_t(\theta,h)}{\partial\theta'}\right)'\frac{\partial\lambda}{\partial\theta'}\Bigg],\\
H_{g\theta\lambda}(\phi,h) &= E\left[\exp\left(\lambda' g_t(\theta,h)\right)\left(\frac{\partial g_t(\theta,h)}{\partial\theta'}\left(I_l+\lambda g_t(\theta,h)'\right) + \left(\frac{\partial\lambda}{\partial\theta'}\right)' g_t(\theta,h)^{\otimes 2}\right)\right],\\
H_{g\lambda\theta}(\theta,\lambda,h) &= H_{g\theta\lambda}(\theta,\lambda,h)', \quad H_{g\lambda\lambda}(\phi,h) = E\left[\exp\left(\lambda' g_t(\theta,h)\right) g_t(\theta,h)^{\otimes 2}\right],
\end{aligned}$$

with

$$\begin{aligned}
\frac{\partial\lambda}{\partial\theta'} &= \frac{\partial\lambda(\theta)}{\partial\theta'} = -\frac{1}{2}E\left[\exp\left(\lambda' g_t(\theta,h)\right) g_t(\theta,h)^{\otimes 2}\right]^{-1}\times\\
&\quad E\left[\exp\left(\lambda' g_t(\theta,h)\right)\left(I_l+\lambda g_t(\theta,h)'\right)\frac{\partial g_t(\theta,h)}{\partial\theta'}\right],
\end{aligned}$$

and

$$\begin{aligned}
\Omega_g(\phi,h) &= \begin{bmatrix} \Omega_{g\theta\theta}(\phi,h) & \Omega_{g\theta\lambda}(\phi,h) \\ \Omega_{g\lambda\theta}(\phi,h) & \Omega_{g\lambda\lambda}(\phi,h) \end{bmatrix}, \\
\Omega_{g\theta\theta}(\phi,h) &= \lim_{T\to\infty} Var\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{\partial \exp(\lambda' g_t(\theta,h))}{\partial\theta}\right), \\
\Omega_{g\theta\lambda}(\phi,h) &= \lim_{T\to\infty} Cov\left[\frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{\partial \exp(\lambda' g_t(\theta,h))}{\partial\theta}, \frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{\partial \exp(\lambda' g_t(\theta,h))}{\partial\lambda}\right], \\
\Omega_{g\lambda\theta}(\phi,h) &= \Omega_{g\theta\lambda}(\theta,\lambda,h)', \\
\Omega_{g\lambda\lambda}(\phi,h) &= \lim_{T\to\infty} Var\left[\frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{\partial \exp(\lambda' g_t(\theta,h))}{\partial\lambda}\right].
\end{aligned}$$

Asymptotically valid (i.e. misspecification robust) inferences on θ_* can be based on the sample analog $\widehat{H}_{g\theta\theta}\left(\widehat{\phi},,\widehat{h}\right)^{-1}$ of $H_{g\theta\theta}(\phi_*,h_*)^{-1}$ and an estimator $\widehat{\Omega}_{g\theta\theta}\left(\widehat{\phi},,\widehat{h}\right)$ of $\Omega_{g\theta\theta}(\phi_*,h_*)$ that can be based on a suitable adaptation (to the semiparametric models considered in this paper) of any of the well known kernel based long run variance estimators typically used in the time series econometric literature, see for example Andrews (1991) or Kitamura & Stutzer (1997).

The following proposition provides an example of how the consistency of $\widehat{H}_g\left(\widehat{\phi},\widehat{h}\right)^{-1}\widehat{\Omega}_g\left(\widehat{\phi},\widehat{h}\right)\times \widehat{H}_g\left(\widehat{\phi},,\widehat{h}\right)^{-1}$ (and hence of $\widehat{H}_{g\theta\theta}\left(\widehat{\phi},,\widehat{h}\right)^{-1}\widehat{\Omega}_{g\theta\theta}\left(\widehat{\phi},\widehat{h}\right)\widehat{H}_{g\theta\theta}\left(\widehat{\phi},,\widehat{h}\right)^{-1}$) can be shown under α mixing (and hence β mixing), using the same blocking technique as that of Kitamura & Stutzer (1997) (see Smith (2011) for a more general but asymptotically equivalent form of smoothing) to construct the estimator $\widehat{\Omega}_g\left(\widehat{\phi},\widehat{h}\right)$. Let

$$\begin{aligned}
b_i(\phi,h) &= \frac{1}{M^{1/2}}\sum_{j=1}^{M}\frac{\partial \exp\left(\lambda' g_{(i-1)+j}(\theta,h)\right)}{\partial\phi}, \quad \bar{b}(\phi,h) = \frac{1}{Q}\sum_{i=1}^{Q} b_i(\phi,h), \qquad (3.3) \\
\widehat{\Omega}_g\left(\widehat{\phi},\widehat{h}\right) &= \frac{1}{Q}\sum_{i=1}^{Q}\left(b_i\left(\widehat{\phi},\widehat{h}\right) - M^{1/2}\bar{b}\left(\widehat{\phi},\widehat{h}\right)\right)^{\otimes 2},
\end{aligned}$$

where $b_i(\phi,h)$ is the $i-$th block of the semiparametric first order conditions, $M =: M(T)\to\infty$ as $T\to\infty$ is the "smoothing" parameter and $Q = \lceil T-M\rceil + 1$ with $\lceil\cdot\rceil$ the integer part function. For a generic class of functions, say $\mathcal{C}$, let $N_{[]}(\varepsilon,\mathcal{C},L_r(P))$ denote its bracketing number, and assume that:

V1 (i) the sequence of random vectors $\{Z_t, t = 0,\pm1,\pm2,...,T\geq 1\}$ is strictly stationary and α mixing with mixing coefficient $\alpha(k)$ satisfying $\sum_{k=1}^{\infty}(k+1)^{p-1}\alpha(k)^{\frac{\delta}{2p+\delta}} < \infty$, (ii) $E\left\|\partial\exp\left(\lambda_*' g_t(\theta_*,h_*)\right)/\partial\phi\right\|^{2p+\delta} < \infty$ for some $p>2$ and $0<\delta\leq 2$, (iii) $M = o(T)$,

V2 (i) the classes of functions $\mathcal{G}_{\partial g}=\{Z_t \to \partial \exp(\lambda' g_t(\theta, h))/\partial\phi, \theta\in\Theta, \lambda\in\Lambda(\theta), h\in\mathcal{H}_\epsilon\}$ and $\mathcal{G}_{\partial^2 g}=\left\{Z_t \to \partial^2 \exp(\lambda' g_t(\theta, h))/(\partial\phi)^{\otimes 2}, \theta\in\Theta, \lambda\in\Lambda(\theta), h\in\mathcal{H}_\epsilon\right\}$ have $N_{[]}(\varepsilon, \mathcal{G}_{\bullet}, L_1(P)) < \infty$, where $\bullet$ is either ∂g or $\partial^2 g$, (ii) $E\sup_{\theta\in\Theta,\lambda\in\Lambda(\theta),h\in\mathcal{H}_\epsilon}\|\partial\exp(\lambda' g_t(\theta,h))/\partial\phi\| < \infty$, $E\sup_{\theta\in\Theta,\lambda\in\Lambda(\theta),h\in\mathcal{H}_\epsilon}\left\|\partial^2\exp(\lambda' g_t(\theta,h))/(\partial\phi)^{\otimes 2}\right\| < \infty$,

V3 (i) $H_g(\phi_*, h_*)$ is nonsingular, (ii) $\Omega_g(\phi_*, h_*)$ is positive definite.

Note that the summability condition of the mixing coefficient $\alpha(k)$ in V1 is satisfied, for example, if $\{Z_t\}$ is m-dependent or it has an exponential decay, or if $\alpha(k) = O\left(k^{-\zeta}\right)$ for $\zeta > p(2p+\delta)/\delta$.

Proposition 1 *Under V1-V3*

$$\left\|\widehat{H}_g\left(\widehat{\phi},\widehat{h}\right)^{-1}\widehat{\Omega}_g\left(\widehat{\phi},\widehat{h}\right)\widehat{H}_g\left(\widehat{\phi},\widehat{h}\right)^{-1} - H_g(\phi_*, h_*)^{-1}\Omega_g(\phi_*, h_*)H_g(\phi_*, h_*)^{-1}\right\| = o_p(1).$$

Standard arguments can then be used to show that the t and Wald statistics based on $\widehat{H}_g\left(\widehat{\phi},\widehat{h}\right)^{-1}\widehat{\Omega}_g\left(\widehat{\phi},\widehat{h}\right)\widehat{H}_g\left(\widehat{\phi},\widehat{h}\right)^{-1}$ have an asymptotic standard normal and chi-squared calibration, respectively.

4 Misspecified semiparametric models selection theory

In this section we consider two types of model selection tests for comparing two possibly misspecified semiparametric moment conditions models. The first type is an extension to the same two-step testing selection procedure (where the first step is based on a pre-test - see Theorem 3 and Remark 1 below) as that proposed by Vuong (1989); we call this extension the naive extension. The second type is an extension to the model selection test statistic proposed by Shi (2015), which we call the uniform extension. As mentioned in the Introduction, both extensions have their own merits, hence we describe them in the following two subsections.

4.1 Naive extension to Vuong's model selection

We assume that the competing semiparametric model is defined through an alternative set of moment functions $f : \mathcal{Z}\times\mathcal{B}\times\mathcal{L} \to \mathbb{R}^s$, where $B\subset\mathbb{R}^b$ $(s\geq b)$ and $\mathcal{L} = \mathcal{L}_1\times ....\times\mathcal{L}_l$ is a pseudo-metric space of functions[4], such that $P_{\beta,l} = \{P|E_P(f(Z_t,\beta,l)) = 0\}$. As in Section 2, we assume that the competing semiparametric model is misspecified, that is $P_0 \notin \mathcal{P}_f$ for all $\beta\in\mathcal{B}$ and $l\in\mathcal{L}$, where $\mathcal{P}_f = \cup_{\beta\in B, l\in\mathcal{L}}P_{\beta,l}$. Then, the two-step semiparametric ET estimator $\widehat{\beta}$ is defined as

$$\left[\widehat{\beta}',\widehat{\gamma}'\right] = \arg\max_{\beta\in B}\arg\min_{\gamma\in\Gamma(B)}\mathcal{N}_T\left(\beta,\widehat{l},\gamma\right),$$

[4]Note that $\mathcal{L}$ could coincide with $\mathcal{H}$, that is the two semiparametric models feature the same infinite dimensional parameter.

where $\mathcal{N}_T\left(\beta,\widehat{l},\gamma\right)=\sum_{t=1}^{T}\exp\left(\gamma' f_t\left(\beta,\widehat{l}\right)\right)/T$.

The comparison criterion we use is the minimum KL divergence, which, given its connection with ET (see (2.3)), amounts to compare $D_{KL}\left(P\left(\theta_*,h_*\right),P_0\right)$ versus $D_{KL}\left(P\left(\beta_*,l_*\right),P_0\right)$. To be specific, following Vuong's (1989) terminology, we say that the semiparametric model $\mathcal{P}_g$ is better (worse) than model $\mathcal{P}_f$ if its KL divergence $D_{KL}\left(P\left(\theta_*,h_*\right),P_0\right)$ is smaller (larger) than $D_{KL}\left(P\left(\beta_*,l_*\right),P_0\right)$. Let $D=\mathcal{M}\left(\theta_*,h_*,\lambda_*\right)-\mathcal{N}\left(\beta_*,l_*,\gamma_*\right)$; the null hypothesis is that the two semiparametric models have the same KL divergence (which implies that we cannot discriminate between them), that is

$$H_0: D=0. \tag{4.1}$$

To test (4.1) we can use the sample analogs $\mathcal{M}_T\left(\widehat{\theta},\widehat{h},\widehat{\lambda}\right)$ and $\mathcal{N}_T\left(\widehat{\beta},\widehat{l},\widehat{\gamma}\right)$, that is the test statistic

$$D_T=\frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(\exp\left(\widehat{\lambda}' g_t\left(\widehat{\theta},\widehat{h}\right)\right)-\exp\left(\widehat{\gamma}' f_t\left(\widehat{\beta},\widehat{l}\right)\right)\right), \tag{4.2}$$

which is a two-step semiparametric extension of the test statistic of Vuong (1989).

