

WHICH MODEL TO MATCH?

Matteo BARIGOZZI^{1,3} Roxana HALBLEIB² David VEREDAS³

January 23, 2014

Abstract

The asymptotic efficiency of indirect estimation methods, such as the efficient method of moments and indirect inference, depends on the choice of the auxiliary model, which is somehow ad hoc and based on an educated guess. We introduce an Akaike-type class of information criteria that helps the user in this choice among nested and non-nested auxiliary models. A Monte Carlo and an empirical illustration show the usefulness of the criteria.

Keywords: Auxiliary model, efficient method of moments, indirect inference, information criteria.

¹London School of Economics and Political Science; email: m.barigozzi@lse.ac.uk.

²University of Konstanz; email: roxana.halbleib@uni-konstanz.de.

³Université libre de Bruxelles, Solvay Brussels School of Economics and Management, ECARES; email: david.veredas@ulb.ac.be.

Corresponding address: David Veredas, ECARES, Solvay Brussels School of Economics and Management, Université libre de Bruxelles, 50 Av F.D. Roosevelt CP114/04, B1050 Brussels, Belgium. Phone: +3226504218. Fax: +3226504475.

This work was written while David Veredas visited the Department of Monetary and Financial Studies at the Banco de España (Madrid, Spain). Roxana Halbleib acknowledges financial support from the European Social Fund and from the Ministry of Science, Research and the Arts Baden-Württemberg. David Veredas acknowledges financial support from the Belgian National Bank and the IAP P6/07 contract, from the IAP programme (Belgian Scientific Policy), 'Economic policy and finance in the global economy'.

We are grateful to the editor, the associate editor, a referee, Francisco Blasques, Christian Brownlees, Richard Davis, Miguel Angel Delgado, Eric Ghysels, Alastair Hall, Marc Hallin, Atsushi Inoue, Trino-Manuel Niguez Grau, Paolo Santucci de Magistris, Javier Perote, Olivier Scaillet and Enrique Sentana for insightful remarks. We also acknowledge the useful feedback of the participants of the International Symposium in Statistics and Financial Time Series (Wakayama, Japan), the Interdisciplinary Workshop on Econometric and Statistical Modelling of Multivariate Time Series (Louvain-la-neuve, Belgium), the DFH Workshop in Econometrics (Königsfeld, Germany), the conference on Statistical Models for Financial Data (Graz, Austria), and the Workshop on Financial Time Series Analysis: High-dimensionality, Non-stationarity and the Financial Crisis (Singapore), as well as the seminar participants at the Universities of Carlos III de Madrid, Manchester, and Pompeu Fabra. Any error and inaccuracy are ours.

1 Introduction

The title of this article intentionally resembles to Gallant and Tauchen (1996)'s *Which moments to match?* They, Gouriéroux et al. (1993), and Smith (1993) introduce an inference method based on matching sample and theoretical features of an auxiliary model in order to estimate the parameters of the model of interest. Gouriéroux et al. (1993) call it Indirect Inference, Smith (1993) Simulated Quasi Maximum Likelihood, and Gallant and Tauchen (1996) the Efficient Method of Moments (EMM). We will denote these methods generically as *indirect methods*.

Because of their flexibility, applications of indirect methods abound, in particular in economics.¹ Provided that the auxiliary estimators converge to some point in their space and that the so-called binding function is injective, then the general theory shows that any auxiliary model is valid. Validity in this context means that the estimated parameters are, under fairly weak regularity conditions, consistent, asymptotically Gaussian, and with a variance-covariance matrix that is a function of the binding function and the Fisher information matrix of the auxiliary model.

In practice, a good auxiliary model is a model that, in some sense, is close to the one of interest. Gallant and Tauchen (1996) introduce the SNP (Semi Non-Parametric) score generator as universal auxiliary model. It is based on Hermite expansions that can, in principle, span the score vector of any model of interest. The SNP score generator is a class of auxiliary models since there are as many models as terms in the expansion; so a choice has to be made. An alternative is to use an auxiliary model that is bigger (and hence with many more parameters) than the model of interest, on the grounds that the bigger the auxiliary model, the better the approximation to the model of interest. In the limit, this leads to the sieve-based approach of Nickl and Potscher (2010) and Blasques (2011). Though theoretically it is a reasonable argument, in practice we face limited information given by the sample size, meaning that the finite sample properties of the estimated parameters may deteriorate sensibly with the dimension of the auxiliary model. This is the classical trade-off between model fit and the number of parameters.

To the best of our knowledge no systematic analysis has been done on how to choose the auxiliary model within the general theory of indirect methods. In this article we propose information criteria that can be used for choosing between a set of nested and non-nested auxiliary models. They are the indirect analogues of the widely used Akaike-type model information criteria (among those, the most widely used are AIC of Akaike (1974), BIC of Schwarz (1978), and HQIC of Hannan and Quinn (1979)). The indirect selection criteria (that we denote with subscript IM) are based on the likelihood of the estimated parameters of the model of interest, which is Gaussian and centered at the true parameter vector regardless of the choice of the auxiliary model. The difference therefore comes in

¹A non-exhaustive list is: discrete stochastic volatility models (Monfardini (1998), Gallant et al. (1997), Lombardi and Calzolari (2009), and Corsi and Reno (2012)), continuous stochastic volatility models (Chernov et al. (2003) and Grassi and de Magistris (2012)), diffusion models (Engle and Lee (1996), Broze et al. (1998)), threshold models (Guay and Scaillet (2003)), heavy tailed distributions (Lombardi and Veredas (2009), Garcia et al. (2011)), DSGE models (Dridi et al. (2007)), survival data models (Jiang and Turnbull (2003)), stochastic agent-based models (Jang and Lux (2012)), labor income risk (Guvenen and Smith (2010)), discrete choice models (Keane and Smith (2003)), heteroskedastic factor models (Calzolari et al. (2008)), learning models (Czellar and Calvet (2011)), and derivative markets (Phillips and Yu (2009)).

terms of precision, i.e. the variance–covariance matrix. But to account for the trade–off between the fit of the auxiliary model to the model of interest and its dimension, we need a penalty that depends on the number of parameters of the auxiliary model.

Our work relates to the existing literature on Generalized Method of Moments (GMM) for choosing the moment conditions (MC). Two issues arise in the GMM context: the choice of MC that ensures consistency of the estimates, and the choice of no redundant moment conditions. The former is related to weak identification, or to the case where there is insufficient information to yield a consistent estimator. Andrews (1999) proposes GMM analogues of the AIC, BIC and HQIC (which we denote with subscript GMM) for consistently selecting the correct MC.² As for the choice of no redundant MC, it relates with moment conditions that provide no incremental information about the parameters of the model of interest. Hall et al. (2007) introduce the Relevant Moment Selection Criteria (RMSC) where the word relevant denotes the subset of MC that are asymptotically efficient but that contain no redundant moment conditions. Though redundant moment conditions have no asymptotic impact, Hall and Peixe (2003) report simulation evidence that their inclusion can lead to a poor approximation of the limiting distribution to the finite sample behavior.

In the framework of indirect methods, the choice of the auxiliary model presents no concerns in terms of consistency, provided the conditions stated above. The closest to our work is therefore Hall et al. (2007). They propose RMSC based on the fact that the entropy of the limiting distribution of the GMM estimator provides a metric for information of the moment conditions. These ideas were also used by Hall et al. (2011) in the context of Dynamic Stochastic General Equilibrium (DSGE) models, where the auxiliary model are impulse response functions (IRF hereafter) stemming from a VAR. This is alike to Dridi et al. (2007) that estimate a DSGE model with sequential partial indirect inference.

We proceed with a thorough Monte Carlo study on a simple, yet important and illustrative, model: a moving average (MA) of order one. In the context of indirect methods, this model has been studied by Gouriéroux et al. (1993), Ghysels et al. (2003), and Genton and Ronchetti (2003). We choose as auxiliary models the $AR(r)$, $ARMA(1,1)$, and the true model. The purpose of this study is to investigate the performance of the criteria when choosing among the true model and nested models.

We conclude the article with an empirical illustration to the volatility of daily financial returns during 2003–2008. The model of interest is a stochastic volatility, which is estimated with indirect methods by Monfardini (1998), Gallant et al. (1997), Andersen et al. (1999), and Calzolari et al. (2004). We consider two auxiliary models: $ARCH(r)$ and $GARCH(1,1)$ on the observations, and $AR(r)$ and $ARMA(1,1)$ on the log of the square of the observations.

The remaining sections are laid out as follows. Section 2 introduces the notation and the basic background of indirect methods while Section 3 describes the criteria and a theoretical example with a Gram–Charlier expansion as auxiliary model. Section 4 and 5 cover the Monte Carlo study and the empirical illustration. Conclusions are touched upon

²See also Andrews and Lu (2001), Kitamura (2002), Hong et al. (2003) and Hong and Preston (2005) for related work on moment selection criteria, and Konishi and Kitagawa (1996) for a generalized information criteria.

in Section 6. The Appendixes contain the technical lemmas and proofs (Appendix A), implementation details (B), and tables (C).

2 Indirect methods

The model of interest We consider a $n \times 1$ random vector \mathbf{Y} that follows a model $\mathcal{M}(\mathbf{y}, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes the p -dimensional vector of unknown parameters, and $\mathbf{y} \in \mathcal{Y} \subset \mathbb{R}^n$ is a realization of \mathbf{Y} . The model is defined by a parametric probability law

$$\mathcal{M}(\mathbf{y}, \boldsymbol{\theta}) = \{P_{\boldsymbol{\theta}}(\mathbf{y}), \mathbf{y} \in \mathcal{Y} \subset \mathbb{R}^n, \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\} \quad (1)$$

that has an associated log-likelihood. The true value of the parameters is denoted by $\boldsymbol{\theta}_0$. We denote by $E_{\mathbf{y}}$ the expectation with respect to the true distribution of \mathbf{Y} and we make the usual assumptions for maximum likelihood estimation:

ASSUMPTION A

- (1) The parameter space Θ is non-empty, compact, and $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$;
- (2) The log-likelihood ℓ is twice-differentiable, i.e. $\ell \in \mathcal{C}^2(\Theta)$;
- (3) The true parameter vector $\boldsymbol{\theta}_0 \in \Theta$ is the unique maximum for $E_{\mathbf{y}}[\ell(\boldsymbol{\theta}; \mathbf{Y})]$;
- (4) The $p \times p$ Fisher information matrix

$$\mathcal{I}(\boldsymbol{\theta}_0) = E_{\mathbf{y}} \left[\begin{array}{cc} \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{Y})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} & \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{Y})}{\partial \boldsymbol{\theta}'} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \end{array} \right],$$

exists, and is finite and positive definite.

We assume to have T realizations of \mathbf{Y} that we denote as $\mathbf{y}_t = (y_{1t} \dots y_{nt})'$, a $n \times 1$ vector of a generic realization with $t = 1, \dots, T$. It is well known that, under the standard regularity conditions in Assumption A, the maximum likelihood (ML) estimator is such that

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_{T,ML} - \boldsymbol{\theta}_0) \sim \mathcal{N}(\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\theta}_0)), \text{ as } T \rightarrow \infty.$$

We are interested in the case in which the likelihood is not available analytically, and therefore $\widehat{\boldsymbol{\theta}}_{T,ML}$ is an unfeasible estimator. This is the case for a broad class of models, including static and dynamic models, with unobserved components, trends, and various innovation terms. Each process and model has its own assumptions. Establishing general hypothesis on \mathbf{Y} and $\mathcal{M}(\mathbf{y}, \boldsymbol{\theta})$ is therefore idle, beyond Assumption A. It is however important that the model is such that, for a given value of the parameter vector, simulated draws of $\mathbf{y}_1(\boldsymbol{\theta}), \dots, \mathbf{y}_T(\boldsymbol{\theta})$ can be obtained. This typically implies that $\mathcal{M}(\mathbf{y}, \boldsymbol{\theta})$ contains at least one *i.i.d.* innovation with known distribution.

The auxiliary model Consider the auxiliary model for \mathbf{Y}

$$\mathcal{M}^a(\mathbf{y}, \boldsymbol{\beta}) = \{P_{\boldsymbol{\beta}}(\mathbf{y}), \mathbf{y} \in \mathcal{Y} \subset \mathbb{R}^n, \boldsymbol{\beta} \in \mathcal{B} \subset \mathbb{R}^q\}$$

that depends on a q -dimensional vector of parameters $\boldsymbol{\beta}$. We assume that the associated log-likelihood $\ell^a(\boldsymbol{\beta}; \mathbf{y}_t)$ is known and available analytically. In a conditional model the

log-likelihood is conditional on other variables and possibly on past information. After denoting by β_0 the true value of the parameters, we make the assumptions:

ASSUMPTION B

- (1) The parameter space \mathcal{B} is non-empty, compact, and $\beta_0 \in \text{int}(\mathcal{B})$;
- (2) The log-likelihood ℓ^a is twice-differentiable, i.e. $\ell^a \in \mathcal{C}^2(\mathcal{B})$;
- (3) The true parameter vector β_0 is the unique maximum of $E_{\mathbf{y}}[\ell^a(\beta; \mathbf{Y})]$;
- (4) The $q \times q$ Fisher information and Hessian matrices

$$\begin{aligned} \mathbf{I}(\beta_0) &= E_{\mathbf{y}} \left[\begin{array}{cc} \frac{\partial \ell^a(\beta; \mathbf{Y})}{\partial \beta} \Big|_{\beta=\beta_0} & \frac{\partial \ell^a(\beta; \mathbf{Y})}{\partial \beta'} \Big|_{\beta=\beta_0} \end{array} \right] \text{ and} \\ \mathbf{J}(\beta_0) &= -E_{\mathbf{y}} \left[\frac{\partial^2 \ell^a(\beta; \mathbf{Y})}{\partial \beta \partial \beta'} \Big|_{\beta=\beta_0} \right]. \end{aligned}$$

exist, and are finite and positive definite.

Under the standard regularity conditions in Assumption B, the Quasi Maximum Likelihood (QML) estimator is such that

$$\sqrt{T}(\hat{\beta}_T - \beta_0) \sim \mathcal{N}(\mathbf{0}, (\mathbf{J}(\beta_0)\mathbf{I}^{-1}(\beta_0)\mathbf{J}(\beta_0))^{-1}), \text{ as } T \rightarrow \infty.$$

For a given value of θ , we can simulate H paths of length T : $\mathbf{y}_1^h(\theta), \dots, \mathbf{y}_T^h(\theta)$, $h = 1, \dots, H$, from the model of interest $\mathcal{M}(\mathbf{y}, \theta)$ and we estimate

$$\hat{\beta}_{HT}(\theta) = \arg \max_{\beta \in \mathcal{B}} \frac{1}{HT} \sum_{h=1}^H \sum_{t=1}^T \ell^a(\beta, \mathbf{y}_t^h(\theta)). \quad (2)$$

Gouriéroux et al. (1993) prove that this is a consistent and asymptotically Gaussian estimator (as $T \rightarrow \infty$) of the so-called binding function $\mathbf{b}(\theta)$ linking θ to β , for which we make the following assumptions:

ASSUMPTION C

- (1) $\mathbf{b}(\theta_0)$ is the unique maximum of $E_{\mathbf{y}}[\ell^a(\mathbf{b}(\theta_0); \mathbf{Y})]$, which implies $\beta_0 = \mathbf{b}(\theta_0)$;
- (2) $\mathbf{b}(\theta)$ is injective and continuously differentiable, i.e. $\mathbf{b} \in \mathcal{C}^1(\Theta)$;
- (3) The $q \times p$ matrix $\frac{\partial \mathbf{b}(\theta_0)}{\partial \theta'}$ has full-column rank p .

Finally, note that we could simulate paths of length $T^* \neq T$ (typically $T^* > T$) and therefore both the estimated auxiliary parameters and their estimated asymptotic variance-covariance matrix will be functions of T^* instead of T . For simplicity of notation we omit this distinction in what follows.

Indirect estimation The indirect estimator is defined as

$$\hat{\theta}_{HT,IM,a,q} = \arg \min_{\theta \in \Theta} \left(\hat{\beta}_{HT}(\theta) - \hat{\beta}_T \right)' \Omega \left(\hat{\beta}_{HT}(\theta) - \hat{\beta}_T \right), \quad (3)$$

where Ω is a $q \times q$ matrix defining the metric. The sub-index IM in $\hat{\theta}_{HT,IM,a,q}$ denotes Indirect Method, and we make explicit the dependence on the auxiliary model a and

the dimension of its parameter vector q . The β -based optimization (3) was introduced by Gouriéroux et al. (1993). Alternatively we may consider a $\mathbf{h}(\beta)$ -based optimization, for an injective and continuously differentiable function with respect to β , and after a suitable re-parametrization of the weighting matrix by means of the delta method. The vector $\mathbf{h}(\beta)$ is of dimension q' that can be larger or equal to q . In the score-based optimization of Gallant and Tauchen (1996) $q' = q$, while in the IRF-based optimization of Hall et al. (2011) $q' > q$. A special case is when the auxiliary model is estimated under constraints on the parameter space, i.e. constrained QML, which entails the estimation of multipliers (Calzolari et al. (2004)). The vector β is expanded to dimension $q' > q$ to include the multipliers. If the constraints are not binding, the multipliers are zero, and inference is unconstrained. If, by contrast, the constraints bind, the bounded parameters contain no information about θ , as it goes through the associated multipliers. In the sequel, we consider the case of the β -based optimization. If, instead, the $\mathbf{h}(\beta)$ -based optimization is used, the criteria we propose are adapted by replacing q by q' .