Let $\psi=\left[\beta',\gamma'\right]'$ and

$$\sigma_*^2=\lim_{T\to\infty}Var\left[\frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(\exp\left(\lambda_*' g_t\left(\theta_*,h_*\right)\right)-\exp\left(\gamma_*' f_t\left(\beta_*,l_*\right)\right)\right)\right],$$

and assume that:

A5 (i) A1(i) holds, (ii) A1(ii)-(v) hold for the competing model P_f, the parameter spaces B and $\Gamma(B)$, the moment indicator $f_t(\beta,l)$ and β_* and γ_*,

A6 A2 holds for $\widehat{\beta}$, $\widehat{\gamma}$, $\widehat{l}$, $\sum_{t=1}^{T}\partial^2\exp\left(\gamma' f_t\left(\beta,l\right)\right)/\left(\partial\psi\right)^{\otimes 2}/T$ and its probability limit $H_f\left(\psi_*,l_*\right)$,

A7 (i) A3 holds for $v_T(h)$ and its analog $v_T(l)$ for $f_t(\beta,l)$; (ii) the empirical processes

$$\begin{aligned} v_T^g(h) &= \frac{1}{T}\sum_{t=1}^{T}\left[\exp\left(\lambda_*' g_t\left(\theta_*,h\right)\right)-E\left(\exp\left(\lambda_*' g_t\left(\theta_*,h\right)\right)\right)\right], \\ v_T^f(l) &= \frac{1}{T}\sum_{t=1}^{T}\left[\exp\left(\gamma_*' f_t\left(\beta_*,l\right)\right)-E\left(\exp\left(\gamma_*' f_t\left(\beta_*,l\right)\right)\right)\right], \end{aligned}$$

satisfy the same stochastic equicontinuity assumption A3(i), and $E\left[\exp\left(\lambda_*' g_t\left(\theta_*,\widehat{h}\right)\right)\right]=o_p\left(T^{-1/2}\right)$ and $E\left[\exp\left(\gamma_*' f_t\left(\beta_*,\widehat{l}\right)\right)\right]=o_p\left(T^{-1/2}\right)$,

A8
$$\frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(\exp\left(\lambda_*' g_t\left(\theta_*,h_*\right)\right)-\exp\left(\gamma_*' f_t\left(\beta_*,l_*\right)\right)\right)\xrightarrow{d}N\left(0,\sigma_*^2\right).$$

Theorem 2 *Assume that $\sigma_*^2>0$. Then under Assumptions A5-A8 and the null hypothesis $H_0: D=0$*

$$D_T\xrightarrow{d}N\left(0,\sigma_*^2\right).$$

Thus the test statistic rejects the null hypothesis if $\left| D_T / \widehat{\sigma}\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right) \right| > z_{1-\alpha/2}$, where $\widehat{\sigma}\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)$ is a consistent estimator of σ_* - see (4.3) below for an example, $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution and α is the nominal size.

The validity of Theorem 2 depends crucially on the assumption that $\sigma_*^2 > 0$, which implies that the two competing semiparametric models are strictly non-nested (or nonoverlapping).

Definition 3 *(i) The two semiparametric models P_g and P_f are strictly non nested if $P_g \cap P_f = \varnothing$. (ii) The two semiparametric models P_g and P_f are overlapping if $P_g \cap P_f \neq \varnothing$ with $P_g \subsetneq P_f$ and $P_f \supsetneq P_g$. (iii) The two semiparametric models P_g and P_f are nested if either $P_g \subset P_f$ (P_g is nested in P_f) or $P_f \subset P_g$ (P_f is nested in P_g).*

Definition 3 extends Vuong's (1989) to semiparametric models. It is important to note that in some cases the non nested or overlapping condition is relatively easy to check: a leading example would be two completely different semiparametric specifications, say a single index versus a partial linear one with different explanatory variables. In general, however, the non nested or overlapping condition can be hard to check in semiparametric moment conditions models, because each model (i.e. each set of probability distributions compatible with the assumed semiparametric specification in terms of the KL divergence) can be rather large (see also Kitamura (2000) for a similar point). For this reason, we propose an additional test statistic for checking the hypothesis[5] that $\sigma_*^2 = 0$ using a variance statistic similar to the $\widehat{\omega}_n^2$ one proposed by Vuong (1989). To be specific, we use the same blocking technique as that described in (3.3) and define

$$\widehat{\sigma}^2\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right) = \frac{1}{Q}\sum_{i=1}^{Q}\left(b_i\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right) - M^{1/2}\overline{b}\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)\right)^2,$$

where

$$b_i\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right) = \frac{1}{M^{1/2}}\sum_{j=1}^{M}\exp\left(\widehat{\lambda}' g_{(i-1)+j}\left(\widehat{\theta}, \widehat{h}\right)\right) - \exp\left(\widehat{\gamma}' f_{(i-1)+j}\left(\widehat{\beta}, \widehat{l}\right)\right). \tag{4.3}$$

Let

$$\begin{aligned}
\Omega_f(\psi, l) &= \lim_{T\to\infty} Var\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{\partial \exp\left(\gamma' f_t(\beta, l)\right)}{\partial \psi}\right), \\
\Omega_{gf}(\phi, h, \psi, l) &= \lim_{T\to\infty} Cov\left[\frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{\partial \exp\left(\lambda' g_t(\theta, h)\right)}{\partial \phi}, \frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{\partial \exp\left(\gamma' f_t(\beta, l)\right)}{\partial \psi},\right],
\end{aligned}$$

[5] It should be noted that there is another situation in which σ_*^2 could be possibly 0, namely if the spectral density at the zero frequency of the time series $\left\{\left(\exp\left(\lambda_*' g_t(\theta_*, h_*)\right) - \exp\left(\gamma_*' f_t(\beta_*, l_*)\right)\right)\right\}_{t=1}^{T}$ is itself 0. This might happen for example if it behaves like a moving average unit root time series, see for example Breitung (2008) for more details on how to test the moving average unit root hypothesis. The result of Theorem 3 implicitly excludes this situation.

$$\Sigma(\phi,h,\psi,l)=\left[\begin{array}{cc}-\Omega_g(\phi,h)H_g(\phi,h)^{-1} & -\Omega_{fg}(\phi,h,\psi,l)H_f(\psi,l)^{-1}\\ \Omega_{gf}(\phi,h,\psi,l)H_g(\phi,h)^{-1} & \Omega_f(\theta,\lambda,h)H_f(\psi,l)^{-1}\end{array}\right],$$

and assume that:

A9

$$\sup_{h\in\mathcal{H}_\epsilon,l\in\mathcal{L}_\epsilon}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\exp\left(\lambda_*'g_t(\theta_*,h)\right)-\exp\left(\gamma_*'f_t(\beta_*,l)\right)\right)\times\right.$$
$$\left.\frac{\partial\left(\exp\left(\lambda_*'g_t(\theta_*,h)\right)-\exp\left(\gamma_*'f_t(\beta_*,l)\right)\right)}{\partial(\phi',\psi')'}-L(\phi_*,h,\psi_*,l)\right\|=o_p(1)$$

with $L(\phi,h,\psi,l)$ being the corresponding finite probability limit,

A10

$$\sup_{\substack{\lambda\in\Lambda_*(\Theta_*),\theta\in\Theta_*,h\in\mathcal{H}_\epsilon\\ \gamma\in\Gamma(B_*),\beta\in B_*,l\in\mathcal{L}_\epsilon}}\left\|\frac{1}{T}\sum_{t=1}^{T}\partial_\bullet\left(\frac{\partial^2\left(\exp\left(\lambda'g_t(\theta,h)\right)-\exp\left(\gamma'f_t(\beta,l)\right)\right)}{\left(\partial(\phi',\psi')'\right)^{\otimes 2}}\right)-P_\bullet(\phi,h,\psi,l)\right\|^2=o_p(1),$$

where "$\bullet$" is either for h or l and both derivatives should be interpreted as functional derivatives with respect to h and l with $P_\bullet(\phi,h,\psi,l)$ being the corresponding finite probability limit.

Theorem 3 *Under A5-A10, $M=o\left(T^{1/2}\right)$ and the null hypothesis $H_0:\sigma_*^2=0$*

$$T\widehat{\sigma}^2\left(\widehat{\phi},\widehat{h},\widehat{\psi},\widehat{l}\right)\xrightarrow{d}\sum_{j=1}^{q}\xi_j^2\chi_j^2(1),\tag{4.4}$$

where, for $j=1,....,q$, $q=l+k+s+b$, ξ_j are the eigenvalues of $\Sigma(\phi_*,h_*,\psi_*,l_*)$ and $\chi_j^2(1)$ are independent chi-squared variables with one degree of freedom.

The test rejects the null hypothesis if $T\widehat{\sigma}^2\left(\widehat{\phi},\widehat{h},\widehat{\psi},\widehat{l}\right)>c_\alpha^\xi$ where c_α^ξ is the upper critical values of the nonstandard distribution (4.4), which can however be easily simulated using consistent estimators $\widehat{\xi}_j^2$ of ξ_j^2 $(j=1,...,q)$. Alternatively we can use the scaled adjusted statistic

$$\widehat{\sigma}_s^2\left(\widehat{\phi},\widehat{h},\widehat{\psi},\widehat{l}\right)=\frac{q}{tr\left(\widehat{\Sigma}\left(\widehat{\phi},\widehat{h},\widehat{\psi},\widehat{l}\right)^2\right)}\widehat{\sigma}^2\left(\widehat{\phi},\widehat{h},\widehat{\psi},\widehat{l}\right),\tag{4.5}$$

which, using the same heuristic argument of Rao & Scott (1981) can be shown to converge to a standard chi-squared distribution, that is $T\widehat{\sigma}_s^2\xrightarrow{d}\chi^2(q)$.

Remark 1 *In practice the naive extension to Vuong's (1989) model selection approach is based on the same two-step procedure suggested by Vuong (1989), as it consists of using the statistic $\widehat{\sigma}^2$ of Theorem 3 (or its modified version $\widehat{\sigma}_s^2$ given in (4.5)) as a pre test to determine whether the variance of Theorem 2 is zero or not. Depending on whether the null hypothesis of the pre test is rejected (or not), model selection is (not) possible with the D_T, and it is the discontinuous change between the asymptotic distributions of $\widehat{\sigma}^2$ and D_T that might result in poor finite sample properties of the two-step procedure.*

4.2 Uniform extension to Vuong's model selection

We begin this section by providing some intuition behind Shi's (2015) modification of Vuong's statistic (4.2). The proof of Theorem 2 shows that

$$\begin{aligned} D_T &= \frac{1}{T^{1/2}} \sum_{t=1}^{T} \left(\exp\left(\lambda_*' g_t\left(\theta_*, h_*\right)\right) - \exp\left(\gamma_*' f_t\left(\beta_*, l_*\right)\right)\right) - \\ & \frac{1}{2T^{1/2}} \left[\left(\widehat{\phi} - \phi_*\right)', \left(\widehat{\psi} - \psi_*\right)'\right] \sum_{t=1}^{T} \frac{\partial^2 \left(\exp\left(\overline{\lambda}' g_t\left(\overline{\theta}, \widehat{h}\right)\right) - \exp\left(\overline{\gamma}' f_t\left(\overline{\beta}, \widehat{l}\right)\right)\right)}{\left(\partial\left(\phi', \psi'\right)'\right)^{\otimes 2}} \times \\ & \left[\left(\widehat{\phi} - \phi_*\right)', \left(\widehat{\psi} - \psi_*\right)'\right]' + o_p(1), \end{aligned} \tag{4.6}$$

where $\overline{\phi}$ is on the line segment between ϕ_* and $\widehat{\phi}$, and similarly for $\overline{\psi}$, and a combination of the central limit theorem, the uniform law of large numbers and the continuous mapping theorem shows that

$$D_T \xrightarrow{d} Z_{\sigma_*^2} - \frac{1}{2T^{1/2}} Z_{\phi,\psi}' H\left(\phi_*, h_*, \psi_*, l_*\right)^{-1} Z_{\phi,\psi}, \tag{4.7}$$

where

$$\left[\begin{array}{c} Z_{\sigma_*^2} \\ Z_{\phi,\psi} \end{array}\right] \sim N\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], \left[\begin{array}{cc} \sigma_*^2 & \Xi\left(\phi_*, h_*, \psi_*, l_*\right)' \\ \Xi\left(\phi_*, h_*, \psi_*, l_*\right) & \Omega\left(\phi_*, h_*, \psi_*, l_*\right) \end{array}\right]\right),$$

$$\begin{aligned} \Xi\left(\phi_*, h_*, \psi_*, l_*\right) &= \lim_{T\to\infty} Cov\left(\sum_{t=1}^{T} \left(\exp\left(\lambda_*' g_t\left(\theta_*, h_*\right)\right) - \exp\left(\gamma_*' f_t\left(\beta_*, l_*\right)\right)\right),\right. \\ & \left. \frac{1}{T^{1/2}} \sum_{t=1}^{T} \frac{\partial\left(\exp\left(\lambda_*' g_t\left(\theta_*, h_*\right)\right) - \exp\left(\gamma_*' f_t\left(\beta_*, l_*\right)\right)\right)}{\partial\left[\phi', \psi'\right]'}\right). \end{aligned}$$

If $\sigma_*^2 > 0$, as $T \to \infty$, the distribution of D_T is well approximated by $Z_{\sigma_*^2}$ as the second term in (4.7) becomes negligible, however if $\sigma_*^2 = 0$ the second term in (4.7) becomes the dominant one, hence the distribution of D_T is better approximated by $Z_{\phi,\psi}' H\left(\phi_*, h_*, \psi_*, l_*\right)^{-1} Z_{\phi,\psi}/2T^{1/2}$. Finally, if $\sigma_*^2 > 0$ but T is not large enough yet, the two terms in (4.7) might be of similar order of magnitude, and thus the asymptotic distribution of D_T might not be well approximated by either of them.