Gouriéroux et al. (1993) prove that under Assumptions A, B, and C and the appropriate model-specific assumptions, the indirect estimator defined in (3) is consistent with limiting distribution

$$\sqrt{T}(\widehat{\theta}_{HT,IM,a,q} - \theta_0) \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_{H,\Omega,a,q}(\theta_0)), \text{ as } T \rightarrow \infty.$$

The asymptotic variance-covariance matrix depends on the true value of the parameters, the number of simulated paths, and the weighting matrix Ω . The optimal choice for the latter is $\Omega^* = \mathbf{J}(\beta)\mathbf{I}^{-1}(\beta)\mathbf{J}(\beta)$ so that

$$\mathbf{W}_{H,\Omega^*,a,q}(\theta) = \left(1 + \frac{1}{H}\right) \left[\frac{\partial \mathbf{b}(\theta)}{\partial \theta} \mathbf{J}(\beta) \mathbf{I}^{-1}(\beta) \mathbf{J}(\beta) \frac{\partial \mathbf{b}(\theta)}{\partial \theta'} \right]^{-1}.$$

Since the value of H is arbitrarily chosen, we consider $H \rightarrow \infty$. In practice, however, H is finite and an educated guess is needed, as explained in detail in the next section. Also, because in the next section we need further notation, hereafter we skip the subindexes H and Ω^* , so $\widehat{\theta}_{HT,IM,a,q}$ becomes $\widehat{\theta}_{T,IM,a,q}$ and $\mathbf{W}_{H,\Omega^*,a,q}(\theta)$ becomes

$$\mathbf{W}_{a,q}(\theta) = \left[\frac{\partial \mathbf{b}(\theta)}{\partial \theta} \mathbf{J}(\beta) \mathbf{I}^{-1}(\beta) \mathbf{J}(\beta) \frac{\partial \mathbf{b}(\theta)}{\partial \theta'} \right]^{-1}. \quad (4)$$

A plethora of theoretical developments and improvements of the original methods have been proposed. What follows is a selection of an extensive and growing literature. Dhaene et al. (1998) introduce a notion of indirect encompassing and non-nested hypotheses tests using indirect methods. That is, the case where not only the relevant binding function does not have closed form, but also at least one of the competing models. Billio and Monfort (2003) propose kernel-based conditional expectations for the binding functions, which have the advantage that no optimization step is involved in the computation of the binding function. They also propose two criteria specific for conditional expectations as binding functions. Another area of research within indirect methods concerns robustification. Genton and Ronchetti (2003) develop robust indirect inference by deriving the influence function of the indirect estimator, as well as the level and power influence function of indirect tests. In a similar spirit, Trojani and Ortelli (2005) focus on robust EMM

estimation of a general parametric time series stationary processes. Czellar and Ronchetti (2010) propose robust over-identification tests that exhibit better finite sample accuracy than classical tests. Partial and semi-nonparametric indirect methods have also been introduced. Dridi et al. (2007) introduce the sequential partial indirect inference approach, in the sense that the model of interest is partially well specified. They show the usefulness of the sequential method on the calibration of DSGE models. Also motivated by DSGE models, Blasques (2011) presents semi-nonparametric indirect methods, i.e. a sieve class of estimator, on the grounds that, at population level, an infinite dimensional auxiliary model may be needed for the indirect estimation of many economic models. Indirect methods have also been used for bias correction (see Gouriéroux et al. (2000) for general time series models, Ghysels et al. (2003) for a MA(1), and Gouriéroux et al. (2010) for dynamic panel data models). The advantage with respect to other bias correction techniques is that indirect methods do not require a given explicit form for the bias function of its expansion, as it is calibrated via simulations. Last, Phillips (2012) provides new limit theory for the delta method and the continuous mapping theorem for cases when the problem involves sample functions that depend on the sample size, or when the quantity of interest appears in an implicit functional form, as it happens in indirect methods.

3 The indirect information criteria

3.1 The principle

Consider two auxiliary models $\mathcal{M}^{a_1}(\mathbf{y}, \boldsymbol{\beta}_1)$ and $\mathcal{M}^{a_2}(\mathbf{y}, \boldsymbol{\beta}_2)$ with $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ of dimension q_1 and q_2 respectively. Under the optimal choice of the weighting matrices and as $T, H \rightarrow \infty$ we have

$$\begin{aligned}\sqrt{T}(\widehat{\boldsymbol{\theta}}_{T,IM,a_1,q_1} - \boldsymbol{\theta}_0) &\sim \mathcal{N}(\mathbf{0}, \mathbf{W}_{a_1,q_1}(\boldsymbol{\theta}_0)) \quad \text{and} \\ \sqrt{T}(\widehat{\boldsymbol{\theta}}_{T,IM,a_2,q_2} - \boldsymbol{\theta}_0) &\sim \mathcal{N}(\mathbf{0}, \mathbf{W}_{a_2,q_2}(\boldsymbol{\theta}_0)).\end{aligned}$$

Both estimators are consistent and the limiting distribution of $\sqrt{T}(\widehat{\boldsymbol{\theta}}_{T,IM,a,q} - \boldsymbol{\theta}_0)$ is centered Gaussian. The difference between them comes in terms of precision, i.e. the variance-covariance matrix. Choices of \mathcal{M}^a and q cover both nested and non-nested auxiliary models. For simplicity of notation we use the index a to indicate a generic auxiliary model \mathcal{M}^a . In particular, if the models are nested $\mathcal{M}^{a_1} \subset \mathcal{M}^{a_2}$ (in the sense that they belong to the same auxiliary class but $q_1 < q_2$) we use the same index $a = a_1 = a_2$ for both. If the models are non-nested, there is no well defined inclusion relation between \mathcal{M}^{a_1} and \mathcal{M}^{a_2} , q_1 and q_2 may or may not be equal, and we say $a_1 \neq a_2$. Finally, we define $\mathcal{A}_{q_{\max}}$ as the set of all possible auxiliary models satisfying Assumptions A, B, and C, where q_{\max} is the largest dimension of the auxiliary parameter we allow for. For the sake of simplicity, in this section we compare two classes of auxiliary models, and for each class we consider the same maximum number of parameters $q_{1,\max} = q_{2,\max} = q_{\max}$. It is however possible that $q_{1,\max} \neq q_{2,\max}$, as in the SV model in Section 5. Each element of $\mathcal{A}_{q_{\max}}$ is made of a couple (a, q) indicating the chosen auxiliary model and the dimension of the corresponding auxiliary parameter.

We say that a matrix \mathbf{A} is larger than another \mathbf{B} , and we denote it as $\mathbf{A} \succeq \mathbf{B}$, when $\|\mathbf{A}\| \geq \|\mathbf{B}\|$, where $\|\cdot\|$ is the Frobenius norm $\|\mathbf{A}\| = \sqrt{\text{Tr}(\mathbf{A}\mathbf{A}')}$. Let us consider two nested (thus belonging to the same class a) auxiliary models with $q_1 < q_2$. Parzen (1959) and Carrasco and Florens (2000) show that

$$\mathbf{W}_{a,q_1}(\boldsymbol{\theta}_0) \succeq \mathbf{W}_{a,q_2}(\boldsymbol{\theta}_0) \succeq \mathbf{W}_{a,q_{\max}}(\boldsymbol{\theta}_0).$$

That is, the more auxiliary parameters, the more information and the smaller is the asymptotic variance–covariance of the estimated parameters. Whatever the value of q_{\max} , the lower bound for the auxiliary models is always larger than the Cramér–Rao lower bound, i.e. $\mathbf{W}_{a,q_{\max}}(\boldsymbol{\theta}_0) \succ \mathcal{I}^{-1}(\boldsymbol{\theta}_0)$ for any a and q_{\max} (Gouriéroux and Monfort (1996)). Therefore, under $\boldsymbol{\theta}_0$ the optimal choice is the largest possible auxiliary model. However, this is not necessarily the case in practice when $\boldsymbol{\theta}_0$ is replaced by the estimator. We have to take into account the increasing estimation error due to the increasing number of estimated parameters.

Simulating the estimated parameters The indirect criteria that we propose below requires realizations of the random vector $\mathbf{X}_{a,q} \equiv \sqrt{T}(\hat{\boldsymbol{\theta}}_{T,IM,a,q} - \boldsymbol{\theta}_0)$ that are not directly available since $\boldsymbol{\theta}_0$ is not observed. However, it is straightforward to simulate from the distribution of $\mathbf{X}_{a,q}$. Indeed, we know that $\mathbf{X}_{a,q} \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_{a,q}(\boldsymbol{\theta}_0))$, as $T \rightarrow \infty$.

Let $\widehat{\mathbf{W}}_{T,a,q}(\hat{\boldsymbol{\theta}}_{T,IM,a,q})$ be an estimator of $\mathbf{W}_{a,q}(\boldsymbol{\theta}_0)$ as defined in Gouriéroux et al. (1993). We then know that

$$p\lim_{T \rightarrow \infty} \widehat{\mathbf{W}}_{T,a,q}(\hat{\boldsymbol{\theta}}_{T,IM,a,q}) = \mathbf{W}_{a,q}(\boldsymbol{\theta}_0).$$

Therefore, we can simulate a sample of arbitrary length N from a p –dimensional centered Gaussian distribution having as variance–covariance matrix $\widehat{\mathbf{W}}_{T,a,q}(\hat{\boldsymbol{\theta}}_{T,IM,a,q})$.

We denote the i –th simulated observation by the p –dimensional vector $\mathbf{x}_{a,q,i} = (x_{1,a,q,i} \dots x_{p,a,q,i})'$, with $i = 1, \dots, N$ and we denote the Gaussian pdf of $\mathbf{X}_{a,q}$ by $\phi(\mathbf{x}_{a,q,i}; \widehat{\mathbf{W}}_{T,a,q}(\hat{\boldsymbol{\theta}}_{T,IM,a,q}))$. In order to define the information criterion below, we also introduce the p –dimensional vector $\mathbf{X}^* \equiv \sqrt{T}(\hat{\boldsymbol{\theta}}_{T,ML} - \boldsymbol{\theta}_0)$.

At this point it is worth emphasizing the different roles and meanings of n , T and N . The first one, n , is the dimension of the observed random vector \mathbf{Y} , while T is the sample length of the observed data, and N is the sample length of the simulated vector $\mathbf{X}_{a,q}$, which can be chosen large enough so the sample distribution of $\mathbf{X}_{a,q}$ converges to its limit.

3.2 The criterion

We want to find the auxiliary model that provides the most efficient indirect estimator (i.e. with the variance-covariance matrix that is the closest to $\mathcal{I}^{-1}(\boldsymbol{\theta}_0)$), while accounting for the estimation error. Given that both indirect methods and Maximum Likelihood provide asymptotically Gaussian estimators, finding such a model intuitively is the same as minimizing the distance between the densities of the vectors $\mathbf{X}_{a,q}$ and \mathbf{X}^* . Given N

observations for both vectors and following Akaike (1974) and White (1982), we choose to minimize the sample Kullback–Leibler information defined as

$$\widehat{\mathcal{D}}_N(\mathcal{I}^{-1}(\boldsymbol{\theta}_0), \mathbf{W}_{a,q}(\boldsymbol{\theta}_0)) = \frac{2}{N} \sum_{i=1}^N \log \frac{\phi(\mathbf{x}_i^*; \mathcal{I}^{-1}(\boldsymbol{\theta}_0))}{\phi(\mathbf{x}_{a,q,i}; \mathbf{W}_{a,q}(\boldsymbol{\theta}_0))}. \quad (5)$$

While we can simulate $\mathbf{x}_{a,q,i}$, we cannot simulate \mathbf{x}_i^* . This is however not a problem as \mathbf{x}_i^* does not depend on the auxiliary model and therefore plays no role in the minimization. Since we do not observe, $\mathbf{W}_{a,q}(\boldsymbol{\theta}_0)$ we have to replace it in (5) by its estimator. Hence we have to minimize

$$\mathbf{E}_w \left[\frac{2}{N} \sum_{i=1}^N \log \frac{\phi(\mathbf{x}_i^*; \mathcal{I}^{-1}(\boldsymbol{\theta}_0))}{\phi(\mathbf{x}_{a,q,i}; \widehat{\mathbf{W}}_{T,a,q}(\widehat{\boldsymbol{\theta}}_{T,IM,a,q}))} \right], \quad (6)$$

where the expectation is taken with respect to the distribution of the vectorized $\widehat{\mathbf{W}}_{T,a,q}$. Notice that, if we define the sample variance–covariance matrix of $\mathbf{X}_{a,q}$ as

$$\widetilde{\mathbf{W}}_{N,a,q} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_{a,q,i} \mathbf{x}'_{a,q,i},$$

then the following holds

$$p \lim_{T \rightarrow \infty} \left(p \lim_{N \rightarrow \infty} \widetilde{\mathbf{W}}_{N,a,q} \right) = p \lim_{T \rightarrow \infty} \left(\widehat{\mathbf{W}}_{T,a,q}(\widehat{\boldsymbol{\theta}}_{T,IM,a,q}) \right) = \mathbf{W}_{a,q}(\boldsymbol{\theta}_0). \quad (7)$$

Therefore, for N and T large enough, in (6) we can use $\widetilde{\mathbf{W}}_{N,a,q}$ instead of $\widehat{\mathbf{W}}_{T,a,q}(\widehat{\boldsymbol{\theta}}_{T,IM,a,q})$.

Before introducing the criterion, we need the following definition.

Definition *The auxiliary model belonging to class \tilde{a} and with auxiliary parameter dimension \tilde{q} is optimal if it minimizes the expected Kullback–Leibler information (6), i.e.*

$$(\tilde{a}, \tilde{q}) = \underset{(a, q) \in \mathcal{A}_{q_{\max}}}{\operatorname{argmin}} \mathbf{E}_w \left[\frac{2}{N} \sum_{i=1}^N \log \frac{\phi(\mathbf{x}_i^*; \mathcal{I}^{-1}(\boldsymbol{\theta}_0))}{\phi(\mathbf{x}_{a,q,i}; \widehat{\mathbf{W}}_{T,a,q}(\widehat{\boldsymbol{\theta}}_{T,IM,a,q}))} \right].$$

And the following proposition shows the indirect Akaike information criterion.

Proposition 1 *Given the models of Section 2, Assumptions A, B, and C above, and Assumptions D and E in Appendix A, minimizing (6) is the same as minimizing the following criterion:*

$$AIC_{IM}(a, q) = -2 \sum_{i=1}^N \log \phi(\mathbf{x}_{a,q,i}; \widetilde{\mathbf{W}}_{N,a,q}) + \frac{q(q+1)}{2}, \quad (8)$$

i.e. the optimal auxiliary model (\tilde{a}, \tilde{q}) is such that:

$$(\tilde{a}, \tilde{q}) = \underset{(a, q) \in \mathcal{A}_{q_{\max}}}{\operatorname{argmin}} AIC_{IM}(a, q).$$

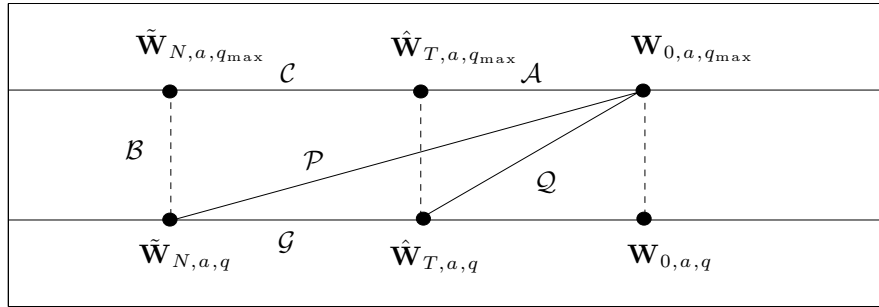
Proof: See Appendix A.

The nested case

What follows is the sketched proof of Proposition 1 for the nested case (i.e. when the class of auxiliary models a is fixed). It has an intuitive geometric interpretation. However, generally AIC_{IM} compares nested and non-nested models, which implies that the best model is the true model, i.e. the one with asymptotic variance covariance matrix $\mathcal{I}^{-1}(\boldsymbol{\theta}_0)$. But since this is never achievable, the geometric argument cannot be made. For this reason a more involving proof than original Akaike's is needed –as it is shown in detail in Appendix B– and that is valid for the nested and non-nested cases.

The diagram in Figure 1 is divided in two rows and three columns. The rows correspond to the largest dimension of the auxiliary model (top row) and a given dimension (bottom). The columns correspond to the variance–covariance matrices in population (right), and the sample and simulated counterparts (middle and left respectively). The calligraphic letters denote vectors. Asymptotically, the auxiliary model that is the closest to the model of interest has the maximum possible dimension with variance–covariance matrix of the estimated parameters equal to $\mathbf{W}_{a,q_{\max}}(\boldsymbol{\theta}_0)$, with q_{\max} large enough. In practice however, $\mathbf{W}_{a,q_{\max}}(\boldsymbol{\theta}_0)$ is unknown and the variance–covariance matrix is estimated for a given $q < q_{\max}$: $\widehat{\mathbf{W}}_{T,a,q}(\widehat{\boldsymbol{\theta}}_{T,IM,a,q})$. The aim is to find the value of q that minimizes the distance between $\mathbf{W}_{a,q_{\max}}(\boldsymbol{\theta}_0)$ and $\widehat{\mathbf{W}}_{T,a,q}(\widehat{\boldsymbol{\theta}}_{T,IM,a,q})$. This value is not necessarily q_{\max} , for fixed T , as when q increases the number of parameters to estimate increases and the estimation error increases accordingly. However, since we use simulated samples $\mathbf{X}_{a,q}$, we need to use asymptotic results for large values of the sample size N . For this reason we have to consider the sample covariance $\widetilde{\mathbf{W}}_{N,a,q}$, and thus the left column in the diagram.

Figure 1: Diagrammatic representation of the nested case



The diagram is divided in two rows and three columns. The rows correspond to the largest dimension of the auxiliary model (top row) and a given dimension (bottom). The columns correspond to the variance–covariance matrices in population (right), and the sample and simulated counterparts (middle and left respectively). Because of representation purposes, the dependence with respect to the parameters is skipped. The calligraphic letters denote vectors.

Let $\mathbf{w} = \text{vech}(\mathbf{W})$ and $\mathcal{Q} = (\mathbf{w}_{0,a,q_{\max}} - \widehat{\mathbf{w}}_{T,a,q})$. Then the distance between $\mathbf{W}_{0,a,q_{\max}}(\boldsymbol{\theta}_0)$ and $\widehat{\mathbf{W}}_{T,a,q}(\widehat{\boldsymbol{\theta}}_{T,IM,a,q})$ can be written as the norm squared

$$\mathcal{Q}^2 = \|\mathbf{w}_{0,a,q_{\max}} - \widehat{\mathbf{w}}_{T,a,q}\|_{\mathcal{J}}^2, \quad (9)$$

where the norm is based on Hessians \mathcal{J} as defined in Appendix B. We want the auxiliary model (a, q) that minimizes the distance \mathcal{Q}^2 . From the geometry of Figure 1, this distance equals

$$\mathcal{Q}^2 = (\mathcal{P} - \mathcal{G})^2 = \mathcal{P}^2 + \mathcal{G}^2 - 2\mathcal{G}\mathcal{P}. \quad (10)$$

From Lemma 1 in Appendix B, the term \mathcal{G}^2 is distributed as $\chi_{\frac{q(q+1)}{2}}^2/N$ so it has expectation $\frac{q(q+1)}{2N}$, while the last term has expectation zero for large N . Finally, by Pythagora's theorem we have

$$\mathcal{P}^2 = (\mathcal{A} + \mathcal{C})^2 + \mathcal{B}^2.$$

Now $\mathcal{A} + \mathcal{C}$ does not depend on q , so it does not play any role in the criterion and we can treat it as a constant, say k . \mathcal{B}^2 gives the log-likelihood term computed in $\widetilde{\mathbf{W}}_{N,a,q}$. Therefore, by letting N diverging and taking expectations of (10), we can write

$$\begin{aligned} \mathbb{E}_w [\mathcal{Q}^2] &\simeq \mathbb{E}_w [(\mathcal{A} + \mathcal{C})^2 + \mathcal{B}^2 + \mathcal{G}^2 - 2\mathcal{G}\mathcal{P}] = \mathbb{E}_w [(\mathcal{A} + \mathcal{C})^2 + \mathcal{B}^2 + \mathcal{G}^2] \\ &\simeq k + \frac{2}{N} \sum_{i=1}^N \log \frac{\phi(\mathbf{x}_{i,a,q_{\max}}, \widetilde{\mathbf{W}}_{N,a,q_{\max}})}{\phi(\mathbf{x}_{i,a,q}, \widetilde{\mathbf{W}}_{N,a,q})} + \frac{q(q+1)}{2N}. \end{aligned} \quad (11)$$

By multiplying the last expression by N and keeping only the terms that depend on (a, q) we have the criterion.

Summing up, \mathcal{Q}^2 depends on i) the distance between $\mathbf{W}_{a,q_{\max}}(\boldsymbol{\theta}_0)$ and its projection in a q dimensional subspace $\mathbf{W}_{a,q}(\boldsymbol{\theta}_0)$, i.e. the term \mathcal{B}^2 , and ii) the distance between $\widetilde{\mathbf{W}}_{N,a,q}$ and the estimated variance-covariance matrix of the auxiliary model $\widehat{\mathbf{W}}_{T,a,q}(\widehat{\boldsymbol{\theta}}_{T,IM,a,q})$, i.e. the term \mathcal{G}^2 . While the first distance decreases as q increases, the second distance is an estimation error –or a penalty– that increases with q .

Three remarks

First, there are important differences between AIC_{IM} and the ML and GMM counterparts. The likelihood term in AIC_{IM} is of the estimated parameters of the model of interest, which is centered Gaussian regardless of the choice of the auxiliary model and the model of interest. By contrast, in AIC_{ML} the likelihood term depends on the model of interest, and in AIC_{GMM} depends on the moment conditions. Another difference is that in AIC_{IM} the penalty term depends on the parameters of the auxiliary model, while in AIC_{ML} it depends on the parameters of the model of interest. Last, the dimension of the parameters of the model of interest is constant over choices of the auxiliary model, while in AIC_{ML} the dimension varies with the model.

Second, as mentioned in previous section, we consider $H \rightarrow \infty$. In practice however, H is finite. There is a trade-off between the theoretical advantage of $H \rightarrow \infty$ and the computational feasibility. If H is the same across auxiliary models it plays no role in the comparison of the models, in the sense that the choice of the auxiliary model for different values of H is coherent. Values for H between 5 and 10 are common in the literature, see e.g. Billio and Monfort (2003), Calzolari et al. (2004), and Garcia et al. (2011). In the Monte Carlo study below we use $H = 1$ and $H = 10$. Results confirm the reasoning.

Finally, similarly to AIC_{ML} and AIC_{GMM} , AIC_{IM} is not consistent when $N \rightarrow \infty$, in the sense that it tends to select large models. Indeed, when $N \rightarrow \infty$ the first term on the

right hand side of (8) diverges to $-\infty$, thus making the penalty term not effective. As a consequence, AIC_{IM} will not necessarily select the true model if included in the set of all possible auxiliary models. In this sense, we say that the criteria is not consistent. In other words, if the true model were among the auxiliary models, a consistent criterion would select it with probability one.

3.3 Consistent criteria

By modifying the penalty, we can generalize AIC_{IM} to obtain consistent criteria. Indeed, we consider penalty functions satisfying the following asymptotic behavior:

ASSUMPTION F The penalty function is such that $p(N, q) = h(q)K_N$, where q is the dimension of the auxiliary parameter, N is the number of simulated observations of $\mathbf{X}_{a,q}$, $h(q)$ strictly increasing, and K_N satisfies

$$\lim_{N \rightarrow \infty} K_N = \infty \quad \text{and} \quad \lim_{N \rightarrow \infty} \frac{K_N}{N} = 0.$$

For the definition of consistency below, consider the couple (a^*, p) corresponding to the true model with p parameters, as defined in (1). We denote the set containing all possible auxiliary models and the true model as $\mathcal{A}^* = \mathcal{A}_{q_{\max}} \cup \{(a^*, p)\}$.

Definition An information criterion is consistent if it selects an auxiliary model belonging to class \check{a} and with auxiliary parameter dimension \check{q} such that $(\check{a}, \check{q}) \in \mathcal{A}^*$ and

$$(\check{a}, \check{q}) \xrightarrow{P} (a^*, p), \quad \text{as } N, T \rightarrow \infty.$$

Note that the limit in probability used in the above definition is a symbolic way of defining consistency. Indeed, \check{a} is an index indicating a particular class of auxiliary models and not an estimated parameter. We could formalize it by considering a selecting vector (consisting of zeros and ones) of the same dimension as the number of models considered, along the lines of Andrews (1999) and Hall et al. (2011). However, we refrain further notation as, once understood the meaning of \check{a} , nothing changes in the following proposition.

Proposition 2 Given the models of Section 2, Assumptions A, B, C, above, Assumptions D and E in Appendix A, and given a penalty function $p(N, q)$ as defined in Assumption F, the criterion

$$IC_{IM}(a, q) = -2 \sum_{i=1}^N \log \phi(\mathbf{x}_{a,q,i}; \widetilde{\mathbf{W}}_{N,a,q}) + p(N, q), \quad (a, q) \in \mathcal{A}^*, \quad (12)$$

is consistent.

Proof: See Appendix A.

The assumed form of the penalty is $p(N, q) = h(q)K_N$, as postulated by Andrews (1999). In $AIC_{IM}(a, q)$ we have $h(q) = q(q+1)/2$ and $K_N = 1$, hence $p(N, q) = p(q)$ and it does not diverge with N . Penalties similar to BIC and $HQIC$ are potential candidates. Andrews (2000) and Hall et al. (2007) report simulation evidence that the BIC-type penalty,

i.e. $K_N = \log N$, works best in their context, since $K_N = \log \log N$ appears to be a too slow rate. Bozdogan (1987) suggests $K_N = (\log N)^b$ for $b \geq 1$ as a generalization of BIC. This penalty seems to perform better, and our preliminary simulations confirm it. In the Monte Carlo study and the empirical applications in the next sections we use $b = 2$.

3.4 A theoretical illustration: the Gram–Charlier auxiliary model

In this section we exemplify i) the gain in efficiency by increasing the size of the auxiliary model, and ii) the trade-off between model fit and number of parameters. As auxiliary model we use a series expansion of a probability distribution. Several expansions are available, such as Gram–Charlier, Edgeworth, and Gallant and Nychka (1987) and Gallant and Tauchen (1996) expansion. We opt for the first, which has pdf:

$$f^a(y_t) = \left[1 + \sum_{i=1}^K \frac{\vartheta_i}{i!} H_i(z_t) \right] \phi(z_t), \quad (13)$$

where $z_t = (y_t - m)/s$, and m and s are the mean and standard deviation. The sequence $H_i(\cdot)$ are Chebyshev–Hermite polynomials defined as $H_i(x) = (-1)^i \phi(x)^{-1} D^i \phi(x)$, with $H_0(x) = 1$, $D = \frac{d}{dx}$, and $\phi(\cdot)$ denotes the standard Gaussian density. The Hermite polynomials have the property that $\vartheta_i = E[H_i(z_t)]$. Since z_t is standardized, $\vartheta_1 = \vartheta_2 = 0$, which implies that, for estimation purposes, the sum in (13) starts at $i = 3$.

The Gaussian parametric term in (13) is informative about the location and dispersion of the true density, while the polynomial expansion (the semi-nonparametric part) accommodates the non-Gaussianity structure. The latter controls the extent to which (13) deviates from Gaussianity. The order K is chosen by the user, meaning that each K implies a different auxiliary model.

The model of interest is the α -stable density, i.e. $Y_t \sim S_\alpha(\beta, \sigma, \mu)$. The parameter $\alpha \in (0, 2]$, often denoted as tail index, measures the thickness of the tails. Asymmetry is captured by $\beta \in [-1, 1]$. The dispersion parameter $\sigma \in \mathbb{R}^+$ expands or contracts the distribution, and the parameter $\mu \in \mathbb{R}$ controls the location of the distribution.

The probability density function of the α -stable distribution does not have a closed form. Since it is a complicated integral, even difficult to evaluate numerically, estimation by ML has often not been considered in applied work. Simulation however is straightforward, see Chambers et al. (1976), making indirect methods a good alternative (see Dominicy and Veredas (2013) and Garcia et al. (2011)). The density (13) is a particularly well suited auxiliary model. Since the Gaussian distribution is nested in the α -stable, the Hermite polynomials should give information about the tail index α and the asymmetry parameter β .

Let $\boldsymbol{\beta} = (m, s, \vartheta_3, \dots, \vartheta_K)$ be the parameters of the density (13). Following Section 2, $\boldsymbol{\beta}$ is estimated under QML (and its asymptotic distribution follows). Let $\phi(z_t) = \phi_t$,

Table 1: AIC and IC for stable density

K	N=100		N=500		N=1000	
	AIC_{IM}	IC_{IM}	AIC_{IM}	IC_{IM}	AIC_{IM}	IC_{IM}
4	-22.0999	-20.0792	-21.6634	-20.9109	-21.6634	-21.1962
5	-24.7701	-21.7389	-24.3735	-23.2448	-24.3785	-23.6778
6	-23.0491	-18.8055	-22.7005	-21.1204	-22.7116	-21.7305
7	-31.4083	-25.7502	-31.1157	-29.0089	-31.1338	-29.8257

The model of interest is the α -stable density function with location 0, scale 1, asymmetry 0 and tail index 1.95. The number of simulated paths H is 1, N equals 100, 500, and 1000, and $T = 1000$. The auxiliary model is the Gram-Charlier density (13). Numbers in bold denote the model chosen by the criteria.

$b_i = \vartheta_i/i!$, $H_{i,t} = H_i(z_t)$, and $P_{K,t} = 1 + \sum_{i=1}^K b_i H_{i,t}$. The score vector is

$$\frac{\partial \ell^a(\boldsymbol{\beta}; \mathbf{Y})}{\partial \boldsymbol{\beta}} = \sum_{t=1}^T \begin{pmatrix} SG_t^m \\ SG_t^s \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \frac{1}{P_{K,t}} \begin{pmatrix} \sum_{i=1}^K b_i \frac{\partial H_{i,t}}{\partial m} \\ \sum_{i=1}^K b_i \frac{\partial H_{i,t}}{\partial s} \\ H_{3,t} \frac{1}{3!} \\ \vdots \\ H_{K,t} \frac{1}{K!} \end{pmatrix},$$

where SG_t^m and SG_t^s are the Gaussian contributions. The distinct elements of the Hessian are

$$\begin{aligned} \frac{\partial^2 \ell^a(\boldsymbol{\beta}; \mathbf{Y})}{\partial m \partial m} &= \sum_{t=1}^T HG_t^m + \frac{1}{P_{K,t}^2} \left(P_{K,t} \left(\sum_{i=1}^K b_i \frac{\partial^2 H_{i,t}}{\partial m \partial m} \right) - \left(\sum_{i=1}^K b_i \frac{\partial H_{i,t}}{\partial m} \right) \frac{\partial P_{K,t}}{\partial m} \right) \\ \frac{\partial^2 \ell^a(\boldsymbol{\beta}; \mathbf{Y})}{\partial s \partial s} &= \sum_{t=1}^T HG_t^s + \frac{1}{P_{K,t}^2} \left(P_{K,t} \left(\sum_{i=1}^K b_i \frac{\partial^2 H_{i,t}}{\partial s \partial s} \right) - \left(\sum_{i=1}^K b_i \frac{\partial H_{i,t}}{\partial s} \right) \frac{\partial P_{K,t}}{\partial s} \right) \\ \frac{\partial^2 \ell^a(\boldsymbol{\beta}; \mathbf{Y})}{\partial m \partial s} &= \sum_{t=1}^T HG_t^{m,s} + \frac{1}{P_{K,t}^2} \left(P_{K,t} \left(\sum_{i=1}^K b_i \frac{\partial^2 H_{i,t}}{\partial m \partial s} \right) - \left(\sum_{i=1}^K b_i \frac{\partial H_{i,t}}{\partial m} \right) \frac{\partial P_{K,t}}{\partial s} \right) \\ \frac{\partial^2 \ell^a(\boldsymbol{\beta}; \mathbf{Y})}{\partial \vartheta_i \partial \vartheta_i} &= \sum_{t=1}^T 0 + \frac{1}{P_{K,t}^2} \left(\frac{\partial H_{i,t}}{\partial \vartheta_i} \frac{P_{K,t}}{i!} - \frac{H_{i,t}^2}{i!^2} \right) \\ \frac{\partial^2 \ell^a(\boldsymbol{\beta}; \mathbf{Y})}{\partial \vartheta_i \partial \vartheta_j} &= \sum_{t=1}^T 0 - \frac{1}{P_{K,t}^2} \left(\frac{H_{i,t} H_{j,t}}{i! j!} \right) \\ \frac{\partial^2 \ell^a(\boldsymbol{\beta}; \mathbf{Y})}{\partial \vartheta_i \partial m} &= \sum_{t=1}^T 0 + \frac{1}{P_{K,t}^2} \left(\frac{\partial H_{i,t}}{\partial m} \frac{P_{K,t}}{i!} - \frac{H_{i,t}}{i!} \frac{\partial P_{K,t}}{\partial m} \right) \\ \frac{\partial^2 \ell^a(\boldsymbol{\beta}; \mathbf{Y})}{\partial \vartheta_i \partial s} &= \sum_{t=1}^T 0 + \frac{1}{P_{K,t}^2} \left(\frac{\partial H_{i,t}}{\partial s} \frac{P_{K,t}}{i!} - \frac{H_{i,t}}{i!} \frac{\partial P_{K,t}}{\partial s} \right) \end{aligned}$$

where HG_t^m , HG_t^s and $HG_t^{m,s}$ are the Gaussian contributions. Though cumbersome expressions, they clearly show the effect of the Hermite polynomial. If $K < 3$, the auxiliary model is Gaussian, and so do the score and Hessian. As K increases the variance of the indirect estimates should approach the Cramér-Rao lower bound (by Parzen (1959))

Table 2: Efficiency gains

	$K = 4$	$K = 5$	$K = 6$	$K = 7$
$Var(\hat{\mu}_{T,IM})$	35.84	4.64	0.87	0.006
$Var(\hat{\sigma}_{T,IM})$	72.84	54.25	0.94	0.65
$Var(\hat{\beta}_{T,IM})$	54.65	2.34	7.40	0.12
$Var(\hat{\alpha}_{T,IM})$	37.75	0.46	1.02	0.31

The model of interest is the α -stable density function with location 0, scale 1, asymmetry 0 and tail index 1.95. The number of simulated paths H is 1, N (first column) equals 100, 500, and 1000, and $T = 1000$. The auxiliary model is the Gram-Charlier density (13).

and Carrasco and Florens (2000)). But because of the estimation uncertainty, the criteria should choose moderate values of K .

To illustrate this, we simulate a series of length $T = 1000$ from the α -stable distribution with location 0 ($\mu = 0$), scale 1 ($\sigma = 1$), asymmetry 0 ($\beta = 0$) and tail index 1.95 ($\alpha = 1.95$). Then we estimate (13) with $K = 4$ (the minimum for identification). We repeat the procedure for $K = 5, 6, 7$. Table 1 provides the values of the AIC and IC for three choices of N and $H = 1$. As expected, by increasing K , the auxiliary model (13) provides a better approximation of the true density and the criteria picks the largest K to be the best choice. As a complement, Table 2 shows the variances of the estimated α -stable parameters. As the terms of the expansion increase, the variances decrease, in line with the intuition.

4 Monte Carlo Study

To investigate the performance of the indirect criteria, we proceed with a Monte Carlo study where the true model is a simple (but widely used) moving average of order one (MA(1)):

$$y_t = \mu + \theta u_{t-1} + u_t, \quad t = 1, \dots, T,$$

where u_t is a standardized Gaussian white noise. The number of parameters is $p = 2$. We set $\mu = 0.1$ and we simulate $R = 500$ series for $T = \{100, 1000, 10000\}$ and $\theta = \{0.2, 0.5, 0.8\}$.