In a similar way, the proof of Theorem 3 shows that if $\sigma_*^2 = 0$

$$T\widehat{\sigma}^2\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right) \xrightarrow{d} Z_{\phi,\psi}' H\left(\phi_*, h_*, \psi_*, l_*\right)^{-1} Z_{\phi,\psi},$$

whereas, as we show in the proof of Theorem 4 below, if $\sigma_*^2 \in (0, \infty)$

$$\begin{aligned} \frac{\widehat{\sigma}^2\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)}{\sigma_T^2\left(\phi_*, h_*, \psi_*, l_*\right)} &\xrightarrow{d} 1 - \frac{2}{\sigma_*} \rho_*' \Delta^{1/2}\left(\phi_*, h_*, \psi_*, l_*\right) H\left(\phi_*, h_*, \psi_*, l_*\right)^{-1} Z_{\phi,\psi} + \\ & \frac{1}{\sigma_*^2} Z_{\phi,\psi}' H\left(\phi_*, h_*, \psi_*, l_*\right)^{-1} \Omega\left(\phi_*, h_*, \psi_*, l_*\right) H\left(\phi_*, h_*, \psi_*, l_*\right)^{-1} Z_{\phi,\psi}, \end{aligned} \tag{4.8}$$

where $\sigma_T^2(\phi_*, h_*, \psi_*, l_*) := \lim_{T\to\infty} Var\left(\sum_{t=1}^T (\exp(\lambda_*' g_t(\theta_*, h_*)) - \exp(\gamma_*' f_t(\beta_*, l_*)))\right)$, and ρ_* and $\Delta(\phi_*, h_*, \psi_*, l_*)$ are defined in (4.9) and (4.10) below, which shows that an appropriately standardized version of $\widehat{\sigma}^2\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)$ converges to a nonstandard distribution, that, as shown by Shi (2015)[Section 3.3] is always nonnegative but can take values close to zero with significant probability. Thus (4.7) and (4.8) imply that the standardized Vuong statistic $D_T/\widehat{\sigma}\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)$ not only is not centered at zero because of the second order bias, but can also be characterized by a fat tail distribution, with obvious negative implications for inference. To overcome these two problems, we follow the same approach of Shi (2015): given a set of distributions $\mathcal{P}_* = \{\mathcal{P}_{g,f}|D = 0\}$ that are consistent with the null hypothesis (4.1), we consider sequences $\{P_T \in P_*, T \geq 1\}$, generalize the dependent structure of the sequence $\{Z_t\}$ to that of a triangular array of α mixing dependent variables $\{Z_{Tt}\}$ and consider sequences $\sigma_T^2(\phi_*, h_*, \psi_*, l_*)$ such that under P_T $\lim_{T\to\infty} T\ \sigma_T^2(\phi_*, h_*, \psi_*, l_*) \to \sigma_*^2 \in [0, \infty]$, where the rate T is chosen so that the resulting local asymptotic distribution represents a smooth transition from a nonstandard one to a standard normal as σ_*^2 ranges from 0 to ∞.

Let $\rho_T(\phi_*, h_*, \psi_*, l_*)$ denote a sequence of "correlation coefficients", such that under P_T $\lim_{T\to\infty} \rho_T(\phi_*, h_*, \psi_*, l_*) \to \rho_*$, where

$$\begin{aligned} \rho_* &= \sigma_*^+ \left(\Delta^{1/2}(\phi_*, h_*, \psi_*, l_*)\right)^+ \Xi(\phi_*, h_*, \psi_*, l_*), \\ \sigma_*^+ &= \begin{cases} 0 \text{ if } \sigma_* = 0, \\ \frac{1}{\sigma_*} \text{ if } \sigma_* > 0, \end{cases} \end{aligned} \tag{4.9}$$

$$\Delta(\phi_*, h_*, \psi_*, l_*) = diag(\Omega(\phi_*, h_*, \psi_*, l_*)), \tag{4.10}$$

" + " denotes the Moore-Penrose inverse of a matrix,

$$\begin{aligned} b_{Ti}(\phi, h, \psi, l) &= \frac{1}{M^{1/2}} \sum_{j=1}^{M} \left(\exp\left(\lambda' g_{T(i-1)+j}(\theta, h)\right) - \exp\left(\gamma' f_{T(1-i)+j}(\beta, l)\right)\right), \\ \widehat{\sigma}^2\left(\widehat{\phi}_T, \widehat{h}, \widehat{\psi}_T, \widehat{l}\right) &= \frac{1}{Q} \sum_{i=1}^{Q} \left(b_{Ti}\left(\widehat{\phi}_T, \widehat{h}, \widehat{\psi}_T, \widehat{l}\right) - M^{1/2}\overline{b}_T\left(\widehat{\phi}_T, \widehat{h}, \widehat{\psi}_T, \widehat{l}\right)\right)^{\otimes 2}, \end{aligned} \tag{4.11}$$

with

$$\left[\widehat{\theta}_T', \widehat{\lambda}_T'\right]' = \arg\max_{\theta\in\Theta} \arg\min_{\lambda\in\Lambda(\Theta)} \frac{1}{T} \sum_{t=1}^{T} \exp\left(\lambda' g_{Tt}\left(\theta, \widehat{h}\right)\right), \tag{4.12}$$

and similarly for the estimators $\left[\widehat{\beta}_T', \widehat{\gamma}_T'\right]'$ of the competing semiparametric model based on $f_{Tt}(\beta, l)$.

Assume that:

U1 The triangular array sequence $\{Z_{Tt}, t = 0, \pm 1, \pm 2, ..., T \geq 1\}$ is α mixing with T-th mixing coefficient $\alpha_T(k)$ satisfying $\sup_T \sum_{k=1}^{\infty} (k+1)^2 \alpha_T(k)^{\frac{\delta}{4+\delta}} < \infty$ for some $\delta > 0$,

U2 (i) A1(ii)-(v) and A5(ii) hold for $g_{Tt}(\theta, h)$ and $f_{Tt}(\beta, l)$, (ii)

$$\begin{aligned} E\left|\exp\left(\lambda_*' g_{Tt}(\theta_*, h_*)\right) - E\left(\exp\left(\lambda_*' g_{Tt}(\theta_*, h_*)\right)\right)\right|^{2(2+\delta)} &< \infty, \\ E\left|\exp\left(\gamma_*' f_{Tt}(\beta_*, l_*)\right) - E\left(\exp\left(\gamma_*' f_{Tt}(\beta_*, l_*)\right)\right)\right|^{2(2+\delta)} &< \infty \end{aligned}$$

for all t and T, (iii) $E\left\|\partial\left(\exp\left(\lambda' g_{Tt}(\theta, h)\right) - \exp\left(\gamma' f_{Tt}(\beta, l)\right)\right)/\partial\left[\phi', \psi'\right]'\right\|^{2(2+\delta)} < \infty$ for all t and T, (iv)

$$E \sup_{\phi\in\Phi, h\in\mathcal{H}_\epsilon, \psi\in\Psi, l\in\mathcal{L}_\epsilon} \left\| \frac{\partial^2\left(\exp\left(\lambda' g_{Tt}(\theta, h)\right) - \exp\left(\gamma' f_{Tt}(\beta, l)\right)\right)}{\left(\partial\left[\phi', \psi'\right]'\right)^{\otimes 2}} \right\| < \infty$$

for all t and T,

U3 (i) $\left\|\widehat{h} - h_\bullet\right\|_{\mathcal{H}} = o_p(1)$, $\left\|\widehat{l} - l_\bullet\right\|_{\mathcal{L}} = o_p(1)$, and $\left\|\widehat{\phi}_T - \phi_*\right\| = o_p(1)$, $\left\|\widehat{\psi}_T - \psi_*\right\| = o_p(1)$, (ii) the empirical processes

$$\begin{aligned} v_T^d(h, l) &= \sum_{t=1}^{T}\left(\exp\left(\lambda_*' g_{Tt}(\theta_*, h)\right) - \exp\left(\gamma_*' f_{Tt}(\beta_*, l)\right) - \right. \\ &\quad \left. E\left(\exp\left(\lambda_*' g_{Tt}(\theta_*, h)\right) - \exp\left(\gamma_*' f_{Tt}(\beta_*, l)\right)\right)\right) \text{ and} \\ v_T^\partial(h, l) &= \frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(\frac{\partial \exp\left(\lambda_*' g_{Tt}(\theta_*, h)\right) - \exp\left(\gamma_*' f_{Tt}(\beta_*, l)\right)}{\partial\left[\phi', \psi'\right]'} - \right. \\ &\quad \left. E\left(\frac{\partial \exp\left(\lambda_*' g_{Tt}(\theta_*, h)\right) - \exp\left(\gamma_*' f_{Tt}(\beta_*, l)\right)}{\partial\left[\phi', \psi'\right]'}\right)\right) \end{aligned}$$

satisfy the stochastic equicontinuity conditions

$$\begin{aligned} \sup_{h\in\mathcal{H}_\epsilon, l\in\mathcal{L}_\epsilon}\left\|v_T^e(h, l) - v_T^e(h_*, l_*)\right\| &= o_p\left(T^{1/2}\sigma_T(\phi_*, h_*, \psi_*, l_*)\right) \text{ and} \\ \sup_{h\in\mathcal{H}_\epsilon, l\in\mathcal{L}_\epsilon}\left\|v_T^\partial(h, l) - v_T^\partial(h_*, l_*)\right\| &= o_p(1) \end{aligned}$$

for all T, (iii) $\left|\widehat{\sigma}^2\left(\phi_*, \widehat{h}, \psi_*, \widehat{l}\right) - \widehat{\sigma}^2(\phi_*, h_*, \psi_*, l_*)\right| = o_p\left(\sigma_T^2(\phi_*, h_*, \psi_*, l_*)\right)$ for all T, (iv)

$$\begin{aligned} E\left[\sum_{t=1}^{T}\exp\left(\lambda_*' g_{Tt}\left(\theta_*, \widehat{h}\right)\right) - \exp\left(\gamma_*' f_{Tt}\left(\beta_*, \widehat{l}\right)\right)\right] &= o_p\left(1/\left(T^{1/2}\sigma_T(\phi_*, h_*, \psi_*, l_*)\right)\right) \text{ and} \\ E\left(\sum_{t=1}^{T}\frac{\partial \exp\left(\lambda_*' g_{Tt}\left(\theta_*, \widehat{h}\right)\right) - \exp\left(\gamma_*' f_{Tt}\left(\beta_*, \widehat{l}\right)\right)}{\partial\left[\phi', \psi'\right]'}\right) &= o_p\left(T^{-1/2}\right). \end{aligned}$$

Assumption U1(i) imposes a standard mixing assumption for triangular arrays of α mixing random vectors, Assumptions U2 and U3 are similar to those assumed in the previous sections: U3(i) can be verified using similar arguments as those used in the proof of Proposition 4 in the supplemental appendix; U3(ii) is a stochastic equicontinuity condition similar to that assumed in

A3(i) and A7(ii), which can be verified under different more primitive conditions, see for example the discussion before Theorem 1 and assumptions E5(iv) and E6(ii)-(iii). Assumption U3(iv) can be verified under standard smoothness conditions (in h and l) of the moments indicators $g_{Tt}(\theta_*, h)$ and $f_{Tt}(\beta_*, l)$ and the convergence rate (in an appropriate norm) of the nonparametric estimators $\widehat{h}$ and $\widehat{l}$, see for example the proof of Proposition 3.

The following theorem establishes the joint local asymptotic distribution of $T^{1/2}D_T$ defined in (4.2) and $T\widehat{\sigma}^2\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)$; let

$$V(\phi_*, h_*, \psi_*, l_*) = diag\left(eig\left(\Omega^{1/2}(\phi_*, h_*, \psi_*, l_*) H(\phi_*, h_*, \psi_*, l_*)^{-1} \Omega^{1/2}(\phi_*, h_*, \psi_*, l_*)\right)\right),$$

denote the diagonal matrix formed by the $l+k+s+b$ eigenvalues (eig) of

$$\Omega^{1/2}(\phi_*, h_*, \psi_*, l_*) H(\phi_*, h_*, \psi_*, l_*) \Omega^{1/2}(\phi_*, h_*, \psi_*, l_*).$$

Theorem 4 *Under U1-U3, $M = o\left(T^{1/2}\right)$ and the null hypothesis $H_0 : D = 0$ (i) if $\sigma_*^2 \in [0, \infty)$*

$$\begin{bmatrix} T^{1/2}D_T \\ T\widehat{\sigma}^2\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} J_D\left(\sigma_*, \rho_*^*, V(\phi_*, h_*, \psi_*, l_*)\right) \\ J_{\sigma_*^2}\left(\sigma_*^2, \rho_*^*, V(\phi_*, h_*, \psi_*, l_*)\right) \end{bmatrix},$$

where

$$\begin{bmatrix} J_D\left(\sigma_*, \rho_*^*, V(\phi_*, h_*, \psi_*, l_*)\right) \\ J_{\sigma_*^2}\left(\sigma_*, \rho_*^*, V(\phi_*, h_*, \psi_*, l_*)\right) \end{bmatrix} = \begin{bmatrix} \sigma_* Z - \frac{1}{2} Z_{\phi,\psi}^{*\prime} V(\phi_*, h_*, \psi_*, l_*) Z_{\phi,\psi}^* \\ \sigma_*^2 - 2\sigma_* \rho_*^{*\prime} V(\phi_*, h_*, \psi_*, l_*) Z_{\phi,\psi}^* - Z_{\phi,\psi}^{*\prime} V(\phi_*, h_*, \psi_*, l_*)^2 Z_{\phi,\psi}^* \end{bmatrix},$$

$$\begin{bmatrix} Z \\ Z_{\phi,\psi}^* \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_*^{*\prime} \\ \rho_*^* & I \end{bmatrix}\right),$$

and ρ_*^* is the solution to the equation $\Omega^{1/2}(\phi_*, h_*, \psi_*, l_*) Q\rho_*^* = \Delta^{1/2}(\phi_*, h_*, \psi_*, l_*) \rho_*$ with Q an orthonormal matrix satisfying

$$QV(\phi_*, h_*, \psi_*, l_*) Q' = \Omega^{1/2}(\phi_*, h_*, \psi_*, l_*) H(\phi_*, h_*, \psi_*, l_*) \Omega^{1/2}(\phi_*, h_*, \psi_*, l_*);$$