The purpose of this study is to investigate the performance of the criteria when choosing among the true and nested auxiliary models.³ Indeed, since the MA(1) can be estimated by maximum likelihood, the criteria should choose it. We consider three classes of auxiliary models. The first is a set of nested autoregressive models (AR) with intercept:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \dots + \phi_r y_{t-r} + \varepsilon_t,$$

for $r = \{1, \dots, 9\}$. The second is an autoregressive moving average model of order one (ARMA(1,1)) with intercept:

$$y_t = \phi_0 + \phi y_{t-1} + \lambda \varepsilon_{t-1} + \varepsilon_t.$$

³We also did a Monte Carlo study with a Stochastic Volatility model. Results are available upon request.

The third is the true model. For all auxiliary models ε_t is a standardized Gaussian white noise. Appendix B shows the step-by-step description of the implementation.

The sample size of the simulated series $\mathbf{x}_{a,q,i}$ is set to $N = \{100, 500, 1000\}$, and $H = \{1, 10\}$. The number of parameters of the auxiliary models is given by $q = 2$ for the MA(1), $q = r + 1$ for the AR(r), and $q = 3$ for the ARMA(1,1).

Appendix C reports the criteria for $T = 100$ (tables 3–5), $T = 1000$ (tables 6–8) and $T = 10000$ (tables 9–11). For each T , the first table shows the results for $\theta = 0.2$, the second for $\theta = 0.5$, and the third for $\theta = 0.8$. Rows-wise, each table is divided first according to N and then to the auxiliary model. Columns-wise, the main division is according to H followed by a sub-division in the two criteria. Then, for each criterion, column Rate shows the percentage of times (out of $R = 500$) that an auxiliary model is chosen, while the columns Mean and Var show the sample mean and variance of the criteria (for the $R = 500$ replications).

Overall, four conclusions can be drawn. First, the criteria are precise. The variance is always small. Second, on average IC_{IM} always chooses an auxiliary AR model with equal or smaller lags than the AIC_{IM} does. This is in line with the simulation evidence reported in Andrews (1999) and Hall et al. (2007). Third, the models that have a MA component are the most chosen. Interestingly enough, the ARMA(1,1) is sometimes preferred to the MA(1), though it is less the case as θ , T , or H increases. However, in almost all cases the values of the criteria for MA(1) and ARMA(1,1) are very close. This is because the AR parameter is not significantly different from zero (results available under request), and hence the only meaningful difference between the MA(1) and ARMA(1,1) is q in the penalty term ($q = 2$ against $q = 3$). Fourth, the combination IC_{IM} and $H = 10$ seems to deliver the best results, in the sense that the model selection is consistent and stable.

Three additional insights are drawn. First, for small T the chosen auxiliary model varies more than for large T , as reflected in the column Rate. Second, in general the higher the persistence of the MA model (i.e the higher θ), the more often the true model is chosen as auxiliary model (in particular for $T = 1000$ and $T = 10000$). In other words, the discrepancy in the criteria between the MA(1) and the nested AR models widens with θ . Indeed, a small persistence in the MA(1) process is easily captured by few AR lags. Third, an increase in H stabilizes the results and increases the quality of the criteria. This is an effect of the increase in the precision of the estimates, leading automatically to an increase in the values of the likelihoods, and to a decrease of the criteria.

5 Empirical Application

We estimate a stochastic volatility (SV) model using log daily returns of IBM for two sample periods. The first, 2003–2007 (1258 days), is pre-crisis and the volatility is relatively low. By contrast, the second, 2004–2008 (1152 days), includes the most turbulent and volatile months of the latest financial crisis. The SV model we estimate is

$$\begin{aligned} y_t &= e^{h_t/2} u_t, & u_t &\sim \mathcal{N}(0, 1), \\ h_t &= \mu + \rho h_{t-1} + v_t, & v_t &\sim \mathcal{N}(0, \sigma^2), & t &= 1 \dots, T. \end{aligned}$$

Following Broto and Ruiz (2004), μ is set to zero. The number of parameters is $p = 2$. Similarly to the Monte Carlo study, we choose $N = \{100, 500, 1000\}$ and $H = \{1, 10\}$.

The first class of auxiliary models we consider are the ARCH(r) models:

$$\begin{aligned} y_t &= d_t^{1/2} z_t, & z_t &\sim \mathcal{N}(0, 1), \\ d_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + \dots + \alpha_r y_{t-r}^2, \end{aligned}$$

for $r = \{1, \dots, 9\}$. The number of auxiliary parameters is $q = r + 1$. This class of auxiliary models was proposed by Engle and Lee (1996) and Calzolari et al. (2008). We also consider a GARCH(1,1) $d_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 d_{t-1}$ for which $q = 3$.

The second class of auxiliary models we consider are AR(r) models on the log of the squared observations:

$$\ln(y_t^2) = \beta_0 + \beta_1 \ln(y_{t-1}^2) + \dots + \beta_r \ln(y_{t-r}^2) + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \varphi)$$

with $r = \{1, \dots, 9\}$. The number of auxiliary parameters is $q = r + 2$. This class of auxiliary models was proposed by Monfardini (1998). We also consider an ARMA(1,1) $\ln(y_t^2) = \beta_0 + \beta_1 \ln(y_{t-1}^2) + \beta_2 \eta_{t-1} + \eta_t$ for which $q = 3$.

Tables 12–13 in Appendix C report the criteria for the two financial periods. We extract four conclusions, in line with the findings of the MA(1) model. First, the criteria always choose the ARMA(1,1) and the GARCH(1,1) to be the best choices within each class of auxiliary model. This is because the average persistence of the SV is high ($\hat{\rho}_{T,IM}$ around 0.85). Second, regardless of the value of H , N , the ARMA model is always preferred (to a large extent) to the GARCH, suggesting that the former is closer to the SV than the latter. Third, the values of the criteria are smaller for the first financial period than for the second. This is due to the fact that the first period is described by a smaller variance of the return data, inducing higher precision of SV estimates that further leads to an increase in the values of the likelihoods, and to a decrease of the criteria.

6 Conclusion

We introduced an Akaike–type class of information criteria that can be used for choosing the auxiliary model within indirect methods. They are easy and are based on the variance–covariance matrix of the estimated parameters and the size of the auxiliary model.

Other matching–based estimation methods, such as GMM, MSM and MSQ, also require the choice of functions. Provided that these functions ensure consistency of the estimates, all methods deliver asymptotically Gaussian estimates; their precision being the only difference. In GMM and MSM q is the number of moment conditions that provide consistent estimators. Then our criteria become similar to the RMSC of Hall et al. (2007). In MSQ the issue of weak identification is not relevant since, in theory, any q –dimensional vector of functions of quantiles provide consistent estimators. We can therefore, in principle, extend the use of our criteria for different matching methods (nested and non–nested models, different moment conditions, and different functions of quantiles), and select the one that provides the smallest indirect criteria. A deeper understanding of the selection process between and within inference methods remains open and deserves further investigation.

References

- Akaike, H. (1974), 'A new look at statistical model identification', *IEEE Transactions on Automatic Control* **19**(6), 716–723.
- Andersen, T. G., Chung, H.-J. and Sorensen, B. E. (1999), 'Efficient method of moments estimation of a stochastic volatility model: a monte carlo study.', *Journal of Econometrics* **91**, 61–87.
- Andrews, D. and Lu, B. (2001), 'Consistent model and moment selection procedures for GMM estimation with applications to dynamic panel data models', *Journal of Econometrics* **101**, 123–164.
- Andrews, D. W. (2000), Consistent moment selection procedures for GMM estimation: strong consistency and simulation results. Discussion paper, Cowles Foundation for Economics.
- Andrews, D. W. K. (1999), 'Consistent moment selection procedures for generalized method of moments estimation', *Econometrica* **67**, 543–564.
- Billio, M. and Monfort, A. (2003), 'Kernel-based indirect inference', *Journal of Financial Econometrics* **1**, 297–326.
- Blasques, F. (2011), *Semi-Nonparametric Indirect Inference*, PhD thesis. University of Maastricht.
- Bozdogan, H. (1987), 'Model selections and akaike's information criteria (AIC): the general theory and its analytical extensions', *Psychometrika* **52**, 345–370.
- Broto, C. and Ruiz, E. (2004), 'Estimation methods for stochastic volatility models: a survey', *Journal of Economic Surveys* **18**, 613–649.
- Broze, L., Scaillet, O. and Zakoian, J. M. (1998), 'Quasi indirect inference for diffusion processes', *Econometric Theory* **14**, 161–186.
- Calzolari, G., Fiorentini, G. and Sentana, E. (2004), 'Constrained indirect estimation', *Review of Economic Studies* **71**, 945–973.
- Calzolari, G., Fiorentini, G. and Sentana, E. (2008), 'Indirect estimation of large conditionally heteroskedastic factor models, with an application to the dow 30 stocks', *Journal of Econometrics* **146**, 10–25.
- Carrasco, M. and Florens, J.-P. (2000), 'Generalization of GMM to a continuum of moment conditions', *Econometric Theory* **16**, 797–834.
- Chambers, J., Mallows, C. and Stuck, B. (1976), 'A method for simulating stable random variables', *Journal of the American Statistical Association* **71**, 340–344.
- Chernov, M., Galland, R., Ghysels, E. and Tauchen, G. (2003), 'Alternative models for stock price dynamics', *Journal of Econometrics* **116**, 225–257.
- Corsi, F. and Reno, R. (2012), 'Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling', *Journal of Business and Economic Statistics* **Forthcoming**.
- Czellar, V. and Calvet, L. (2011), Indirect inference for learning models. HEC Mimeo.

- Czellar, V. and Ronchetti, E. (2010), ‘Accurate and robust tests for indirect inference’, *Biometrika* **97**, 621–630.
- Dhaene, G., Gouriéroux, C. and Scaillet, O. (1998), ‘Instrumental models and indirect encompassing’, *Econometrica* **66**, 673–688.
- Dominicy, Y. and Veredas, D. (2013), ‘The method of simulated quantiles’, *Journal of Econometrics* **172**(2), 235–247.
- Dridi, R., Guay, A. and Renault, E. (2007), ‘Indirect inference and calibration of dynamic stochastic general equilibrium models’, *Journal of Econometrics* **136**, 397–430.
- Engle, R. F. and Lee, G. (1996), Estimating diffusions models of stochastic volatility, in P. E. Rossi, ed., ‘Modelling Stock market Volatility: Bridging the Gap to Continuous Time’, San Diego: Academic Press, pp. 333–335.
- Gallant, A. and Nychka, D. (1987), ‘Seminonparametric maximum likelihood estimation’, *Econometrica* **55**, 363–390.
- Gallant, A. R. and Tauchen, G. (1996), ‘Which moments to match?’, *Econometric Theory* **12**, 657–681.
- Gallant, R., Hsieh, D. and Tauchen, G. (1997), ‘Estimation of stochastic volatility models with diagnostics’, *Journal of Econometrics* **81**, 159–192.
- Garcia, R., Renault, E. and Veredas, D. (2011), ‘Estimation of stable distributions with indirect inference’, *Journal of Econometrics* **161**, 325–337.
- Genton, M. G. and Ronchetti, E. (2003), ‘Robust indirect inference’, *Journal of the American Statistical Association* **98**, 67–76.
- Ghysels, E., Khalaf, L. and Vodounou, C. (2003), ‘Simulation based inference in moving average models’, *Annales d’Economie et Statistique* **69**, 85–99.
- Gouriéroux, C. and Monfort, A. (1996), *Simulation-based Estimation Methods*, CORE Lectures, Oxford University Press.
- Gouriéroux, C., Monfort, A. and Renault, E. (1993), ‘Indirect inference’, *Journal of Applied Econometrics* **8**, S85–S118.
- Gouriéroux, C., Phillips, P. and Yu, J. (2010), ‘Indirect inference for dynamic panel models’, *Journal of Econometrics* **157**, 68–77.
- Gourieroux, C., Renault, E. and Touzi, N. (2000), Calibration by simulation for small sample bias correction, in R. Mariano, T. Schuerman and M. Weeks, eds, ‘Simulation-based Inference in Econometrics: Methods and Applications’, Cambridge University Press, chapter 13, pp. 328–358.
- Grassi, S. and de Magistris, P. S. (2012), A dynamic multi factor model for high and low frequency volatility activity. Mimeo.
- Guay, A. and Scaillet, O. (2003), ‘Indirect inference, nuisance parameter and threshold moving average models’, *Journal of Business and Economic Statistics* **21**, 122–132.

- Guvenen, F. and Smith, A. (2010), ‘Inferring labor income risk from economic choices: an indirect inference approach’, *Federal research bank of Minneapolis. Research department staff report 450*.
- Hall, A., Inoue, A., Jana, K. and Shin, C. (2007), ‘Information in generalized method of moments estimation and entropy-based moment selection’, *Journal of Econometrics* **138**, 488–512.
- Hall, A., Inoue, A., Nason, J. M. and Rossi, B. (2011), ‘Information criteria for impulse response function matching estimation of DSGE models’, *Journal of Econometrics* **Forthcoming**.
- Hall, A. and Peixe, F. (2003), ‘A consistent method for the selection of relevant instruments’, *Econometric Reviews* **22**, 269–288.
- Hannan, E. and Quinn, B. (1979), ‘The determination of order of an autoregression’, *Journal of the Royal Statistical Society Series B* **41** (2), 190–195.
- Hong, H. and Preston, B. (2005), Non-nested model selection criteria. Mimeo.
- Hong, H., Preston, B. and Shum, B. (2003), ‘Generalized empirical likelihood-based model selection criteria for moment condition moments’, *Econometric Theory* **19**, 923–943.
- Jang, T.-S. and Lux, T. (2012), Identification of social interaction effects in financial data: inference of herd behaviour via Monte Carlo simulations. Mimeo.
- Jiang, W. and Turnbull, B. (2003), ‘Indirect inference for survival data’, *Statistics and Operations Research Transactions* **27**, 79–94.
- Keane, M. and Smith, A. (2003), Generalized indirect inference for discrete choice models. Yale University, Department of Economics, mimeo.
- Kitamura, Y. (2002), A likelihood-based approach to the analysis of a class of nested and non-nested models. Mimeo department of economics, Yale University.
- Konishi, S. and Kitagawa, G. (1996), ‘Generalized information criteria in model selection’, *Biometrika* **83**, 875–890.
- Lombardi, M. and Calzolari, G. (2009), ‘Indirect estimation of α -stable stochastic volatility models’, *Computational Statistics and Data Analysis* **53**(6), 2298–2308.
- Lombardi, M. and Veredas, D. (2009), ‘Indirect estimation of elliptical stable distributions’, *Computational Statistics and Data Analysis* **53**, 2309–2324.
- Monfardini, C. (1998), ‘Estimating stochastic volatility models through indirect inference’, *Econometrics Journal* **1**, 113–128.
- Nickl, R. and Potscher, B. (2010), ‘Efficient simulation-based minimum distance estimation’, *Mathematical Methods of Statistics* **19**, 327–364.
- Parzen, E. (1959), Statistical inference on time series by Hilbert space methods. Technical report 23, Applied Mathematics and Statistics Laboratory, Stanford.
- Phillips, P. C. (2012), ‘Folklore theorems, implicit maps, and indirect inference’, *Econometrica* **80**, 425–454.

- Phillips, P. C. and Yu, J. (2009), 'Simulation-based estimation of contingent-claims prices', *Review of Financial Studies* **22**, 3669–3705.
- Schwarz, G. (1978), 'Estimating the dimension of a model', *Annals of Statistics* **6**, 461–464.
- Smith, A. (1993), 'Estimating non-linear time series models using simulated vector autoregressions: two approaches', *Journal of Applied Econometrics* **8**, 63 – 84.
- Trojani, F. and Ortelli, C. (2005), 'Robust efficient method of moments estimation', *Journal of Econometrics* **128**, 69–97.
- White, H. (1982), 'Maximum likelihood estimation of misspecified models', *Econometrica* **50**, 1–25.

A Technical appendix

We first define some quantities needed for the proofs, followed by two preliminary lemmas, and then the proofs of Propositions 1 and 2.

Definitions and preliminary results

We simplify the notation using the following definitions of the indirect methods and ML asymptotic variance–covariance matrices

$$\mathbf{W}_{0,a,q} \equiv \mathbf{W}_{a,q}(\boldsymbol{\theta}_0), \quad \mathbf{W}_{a,q} \equiv \mathbf{W}_{a,q}(\boldsymbol{\theta}), \quad \mathbf{W}_0^* \equiv \mathcal{I}^{-1}(\boldsymbol{\theta}_0), \quad \text{and} \quad \mathbf{W}^* \equiv \mathcal{I}^{-1}(\boldsymbol{\theta}). \quad (14)$$

Moreover, analogously to the q dimensional simulated vector $\mathbf{X}_{a,q} \equiv \sqrt{T}(\widehat{\boldsymbol{\theta}}_{T,IM,a,q} - \boldsymbol{\theta}_0)$, we define the p dimensional random vector $\mathbf{X}^* \equiv \sqrt{T}(\widehat{\boldsymbol{\theta}}_{T,ML} - \boldsymbol{\theta}_0)$. For $\mathbf{X}_{a,q}$ and \mathbf{X}^* we can compute their sample variance–covariance matrices $\widetilde{\mathbf{W}}_{N,a,q}$ and $\widetilde{\mathbf{W}}_N^*$ respectively. These matrices are the consistent ML estimators of $\widehat{\mathbf{W}}_{T,a,q} \equiv \widehat{\mathbf{W}}_{T,a,q}(\widehat{\boldsymbol{\theta}}_{T,IM,a,q})$ and $\widehat{\mathbf{W}}_T^* \equiv \widehat{\mathbf{W}}_T^*(\widehat{\boldsymbol{\theta}}_{T,ML})$, which in turn are estimators of $\mathbf{W}_{0,a,q}$ and \mathbf{W}_0^* . For any of the matrices \mathbf{W} defined above, we have a vectorized form, generically denoted as $\mathbf{w} \equiv \text{vech}(\mathbf{W})$.