(ii) if $\sigma_*^2 = \infty$

$$\frac{D_T}{\sigma_T(\phi_*, h_*, \psi_*, l_*)} \xrightarrow{d} Z \text{ and } \frac{\widehat{\sigma}\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)}{\sigma_T(\phi_*, h_*, \psi_*, l_*)} \xrightarrow{p} 1.$$

Theorem 4 and a straightforward application of the continuous mapping theorem imply that the local asymptotic distribution of the standardized Vuong statistic $D_T/\widehat{\sigma}\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)$ is

$$\frac{D_T}{\widehat{\sigma}\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)} \xrightarrow{d} \frac{J_D\left(\sigma_*, \rho_*^*, V(\phi_*, h_*, \psi_*, l_*)\right)}{J_{\sigma_*^2}^{1/2}\left(\sigma_*^2, \rho_*^*, V(\phi_*, h_*, \psi_*, l_*)\right)},$$

which can be quite different from a standard normal, because of the asymptotic bias in the numerator and the random denominator, as discussed at the beginning of this section. Thus, as suggested by Shi (2015), we propose a modified standardized Vuong statistic that addresses both problems. To be specific, let

$$D_T^m(c) = \frac{D_T + tr\left(\widehat{V}\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)\right)/2T^{1/2}}{\left(\widehat{\sigma}^2\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right) + ctr\left(\widehat{V}^2\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)\right)/T\right)^{1/2}}, \tag{4.13}$$

denote the modified standardized Vuong statistic. Similar arguments as those used in the proof of Proposition 1 can be used to show that $tr\left(\widehat{V}\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)\right) \xrightarrow{p} tr\left(V\left(\phi_*, h_*, \psi_*, l_*\right)\right)$ and $tr\left(\widehat{V}^2\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)\right) \xrightarrow{p} tr\left(V^2\left(\phi_*, h_*, \psi_*, l_*\right)\right)$, thus, again by a straightforward application of the continuous mapping theorem, the local asymptotic distribution of (4.13) is given by

$$D_T^m(c) \xrightarrow{d} \frac{J_D\left(\sigma_*, \rho_*^*, V\left(\phi_*, h_*, \psi_*, l_*\right)\right) + tr\left(V\left(\phi_*, h_*, \psi_*, l_*\right)\right)/2}{\left(J_{\sigma_*^2}\left(\sigma_*, \rho_*^*, V\left(\phi_*, h_*, \psi_*, l_*\right)\right) + ctr\left(V^2\left(\phi_*, h_*, \psi_*, l_*\right)\right)\right)^{1/2}} := J\left(\sigma_*, \rho_*^*, V\left(\phi_*, h_*, \psi_*, l_*\right), c\right), \tag{4.14}$$

which is still not a standard normal, implying that its critical values can be obtained by simulation. Let

$$cv\left(1-\alpha, V\left(\phi_*, h_*, \psi_*, l_*\right), c\right) = \sup_{\sigma_* \in [0,\infty], \rho_*^* : \|\rho_*^*\| \leq 1} F^{-1}_{J\left(\sigma_*, \rho_*^*, V\left(\phi_*, h_*, \psi_*, l_*\right), c\right)}\left(1-\alpha\right) \tag{4.15}$$

denote the critical value of $J\left(\sigma_*, \rho_*^*, V\left(\phi_*, h_*, \psi_*, l_*\right), c\right)$, which is worth noting to be weakly larger than the corresponding $1-\alpha$ quantile of a standard normal, as $J\left(\infty, \rho_*^*, V\left(\phi_*, h_*, \psi_*, l_*\right), c\right) \sim N(0,1)$ for any ρ_*^*, $V\left(\phi_*, h_*, \psi_*, l_*\right)$ and c. In the online supplemental Appendix we provide details on how to simulate (4.15), including how to choose the crucial parameter c, using essentially the same approach proposed by Shi (2015). The following theorem extends the uniformity property of the model selection test of Shi (2015) to possibly misspecified semiparametric models with α mixing observations.

Theorem 5 *Under the same assumptions of Theorem 4, for any $c \geq 0$*

$$\lim_{T\to\infty} \sup_{P \in \mathcal{P}_*} \Pr\left(\left|D_T^m(c)\right| \geq cv\left(1-\alpha, \widehat{V}\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right), c\right)\right) \leq \alpha.$$

Remark 2 *If the two competing semiparametric models P_g and P_f are known to be nested, say $P_g \subset P_f$, then $\sigma_*^2 = 0$, hence σ_* and ρ_*^* are not required for the computation of (4.15); furthermore a one sided test is more useful than a two sided one because the nested model cannot be closer than the nesting model (in terms of KL divergence) to the true one. Note also that because in the nested case nondegeneracy does not occur, one could simply set $c = 0$, further simplifying the computation of (4.15).*

5 An example and Monte Carlo evidence

In this section we consider an example to illustrate both how the high level assumptions of the previous section can be verified under more primitive conditions and how the proposed model comparison tests perform in finite samples.

5.1 Misspecified instrumental variables partial linear model

We consider the following model

$$r_t(\theta, s) = q(V_t)(Y_t - X_{1t}'\theta - s(X_{2t})),$$

where $q(V_t) : \mathcal{V} \to \mathbb{R}^l$ is a vector of known functions of the instruments V_t (which might contain elements of X_{jt} $(j = 1, 2)$), $X_{1t} \in \mathbb{R}^k$ $(k < l)$, $X_{2t} \in \mathbb{R}$, $s : \mathcal{X}_2 \to \mathbb{R}$ is an unknown function, and let

$$g_t(\theta, h) = q(V_t)\left(Y_t - E(Y_t|X_{2t}) - (X_{1t} - E(X_{1t}|X_{2t}))'\theta\right) \tag{5.1}$$

denote the profile moment indicator, where $h = [h_1, h_2']' = \left[E(Y_t|X_{2t}), E(X_{1t}|X_{2t})'\right]'$.

Let $\widehat{h} = \left[\widehat{E}(Y_t|X_{2t}), \widehat{E}(X_{1t}|X_{2t})\right]'$, where $\widehat{E}(Y_t|X_{2t}) = \sum_{j\neq t=1}^{T} Y_j \omega_{jb}(X_{2t})$, $\omega_{jb}(X_{2t}) = K((X_{2j} - X_{2t})/b) / \sum_{k=1}^{T} K((X_{2k} - X_{2j})/b)$ and $b := b(T)$ is a bandwidth, and similarly for $\widehat{E}(X_{1t}|X_{2t})$, $\overline{Y}_t = Y_t - E(Y_t|X_{2t})$ and similarly for $\overline{X}_{1t}$, and $\mathcal{W}_{q,k}(f) = \left|\sum_{j=0}^{k} \int |d^j f|^q\right|^{1/q}$ denote the Sobolev space of functions.

Assume that:

E1 (i) the sequence of random vectors $\{Z_t : t = 0, \pm 1, \pm 2, ...\}$, where $Z_t = [Y_t, X_t', V_t']'$, is strictly stationary α-mixing with mixing coefficient $\alpha(k)$ satisfying $\sum_{k=1}^{\infty} (k+1)^2 \alpha(k)^{\delta/(\delta+4)} < \infty$ for some finite $\delta > 0$, (ii) the marginal density f_{X_2} of X_{2t} is twice continuously differentiable on $\mathcal{X}_2$, $\inf_{X_{2t}\in\mathcal{X}_2} |f_{X_2}| > 0$, and $\mathcal{X}_2$ is a compact set, (iii) $rank\left(E\left(q(V_t)\overline{X}_{1t}'\right)\right) = k$, (iv) the parameter spaces Θ and $\Lambda(\Theta)$ are compact subsets of $\mathbb{R}^k$ and $\mathbb{R}^l$, respectively, and $\mathcal{H} = \mathcal{W}_{q,k}(h) \leq C$ for some finite $C > 0$ - the Sobolev ball with radius C, (v) $\theta_* \in int(\Theta)$, $\lambda_* \in int(\Lambda(\Theta))$ and $h_* \in \mathcal{H}$,

E2 (i) $E\sup_{\lambda\in\Lambda(\Theta),\theta\in\Theta,h\in\mathcal{H}_\epsilon} \left|\exp\left(\lambda' q(V_t)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right)\right| < \infty$, (ii)

$$E \sup_{\lambda\in\Lambda(\Theta),\theta\in\Theta,h\in\mathcal{H}_\epsilon} \left\|\exp\left(\lambda' q(V_t)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right) q(V_t)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right\| < \infty,$$

$$E \sup_{\lambda\in\Lambda_*(\Theta_*),\theta\in\Theta_*,h\in\mathcal{H}_\epsilon} \left\|\exp\left(\lambda' q(V_t)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right) \lambda' q(V_t) \begin{bmatrix} -1 \\ -\theta \end{bmatrix}\right\| < \infty,$$

where $\mathcal{H}_\epsilon = \{h : \|h - h_\bullet\|_{\mathcal{H}} \leq \epsilon\}$ for some $\epsilon > 0$, (iii) $E\sup_{\lambda\in\Lambda_*(\Theta_*),\theta\in\Theta_*,h\in\mathcal{H}_\epsilon} \left\|\partial^2 \exp\left(\lambda' q(V_t)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right) / (\partial\phi)^{\otimes 2}\right\| < \infty$ where $\Lambda_*(\Theta_*)$ and

Θ_* are neighborhoods of λ_* and θ_*, and

$$\left\| \frac{\partial^2 \exp\left(\lambda' g_t\left(\theta, h\right)\right)}{\left(\partial \phi\right)^{\otimes 2}} - \frac{\partial^2 \exp\left(\lambda_*' g_t\left(\theta_*, h_*\right)\right)}{\left(\partial \phi\right)^{\otimes 2}} \right\| \leq B_{\phi,h}\left(Z_t\right) \left(\left\| \widehat{\phi} - \phi_* \right\| + \left\| \widehat{h} - h_* \right\|_{\mathcal{H}_\epsilon} \right)$$

with $E\left(B_{\phi,h}\left(Z_t\right)\right) < \infty$, (iv) $H_g\left(\theta_*, \lambda_*, h_*\right)$ is nonsingular, where

$$H_{g\theta\theta}\left(\phi, h\right) = E\left[\exp\left(\lambda' q\left(V_t\right)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right)\left(\left(q\left(V_t\right)\overline{X}_{1t}'\right)'\lambda\right)^{\otimes 2} + \right. \tag{5.2}$$

$$\frac{\left(q\left(V_t\right)\overline{X}_{1t}'\right)'}{2}\left[E\left(\exp\left(\lambda' q\left(V_t\right)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right)\left(q\left(V_t\right)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right)^{\otimes 2}\right)\right]^{-1} \times$$

$$\left. E\left(\left(I_l + q\left(V_t\right)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\lambda'\right) q\left(V_t\right)\overline{X}_{1t}'\right)\right],$$

$$H_{g\theta\lambda}\left(\phi, h\right) = E\left[\exp\left(\lambda' q\left(V_t\right)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right)\left(\left(\overline{X}_{1t} q\left(V_t\right)' + \left(q\left(V_t\right)\overline{X}_{1t}'\right)'\lambda q\left(V_t\right)'\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right.\right.\right.$$

$$-\frac{1}{2}E\left(\overline{X}_{1t} q\left(V_t\right)'\left(I_l + q\left(V_t\right)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\lambda'\right)'\right) \times$$

$$\left.\left[E\left(\exp\left(\lambda' q\left(V_t\right)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right)\left(q\left(V_t\right)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right)^{\otimes 2}\right)\right]^{-1}\left(q\left(V_t\right)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right)^{\otimes 2}\right)\right],$$

$$H_{g\lambda\theta}\left(\phi, h\right) = H_{g\theta\lambda}\left(\phi, h\right)',$$

$$H_{g\lambda\lambda}\left(\phi, h\right) = E\left[\exp\left(\lambda' q\left(V_t\right)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right)\left(q\left(V_t\right)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right)^{\otimes 2}\right],$$

E3 (i)

$$\left\| \frac{\partial \exp\left(\lambda_*' g_t\left(\theta_*, h\right)\right)}{\partial \phi} - \frac{\partial \exp\left(\lambda_*' g_t\left(\theta_*, h_*\right)\right)}{\partial \phi} \right\| \leq B_h\left(Z_t\right)\left\| \widehat{h} - h_* \right\|_{\mathcal{H}_\epsilon}$$

with $E\left(B_h\left(Z_t\right)\right) < \infty$ and, for $j = 1, ..., k+1$,

$$E \sup_{h \in \mathcal{H}_\epsilon}\left\| \exp\left(\lambda_*' q\left(V_t\right)\left(\overline{Y}_t - \overline{X}_{1t}'\theta_*\right)\right)\left(\lambda_*' q\left(V_t\right)\right)^2 \begin{bmatrix} 1 \\ \theta_* \end{bmatrix}^{\otimes 2} v_j \right\| < \infty$$

where $v_1 = 1$ and $v_j = \theta_{j-1}$ for $j = 2, ..., k+1$, (ii)

$$E\left\| \begin{array}{c} \exp\left(\lambda_*' q\left(V_t\right)\left(\overline{Y}_t - \overline{X}_{1t}'\theta_*\right)\right)\left(q\left(V_t\right)\overline{X}_{1t}'\right)'\lambda_* \\ \exp\left(\lambda_*' q\left(V_t\right)\left(\overline{Y}_t - \overline{X}_{1t}'\theta_*\right) q\left(V_t\right)\left(\overline{Y}_t - \overline{X}_{1t}'\theta_*\right)\right) \end{array} \right\|^{2+\delta} < \infty,$$

(iii) $\Omega_g(\theta_*, \lambda_*, h_*)$ is positive definite with

$$\begin{aligned}
\Omega_{g\theta\theta}(\theta,\lambda,h) &= \lim_{T\to\infty} Var\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T}\exp\left(\lambda' q(V_t)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right)\left(q(V_t)\overline{X}_{1t}'\right)'\lambda\right), \quad (5.3)\\
\Omega_{g\theta\lambda}(\theta,\lambda,h) &= \lim_{T\to\infty} Cov\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T}\exp\left(\lambda' q(V_t)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right)\left(q(V_t)\overline{X}_{1t}'\right)'\lambda,\right.\\
&\qquad \left.\frac{1}{T^{1/2}}\sum_{t=1}^{T}\exp\left(\lambda' q(V_t)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right)q(V_t)'\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right),\\
\Omega_{g\lambda\theta}(\theta,\lambda,h) &= \Omega_{\theta\lambda}(\theta,\lambda,h)',\\
\Omega_{g\lambda\lambda}(\theta,\lambda,h) &= \lim_{T\to\infty} Var\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T}\exp\left(\lambda' q(V_t)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)q(V_t)\left(\overline{Y}_t - \overline{X}_{1t}'\theta\right)\right)\right).
\end{aligned}$$

E4 (i) the kernel function $K : \mathcal{X}_2 \to \mathbb{R}$, is symmetric with compact support and $\sup_{X_{2t}\in\mathcal{X}_2} |d^j K|^q < \infty$ $j = 0, ..., k$, (ii) the bandwidth b is such that $b \to 0$, $Tb \to \infty$ and $T^{1/2}b^4 \to 0$ as $T \to \infty$.

Assumption E1(i) specifies the dependence structure of Z_t in terms of α-mixing with a summability assumption on the mixing coefficient $\alpha(k)$ that is fairly standard, see for example Doukhan (1994). Assumption E1(ii) is standard in the nonparametric estimation literature; note, however, that the compactness assumption on the support $\mathcal{X}_2$ could be relaxed, if a trimming function is used, see for example Andrews (1994) and Bravo et al. (2017). Assumption E1(iii) is an identification condition that implies A1(ii) (see the proof of Proposition 2 in the supplemental Appendix for more details). Assumptions E1(iv)-(v) specify that the unknown infinite dimensional parameter h_* belongs to a Sobolev ball, which is implied by the more primitive assumption that h is sufficiently smooth with derivatives satisfying $\sup_{X_{2t}\in\mathcal{X}_2} \left|d^j h/dX_{2t}^j\right|^q < \infty$. Assumption E3(i) is a mild Lipschitz condition, while E3(ii) together with E1(i) ensures that a central limit theorem for α-mixing processes applies to the random vector

$$T^{-1/2}\begin{bmatrix} \sum_{t=1}^{T}\exp\left(\lambda_*' q(V_t)\left(\overline{Y}_t - \overline{X}_{1t}'\theta_*\right)\right)\left(q(V_t)\overline{X}_{1t}'\right)'\lambda_* \\ \sum_{t=1}^{T}\exp\left(\lambda_*' q(V_t)\left(\overline{Y}_t - \overline{X}_{1t}'\theta_*\right)\right)q(V_t)\left(\overline{Y}_t - \overline{X}_{1t}'\theta_*\right)\end{bmatrix},$$

see for example Doukhan (1994). Finally E4(i) is a standard smoothness assumption on the kernel function K, whereas E4(ii) assumes undersmoothing, which again is typical in the semiparametric estimation literature.

Proposition 2 *Under E1-E4, Theorem 1 holds with $H_g(\phi_*, h_*)$ given in (5.2) and $\Omega_g(\phi_*, h_*)$ given in (5.3).*

In order to consider the uniform results of Section 4.2, let $\{Z_{Tt}\}$ denote a sequence of triangular arrays α mixing random vectors,

$$\begin{aligned}
g_{Tt}(\theta,h) &= q(V_{1Tt})\left(Y_{Tt} - E(Y_{Tt}|X_{2Tt}) - (X_{1Tt} - E(X_{1Tt}|X_{2Tt}))'\theta\right) \quad (5.4)\\
f_{Tt}(\beta,l) &= q(V_{2Tt})\left(Y_{Tt} - E(Y_{Tt}|X_{4Tt}) - (X_{3Tt} - E(X_{3Tt}|X_{4Tt}))'\beta\right)
\end{aligned}$$

denote the corresponding profile moment indicators for the two competing semiparametric models, and let

$$\left[\widehat{\theta}_T', \widehat{\lambda}_T'\right]' = \arg\max_{\theta\in\Theta}\arg\min_{\lambda\in\Lambda(\Theta)} \frac{1}{T}\sum_{t=1}^{T}\exp\left(\lambda' g_{Tt}\left(\theta, \widehat{h}\right)\right),$$
$$\left[\widehat{\beta}_T', \widehat{\gamma}_T'\right]' = \arg\max_{\beta\in B}\arg\min_{\gamma\in\Gamma(B)} \frac{1}{T}\sum_{t=1}^{T}\exp\left(\gamma' f_{Tt}\left(\beta, \widehat{l}\right)\right)$$

where $\widehat{h} = \left[\widehat{E}\left(Y_{Tt}|X_{2Tt}\right), \widehat{E}\left(X_{1Tt}|X_{2Tt}\right)\right]'$ and $\widehat{l} = \left[\widehat{E}\left(Y_{Tt}|X_{4Tt}\right), \widehat{E}\left(X_{3Tt}|X_{4Tt}\right)\right]'$. We slightly strengthen E1-E4 to account for the triangular array structure of $\{Z_{Tt}\}$ and assume that:

E5 (i) The triangular array sequence $\{Z_{Tt}, t = 0, \pm1, \pm2, ..., T \geq 1\}$ with with $Z_{Tt} = [Y_{Tt}, X_{Tt}', V_{Tt}']'$, $X_{Tt}' = [X_{1Tt}', X_{2Tt}, X_{3Tt}', X_{4Tt}]$ and $V_{Tt} = [V_{1Tt}', V_{2Tt}']'$ is α mixing with the T-th mixing coefficient $\alpha_T(k)$ satisfying $\sup_T \sum_{k=1}^{\infty}(k+1)^2 \alpha_T(k)^{\frac{\delta}{4+\delta}} < \infty$ for some $\delta > 0$, (ii) the marginal densities $f_{X_{2T}}$ and $f_{X_{4T}}$ of X_{2Tt} and of X_{4Tt} are twice continuously differentiable on $\mathcal{X}_2$ and $\mathcal{X}_4$, $\inf_{X_{2Tt}\in\mathcal{X}_2}|f_{X_{2T}}| > 0$ and $\inf_{X_{4Tt}\in\mathcal{X}_4}|f_{X_{4T}}| > 0$ for all t and T, and $\mathcal{X}_2$ and $\mathcal{X}_4$ are compact sets, (iii) $rank\left(\lim_{T\to\infty} E\left(q\left(V_{1Tt}\right)\overline{X}_{1Tt}'\right)/T\right) = k$ and $rank\left(\lim_{T\to\infty} E\left(q\left(V_{2Tt}\right)\overline{X}_{3Tt}'\right)/T\right) = b$, (iv) the parameter spaces Θ, B, $\Lambda(\Theta)$ and $\Gamma(B)$ are compact subsets of $\mathbb{R}^k$, $\mathbb{R}^l$, $\mathbb{R}^b$ and $\mathbb{R}^s$ respectively, and $\mathcal{H} = \mathcal{W}_{q_1,k_1}(h) \leq C_1$, $\mathcal{L} = \mathcal{W}_{q_2,k_2}(l) \leq C_2$ for some finite $C_1, C_2 > 0$, (v) $\theta_* \in int(\Theta)$, $\lambda_* \in int(\Lambda(\Theta))$, $h_* \in \mathcal{H}$, and $\beta_* \in int(B)$, $\gamma_* \in int(\Gamma(B))$, $l_* \in \mathcal{L}$,

E6 (i) E2 holds for $g_{Tt}(\theta, h)$ and $f_{Tt}(\beta, l)$ defined in (5.4) for all t and T with $B_{\phi,h}(Z_{Tt})$ and $B_{\psi,l}(Z_{Tt})$ satisfying $\lim_{T\to\infty}\sum_{t=1}^{T} E\left(B_{\phi,h}(Z_{Tt})\right)^{2+\delta}/T < \infty$ and $\lim_{T\to\infty}\sum_{t=1}^{T} E\left(B_{\psi,l}(Z_{Tt})\right)^{2+\delta}/T < \infty$, (ii) $E\left|\exp\left(\lambda_*' g_{Tt}(\theta_*, h)\right)\right|^{2(2+\delta)} < \infty$ and $E\left|\exp\left(\gamma_*' f_{Tt}(\beta_*, l)\right)\right|^{2(2+\delta)} < \infty$, $E\left\|\partial\exp\left(\lambda_*' g_{Tt}(\theta_*, h)\right)/\partial\phi\right\|^{2(2+\delta)} < \infty$ and $E\left\|\partial\exp\left(\gamma_*' f_{Tt}(\beta_*, l)\right)/\partial\psi\right\|^{2(2+\delta)} < \infty$ for all $h \in \mathcal{H}_\varepsilon, l \in \mathcal{L}_\varepsilon$, t and T, (iii) there exists a real valued function $b(Z_{Tt})$ such that $E\left(|b(Z_{Tt})|^{2(2+\delta)}\right)^{1/2} < \infty$ for all t and T,

E7 (i)

$$E\left|\exp\left(\lambda_*' q\left(V_{1Tt}\right)\left(\overline{Y}_{Tt} - \overline{X}_{1Tt}'\theta_*\right)\right) - \exp\left(\gamma_*' q\left(V_{2Tt}\right)\left(\overline{Y}_{Tt} - \overline{X}_{3Tt}'\beta_*\right)\right)\right|^{2+\delta} < \infty,$$
$$E\left\|\frac{\partial\exp\left(\lambda_*' q\left(V_{1Tt}\right)\left(\overline{Y}_{Tt} - \overline{X}_{1Tt}'\theta_*\right)\right)}{\partial\phi} - \frac{\partial\exp\left(\gamma_*' q\left(V_{2Tt}\right)\left(\overline{Y}_{Tt} - \overline{X}_{3Tt}'\beta_*\right)\right)}{\partial\psi}\right\|^{2+\delta} < \infty,$$

for all t and T, (ii)

$$\lim_{T\to\infty} Var\left[\begin{array}{c}\sum_{t=1}^{T}\left(\exp\left(\lambda_*' q\left(V_{1Tt}\right)\left(\overline{Y}_{Tt} - \overline{X}_{1Tt}'\theta_*\right)\right) - \exp\left(\gamma_*' q\left(V_{2Tt}\right)\left(\overline{Y}_{Tt} - \overline{X}_{3Tt}'\beta_*\right)\right)\right) \\ \frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{\partial\left(\exp\left(\lambda_*' q(V_{1Tt})\left(\overline{Y}_{Tt} - \overline{X}_{1Tt}'\theta_*\right)\right) - \exp\left(\gamma_*' q(V_{2Tt})\left(\overline{Y}_{Tt} - \overline{X}_{3Tt}'\beta_*\right)\right)\right)}{\partial[\phi', \psi']'}\end{array}\right]$$

is positive definite, (iii)

$$\begin{aligned}
\left\| \frac{\partial^2 \exp\left(\lambda' g_{Tt}\left(\theta, h\right)\right)}{\left(\partial\phi\right)^{\otimes 2}} - \frac{\partial \exp\left(\lambda_*' g_{Tt}\left(\theta_*, h_*\right)\right)}{\left(\partial\phi\right)^{\otimes 2}} \right\| &\leq B_g\left(Z_{Tt}\right)\left(\left\|\phi - \phi_*\right\| + \left\|h - h_*\right\|_{\mathcal{H}_\varepsilon}\right), \\
\left\| \frac{\partial^2 \exp\left(\gamma_*' f_{Tt}\left(\beta, l\right)\right)}{\left(\partial\psi\right)^{\otimes 2}} - \frac{\partial \exp\left(\gamma_*' f_{Tt}\left(\beta_*, l_*\right)\right)}{\left(\partial\psi\right)^{\otimes 2}} \right\| &\leq B_f\left(Z_{Tt}\right)\left(\left\|\psi - \psi_*\right\| + \left\|l - l_*\right\|_{\mathcal{L}_\varepsilon}\right)
\end{aligned}$$