Finally, notice that, since ML estimation of the model of interest is unfeasible, we cannot compute $\widehat{\boldsymbol{\theta}}_{T,ML}$ and therefore \mathbf{X}^* cannot be simulated. This, however, is an object needed only for the proof and plays no role in the final formulation of the criteria.

Given consistency of the sample variance–covariance matrices we have, as $N \rightarrow \infty$,

$$\sqrt{N}(\widetilde{\mathbf{w}}_{N,a,q} - \widehat{\mathbf{w}}_{T,a,q}) \sim \mathcal{N}(\mathbf{0}, \mathcal{J}_{a,q}^{-1}(\widehat{\mathbf{w}}_{T,a,q})) \quad (15)$$

and

$$\sqrt{N}(\widetilde{\mathbf{w}}_N^* - \widehat{\mathbf{w}}_T^*) \sim \mathcal{N}(\mathbf{0}, \mathcal{J}^{*-1}(\widehat{\mathbf{w}}_T^*)), \quad (16)$$

where the asymptotic variance–covariance matrices have now size $\frac{p(p+1)}{2} \times \frac{p(p+1)}{2}$ and are defined as

$$\begin{aligned} \mathcal{J}_{a,q}(\widehat{\mathbf{w}}_{T,a,q}) &= -\mathbb{E}_{\mathbf{x}_{a,q}} \left[\frac{\partial^2 \log \phi(\mathbf{X}_{a,q}; \mathbf{w}_{a,q})}{\partial \mathbf{w}_{a,q} \partial \mathbf{w}'_{a,q}} \Big|_{\mathbf{w}_{a,q} = \widehat{\mathbf{w}}_{T,a,q}} \right] \text{ and} \\ \mathcal{J}^*(\widehat{\mathbf{w}}_T^*) &= -\mathbb{E}_{\mathbf{x}^*} \left[\frac{\partial^2 \log \phi(\mathbf{X}^*; \mathbf{w}^*)}{\partial \mathbf{w}^* \partial \mathbf{w}^{*'}} \Big|_{\mathbf{w}^* = \widehat{\mathbf{w}}_T^*} \right]. \end{aligned}$$

The sample analogues of the two previous expressions are:

$$\begin{aligned} \widetilde{\mathcal{J}}_{N,a,q}(\widehat{\mathbf{w}}_{T,a,q}) &= -\frac{1}{N} \sum_{i=1}^N \frac{\partial^2 \log \phi(\mathbf{x}_{a,q,i}; \mathbf{w}_{a,q})}{\partial \mathbf{w}_{a,q} \partial \mathbf{w}'_{a,q}} \Big|_{\mathbf{w}_{a,q} = \widehat{\mathbf{w}}_{T,a,q}} \quad \text{and} \\ \widetilde{\mathcal{J}}_N^*(\widehat{\mathbf{w}}_T^*) &= -\frac{1}{N} \sum_{i=1}^N \frac{\partial^2 \log \phi(\mathbf{x}_i^*; \mathbf{w}^*)}{\partial \mathbf{w}^* \partial \mathbf{w}^{*'}} \Big|_{\mathbf{w}^* = \widehat{\mathbf{w}}_T^*}. \end{aligned}$$

We make the following assumption:

ASSUMPTION D The Hessian matrices $\mathcal{J}_{a,q}(\widehat{\mathbf{w}}_{T,a,q})$ and $\mathcal{J}^*(\widehat{\mathbf{w}}_T^*)$ exist, and are finite and positive definite.

In order to prove the following lemmas we omit all indexes referring to the class of auxiliary models a , the size of the auxiliary parameter q , and the sample size T . We also omit any reference to the value of the parameters in which the matrices defined below are computed and we implicitly assume that everything is computed in the true value of the parameters θ_0 and β_0 . We re-write the asymptotic variance-covariance matrix given in (4) as

$$\mathbf{W} = (\mathbf{B}'\mathbf{V}\mathbf{B})^{-1}, \quad (17)$$

where $\mathbf{V} = \mathbf{J}(\beta_0)\mathbf{I}^{-1}(\beta_0)\mathbf{J}(\beta_0)$ and $\mathbf{B} = \left. \frac{\partial \mathbf{b}(\theta)}{\partial \theta'} \right|_{\theta=\theta_0}$. Recall that \mathbf{V} is $q \times q$, \mathbf{B} is $q \times p$ with full column rank p (see assumption A7). We define the $q \times p$ right-inverse matrix of \mathbf{B}'

$$(\mathbf{B}')^+ = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}, \quad (18)$$

such that $\mathbf{B}'(\mathbf{B}')^+ = \mathbf{I}_p$. Notice that this matrix always exists as long as \mathbf{B}' has full column rank p . Then, consider the $q \times q$ matrix $\mathbf{B}\mathbf{B}'$ and its eigenvalue decomposition $\mathbf{B}\mathbf{B}' = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$, where the columns of \mathbf{P} are the eigenvectors ($\mathbf{P}\mathbf{P}' = \mathbf{I}_q$) and $\mathbf{\Lambda}$ is a $q \times q$ diagonal matrix of eigenvalues. Since from assumption A.6 we have $q \geq p$, $\mathbf{\Lambda}$ has entries $\lambda_i > 0$ for $i = 1, \dots, p$ and $\lambda_i = 0$ for $i = p + 1, \dots, q$. Then, we define the generalized inverse matrix

$$(\mathbf{B}\mathbf{B}')^+ = \mathbf{P}\mathbf{\Lambda}^+\mathbf{P}', \quad (19)$$

where

$$\mathbf{\Lambda}^+ = \left(\begin{array}{c|c} \mathbf{L}^{-1} & \mathbf{0}_{p \times (q-p)} \\ \hline \mathbf{0}_{(q-p) \times p} & \mathbf{0}_{(q-p) \times (q-p)} \end{array} \right) \text{ and } \mathbf{L}^{-1} = \text{diag}(\lambda_1^{-1} \dots \lambda_p^{-1}). \quad (20)$$

Using these definitions we define a $q \times p$ matrix

$$\mathbf{B}^\dagger = \mathbf{B}'(\mathbf{B}\mathbf{B}')^+. \quad (21)$$

Finally, for the $q \times q$ matrix \mathbf{V}^{-1} , we define the duplication matrix, \mathbf{D}_q of dimension $q^2 \times \frac{q(q+1)}{2}$, while for the $p \times p$ matrix \mathbf{W} , we define the duplication matrix \mathbf{D}_p of dimension $p^2 \times \frac{p(p+1)}{2}$ which has a left-inverse $\mathbf{D}_p^+ = (\mathbf{D}_p'\mathbf{D}_p)^{-1}\mathbf{D}_p'$, such that $\mathbf{D}_p^+\mathbf{D}_p = \mathbf{I}_{\frac{p(p+1)}{2}}$. These matrices are such that

$$\text{vec}(\mathbf{V}^{-1}) = \mathbf{D}_q \text{vech}(\mathbf{V}^{-1}), \quad \text{vech}(\mathbf{W}) = \mathbf{D}_p^+ \text{vec}(\mathbf{W}). \quad (22)$$

The following Lemma decomposes the asymptotic variance-covariance matrix \mathbf{W} into the binding function and the information matrix of the auxiliary model.

Lemma 1 *Let $\mathbf{w} = \text{vech}(\mathbf{W})$ be a $\frac{p(p+1)}{2}$ -dimensional vector and \mathbf{W} the asymptotic variance-covariance matrix (4). Then $\mathbf{w} = \mathbf{A}\mathbf{v}$, where \mathbf{A} is a $\frac{p(p+1)}{2} \times \frac{q(q+1)}{2}$ matrix defined as $\mathbf{A} = \mathbf{D}_p^+(\mathbf{B}^\dagger \otimes \mathbf{B}^+)\mathbf{D}_q$, and \mathbf{v} is a $\frac{q(q+1)}{2} \times 1$ vector defined as $\mathbf{v} = \text{vech}(\mathbf{V}^{-1})$.*

Proof. From (17), we have

$$\mathbf{W} = (\mathbf{B}'\mathbf{V}\mathbf{B})^{-1} = \mathbf{C}_1\mathbf{V}^{-1}\mathbf{C}_2,$$

where \mathbf{C}_1 and \mathbf{C}_2 must be such that

$$\mathbf{B}'\mathbf{V}\mathbf{B}\mathbf{C}_1\mathbf{V}^{-1}\mathbf{C}_2 = \mathbf{I}_p, \quad (23)$$

and

$$\mathbf{C}_1\mathbf{V}^{-1}\mathbf{C}_2\mathbf{B}'\mathbf{V}\mathbf{B} = \mathbf{I}_p. \quad (24)$$

We first proof the equality in (23). By substituting in (23) $\mathbf{C}_1 = \mathbf{B}^\dagger$ (see (21)) and $\mathbf{C}_2 = (\mathbf{B}')^+$ (see (18)), and using (19) and (20), we have

$$\begin{aligned}
\mathbf{B}'\mathbf{V}\mathbf{B}\mathbf{B}^\dagger\mathbf{V}^{-1}\mathbf{B}^+ &= \mathbf{B}'\mathbf{V}\mathbf{B}\mathbf{B}'(\mathbf{B}\mathbf{B}')^+\mathbf{V}^{-1}\mathbf{B}^+ \\
&= \mathbf{B}'\mathbf{V}\mathbf{P}\mathbf{\Lambda}\mathbf{P}'\mathbf{P}\mathbf{\Lambda}^+\mathbf{P}'\mathbf{V}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \\
&= \mathbf{B}'\mathbf{V}\mathbf{P} \begin{pmatrix} \mathbf{I}_p & \mathbf{0}_{p \times (q-p)} \\ \mathbf{0}_{(q-p) \times p} & \mathbf{0}_{(q-p) \times (q-p)} \end{pmatrix} \mathbf{P}'\mathbf{V}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \\
&= \mathbf{B}'\mathbf{V}\mathbf{P}^{(p)}\mathbf{P}^{(p)T}\mathbf{V}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \\
&= \mathbf{B}'\mathbf{V}\mathbf{V}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \\
&= \mathbf{B}'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \\
&= \mathbf{I}_p,
\end{aligned}$$

where $\mathbf{P}^{(p)}$ is a $q \times p$ matrix with just the first p columns of \mathbf{P} . The equality in (24) is proven analogously.

Then

$$\mathbf{W} = \mathbf{B}^\dagger\mathbf{V}^{-1}(\mathbf{B}')^+ = \mathbf{B}^\dagger\mathbf{V}^{-1}(\mathbf{B}^+)',$$

and, from the properties of the $\text{vec}(\cdot)$ operator,

$$\text{vec}(\mathbf{W}) = \text{vec}(\mathbf{B}^\dagger\mathbf{V}^{-1}(\mathbf{B}^+)') = (\mathbf{B}^\dagger \otimes \mathbf{B}^+) \text{vec}(\mathbf{V}^{-1}). \quad (25)$$

Finally, from (25), and using the duplication matrices defined in (22), we have

$$\begin{aligned}
\mathbf{w} &= \text{vech}(\mathbf{W}) = \mathbf{D}_p^+ \text{vec}(\mathbf{W}) \\
&= \mathbf{D}_p^+ (\mathbf{B}^\dagger \otimes \mathbf{B}^+) \text{vec}(\mathbf{V}^{-1}) \\
&= \mathbf{D}_p^+ (\mathbf{B}^\dagger \otimes \mathbf{B}^+) \mathbf{D}_q \text{vech}(\mathbf{V}^{-1}) \\
&= \mathbf{A}\mathbf{v}.
\end{aligned}$$

□

In order to estimate the binding function, we have to simulate H paths $(\mathbf{y}_1^h(\boldsymbol{\theta}), \dots, \mathbf{y}_{T^*}^h(\boldsymbol{\theta}))$ of length T^* not necessarily equal to T . Thus, $\mathbf{b}(\boldsymbol{\theta})$ is estimated using $\widehat{\boldsymbol{\beta}}_{HT^*}(\boldsymbol{\theta})$ as defined in (2). The error of this estimator can be made as small as we wish by taking $H \rightarrow \infty$ and $T^* \rightarrow \infty$. Moreover, we estimate derivatives numerically as

$$\frac{\partial \widehat{\boldsymbol{\beta}}_{HT^*}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \simeq \frac{-\widehat{\boldsymbol{\beta}}_{HT^*}(\boldsymbol{\theta} + 2\Delta) + 8\widehat{\boldsymbol{\beta}}_{HT^*}(\boldsymbol{\theta} + \Delta) - 8\widehat{\boldsymbol{\beta}}_{HT^*}(\boldsymbol{\theta} - \Delta) + \widehat{\boldsymbol{\beta}}_{HT^*}(\boldsymbol{\theta} - 2\Delta)}{12\Delta}$$

for some arbitrarily small Δ . Thus,

$${}_p \lim_{H, T^* \rightarrow \infty} \left(\lim_{\Delta \rightarrow 0} \frac{\partial \widehat{\boldsymbol{\beta}}_{HT^*}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) = \frac{\partial \mathbf{b}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{B}.$$

Since Δ , H , and T^* can be arbitrarily fixed, in simulations we can neglect the error made in estimating \mathbf{B} and thus consider \mathbf{A} , which involves \mathbf{B} , as non random. Intuitively, \mathbf{A} does not involve any additional information with respect to the information contained in the elements of the asymptotic variance–covariance matrix of the auxiliary model \mathbf{V}^{-1} .

We now re-establish the complete notation, and the estimator of the asymptotic variance–covariance matrix is such that

$$\widehat{\mathbf{w}}_{T,a,q} = \mathbf{A}\widehat{\mathbf{v}}_{T,a,q}.$$

Recall that for the simulated vector $\mathbf{X}_{a,q} \equiv \sqrt{T}(\widehat{\boldsymbol{\theta}}_{T,IM,a,q} - \boldsymbol{\theta}_0)$ we have a sample covariance matrix $\widetilde{\mathbf{w}}_{N,a,q}$. If we define a new vector $\mathbf{Y}_{a,q} \equiv \sqrt{T}(\widehat{\boldsymbol{\beta}}_{T,IM,a,q} - \boldsymbol{\beta}_0)$, then the sample covariance matrix of $\mathbf{Y}_{a,q}$ is $\widetilde{\mathbf{v}}_{N,a,q}$ such that

$$\widetilde{\mathbf{w}}_{N,a,q} = \mathbf{A}\widetilde{\mathbf{v}}_{N,a,q}. \quad (26)$$

The following Lemma shows the asymptotic distribution of $\widetilde{\mathbf{v}}_{N,a,q}$.

Lemma 2 *For any given auxiliary model a with parameters of dimension q , we have, for $N \rightarrow \infty$,*

$$\sqrt{N}(\widetilde{\mathbf{v}}_{N,a,q} - \widehat{\mathbf{v}}_{T,a,q}) \sim \mathcal{N}\left(\mathbf{0}, \mathbf{A}^\dagger \mathcal{J}_{a,q}^{-1}(\widehat{\mathbf{w}}_{T,a,q}) \mathbf{A}^{\dagger'}\right),$$

where N is the size of the simulated process $(\mathbf{x}_{a,q,1}, \dots, \mathbf{x}_{a,q,N})$,

$$\mathcal{J}_{a,q}(\widehat{\mathbf{w}}_{T,a,q}) = -\mathbb{E}_{\mathbf{x}_{a,q}} \left[\frac{\partial^2 \log \phi(\mathbf{X}_{a,q}; \mathbf{w}_{a,q})}{\partial \mathbf{w}_{a,q} \partial \mathbf{w}'_{a,q}} \Big|_{\mathbf{w}_{a,q} = \widehat{\mathbf{w}}_{T,a,q}} \right],$$

and $\mathbf{A}^\dagger = (\mathbf{A}'\mathbf{A})^+ \mathbf{A}$ with $(\mathbf{A}'\mathbf{A})^+$ defined analogously as in (19), but this time with $\frac{p(p+1)}{2}$ non-zero eigenvalues, thus $\mathbf{A}^\dagger \mathbf{A} = \mathbf{I}_{\frac{q(q+1)}{2}}$.

Proof. We defined $\widetilde{\mathbf{w}}_{N,a,q}$ to be the vector containing the entries of the sample covariance matrix of $\mathbf{X}_{a,q}$ which in turn is distributed as a Gaussian with zero mean and covariance matrix $\widehat{\mathbf{W}}_{T,a,q}$. Then by defining $\widehat{\mathbf{w}}_{T,a,q} = \text{vech}(\widehat{\mathbf{W}}_{T,a,q})$, we have, as $N \rightarrow \infty$,

$$\sqrt{N}(\widetilde{\mathbf{w}}_{N,a,q} - \widehat{\mathbf{w}}_{T,a,q}) \sim \mathcal{N}\left(\mathbf{0}, \mathcal{J}_{a,q}^{-1}(\widehat{\mathbf{w}}_{T,a,q})\right).$$

Using Lemma 1, the estimator in (26), and the delta method, we obtain the result. \square

Proof of Proposition 1

In this and the following proofs, we reintroduce those indexes referring to the class of auxiliary models a , the size of the auxiliary parameter q , and the observed and simulated samples sizes T and N . We also indicate the value of the parameters in which the matrices are computed.

We make the following assumption:

ASSUMPTION E The function $AIC_{IM}(a, q)$ has a unique minimum over $\mathcal{A}_{q_{\max}}$.