with $\lim_{T\to\infty} \sum_{t=1}^{T} EB_g\left(Z_{Tt}\right)^{2+\delta}/T < \infty$ and $\lim_{T\to\infty} \sum_{t=1}^{T} EB_f\left(Z_{Tt}\right)^{2+\delta}/T < \infty$, (iv) $E\sup_{\lambda\in\Lambda_*(\Theta_*),\theta\in\Theta_*,h\in\mathcal{H}_\epsilon} \left\|\partial^2 \exp\left(\lambda' q\left(V_{1Tt}\right)\left(\overline{Y}_{Tt} - \overline{X}_{1Tt}'\theta\right)\right)/\left(\partial\phi\right)^{\otimes 2}\right\|^{2+\delta} < \infty$ for all t and T, where $\Lambda_*\left(\Theta_*\right)$ and Θ_* are neighborhoods of λ_* and θ_* and $\mathcal{H}_\epsilon = \left\{h : \left\|h - h_\bullet\right\|_{\mathcal{H}} \leq \epsilon\right\}$ for some $\epsilon > 0$, $E\sup_{\gamma\in\Gamma_*(B_*),\beta\in B_*,l\in\mathcal{L}_\epsilon} \left\|\partial^2 \exp\left(\gamma' q\left(V_{2Tt}\right)\left(\overline{Y}_{Tt} - \overline{X}_{3Tt}'\beta\right)\right)/\left(\partial\psi\right)^{\otimes 2}\right\|^{2+\delta} < \infty$ for all t and T, where $\Gamma_*\left(B_*\right)$ and B_* are neighborhoods of γ_* and β_* and $\mathcal{L}_\epsilon = \left\{l : \left\|l - l_\bullet\right\|_{\mathcal{L}} \leq \epsilon\right\}$ for some $\epsilon > 0$, (v)

$$\begin{aligned}
\lim_{T\to\infty} E\left[T^{-1} \sum_{t=1}^{T} \frac{\partial^2 \exp\left(\lambda_*' q\left(V_{1Tt}\right)\left(\overline{Y}_{Tt} - \overline{X}_{1Tt}'\theta_*\right)\right)}{\left(\partial\phi\right)^{\otimes 2}}\right] &= H_g\left(\phi_*, h_*\right), \\
\lim_{T\to\infty} E\left[T^{-1} \sum_{t=1}^{T} \frac{\partial^2 \exp\left(\gamma_*' q\left(V_{2Tt}\right)\left(\overline{Y}_{Tt} - \overline{X}_{3Tt}'\beta_*\right)\right)}{\left(\partial\psi\right)^{\otimes 2}}\right] &= H_f\left(\psi_*, l_*\right)
\end{aligned}$$

are nonsingular, (vi)

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E\sup_{h\in\mathcal{H}} \left\| \exp\left(\lambda_*' q\left(V_{1Tt}\right)\left(\overline{Y}_{Tt} - \overline{X}_{1Tt}'\theta_*\right)\right) \begin{bmatrix} 1 \\ \theta_* \end{bmatrix}^{\otimes 2} v_{1j}^{k_1} \right\| &< \infty, \\
\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E\sup_{l\in\mathcal{L}} \left\| \exp\left(\gamma_*' q\left(V_{2Tt}\right)\left(\overline{Y}_{Tt} - \overline{X}_{3Tt}'\beta_*\right)\right) \begin{bmatrix} 1 \\ \beta_* \end{bmatrix}^{\otimes 2} v_{2j}^{k_1} \right\| &< \infty,
\end{aligned}$$

where, for $k_1 = 0, 1$, $v_{11}^{k_1} = 1$, $v_{1j}^{k_1} = \theta_{*j-1}$ for $j = 2, ...k+1$, and similarly for $v_{2j}^{k_1}$ with β_{*j-1} replacing θ_{*j-1},

E8 (i) the kernel functions $K_1 : \mathcal{X}_2 \to \mathbb{R}$ and $K_2 : \mathcal{X}_4 \to \mathbb{R}$ are symmetric with compact support, and $\sup_{X_{2Tt}\in\mathcal{X}_2} \left|d^{j_1} K_1\right|^{q_1} < \infty$ $j_1 = 0, ..., k_1$, $\sup_{X_{4Tt}\in\mathcal{X}_2} \left|d^{j_2} K_2\right|^{q_2} < \infty$ $j_2 = 0, ..., k_2$ for all T, (ii) the bandwidths b_j are such that $b_j \to 0$, $Tb_j \to \infty$ and $T^{1/2} b_j^4 \to 0$ $(j = 1, 2)$ as $T \to \infty$.

Proposition 3 *Under E5-E9, the conclusion of Theorem 5 holds for*

$$D_T^m\left(c\right) = \frac{1}{T^{1/2}} \sum_{t=1}^{T} \frac{\left(\exp\left(\widehat{\lambda}' q\left(V_{1Tt}\right)\left(\widehat{Y}_{1Tt} - \widehat{X}_{1Tt}'\widehat{\theta}\right)\right) - \right.}{\left(\widehat{\sigma}^2\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right) + ctr\left(\widehat{V}^2\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)\right)/T\right)^{1/2}}$$

$$\left. \exp\left(\widehat{\gamma}' q\left(V_{2Tt}\right)\left(\widehat{Y}_{2Tt} - \widehat{X}_{3Tt}'\widehat{\beta}\right)\right) + tr\left(\widehat{V}\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)\right)/2T^{1/2}\right)$$

with $\widehat{Y}_{1Tt} = Y_{Tt} - \widehat{E}(Y_{Tt}|X_{2Tt})$, $\widehat{Y}_{2Tt} = Y_{Tt} - \widehat{E}(Y_{Tt}|X_{4Tt})$, $\widehat{X}_{1Tt} = X_{1Tt} - \widehat{E}(X_{1Tt}|X_{2Tt})$, $\widehat{X}_{2Tt} = X_{3Tt} - \widehat{E}(X_{3Tt}|X_{4Tt})$, $\widehat{\sigma}^2\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)$ *is the same blocking estimator as given in* (4.11) *with*

$$\begin{aligned} b_{Ti}(\phi, h, \psi, l) &= \frac{1}{M^{1/2}} \sum_{j=1}^{M} \left(\exp\left(\lambda' q\left(V_{1T(i-1)+j}\right) \left(\widehat{Y}_{1T(i-1)+j} - \widehat{X}'_{1T(i-1)+j} \widehat{\theta} \right) \right) - \right. \\ &\quad \left. \exp\left(\gamma' q\left(V_{1T(i-1)+j}\right) \left(\widehat{Y}_{2T(i-1)+j} - \widehat{X}'_{3T(i-1)+j} \widehat{\beta} \right) \right) \right) \end{aligned}$$

and $\widehat{V}\left(\widehat{\phi}, \widehat{h}, \widehat{\psi}, \widehat{l}\right)$ *is based on the sample analog of* (5.2).

5.2 Monte Carlo results

We now investigate the finite sample properties of the proposed model selection tests by considering two examples. In the first one, we consider

$$Y_t = 1 + \tau \sum_{j=1}^{2} X_{1jt} + (1-\tau) \sum_{j=1}^{2} X_{2jt} + h_0(X_{3t}, X_{4t}) + \varepsilon_t, \ \ \tau \in \mathbb{R}, \tag{5.5}$$

where $X_{12t} = \alpha_1 W_{1t} + u_{1t}$ and $X_{21t} = \alpha_2 W_{2t} + u_{2t}$ with u_{jt} $(j = 1, 2)$ correlated with ε_t, and two competing misspecified semiparametric moment conditions models

$$\begin{aligned} M_1 &: g_t(\theta, h) = V_{1t}(Y_t - \theta_1 - \theta_2 X_{12t} - h(X_{3t})), \\ M_2 &: f_t(\beta, h) = V_{2t}(Y_t - \beta_1 - \beta_2 X_{21t} - h(X_{4t})), \end{aligned} \tag{5.6}$$

where $V_{1t} = [1, X_{11t}, W_{1t}]'$ and $V_{2t} = [1, X_{22t}, W_{2t}]'$ are two sets of instruments, which are clearly misspecified and non nested. We specify (5.5) and (5.6) as follows: the exogenous regressors X_{jjt} $(j = 1, 2)$ in (5.5) are $N(0, 1)$, the additional variables W_{jt} $(j = 1, 2)$ are either $N(0, 4)$ or $W_{jt} = \alpha_3 W_{jt-1} + u_{3t}$, the unknown function h_0 is $h_0(X_{3t}, X_{4t}) = \exp(-(X_{3t} + X_{4t})/2)$ with $X_{jt} \sim U(0, 2)$ $(j = 3, 4)$, $h(X_{jt}) = \exp(-X_{jt}/2)$ $(j = 3, 4)$, the unobservable errors ε_t, u_{jt} $(j = 1, 2)$ and u_{3t} are

$$\begin{bmatrix} \varepsilon_t \\ u_{1t} \\ u_{2t} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{\varepsilon u_1} & \rho_{\varepsilon u_2} \\ \rho_{\varepsilon u_1} & 1 & 0 \\ \rho_{\varepsilon u_2} & 0 & 1 \end{bmatrix} \right),$$

and $u_{3t} \sim N(0, 1)$ independent of both ε_t, u_{jt}. In the simulations, we consider two values of τ, $\tau = 0.5$ and $\tau = 1$, corresponding, respectively, to the case where we cannot discriminate between M_1 and M_2 (that is, the null hypothesis (4.1) is true) and to the case where discrimination is possible and M_1 should be chosen. Estimation of the parameters in (5.6) is carried out using the profile moment indicator $g_t\left(\theta, \widehat{h}\right) = V_{1t}\left(\widehat{Y}_t - \widehat{X}'_{1t}\theta\right)$, where $\widehat{Y}_t = Y_t - \widehat{E}(Y_t|X_{2t})$,

$\widehat{X}'_{1t} = X'_{1t} - \widehat{E}(X_{1t}|X_{2t})$ with the kernel function K specified as $K(u) = (3/4)(1-u^2)$ for $|u| \leq 1$ (that is the univariate Epanechnikov kernel), and similarly for $f_t\left(\beta, \widehat{l}\right)$. We consider three statistics: D_T, $D_T^m(c)$ and V_T, which corresponds, respectively, to the standardized D_T statistic of Theorem 2, the modified Vuong standardized statistic $D_T^m(c)$ of Theorem 5, and Vuong's (1989) two-step procedure based on the scale adjusted statistic $\widehat{\sigma}_s^2$ (4.5) used as a pretest and then the standardized D_T statistic, to test whether discrimination is possible between M_1 and M_2. Tables 1-2 report the rejection probabilities (size) of the three statistics for the null hypothesis $H_0 : D = 0$ with $\gamma = 1/2$, that is the models cannot be discriminated, using 5000 replications and two sample sizes $T = 100$ and $T = 400$. The rejection probabilities are calculated at a 5% nominal level. The simulations are based on two values of the parameters α_j and $\rho_{\varepsilon u_j}$ $(j = 1, 2)$, $\alpha_1 = \alpha_2$ either 0.2 or 0.8, corresponding to low and high persistence, and $\rho_{\varepsilon u_1} = \rho_{\varepsilon u_2}$ either 0.3 and 0.7, corresponding to a relatively weak and strong instruments, whereas $\alpha_3 = 0.5$.

Tables 1 and 2 approx. here

Tables 1 and 2 show that both the D_T and $D_T^m(c)$ statistics are slightly oversized, whereas the V_T statistic is slightly undersized. Among the D_T and $D_T^m(c)$ statistics, the latter is characterized by the smallest size distortion, across the different degrees of persistency and the strength of the instruments. Tables 3 and 4 report the rejection probabilities (power) of the three statistics for the null hypothesis $H_0 : D = 0$ with $\gamma = 1$, that is the models can be discriminated (and M2 should be preferred), using 5000 replications, two sample sizes $T = 100$ and $T = 400$ and the same specifications of the parameters α_j,$\rho_{\varepsilon u_j}$ $(j = 1, 2)$ and α_3 as those used in Tables 1 and 2.

Tables 3 and 4 approx. here

Tables 3 and 4 confirm the results of Tables 1 and 2, in the sense that they show that all the test statistics perform reasonably well, correctly rejecting the null hypothesis more than 90% (or above) of times across the different Monte Carlo designs. In this case, the statistic characterized by the best finite sample performance is D_T, which is not surprising since the models are nonoverlapping. It is also important to note that both the D_T and $D_T^m(c)$ statistics are slightly oversized, hence their power is bound to be larger than that of V_T, since the latter is undersized. Figure 1 shows the finite sample (size unadjusted) power curves of the four statistics for the case $\alpha_1 = \alpha_2 = 0.2$, $\rho_{\varepsilon u_1} = \rho_{\varepsilon u_2} = 0.7$ and W_{jt} i.i.d. with the two sample sizes $T = 100$ and $T = 400$. As expected, D_T shows a (slightly) higher power curve as a function of τ, but it is worth pointing out that V_T performs rather well, particularly in view of the fact that it is the only one undersized.

Figure 1 approx. here

In the second example, we extend the same example used by both Shi (2015) and Schennach

& Wilhelm (2017); the model is

$$Y_t = 1 + \tau \left(\frac{1}{J} \sum_{j=1}^{J} X_{1jt} + h_{10}(X_{2t}) \right) + \tau \left(X_{1J+1t} + h_{20}(X_{3t}) \right) + \varepsilon_t, \ \tau \in \mathbb{R},$$

where $X_{jt} \sim N(0,1)$ $(j = 1, ..., J+3)$, $h_{10}(X_{2t}) = E(X_{1J+2t}|X_{2t}) = \sin(2\pi X_{2t})$, $h_{20}(X_{4t}) = E(X_{1J+3t}|X_{3t}) = \cos(\pi X_{3t})$ and $\varepsilon_t = \alpha_4 \varepsilon_{t-1} + u_t$ with $u_t \sim N(0,1)$. The competing misspecified semiparametric models are

$$\begin{aligned} M_3 &: \ g_t(\theta, h) = [1, X'_{1t}]' \left(Y_t - \theta_1 - \sum_{j=1}^{J} \theta_{2j} X_{1jt} - h_2(X_{3t}) \right), \\ M_4 &: \ f_t(\beta, h) = \left[1, X'_{1J+1t}\right]' \left(Y_t - \beta_1 - \beta_2 X_{1J+1t} - h_1(X_{2t}) \right), \end{aligned} \tag{5.7}$$

$X'_{1t} = [X_{11t}, ... X_{1Jt}]$, which, for any value $\tau \neq 0$, have the same distance to the true model, but are both misspecified. We estimate both $h_1(X_{2t})$ and $h_2(X_{3t})$ with the same Epanechnikov kernel used in the previous example, and, as in the previous example, we consider the same three statistics D_T, $D_T^m(c)$ and V_T. Figure 2 below shows the rejection probabilities (size) calculated at the 5% nominal level for $J = 2$ (top figure) and $J = 9$ (bottom figure) using 5000 replications for different values of τ with two sample sizes $T = 100$ and $T = 400$.

Figure 2 approx. here

Figure 2 shows that, as mentioned in the Introduction, the difference in the dimension of the two competing models has some bearings on the finite sample performances of all of the three statistics considered, however, among them, V_T is clearly the most affected in the sense of being rather undersized when the difference between the two competing models is substantial ($J = 9$), and this has some obvious implication for its power, as documented in Shi (2015) and Schennach & Wilhelm (2017). Note also that the modified Vuong statistic $D_T(c)$ seems to be the most robust in terms of size control, which confirms the theoretical results of Section 4.2.

6 Empirical application

As an application of the estimation and model selection procedures of this paper, we consider two alternative specifications of the SDF, which is a positive random variable representing time discount and risk adjustment in the pricing of future risky assets. In Fama & French's (1993) three factor model, the SDF is linearly related to three observed risk factors: the market excess returns, the performance of small firms compared to big firms and the performance of high to book value companies compared to low to book ones. Let $R_t = [R_{1t}, ..., R_{Nt}]'$ denote a vector of

N gross returns between t and $t+1$ and let m_t denote the (admissible) SDF that satisfies the conditional no arbitrage condition

$$E\left(m_{t+1}R_{t+1} - 1_N|U_t\right) = 0_N, \tag{6.1}$$

where 1_N and 0_N are N dimensional vectors of ones and zeroes and U_t is a possibly d_U dimensional random state vector representing the information available at time t. We consider two specification of m_t: a fully parametric one, that is $m_t = m_t(\theta) = \theta_1 + \theta_2 F_{1t} + \theta_3 F_{2t} + \theta_4 F_{3t}$, where F_{jt} $(j = 1, 2, 3)$ are the three observed risk factors, and a semiparametric one $m_t = m_t(\theta, h) = \theta_1 + \theta_2 F_{1t} + \theta_3 F_{2t} + h(U_t)$, where the third factor is replaced by an unknown function of a specific state variable described below[6]. Under correct specification of m_t, (6.1) holds for a unique $\theta_0 = [\theta_{10}, \theta_{20}, \theta_{30}, \theta_{40}]'$ or a unique $\theta_0 = [\theta_{10}, \theta_{20}, \theta_{30}]'$ and h_0. Let $g_t(\theta) := [1, U_t']' \otimes (R_{t+1}(\theta_1 + \theta_2 F_{1t+1} + \theta_3 F_{2t+1} + \theta_4 F_{3t+1}) - 1_N)$ and $f_t(\beta, h) = \overline{R}_{t+1}\beta'\overline{F}_{t+1} - 1_N$ where $\overline{R}_{t+1} = R_{t+1} - E(R_{t+1}|U_t)$, $\overline{F}_{t+1} = F_{t+1} - E(F_{t+1}|U_t)$ and $F_t = [F_{1t}, F_{2t}]'$. The data used in the estimation are the monthly returns on 10 size-sorted portfolios for US equities from Kenneth French's data library[7] as risky assets R_t; the excess returns are computed over the one-month Treasury bill yield obtained from the Center for Research in Securities Prices (CRSP). The state variable U_t is chosen to be the BAA corporate bond yield relative to the constant maturity ten-year Treasury yield. The latter serves as a proxy for the default risk, and it was used in the consumption-based CAPM model of Jagannathan & Wang (1996). The corporate bond spread is obtained from the Federal Reserve Economic Data. The sample period is $1964:01-2018:12$ minus the $2008:01-2009:04$ financial crisis, for a total of $T = 641$ observations.

To estimate the two different specifications of the SDF, we first use blockwise ET, as described in Section 4, and test for the correct specification of both models using the ET based test statistics $2T\log\left(\sum_{i=1}^{Q} b_i\left(\widehat{\lambda}, \widehat{\theta}\right)/Q\right)$ for $g_t(\theta)$ and $2T\log\left(\sum_{i=1}^{Q} b_i\left(\widehat{\gamma}, \widehat{\beta}, \widehat{h}\right)/Q\right)$ for $f_t(\beta, h)$, where $\widehat{h} = \left[\widehat{E}(R_{t+1}|U_t), \widehat{E}(F_{t+1}|U_t)\right]'$ are the same kernel estimators defined in the previous section. Under the null hypothesis of correct specification, both statistics converge to a chi-squared random variable with degrees of freedom depending on the degree of overidentification, see for example Kitamura & Stutzer (1997). Both test statistics reject the null hypothesis of correct specification at the 5% significance level, with p-values of 0.003 and 0.012, respectively, indicating therefore that both models are misspecified. We re-estimated the models using the method of this paper and then tested whether discrimination is possible. Table 5 reports, respectively, the estimated parameters with relative standard errors and the values of the two

[6] It should be noted that the three factor model has no real theoretical foundation as such; rather it was based as a response to the well documented empirical shortcomings of the traditional CAPM model, see for example Fama & French (2004).

[7] Available at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

statistics $D_T^m(c)$ and V_T with associated p-values.

Table 5 approx. here

Table 5 indicates that model comparison is possible, albeit for the V_T not at the standard 0.05 significance level. The sample value of the V_T statistic is 1.646 with a p-value based on a two sided alternative of 0.099. Using a one-sided alternative, that is that the empirical value of the KL divergence for $g_t(\theta)$ is statistically significantly larger than that of the $f_t(\beta, l)$ specification (that is the latter specification is better) the p-value of the V_T is 0.049, which seems to indicate that the semiparametric specification should be preferred. The sample value of the $D_T^m(c)$ statistic is 2.212 with p-values 0.033 (two-sided alternative) and 0.016 (one-sided alternative), which strengthens the finding of the V_T statistic that model selection is possible and the semiparametric specification should be chosen.

7 Conclusion

In this paper, we propose a general approach that can be used to estimate and, more importantly, discriminate between possibly misspecified semiparametric moment conditions models with weakly dependent observations. The proposed tests are extensions of those proposed by Vuong (1989) and Shi (2015) in the context of parametric likelihood models, and have a similar information theoretic interpretation through the connection between ET and KL divergence. The results are rather general as they are obtained under a set of high level conditions that can be verified under typically mild primitive assumptions on the infinite dimensional parameter and mixing conditions, as illustrated in the example of Section 5. A simulation study shows that the proposed test statistics have competitive finite sample properties, and that both the naive and the uniform extensions to Vuong's (1989) model selection theory can be used in the context of possibly misspecified semiparametric models. To further illustrate the applicability of the proposed estimator and test statistics, we show that model selection is possible between two different specifications of Fama & French's (1993) three factor model.

References

Ai, C. & Chen, X. (2007), 'Estimation of possibly misspecified semiparametric conditional moment restrictions models with different conditioning variables', *Journal of Econometrics* **141**, 5–43.

Akaike, H. (1973), Information theory and an extension of the likelihood principle, *in* B. Petrov & F. Csaki, eds, 'Proceedings of the Second International Symposium of Information Theory'.

Andrews, D. (1991), 'Heteroskedasticy and autocorrelation consistent covariance matrix estimation', *Econometrica* **59**, 817–858.

Andrews, D. (1994), 'Asymptotics for semiparametric econometric models via stochastic equicontinuity', *Econometrica* **62**, 43–72.

Andrews, D. & Pollard, D. (1994), 'An introduction to functional central limit theorems for dependent stochastic processes', *International Statistical Review* **62**, 119–132.

Arcones, M. & Yu, B. (1994), 'Central limit theorems for empirical and U processes of stationary mixing sequences', *Journal of Theoretical Probability* **7**, 47–71.

Bradley, R. (2005), 'Basic properties of strong mixing conditions. A survey and some open questions', *Probability Surveys* **2**, 107–144.

Bravo, F. (2020), 'Two-step combined nonparametric likelihood estimation of misspecified semiparametric models', *Journal of Nonparametric Statistics* **32**, 769–792.

Bravo, F., Chu, B. & Jacho-Chavez, D. (2017), 'Semiparametric estimation of moment conditions models with weakly dependent data', *Journal of Nonparametric Statistics* **29**, 108–136.

Breitung, J. (2008), 'Simple tests of the moving average hypothesis', *Journal of Time Series Analysis* **15**, 357–370.

Carrasco, M. & Chen, X. (2002), 'Mixing and moment properties of various GARCH and stochastic volatility models', *Econometric Theory* **18**, 17–39.

Carroll, R. & van Keilegom, I. (2007), 'Backfitting versus profiling in general criterion functions', *Statistica Sinica* **17**, 797–816.

Chen, X., Hong, H. & Shum, M. (2007), 'Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models', *Journal of Econometrics* **141**, 109–140.

Chen, X. & Liao, Z. (2015), 'Sieve semiparametric two-step GMM under weak dependence', *Journal of Econometrics* **189**, 163–186.

Chen, X., Liao, Z. & Sun, Y. (2014), 'Sieve inference on possibly misspecified semi-nonparametric time series models', *Journal of Econometrics* **178**, 639–658.

Christoffersen, P., Hahn, J. & Inoue, A. (2001), 'Testing and comparing value-at-risk measures', *Journal of Empirical Finance* **8**, 325–342.

Dahlhaus, R. & Wefelmeyer, W. (1996), 'Asymptotically optimal estimation in misspecified time series models', *Annals of Statistics* **24**, 952–972.

Doukhan, P. (1994), *Mixing: Properties and Examples*, Vol. 85, New York: Springer and Verlag. Lecture Notes in Statistics.

Doukhan, P., Massart, P. & Rio, E. (1995), 'Invariance principles for absolutely regular empirical processes', *Annales de l'Institute Henri Poincare Probababilites et Statististiques* **31**, 393–427.

Fama, E. & French, K. (1993), 'Common risks factors in the returns on stock and bonds', *Journal of Financial Economics* **33**, 3–56.

Fama, E. & French, K. (2004), 'The capital asset pricing model: theory and evidence', *Journal of Economic Perspectives* **18**, 25–46.

Ghosh, A., Julliard, M. & Taylor, A. (2017), 'What is the consumption-CAPM missing? An information-theoretic framework for the analysis of asset pricing models', *The Review of Financial Studies* **30**, 442–504.

Gospodinov, N., Kan, R. & Robotti, C. (2013), 'Chi-squared tests for evaluation and comparison of asset pricing models', *Journal of Econometrics* **173**, 108–125.

Gospodinov, N., Kan, R. & Robotti, C. (2014), 'Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors', *Review of Financial Studies* **27**, 2139–2170.

Hall, A. & Inoue, A. (2003), 'The large sample behaviour of the generalized method of moments estimator in misspecified models', *Journal of Econometrics* **114**, 361–394.

Hall, A. & Pelletier, D. (2011), 'On non-nested testing in models estimated by generalized methods of moments', *Econometric Theory* **27**, 443–456.

Hansen, L. & Jagannathan, R. (1997), 'Assessing specification errors in stochastic discount factors models', *Journal of Finance* **52**, 557–590.