Once $\mathbf{X}_{a,q}$ is given by simulations, we have to find an auxiliary model with asymptotic variance–covariance matrix $\mathbf{W}_{a,q}$ that better approximates (in terms of information) the best achievable value \mathbf{W}_0^* . In order to do this, we minimize the Kullback–Leibler information (or negentropy)

$$\mathcal{D}(\mathbf{w}_0^*, \mathbf{w}_{a,q}) = 2 \mathbb{E}_{\mathbf{x}^*} \left[\log \frac{\phi(\mathbf{X}^*; \mathbf{w}_0^*)}{\phi(\mathbf{X}_{a,q}; \mathbf{w}_{a,q})} \right], \quad (27)$$

If we replace the expectation by its sample counterpart, we have to minimize

$$\widehat{\mathcal{D}}_N(\mathbf{w}_0^*, \mathbf{w}_{a,q}) = \frac{2}{N} \sum_{i=1}^N \log \frac{\phi(\mathbf{x}_i^*; \mathbf{w}_0^*)}{\phi(\mathbf{x}_{a,q,i}; \mathbf{w}_{a,q})}. \quad (28)$$

Since we can only estimate $\mathbf{w}_{a,q}$, the best we can do is to use its indirect estimator $\widehat{\mathbf{w}}_{T,a,q}$. Therefore, instead of minimizing (28) we look for the auxiliary model defined by the couple (a, q) that minimizes

$$\mathbb{E}_{\mathbf{w}} \left[\widehat{\mathcal{D}}_N(\mathbf{w}_0^*, \widehat{\mathbf{w}}_{T,a,q}) \right] = \mathbb{E}_{\mathbf{w}} \left[\frac{2}{N} \sum_{i=1}^N \log \frac{\phi(\mathbf{x}_i^*; \mathbf{w}_0^*)}{\phi(\mathbf{x}_{a,q,i}; \widehat{\mathbf{w}}_{T,a,q})} \right]. \quad (29)$$

We decompose the objective function as the sum of two terms

$$\mathcal{Q} \equiv \widehat{\mathcal{D}}_N(\mathbf{w}_0^*, \widehat{\mathbf{w}}_{T,a,q}) = \underbrace{\widehat{\mathcal{D}}_N(\mathbf{w}_0^*, \widetilde{\mathbf{w}}_{N,a,q})}_{\mathcal{P}} + \underbrace{\widehat{\mathcal{D}}_N(\widetilde{\mathbf{w}}_{N,a,q}, \widehat{\mathbf{w}}_{T,a,q})}_{\mathcal{G}}. \quad (30)$$

Term \mathcal{P} of (30) can be written as

$$\mathcal{P} = \widehat{\mathcal{D}}_N(\mathbf{w}_0^*, \widetilde{\mathbf{w}}_{N,a,q}) = \underbrace{\widehat{\mathcal{D}}_N(\mathbf{w}_0^*, \widehat{\mathbf{w}}_T^*)}_{\mathcal{A}} + \underbrace{\widehat{\mathcal{D}}_N(\widehat{\mathbf{w}}_T^*, \widetilde{\mathbf{w}}_N^*)}_{\mathcal{C}} + \underbrace{\widehat{\mathcal{D}}_N(\widetilde{\mathbf{w}}_N^*, \widetilde{\mathbf{w}}_{N,a,q})}_{\mathcal{B}}.$$

So (30) can be written as

$$\mathcal{Q} = \mathcal{A} + \mathcal{C} + \mathcal{B} + \mathcal{G}. \quad (31)$$

Both term \mathcal{A} and \mathcal{C} do not depend on the chosen auxiliary model thus it can be treated as constants in the criterion. Let us consider the other two terms \mathcal{B} and \mathcal{G} separately.

\mathcal{B} : This term is

$$\mathcal{B} = \frac{2}{N} \sum_{i=1}^N \log \frac{\phi(\mathbf{x}_i^*; \widetilde{\mathbf{w}}_N^*)}{\phi(\mathbf{x}_{a,q,i}; \widetilde{\mathbf{w}}_{N,a,q})} = \frac{2}{N} \sum_{i=1}^N \{ \log \phi(\mathbf{x}_i^*; \widetilde{\mathbf{w}}_N^*) - \log \phi(\mathbf{x}_{a,q,i}; \widetilde{\mathbf{w}}_{N,a,q}) \}.$$

Only the second term on the right hand side depends on the chosen auxiliary model, (a, q) , while the first term can be treated as a constant.

\mathcal{G} : This term can be approximated using a Taylor expansion in a neighborhood of $\widetilde{\mathbf{w}}_{N,a,q}$:

$$\begin{aligned} \mathcal{G} &\simeq -(\widehat{\mathbf{w}}_{T,a,q} - \widetilde{\mathbf{w}}_{N,a,q})' \frac{1}{N} \sum_{i=1}^N \frac{\partial^2 \log \phi(\mathbf{x}_{i,a,q}; \mathbf{w}_{a,q})}{\partial \mathbf{w}_{a,q} \partial \mathbf{w}'_{a,q}} \Big|_{\mathbf{w}_{a,q} = \widetilde{\mathbf{w}}_{N,a,q}} (\widehat{\mathbf{w}}_{T,a,q} - \widetilde{\mathbf{w}}_{N,a,q}) \\ &= (\widehat{\mathbf{w}}_{T,a,q} - \widetilde{\mathbf{w}}_{N,a,q})' \widetilde{\mathcal{J}}_{N,a,q}(\widetilde{\mathbf{w}}_{N,a,q}) (\widehat{\mathbf{w}}_{T,a,q} - \widetilde{\mathbf{w}}_{N,a,q}). \end{aligned}$$

Since using (7) we can show that

$$p \lim_{N \rightarrow \infty} \widetilde{\mathcal{J}}_{N,a,q}(\widetilde{\mathbf{w}}_{N,a,q}) = \mathcal{J}_{a,q}(\widehat{\mathbf{w}}_{T,a,q}),$$

and given (15) and lemma 2, we have for $N \rightarrow \infty$

$$\begin{aligned} N\mathcal{G} &\simeq N(\widehat{\mathbf{w}}_{T,a,q} - \widetilde{\mathbf{w}}_{N,a,q})' \mathcal{J}_{a,q}(\widehat{\mathbf{w}}_{T,a,q}) (\widehat{\mathbf{w}}_{T,a,q} - \widetilde{\mathbf{w}}_{N,a,q}) \\ &= N(\widehat{\mathbf{v}}_{T,a,q} - \widetilde{\mathbf{v}}_{N,a,q})' \mathbf{A}^\dagger \mathcal{J}_{a,q}(\widehat{\mathbf{w}}_{T,a,q}) \mathbf{A}^\dagger (\widehat{\mathbf{v}}_{T,a,q} - \widetilde{\mathbf{v}}_{N,a,q}) \sim \chi_{\frac{q(q+1)}{2}}^2. \end{aligned}$$

By combining the previous results, we have that minimizing (29) is equivalent to minimize

$$\begin{aligned} \mathbb{E}_{\mathbf{w}} [N\mathcal{Q}] &= \mathbb{E}_{\mathbf{w}} [N\mathcal{A} + N\mathcal{C} + N\mathcal{B} + N\mathcal{G}] \\ &= k - 2 \sum_{i=1}^N \log \phi(\mathbf{x}_{a,q,i}; \widetilde{\mathbf{w}}_{N,a,q}) + \frac{q(q+1)}{2}. \end{aligned}$$

where the first term k on the right hand side contains \mathcal{A} , \mathcal{C} , and the first term of \mathcal{B} , therefore it does not depend on the chosen auxiliary model, (a, q) . Thus, minimizing (29) is equivalent to minimize with respect to all possible auxiliary models (a, q) the following criterion

$$AIC_{IM}(a, q) = -2 \sum_{i=1}^N \log \phi(\mathbf{x}_{a,q,i}; \tilde{\mathbf{w}}_{N,a,q}) + \frac{q(q+1)}{2},$$

where $\tilde{\mathbf{w}}_{N,a,q} = \text{vech}(\tilde{\mathbf{W}}_{N,a,q})$ is the vector containing the entries of the sample variance–covariance matrix of the simulated vector $\mathbf{X}_{a,q}$. \square

Proof of Proposition 2

For each model $(a, q) \in \mathcal{A}^*$, we have the criterion

$$IC_{IM}(a, q) = -2 \sum_{i=1}^N \log \phi(\mathbf{x}_{a,q,i}; \tilde{\mathbf{w}}_{N,a,q}) + h(q)K_N,$$

If we could compute the criterion in correspondence of the true model, we would have

$$IC_{IM}(a^*, p) = -2 \sum_{i=1}^N \log \phi(\mathbf{x}_i^*; \tilde{\mathbf{w}}_N^*) + h(p)K_N.$$

Therefore,

$$\begin{aligned} \frac{1}{N} [IC_{IM}(a, q) - IC_{IM}(a^*, p)] &= \frac{2}{N} \sum_{i=1}^N \log \frac{\phi(\mathbf{x}_i^*; \tilde{\mathbf{w}}_N^*)}{\phi(\mathbf{x}_{a,q,i}; \tilde{\mathbf{w}}_{N,a,q})} + (h(q) - h(p)) \frac{K_N}{N} \\ &= \hat{\mathcal{D}}_N(\tilde{\mathbf{w}}_N^*, \tilde{\mathbf{w}}_{N,a,q}) + (h(q) - h(p)) \frac{K_N}{N}. \end{aligned} \quad (32)$$

The first term on the right–hand–side of (32) can be decomposed as

$$\begin{aligned} \hat{\mathcal{D}}_N(\tilde{\mathbf{w}}_N^*, \tilde{\mathbf{w}}_{N,a,q}) &= \hat{\mathcal{D}}_N(\tilde{\mathbf{w}}_N^*, \hat{\mathbf{w}}_T^*) + \hat{\mathcal{D}}_N(\hat{\mathbf{w}}_T^*, \mathbf{w}_0^*) + \hat{\mathcal{D}}_N(\mathbf{w}_0^*, \mathbf{w}_{0,a,q}) + \\ &\quad + \hat{\mathcal{D}}_N(\mathbf{w}_{0,a,q}, \hat{\mathbf{w}}_{T,a,q}) + \hat{\mathcal{D}}_N(\hat{\mathbf{w}}_{T,a,q}, \tilde{\mathbf{w}}_{N,a,q}). \end{aligned} \quad (33)$$

From (15) and (16), and using Taylor expansions, we have that as $N \rightarrow \infty$ the first and last term of (33) are $O_p(N^{-1})$. In a similar way, using Taylor approximations we write

$$\hat{\mathcal{D}}_N(\mathbf{w}_{0,a,q}, \hat{\mathbf{w}}_{T,a,q}) \simeq (\mathbf{w}_{0,a,q} - \hat{\mathbf{w}}_{T,a,q})' \mathcal{J}_{a,q}(\hat{\mathbf{w}}_{T,a,q})(\mathbf{w}_{0,a,q} - \hat{\mathbf{w}}_{T,a,q}).$$

From Gouriéroux et al. (1993) we have, as $T \rightarrow \infty$,

$$p \lim_{T \rightarrow \infty} \hat{\mathbf{w}}_{T,a,q} = \mathbf{w}_{0,a,q},$$

which, provided that $\mathcal{J}_{a,q}$ is bounded by Assumption D, implies $\|\mathbf{w}_{0,a,q} - \hat{\mathbf{w}}_{T,a,q}\| = O_p(T^{-1/2})$ and therefore the fourth term of (33) is $O_p(T^{-1})$. Analogously from Maximum Likelihood theory we have that the second term in (33) can be approximated as

$$\hat{\mathcal{D}}_N(\hat{\mathbf{w}}_T^*, \mathbf{w}_0^*) \simeq (\mathbf{w}_0^* - \hat{\mathbf{w}}_T^*)' \mathcal{J}^*(\hat{\mathbf{w}}_T^*)(\mathbf{w}_0^* - \hat{\mathbf{w}}_T^*),$$

which, given consistency of $\widehat{\mathbf{w}}_T^*$ as $T \rightarrow \infty$, and boundedness of \mathcal{J}^* (see Assumption D), is also $O_p(T^{-1})$.

Therefore, (32) becomes

$$\frac{1}{N}[IC_{IM}(a, q) - IC_{IM}(a^*, p)] = \widehat{\mathcal{D}}_N(\mathbf{w}_0^*, \mathbf{w}_{0,a,q}) + O_p(N^{-1}) + O_p(T^{-1}) + (h(q) - h(p))\frac{K_N}{N}.$$

And, since $K_N/N = o(1)$,

$$\begin{aligned} p \lim_{N \rightarrow \infty} \left\{ \frac{1}{N}[IC_{IM}(a, q) - IC_{IM}(a^*, p)] \right\} &= p \lim_{N \rightarrow \infty} \left\{ \widehat{\mathcal{D}}_N(\mathbf{w}_0^*, \mathbf{w}_{0,a,q}) \right\} + O_p(T^{-1}) \\ &= \mathcal{D}(\mathbf{w}_0^*, \mathbf{w}_{0,a,q}) + O_p(T^{-1}). \end{aligned}$$

If we also let $T \rightarrow \infty$, we have

$$p \lim_{N, T \rightarrow \infty} \left\{ \frac{1}{N}[IC_{IM}(a, q) - IC_{IM}(a^*, p)] \right\} = \mathcal{D}(\mathbf{w}_0^*, \mathbf{w}_{0,a,q}) \geq 0, \quad (34)$$

where the last inequality follows from the definition of Kullback–Leibler information which would be zero when $\mathbf{w}_{0,a,q} = \mathbf{w}_0^*$.

From (34), we have that, as $N, T \rightarrow \infty$,

$$IC_{IM}(a^*, p) \leq IC_{IM}(a, q), \text{ for any } (a, q) \in \mathcal{A}^*,$$

with probability one. Moreover, given the definition of (\tilde{a}, \tilde{q}) and A.8, we also have

$$IC_{IM}(\tilde{a}, \tilde{q}) < IC_{IM}(a, q), \text{ for any } (a, q) \in \mathcal{A}_{q_{\max}}.$$

Therefore, as $N, T \rightarrow \infty$, we have that with probability one and when considering all models in \mathcal{A}^* , the models (\tilde{a}, \tilde{q}) and (a^*, p) will coincide. In other words we would select the true model if it is included in the set of models we are considering. \square

B Step-by-step implementation

Step 1: Choose an auxiliary model $\mathcal{M}^a(\mathbf{y}, \boldsymbol{\beta})$ and estimate $\boldsymbol{\beta}$ from the observations to obtain $\widehat{\boldsymbol{\beta}}_T$.

Step 2: Find the Indirect Inference estimator of $\boldsymbol{\theta}$:

$$\widehat{\boldsymbol{\theta}}_{HT,IM,a,q} = \arg \min_{\boldsymbol{\theta} \in \Theta} \left(\widehat{\boldsymbol{\beta}}_{HT}(\boldsymbol{\theta}) - \widehat{\boldsymbol{\beta}}_T \right)' \boldsymbol{\Omega}^* \left(\widehat{\boldsymbol{\beta}}_{HT}(\boldsymbol{\theta}) - \widehat{\boldsymbol{\beta}}_T \right),$$

where $\boldsymbol{\Omega}^*$ is the optimal weighting matrix. The algorithm for the minimization starts at some initial value $\boldsymbol{\theta}^0$. The corresponding simulated H paths of length T are $\mathbf{y}_1^h(\boldsymbol{\theta}^0), \dots, \mathbf{y}_T^h(\boldsymbol{\theta}^0)$, $h = 1, \dots, H$, and the estimates $\widehat{\boldsymbol{\beta}}_{HT}(\boldsymbol{\theta}^0)$. Note that the simulated paths can be of size $T^* > T$.

Step 3: Let $\mathbf{X}_{a,q} \equiv \sqrt{T}(\widehat{\boldsymbol{\theta}}_{T,IM,a,q} - \boldsymbol{\theta}_0)$. Simulate a sample of length N from the $\mathcal{N}(\mathbf{0}, \widehat{\mathbf{W}}_{T,a,q}(\widehat{\boldsymbol{\theta}}_{T,IM,a,q}))$. Denote the i -th simulated observation of $\mathbf{X}_{a,q}$ as the p -dimensional vector $\mathbf{x}_{a,q,i} = (x_{1,a,q,i} \dots x_{p,a,q,i})'$.

Step 5: From the N simulated series $\mathbf{x}_{a,q,i}$ compute the sample variance-covariance matrix $\widetilde{\mathbf{W}}_{N,a,q}$ and the preferred information criterion given in (8) or (12).

Step 6: Select another (nested or non-nested) auxiliary model and repeat steps 1–5.

Step 7: Choose the auxiliary model that provides the smallest value for the preferred information criterion.