Hansen, L. & Singleton, K. (1982), 'Generalized instrumental variables estimation of nonlinear rational expectations models', *Econometrica* **50**, 1269–1286.

Hsu, Y. & Shi, X. (2017), 'Model selection test for conditional moment inequalities models', *Econometrics Journal* **20**, 52–85.

Jagannathan, R. & Wang, Z. (1996), 'The conditional CAPM and cross-section of expected returns', *Journal of Finance* **51**, 3–53.

Kitamura, Y. (2000), Comparing misspecified dynamic econometric models using nonparametric likelihood, Technical report, University of Pennsylvania.

Kitamura, Y. & Stutzer, M. (1997), 'An information theoretic alternative to generalized method of moments estimation', *Econometrica* **65**, 861–874.

Kitamura, Y. & Stutzer, M. (2002), 'Connections between entropic and linear projections in asset pricing estimation', *Journal of Econometrics* **107**, 159–174.

Kunitomo, N. & Yamamoto, T. (1985), 'Properties of predictors in missspecified time series models', *Journal of the American Statistical Association* **80**, 941–950.

Li, T. (2009), 'Simulation based selection of competing structural econometric models', *Journal of Econometrics* **148**, 114–123.

Liang, K. & Zeger, S. (1986), 'Longitudinal data analysis using generalised linear models', *Biometrika* **73**, 12–22.

Liao, Z. & Shi, X. (2020), 'A uniform vuong test for semi/nonparametric models', *Quantitative Economics* **11**, 983–1017.

Masry, E. (1996), 'Multivariate local polynomial regression for time series: Uniform strong consistency and rates', *Journal of Time Series Analysis* **17**, 571–599.

McElroy, T. (2016), 'Nonnested model comparisons for time series models', *Biometrika* **103**, 905–914.

Mohr, M. (2020), 'A weak convergence result for sequential empirical processes under weak dependence', *Stochastics* **92**, 140–164.

Morgan, J. P. (1996), Riskmetrics, Technical report, New York NY.

Newey, W. (1997), 'Convergence rates and asymptotic normality for series estimators', *Journal of Econometrics* **79**, 147–168.

Qu, A., Lindsay, B. G. & Li, B. (2000), 'Improving generalised estimating equations using quadratic inference functions', *Biometrika* **87**(4), 823–836.

Rao, J. & Scott, A. (1981), 'The analysis of categorical data from complex sampling surveys: Chi-squared tests for goodness of fit and independence in two-way tables', *Journal of the American Statistical Association* **76**, 221–230.

Rivers, D. & Vuong, Q. (2002), 'Model selection tests for nonlinear dynamic models', *Econometrics Journal* **5**, 1–39.

Schennach, S. (2007), ‘Point estimation with exponentially tilted empirical likelihood’, *Annals of Statistics* **35**, 634–672.

Schennach, S. & Wilhelm, D. (2017), ‘A simple parametric model selection test’, *Journal of the American Statistical Association* **112**, 1663–1674.

Shi, X. (2015), ‘A non-degenerate vuong test’, *Quantitative Economics* **6**, 85–121.

Smith, R. (2011), ‘GEL criteria for moment conditions models’, *Econometric Theory* **27**, 1192–1235.

Stutzer, M. (1995), ‘A Bayesian approach to diagnosis of asset pricing models’, *Journal of Econometrics* **68**, 367–397.

Stutzer, M. (1996), ‘A simple nonparametric approach to derivative security valuation’, *Journal of Finance* **51**, 1633–1652.

Van der Vaart, A. (1998), *Asymptotic statistics*, Cambridge University Press.

Van der Vaart, A. & Wellner, J. (1996), *Weak Convergence and Empirical Processes*, Springer, New York.

Volkonskii, V. & Rozanov, Y. (1959), ‘Some limit theorems for random functions I’, *Theory of Probability and its Applications* **4**, 1978–197.

Vuong, Q. (1989), ‘Likelihood ratio tests for model selection and non nested hypothesis’, *Econometrica* **57**, 307–333.

White, H. (1982), ‘Maximum likelihoodestimation of misspecified models’, *Econometrica* **50**, 1–25.

Yu, B. (1994), ‘Rates of convergence for empirical processes of stationary mixing sequences’, *Annals of Probability* **22**, 94–116.

8 Tables and figures

Table 1. Finite sample rejection percentages for (5.6) under $\tau = 1/2$ and i.i.d. W_{jt}

		D_T	$D_T^m(c)$	V_T
T	α^a		$\rho_{\varepsilon u}^b = 0.3$	
	0.2	0.061	0.055	0.047
100	0.8	0.067	0.057	0.045
	α^a		$\rho_{\varepsilon u}^b = 0.7$	
	0.2	0.062	0.055	0.044
100	0.8	0.068	0.061	0.042
T	α^a		$\rho_{\varepsilon u}^b = 0.3$	
	0.2	0.056	0.052	0.046
400	0.8	0.063	0.054	0.047
	α^a		$\rho_{\varepsilon u}^b = 0.7$	
	0.2	0.058	0.053	0.046
400	0.8	0.063	0.055	0.044

a α_1=α_2, b $\rho_{\varepsilon u_1}$=$\rho_{\varepsilon u_2}$

Table 2. Finite sample rejection percentages for (5.6) under $\tau = 1/2$ and dependent W_{jt}

		D_T	$D_T^m(c)$	V_T
T	α^a		$\rho_{\varepsilon u}^b = 0.3$	
	0.2	0.062	0.054	0.044
100	0.8	0.069	0.055	0.044
	α^a		$\rho_{\varepsilon u}^b = 0.7$	
	0.2	0.063	0.054	0.047
100	0.8	0.068	0.057	0.046
T	α^a		$\rho_{\varepsilon u}^b = 0.3$	
	0.2	0.058	0.054	0.043
400	0.8	0.060	0.055	0.042
	α^a		$\rho_{\varepsilon u}^b = 0.7$	
	0.2	0.057	0.052	0.045
400	0.8	0.061	0.054	0.043

a α_1=α_2, b $\rho_{\varepsilon u_1}$=$\rho_{\varepsilon u_2}$

Table 3. Finite sample rejection percentages for (5.6) under $\tau = 1$ and i.i.d. W_{jt}

		D_T	$D_T^m(c)$	V_T
T	α^a		$\rho_{\varepsilon u}^b = 0.3$	
	0.2	0.942	0.938	0.908
100	0.8	0.938	0.933	0.907
	α^a		$\rho_{\varepsilon u}^b = 0.7$	
	0.2	0.944	0.940	0.906
100	0.8	0.941	0.938	0.909
T	α^a		$\rho_{\varepsilon u}^b = 0.3$	
	0.2	0.941	0.945	0.921
400	0.8	0.945	0.942	0.919
	α^a		$\rho_{\varepsilon u}^b = 0.7$	
	0.2	0.946	0.944	0.919
400	0.8	0.944	0.940	0.917

a α_1=α_2, b $\rho_{\varepsilon u_1}$=$\rho_{\varepsilon u_2}$

Table 4. Finite sample rejection percentages for (5.6) under $\tau = 1$ and dependent W_{jt}

		D_T	$D_T^m(c)$	V_T
T	α^a		$\rho_{\varepsilon u}^b = 0.3$	
	0.2	0.940	0.937	0.911
100	0.8	0.938	0.936	0.909
	α^a		$\rho_{\varepsilon u}^b = 0.7$	
	0.2	0.930	0.948	0.918
100	0.8	0.934	0.947	0.920
T	α^a		$\rho_{\varepsilon u}^b = 0.3$	
	0.2	0.951	0.953	0.926
400	0.8	0.948	0.949	0.924
	α^a		$\rho_{\varepsilon u}^b = 0.7$	
	0.2	0.947	0.947	0.928
400	0.8	0.946	0.945	0.920

a α_1=α_2, b $\rho_{\varepsilon u_1}$=$\rho_{\varepsilon u_2}$

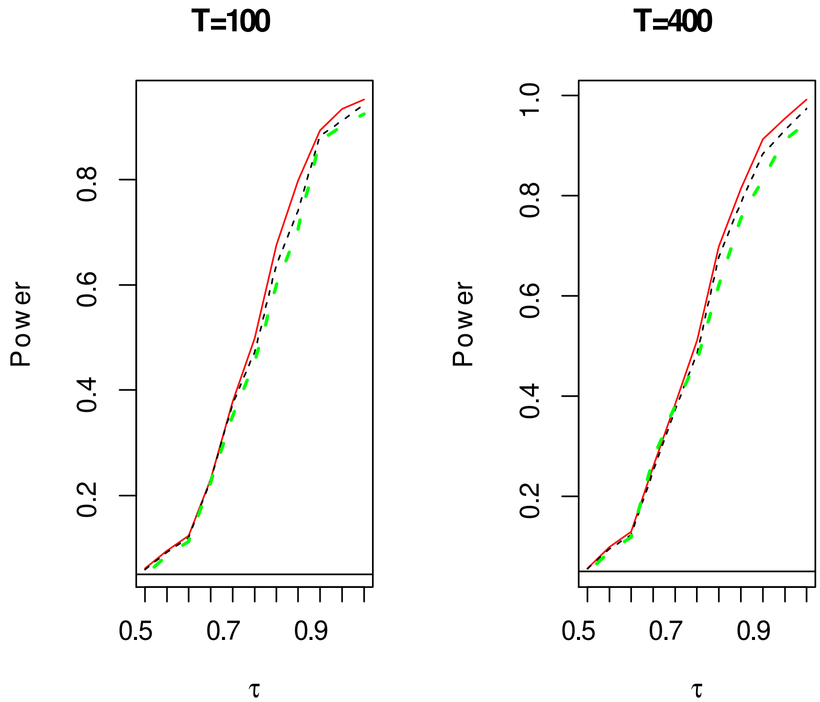

Figure 1: Finite sample power of the three statistics D_T (solid line), $D_T^m(c)$ (dashed line) and V_T (dot-dashed line). The horizontal line represents the 0.05 nominal level.

Table 5. Estimated coefficients, standard errors of the two competing models with the $D_T^m(c)$ and V_T statistics.

$\widehat{\theta}_1$	0.003 (0.015)^a	$\widehat{\beta}_1$	0.007 (0.006)^a
$\widehat{\theta}_2$	0.032 (0.011)^a	$\widehat{\beta}_2$	0.037 (0.011)^a
$\widehat{\theta}_3$	0.024 (0.014)^a	$\widehat{\beta}_3$	0.031 (0.012)^a
$\widehat{\theta}_4$	0.066 (0.038)^a	–	–
$D_T^m(c)$	2.121 (0.033, 0.016)^b		
V_T	1.646 (0.099, 0.049)^b		

a standard errors, *b* p-values (2-sided and 1-sided alternatives)

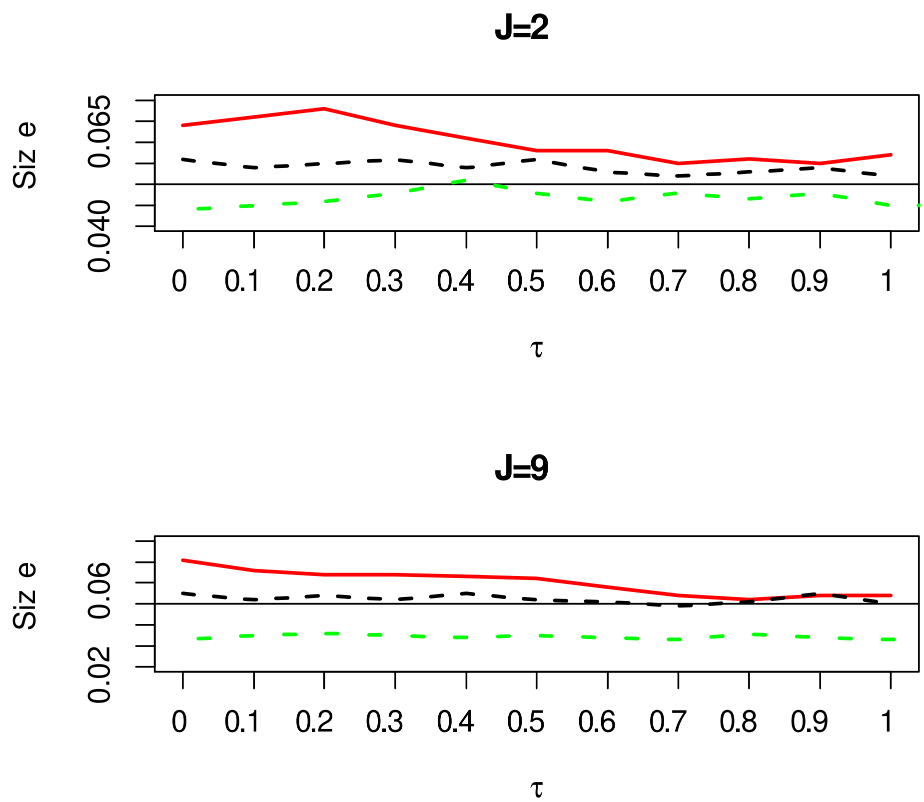

Figure 2: Finite sample size of the three statistics D_T (solid line), $D_T^m(c)$ (dashed line) and V_T (dot-dashed line). The horizontal line represents the 0.05 nominal level.