C Tables

Table 3: Monte Carlo results for $\theta = 0.2$ and $T = 100$

N	Aux. Model	$H = 1$						$H = 10$					
		AIC_{IM}			IC_{IM}			AIC_{IM}			IC_{IM}		
		Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var
100	MA(1)	17.26	14.0973	0.3205	34.32	<i>14.7035</i>	0.3205	14.08	12.6397	0.3389	46.76	<i>13.2460</i>	0.3389
	ARMA(1,1)	19.37	14.8071	3.5881	12.84	16.0195	3.5881	49.86	11.8871	4.6739	42.82	13.0995	4.6739
	AR(1)	28	<i>13.9719</i>	0.0786	52.42	14.5782	0.0786	0	13.0837	0.1062	8.73	13.6899	0.1062
	AR(2)	2.74	14.0060	0.1038	0.21	15.2185	0.1038	7.61	12.5216	0.0840	1.69	13.7340	0.0840
	AR(3)	20.84	13.8557	0.1847	0.21	15.8765	0.1847	7.61	<i>12.4187</i>	0.1371	0	14.4395	0.1371
	AR(4)	5.47	14.1106	0.2586	0	17.1418	0.2586	3.38	12.5417	0.1660	0	15.5728	0.1660
	AR(5)	1.05	14.4079	0.2430	0	18.6515	0.2430	0.85	12.5933	0.1492	0	16.8369	0.1492
	AR(6)	1.47	14.4145	0.2732	0	20.0726	0.2732	3.94	12.5381	0.1663	0	18.1962	0.1663
	AR(7)	0.42	14.7021	0.3171	0	21.9768	0.3171	1.69	12.6004	0.1821	0	19.8751	0.1821
	AR(8)	0.42	14.5790	0.3118	0	23.6724	0.3118	3.1	12.5393	0.1882	0	21.6328	0.1882
AR(9)	2.95	14.5624	0.3490	0	25.6766	0.3490	7.89	12.5510	0.1912	0	23.6652	0.1912	
500	MA(1)	15.37	14.1607	0.3205	29.26	<i>14.3864</i>	0.3205	9.86	12.7031	0.3389	28.73	<i>12.9288</i>	0.3389
	ARMA(1,1)	19.58	14.8464	3.5881	18.32	15.2979	3.5881	45.07	11.9264	4.6739	52.96	12.3779	4.6739
	AR(1)	23.79	<i>14.0353</i>	0.0786	43.79	14.2610	0.0786	0	13.1470	0.1062	0	13.3728	0.1062
	AR(2)	1.89	14.0454	0.1038	2.95	14.4969	0.1038	0.85	12.5609	0.0840	15.49	13.0124	0.0840
	AR(3)	20.63	13.8631	0.1847	5.05	14.6155	0.1847	3.66	12.4261	0.1371	2.82	13.1785	0.1371
	AR(4)	4.84	14.0780	0.2586	0.63	15.2066	0.2586	1.97	12.5091	0.1660	0	13.6377	0.1660
	AR(5)	1.68	14.3273	0.2430	0	15.9074	0.2430	0.85	12.5127	0.1492	0	14.0928	0.1492
	AR(6)	1.89	14.2779	0.2732	0	16.3847	0.2732	1.69	12.4015	0.1663	0	14.5082	0.1663
	AR(7)	0.63	14.5014	0.3171	0	17.2102	0.3171	0.85	12.3997	0.1821	0	15.1085	0.1821
	AR(8)	1.05	14.3064	0.3118	0	17.6923	0.3118	6.48	12.2667	0.1882	0	15.6526	0.1882
AR(9)	8.63	14.2098	0.3490	0	18.3481	0.3490	28.73	<i>12.1983</i>	0.1912	0	16.3367	0.1912	
1000	MA(1)	15.16	14.1852	0.3205	25.47	<i>14.3253</i>	0.3205	7.61	12.7276	0.3389	23.94	12.8677	0.3389
	ARMA(1,1)	19.16	14.8679	3.5881	19.37	15.1483	3.5881	43.94	11.9479	4.6739	53.52	12.2282	4.6739
	AR(1)	23.37	<i>14.0598</i>	0.0786	37.68	14.1999	0.0786	0	13.1716	0.1062	0	13.3117	0.1062
	AR(2)	1.68	14.0669	0.1038	2.53	14.3472	0.1038	0.85	12.5824	0.0840	14.08	<i>12.8627</i>	0.0840
	AR(3)	20.63	13.8806	0.1847	12.63	14.3477	0.1847	3.66	12.4436	0.1371	5.92	12.9108	0.1371
	AR(4)	4.84	14.0905	0.2586	2.11	14.7912	0.2586	1.69	12.5216	0.1660	2.25	13.2223	0.1660
	AR(5)	1.47	14.3338	0.2430	0.21	15.3148	0.2430	0.85	12.5192	0.1492	0	13.5002	0.1492
	AR(6)	1.89	14.2774	0.2732	0	15.5854	0.2732	1.13	12.4010	0.1663	0	13.7090	0.1663
	AR(7)	0.63	14.4929	0.3171	0	16.1747	0.3171	1.13	12.3912	0.1821	0.28	14.0730	0.1821
	AR(8)	1.05	14.2889	0.3118	0	16.3911	0.3118	7.04	12.2492	0.1882	0	14.3515	0.1882
AR(9)	10.11	14.1823	0.3490	0	16.7517	0.3490	32.11	<i>12.1708</i>	0.1912	0	14.7403	0.1912	

The model of interest is a $MA(1)$ with intercept 0.1, moving average parameter 0.2 and variance of the errors equal to 1. The total number of replications is $R = 500$ and the simulated time series contains $T = 100$ observations. The number of simulated paths H is 1 and 10, and N (first column) equals 100, 500, and 1000. The auxiliary model is either the model of interest itself or an $AR(r)$ or the $ARMA(1,1)$ (column Aux. Model). The table reports the percentage of choosing each model as the best (column Rate), the average value and the variance of the criteria (columns Mean and Var respectively) over the 500 replications. Numbers in bold denote the model chosen by the criteria. Numbers in italics denote the second best model chosen by the criteria.

Table 4: Monte Carlo results for $\theta = 0.5$ and $T = 100$

N	Aux. Model	$H = 1$						$H = 10$					
		AIC_{IM}			IC_{IM}			AIC_{IM}			IC_{IM}		
		Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var
100	MA(1)	29.05	13.0701	0.5173	78.84	13.6763	0.5173	25.92	<i>11.3202</i>	0.6659	63.28	11.9264	0.6659
	ARMA(1,1)	18.26	13.5805	1.1093	17.84	<i>14.7929</i>	1.1093	47.73	11.0355	1.1745	35.85	<i>12.2479</i>	1.1745
	AR(1)	0	14.5374	0.3582	1.87	15.1436	0.3582	0	13.8520	0.6153	0	14.4582	0.6153
	AR(2)	0.62	13.8127	0.2550	0.83	15.0252	0.2550	0	12.3763	0.1415	0.65	13.5887	0.1415
	AR(3)	7.68	<i>13.0648</i>	0.1401	0.62	15.0855	0.1401	1.08	11.7463	0.0986	0.22	13.7670	0.0986
	AR(4)	19.92	12.9058	0.2839	0	15.9369	0.2839	2.59	11.5580	0.1774	0	14.5891	0.1774
	AR(5)	3.73	13.1786	0.4104	0	17.4222	0.4104	3.24	11.6047	0.2442	0	15.8483	0.2442
	AR(6)	12.86	13.1249	0.5295	0	18.7830	0.5295	7.13	11.5180	0.3011	0	17.1761	0.3011
	AR(7)	2.7	13.4825	0.5663	0	20.7572	0.5663	1.94	11.6115	0.3035	0	18.8862	0.3035
	AR(8)	2.07	13.5173	0.5215	0	22.6107	0.5215	3.67	11.5493	0.3243	0	20.6428	0.3243
AR(9)	3.11	13.5119	0.4611	0	24.6261	0.4611	6.7	11.5599	0.3511	0	22.6741	0.3511	
500	MA(1)	24.48	13.1334	0.5173	54.77	13.3592	0.5173	22.46	11.3835	0.6659	42.76	<i>11.6093</i>	0.6659
	ARMA(1,1)	16.18	13.6198	1.1093	21.37	14.0713	1.1093	42.33	11.0748	1.1745	50.97	11.5263	1.1745
	AR(1)	0	14.6008	0.3582	0.41	14.8265	0.3582	0	13.9154	0.6153	0	14.1411	0.6153
	AR(2)	0.62	13.8521	0.2550	1.45	14.3035	0.2550	0	12.4156	0.1415	0.43	12.8671	0.1415
	AR(3)	4.77	13.0721	0.1401	12.45	<i>13.8246</i>	0.1401	0.65	11.7536	0.0986	3.67	12.5061	0.0986
	AR(4)	18.88	12.8731	0.2839	8.71	14.0018	0.2839	1.73	11.5253	0.1774	1.94	12.6540	0.1774
	AR(5)	4.15	13.0979	0.4104	0.62	14.6780	0.4104	1.94	11.5241	0.2442	0.22	13.1042	0.2442
	AR(6)	15.15	<i>12.9882</i>	0.5295	0.21	15.0950	0.5295	4.54	11.3814	0.3011	0	13.4882	0.3011
	AR(7)	3.53	13.2819	0.5663	0	15.9906	0.5663	1.73	11.4108	0.3035	0	14.1196	0.3035
	AR(8)	3.53	13.2446	0.5215	0	16.6305	0.5215	6.26	11.2767	0.3243	0	14.6626	0.3243
AR(9)	8.71	13.1593	0.4611	0	17.2976	0.4611	18.36	<i>11.2073</i>	0.3511	0	15.3457	0.3511	
1000	MA(1)	23.24	13.1579	0.5173	46.68	13.2981	0.5173	21.38	11.4080	0.6659	38.66	<i>11.5482</i>	0.6659
	ARMA(1,1)	15.98	13.6413	1.1093	20.95	13.9216	1.1093	41.04	11.0964	1.1745	51.62	11.3767	4.6739
	AR(1)	0	14.6253	0.3582	0	14.7654	0.3582	0	13.9399	0.6153	0	14.0800	0.1062
	AR(2)	0.62	13.8736	0.2550	1.66	14.1539	0.2550	0	12.4371	0.1415	0	12.7174	0.0840
	AR(3)	4.36	13.0897	0.1401	13.49	<i>13.5568</i>	0.1401	0.65	11.7711	0.0986	3.02	12.2383	0.1371
	AR(4)	19.71	12.8856	0.2839	12.86	13.5864	0.2839	1.51	11.5378	0.1774	4.32	12.2386	0.1660
	AR(5)	4.15	13.1044	0.4104	1.45	14.0855	0.4104	1.94	11.5306	0.2442	1.08	12.5117	0.1492
	AR(6)	15.56	<i>12.9877</i>	0.5295	2.28	14.2958	0.5295	4.54	11.3809	0.3011	0.65	12.6890	0.1663
	AR(7)	3.53	13.2734	0.5663	0.62	14.9552	0.5663	1.51	11.4023	0.3035	0.65	13.0841	0.1821
	AR(8)	3.32	13.2271	0.5215	0	15.3294	0.5215	6.7	11.2592	0.3243	0	13.3615	0.1882
AR(9)	9.54	13.1318	0.4611	0	15.7012	0.4611	20.73	<i>11.1798</i>	0.3511	0	13.7492	0.1912	

The model of interest is a $MA(1)$ with intercept 0.1, moving average parameter 0.5 and variance of the errors equal to 1. The total number of replications is $R = 500$ and the simulated time series contains $T = 100$ observations. The number of simulated paths H is 1 and 10, and N (first column) equals 100, 500, and 1000. The auxiliary model is either the model of interest itself or an $AR(r)$ or the $ARMA(1,1)$ (column Aux. Model). The table reports the percentage of choosing each model as the best (column Rate), the average value and the variance of the criteria (columns Mean and Var respectively) over the 500 replications. Numbers in bold denote the model chosen by the criteria. Numbers in italics denote the second best model chosen by the criteria.

Table 5: Monte Carlo results for $\theta = 0.8$ and $T = 100$

N	Aux. Model	$H = 1$						$H = 10$					
		AIC_{IM}			IC_{IM}			AIC_{IM}			IC_{IM}		
		Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var
100	MA(1)	70.08	10.8911	1.9758	87.72	11.4973	1.9758	0	16.8842	0.0295	0	17.4904	0.0295
	ARMA(1,1)	15.6	12.6694	4.5505	11.76	<i>13.8818</i>	4.5505	0	17.2052	0.0716	0	18.4176	0.0716
	AR(1)	0	15.2524	0.5712	0	15.8586	0.5712	0	14.7601	1.0449	2.35	15.3664	1.0449
	AR(2)	0	14.5680	0.9203	0.26	15.7805	0.9203	0	13.2518	0.8543	27.94	14.4643	0.8543
	AR(3)	0	13.2360	0.3697	0	15.2568	0.3697	0.59	11.9620	0.4255	60.88	13.9827	0.4255
	AR(4)	0.51	12.4958	0.2334	0.26	15.5269	0.2334	0.29	11.2841	0.3663	8.82	<i>14.3153</i>	0.3663
	AR(5)	0.51	12.4400	0.2977	0	16.6836	0.2977	1.18	10.9924	0.2773	0	15.2360	0.2773
	AR(6)	11.25	<i>11.9672</i>	0.2530	0	17.6253	0.2530	20.59	10.6026	0.2063	0	16.2608	0.2063
	AR(7)	1.28	12.2654	0.3025	0	19.5402	0.3025	5.88	10.6000	0.2162	0	17.8748	0.2162
	AR(8)	0.77	12.4884	0.3713	0	21.5818	0.3713	17.94	<i>10.4697</i>	0.2017	0	19.5631	0.2017
AR(9)	0	12.7738	0.4273	0	23.8879	0.4273	53.53	10.3999	0.2307	0	21.5140	0.2307	
500	MA(1)	65.98	10.9544	1.9758	80.82	11.1802	1.9758	0	16.9476	0.0295	0	17.1733	0.0295
	ARMA(1,1)	14.32	12.7087	4.5505	16.62	<i>13.1602</i>	4.5505	0	17.2445	0.0716	0	17.6960	0.0716
	AR(1)	0	15.3158	0.5712	0	15.5415	0.5712	0	14.8235	1.0449	0	15.0492	1.0449
	AR(2)	0	14.6074	0.9203	0	15.0589	0.9203	0	13.2912	0.8543	1.76	13.7426	0.8543
	AR(3)	0	13.2434	0.3697	0	13.9958	0.3697	0	11.9693	0.4255	14.71	12.7218	0.4255
	AR(4)	0.51	12.4631	0.2334	2.05	13.5918	0.2334	0.29	11.2515	0.3663	52.94	12.3801	0.3663
	AR(5)	0.51	12.3593	0.2977	0	13.9394	0.2977	0.29	10.9118	0.2773	15.88	<i>12.4919</i>	0.2773
	AR(6)	13.55	<i>11.8306</i>	0.2530	0.51	13.9373	0.2530	6.18	10.4660	0.2063	14.41	12.5728	0.2063
	AR(7)	2.56	12.0648	0.3025	0	14.7735	0.3025	3.24	10.3994	0.2162	0.29	13.1081	0.2162
	AR(8)	1.53	12.2158	0.3713	0	15.6017	0.3713	11.47	<i>10.1970</i>	0.2017	0	13.5829	0.2017
AR(9)	1.02	12.4211	0.4273	0	16.5595	0.4273	78.53	10.0472	0.2307	0	14.1856	0.2307	
1000	MA(1)	65.47	10.9789	1.9758	78.77	11.1191	1.9758	0	16.9721	0.0295	0	17.1122	0.0295
	ARMA(1,1)	14.32	12.7302	4.5505	16.88	<i>13.0105</i>	4.5505	0	17.2660	0.0716	0	17.5463	0.0716
	AR(1)	0	15.3403	0.5712	0	15.4804	0.5712	0	14.8480	1.0449	0	14.9882	1.0449
	AR(2)	0	14.6289	0.9203	0	14.9092	0.9203	0	13.3127	0.8543	0	13.5930	0.8543
	AR(3)	0	13.2609	0.3697	0	13.7281	0.3697	0	11.9868	0.4255	4.71	12.4540	0.4255
	AR(4)	0.51	12.4756	0.2334	1.02	13.1764	0.2334	0	11.2640	0.3663	23.24	11.9648	0.3663
	AR(5)	0.51	12.3658	0.2977	0	13.3469	0.2977	0.29	10.9183	0.2773	15	<i>11.8994</i>	0.2773
	AR(6)	13.81	<i>11.8301</i>	0.2530	3.32	13.1381	0.2530	5	10.4655	0.2063	50	11.7736	0.2063
	AR(7)	2.56	12.0563	0.3025	0	13.7381	0.3025	2.65	10.3909	0.2162	4.12	12.0727	0.2162
	AR(8)	1.79	12.1983	0.3713	0	14.3006	0.3713	10.29	<i>10.1795</i>	0.2017	1.76	12.2818	0.2017
AR(9)	1.02	12.3936	0.4273	0	14.9631	0.4273	81.76	10.0197	0.2307	1.18	12.5892	0.2307	

The model of interest is a $MA(1)$ with intercept 0.1, moving average parameter 0.8 and variance of the errors equal to 1. The total number of replications is $R = 500$ and the simulated time series contains $T = 100$ observations. The number of simulated paths H is 1 and 10, and N (first column) equals 100, 500, and 1000. The auxiliary model is either the model of interest itself or an $AR(r)$ or the $ARMA(1,1)$ (column Aux. Model). The table reports the percentage of choosing each model as the best (column Rate), the average value and the variance of the criteria (columns Mean and Var respectively) over the 500 replications. Numbers in bold denote the model chosen by the criteria. Numbers in italics denote the second best model chosen by the criteria.

Table 6: Monte Carlo results for $\theta = 0.2$ and $T = 1000$

N	Aux. Model	$H = 1$						$H = 10$					
		AIC_{TM}			IC_{TM}			AIC_{TM}			IC_{TM}		
		Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var
100	MA(1)	0	<i>18.3589</i>	0.0315	51	18.9652	0.0315	2	<i>17.2088</i>	0.0310	87.2	17.8151	0.0310
	ARMA(1,1)	85.4	17.7696	0.3599	48.8	<i>18.9820</i>	0.3599	61.8	17.0790	0.2859	12.2	18.2914	0.2859
	AR(1)	0	18.8423	0.0084	0.2	19.4485	0.0084	0	17.6215	0.0083	0.6	<i>18.2277</i>	0.0083
	AR(2)	0	18.5198	0.0098	0	19.7323	0.0098	30.8	17.2479	0.0111	0	18.4603	0.0111
	AR(3)	10.6	18.4211	0.0127	0	20.4418	0.0127	4.6	17.2926	0.0150	0	19.3134	0.0150
	AR(4)	4	18.4259	0.0147	0	21.4570	0.0147	0.8	17.3543	0.0157	0	20.3855	0.0157
	AR(5)	0	18.5322	0.0151	0	22.7758	0.0151	0	17.4110	0.0159	0	21.6546	0.0159
	AR(6)	0	18.5816	0.0158	0	24.2397	0.0158	0	17.4805	0.0161	0	23.1386	0.0161
	AR(7)	0	18.7730	0.0162	0	26.0478	0.0162	0	17.5619	0.0165	0	24.8366	0.0165
	AR(8)	0	18.8083	0.0160	0	27.9017	0.0160	0	17.6617	0.0165	0	26.7551	0.0165
AR(9)	0	18.9610	0.0163	0	30.0752	0.0163	0	17.7602	0.0167	0	28.8744	0.0167	
500	MA(1)	0	18.4223	0.0315	20.2	<i>18.6480</i>	0.0315	0	<i>17.2722</i>	0.0310	52.6	17.4979	0.0310
	ARMA(1,1)	82.8	17.8089	0.3599	79.4	18.2604	0.3599	62.6	17.1183	0.2859	43.8	<i>17.5698</i>	0.2859
	AR(1)	0	18.9057	0.0084	0.2	19.1314	0.0084	0	17.6849	0.0083	0.4	17.9106	0.0083
	AR(2)	0	18.5592	0.0098	0.2	19.0106	0.0098	23.6	17.2872	0.0111	3.2	17.7387	0.0111
	AR(3)	0.4	18.4284	0.0127	0	19.1809	0.0127	7.8	17.3000	0.0150	0	18.0524	0.0150
	AR(4)	15.6	<i>18.3932</i>	0.0147	0	19.5219	0.0147	2.4	17.3217	0.0157	0	18.4503	0.0157
	AR(5)	0.2	18.4516	0.0151	0	20.0317	0.0151	2	17.3304	0.0159	0	18.9105	0.0159
	AR(6)	0.6	18.4449	0.0158	0	20.5517	0.0158	0.8	17.3438	0.0161	0	19.4506	0.0161
	AR(7)	0	18.5724	0.0162	0	21.2811	0.0162	0.6	17.3613	0.0165	0	20.0700	0.0165
	AR(8)	0.4	18.5356	0.0160	0	21.9215	0.0160	0	17.3891	0.0165	0	20.7750	0.0165
AR(9)	0	18.6084	0.0163	0	22.7467	0.0163	0.2	17.4076	0.0167	0	21.5459	0.0167	
1000	MA(1)	0	18.4468	0.0315	14.2	<i>18.5869</i>	0.0315	0	<i>17.2967</i>	0.0310	34.4	<i>17.4369</i>	0.0310
	ARMA(1,1)	82.6	17.8304	0.3599	83.6	18.1107	0.3599	62.2	17.1398	0.2859	52	17.4201	0.2859
	AR(1)	0	18.9302	0.0084	0	19.0703	0.0084	0	17.7094	0.0083	0	17.8495	0.0083
	AR(2)	0	18.5807	0.0098	2	18.8610	0.0098	21.2	17.3087	0.0111	13.6	17.5890	0.0111
	AR(3)	0	18.4459	0.0127	0.2	18.9131	0.0127	7.4	17.3175	0.0150	0	17.7847	0.0150
	AR(4)	15.6	<i>18.4057</i>	0.0147	0	19.1065	0.0147	2.8	17.3342	0.0157	0	18.0350	0.0157
	AR(5)	0	18.4581	0.0151	0	19.4391	0.0151	3	17.3369	0.0159	0	18.3179	0.0159
	AR(6)	1.4	18.4444	0.0158	0	19.7525	0.0158	1.4	17.3433	0.0161	0	18.6514	0.0161
	AR(7)	0	18.5639	0.0162	0	20.2457	0.0162	1	17.3528	0.0165	0	19.0346	0.0165
	AR(8)	0.4	18.5181	0.0160	0	20.6204	0.0160	0.4	17.3716	0.0165	0	19.4739	0.0165
AR(9)	0	18.5809	0.0163	0	21.1503	0.0163	0.6	17.3801	0.0167	0	19.9495	0.0167	

The model of interest is a $MA(1)$ with intercept 0.1, moving average parameter 0.2 and variance of the errors equal to 1. The total number of replications is $R = 500$ and the simulated time series contains $T = 100$ observations. The number of simulated paths H is 1 and 10, and N (first column) equals 100, 500, and 1000. The auxiliary model is either the model of interest itself or an $AR(r)$ or the $ARMA(1,1)$ (column Aux. Model). The table reports the percentage of choosing each model as the best (column Rate), the average value and the variance of the criteria (columns Mean and Var respectively) over the 500 replications. Numbers in bold denote the model chosen by the criteria. Numbers in italics denote the second best model chosen by the criteria.

Table 7: Monte Carlo results for $\theta = 0.5$ and $T = 1000$

N	Aux. Model	$H = 1$						$H = 10$					
		AIC_{IM}			IC_{IM}			AIC_{IM}			IC_{IM}		
		Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var
100	MA(1)	5.02	<i>17.2099</i>	0.0462	92.37	17.8162	0.0462	35.07	<i>16.0321</i>	0.0486	99.8	16.6383	0.0486
	ARMA(1,1)	88.96	16.9195	0.1483	7.63	<i>18.1319</i>	0.1483	57.72	15.9898	0.1376	0.2	<i>17.2022</i>	0.1376
	AR(1)	0	19.8801	0.1290	0	20.4863	0.1290	0	18.6136	0.1245	0	19.2198	0.1245
	AR(2)	0	18.4603	0.0158	0	19.6727	0.0158	0	17.0471	0.0101	0	18.2596	0.0101
	AR(3)	0	17.8040	0.0074	0	19.8248	0.0074	0	16.5178	0.0082	0	18.5386	0.0082
	AR(4)	1.2	17.5148	0.0123	0	20.5459	0.0123	1	16.3317	0.0142	0	19.3628	0.0142
	AR(5)	0.4	17.4849	0.0189	0	21.7285	0.0189	2.2	16.2873	0.0211	0	20.5309	0.0211
	AR(6)	4.22	17.4521	0.0241	0	23.1103	0.0241	3.01	16.3085	0.0256	0	21.9666	0.0256
	AR(7)	0	17.6483	0.0277	0	24.9230	0.0277	0.6	16.3666	0.0286	0	23.6413	0.0286
	AR(8)	0.2	17.6907	0.0286	0	26.7841	0.0286	0.2	16.4549	0.0302	0	25.5484	0.0302
AR(9)	0	17.8780	0.0288	0	28.9921	0.0288	0.2	16.5505	0.0311	0	27.6647	0.0311	
500	MA(1)	0	<i>17.2733</i>	0.0462	35.14	<i>17.4990</i>	0.0462	10.22	<i>16.0954</i>	0.0486	76.55	16.3211	0.0486
	ARMA(1,1)	82.73	16.9589	0.1483	64.86	17.4103	0.1483	61.12	16.0291	0.1376	23.45	<i>16.4806</i>	0.1376
	AR(1)	0	19.9435	0.1290	0	20.1692	0.1290	0	18.6769	0.1245	0	18.9027	0.1245
	AR(2)	0	18.4996	0.0158	0	18.9511	0.0158	0	17.0865	0.0101	0	17.5379	0.0101
	AR(3)	0	17.8114	0.0074	0	18.5638	0.0074	0	16.5252	0.0082	0	17.2776	0.0082
	AR(4)	0.2	17.4821	0.0123	0	18.6108	0.0123	0.2	16.2990	0.0142	0	17.4277	0.0142
	AR(5)	0.4	17.4043	0.0189	0	18.9844	0.0189	2.81	16.2067	0.0211	0	17.7868	0.0211
	AR(6)	15.26	17.3155	0.0241	0	19.4223	0.0241	7.62	16.1718	0.0256	0	18.2786	0.0256
	AR(7)	0	17.4476	0.0277	0	20.1564	0.0277	9.82	16.1659	0.0286	0	18.8747	0.0286
	AR(8)	1	17.4181	0.0286	0	20.8040	0.0286	5.01	16.1823	0.0302	0	19.5682	0.0302
AR(9)	0.4	17.5253	0.0288	0	21.6637	0.0288	3.21	16.1979	0.0311	0	20.3362	0.0311	
1000	MA(1)	0	<i>17.2978</i>	0.0462	23.09	<i>17.4380</i>	0.0462	5.61	<i>16.1199</i>	0.0486	60.32	16.2601	0.0486
	ARMA(1,1)	81.33	16.9804	0.1483	76.91	17.2607	0.1483	59.52	16.0506	0.1376	39.68	<i>16.3309</i>	0.2859
	AR(1)	0	19.9680	0.1290	0	20.1081	0.1290	0	18.7014	0.1245	0	18.8416	0.0083
	AR(2)	0	18.5211	0.0158	0	18.8014	0.0158	0	17.1080	0.0101	0	17.3883	0.0111
	AR(3)	0	17.8289	0.0074	0	18.2960	0.0074	0	16.5427	0.0082	0	17.0098	0.0150
	AR(4)	0.2	17.4946	0.0123	0	18.1954	0.0123	0.2	16.3115	0.0142	0	17.0123	0.0157
	AR(5)	0.4	17.4108	0.0189	0	18.3918	0.0189	2.61	16.2132	0.0211	0	17.1942	0.0159
	AR(6)	16.27	17.3150	0.0241	0	18.6231	0.0241	6.01	16.1714	0.0256	0	17.4794	0.0161
	AR(7)	0	17.4391	0.0277	0	19.1210	0.0277	11.62	16.1574	0.0286	0	17.8392	0.0165
	AR(8)	1.41	17.4006	0.0286	0	19.5029	0.0286	7.01	16.1648	0.0302	0	18.2671	0.0165
AR(9)	0.4	17.4978	0.0288	0	20.0673	0.0288	7.41	16.1704	0.0311	0	18.7398	0.0167	

The model of interest is a $MA(1)$ with intercept 0.1, moving average parameter 0.5 and variance of the errors equal to 1. The total number of replications is $R = 500$ and the simulated time series contains $T = 100$ observations. The number of simulated paths H is 1 and 10, and N (first column) equals 100, 500, and 1000. The auxiliary model is either the model of interest itself or an $AR(r)$ or the $ARMA(1,1)$ (column Aux. Model). The table reports the percentage of choosing each model as the best (column Rate), the average value and the variance of the criteria (columns Mean and Var respectively) over the 500 replications. Numbers in bold denote the model chosen by the criteria. Numbers in italics denote the second best model chosen by the criteria.

Table 8: Monte Carlo results for $\theta = 0.8$ and $T = 1000$

N	Aux. Model	$H = 1$						$H = 10$					
		AIC_{IM}			IC_{IM}			AIC_{IM}			IC_{IM}		
		Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var
100	MA(1)	11.72	<i>14.7006</i>	0.1140	95.31	15.3068	0.1140	34.38	<i>13.3931</i>	0.1206	100	13.9993	0.1206
	ARMA(1,1)	88.28	14.4805	0.2076	4.69	<i>15.6929</i>	0.2076	65.63	13.3335	0.2088	0	<i>14.5459</i>	0.2088
	AR(1)	0	21.5170	0.2399	0	22.1233	0.2399	0	20.4488	0.4082	0	21.0550	0.4082
	AR(2)	0	20.3743	0.4245	0	21.5867	0.4245	0	18.6657	0.4911	0	19.8782	0.4911
	AR(3)	0	19.1679	0.4135	0	21.1887	0.4135	0	17.4481	0.2960	0	19.4688	0.2960
	AR(4)	0	17.8598	0.1104	0	20.8909	0.1104	0	16.6572	0.1755	0	19.6884	0.1755
	AR(5)	0	17.2093	0.0574	0	21.4529	0.0574	0	16.0043	0.0806	0	20.2479	0.0806
	AR(6)	0	16.5968	0.0130	0	22.2549	0.0130	0	15.5338	0.0295	0	21.1919	0.0295
	AR(7)	0	16.4376	0.0106	0	23.7123	0.0106	0	15.1952	0.0140	0	22.4699	0.0140
	AR(8)	0	16.2016	0.0088	0	25.2950	0.0088	0	14.9810	0.0096	0	24.0744	0.0096
AR(9)	0	16.2416	0.0092	0	27.3558	0.0092	0	14.8181	0.0101	0	25.9323	0.0101	
500	MA(1)	8.59	<i>14.7639</i>	0.1140	44.53	<i>14.9897</i>	0.1140	28.65	<i>13.4564</i>	0.1206	79.69	13.6822	0.1206
	ARMA(1,1)	91.41	14.5198	0.2076	55.47	14.9713	0.2076	71.35	13.3728	0.2088	20.31	<i>13.8243</i>	0.2088
	AR(1)	0	21.5804	0.2399	0	21.8061	0.2399	0	20.5121	0.4082	0	20.7379	0.4082
	AR(2)	0	20.4136	0.4245	0	20.8651	0.4245	0	18.7051	0.4911	0	19.1566	0.4911
	AR(3)	0	19.1753	0.4135	0	19.9277	0.4135	0	17.4554	0.2960	0	18.2079	0.2960
	AR(4)	0	17.8272	0.1104	0	18.9558	0.1104	0	16.6246	0.1755	0	17.7532	0.1755
	AR(5)	0	17.1286	0.0574	0	18.7087	0.0574	0	15.9237	0.0806	0	17.5038	0.0806
	AR(6)	0	16.4602	0.0130	0	18.5670	0.0130	0	15.3972	0.0295	0	17.5040	0.0295
	AR(7)	0	16.2369	0.0106	0	18.9457	0.0106	0	14.9945	0.0140	0	17.7033	0.0140
	AR(8)	0	15.9290	0.0088	0	19.3149	0.0088	0	14.7083	0.0096	0	18.0942	0.0096
AR(9)	0	15.8890	0.0092	0	20.0273	0.0092	0	14.4655	0.0101	0	18.6038	0.0101	
1000	MA(1)	8.59	<i>14.7884</i>	0.1140	27.34	<i>14.9286</i>	0.1140	28.65	<i>13.4809</i>	0.1206	62.5	13.6211	0.1206
	ARMA(1,1)	91.41	14.5413	0.2076	72.66	14.8216	0.2076	71.35	13.3943	0.2088	37.5	<i>13.6747</i>	0.2088
	AR(1)	0	21.6049	0.2399	0	21.7451	0.2399	0	20.5366	0.4082	0	20.6768	0.4082
	AR(2)	0	20.4351	0.4245	0	20.7154	0.4245	0	18.7266	0.4911	0	19.0069	0.4911
	AR(3)	0	19.1928	0.4135	0	19.6600	0.4135	0	17.4730	0.2960	0	17.9401	0.2960
	AR(4)	0	17.8397	0.1104	0	18.5404	0.1104	0	16.6371	0.1755	0	17.3378	0.1755
	AR(5)	0	17.1351	0.0574	0	18.1162	0.0574	0	15.9302	0.0806	0	16.9112	0.0806
	AR(6)	0	16.4597	0.0130	0	17.7678	0.0130	0	15.3967	0.0295	0	16.7047	0.0295
	AR(7)	0	16.2284	0.0106	0	17.9102	0.0106	0	14.9860	0.0140	0	16.6679	0.0140
	AR(8)	0	15.9115	0.0088	0	18.0138	0.0088	0	14.6908	0.0096	0	16.7931	0.0096
AR(9)	0	15.8615	0.0092	0	18.4309	0.0092	0	14.4380	0.0101	0	17.0074	0.0101	

The model of interest is a $MA(1)$ with intercept 0.1, moving average parameter 0.8 and variance of the errors equal to 1. The total number of replications is $R = 500$ and the simulated time series contains $T = 100$ observations. The number of simulated paths H is 1 and 10, and N (first column) equals 100, 500, and 1000. The auxiliary model is either the model of interest itself or an $AR(r)$ or the $ARMA(1,1)$ (column Aux. Model). The table reports the percentage of choosing each model as the best (column Rate), the average value and the variance of the criteria (columns Mean and Var respectively) over the 500 replications. Numbers in bold denote the model chosen by the criteria. Numbers in italics denote the second best model chosen by the criteria.

Table 9: Monte Carlo results for $\theta = 0.2$ and $T = 10000$

N	Aux. Model	$H = 1$						$H = 10$					
		AIC_{TM}			IC_{TM}			AIC_{TM}			IC_{TM}		
		Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var
100	MA(1)	45.6	23.0317	0.0031	100	23.6380	0.0031	31.6	<i>21.8262</i>	0.0032	100	22.4325	0.0032
	ARMA(1,1)	44.6	<i>23.0456</i>	0.0345	0	24.2581	0.0345	62	21.7748	0.0349	0	22.9872	0.0349
	AR(1)	0	23.3737	0.0006	0	23.9799	0.0006	0	22.1877	0.0007	0	22.7939	0.0007
	AR(2)	9.6	23.0902	0.0011	0	24.3027	0.0011	5.2	21.8916	0.0010	0	23.1041	0.0010
	AR(3)	0.2	23.1122	0.0014	0	25.1330	0.0014	1.2	21.8989	0.0014	0	23.9197	0.0014
	AR(4)	0	23.1597	0.0014	0	26.1908	0.0014	0	21.9638	0.0014	0	24.9950	0.0014
	AR(5)	0	23.2123	0.0014	0	27.4559	0.0014	0	22.0286	0.0014	0	26.2722	0.0014
	AR(6)	0	23.2663	0.0014	0	28.9244	0.0014	0	22.0904	0.0014	0	27.7486	0.0014
	AR(7)	0	23.3811	0.0015	0	30.6558	0.0015	0	22.1703	0.0014	0	29.4450	0.0014
	AR(8)	0	23.4553	0.0014	0	32.5488	0.0014	0	22.2574	0.0014	0	31.3508	0.0014
AR(9)	0	23.5387	0.0015	0	34.6529	0.0015	0	22.3560	0.0014	0	33.4702	0.0014	
500	MA(1)	20.4	<i>23.0951</i>	0.0031	93.6	23.3208	0.0031	0.2	<i>21.8896</i>	0.0032	85.2	22.1153	0.0032
	ARMA(1,1)	50.8	23.0850	0.0345	6.4	<i>23.5364</i>	0.0345	68.4	21.8142	0.0349	14.8	<i>22.2656</i>	0.0349
	AR(1)	0	23.4371	0.0006	0	23.6628	0.0006	0	22.2510	0.0007	0	22.4767	0.0007
	AR(2)	4.4	23.1296	0.0011	0	23.5811	0.0011	0	21.9310	0.0010	0	22.3824	0.0010
	AR(3)	22.6	23.1196	0.0014	0	23.8720	0.0014	31.4	21.9063	0.0014	0	22.6587	0.0014
	AR(4)	1.2	23.1270	0.0014	0	24.2557	0.0014	0	21.9312	0.0014	0	23.0598	0.0014
	AR(5)	0	23.1316	0.0014	0	24.7117	0.0014	0	21.9480	0.0014	0	23.5281	0.0014
	AR(6)	0.6	23.1296	0.0014	0	25.2364	0.0014	0	21.9538	0.0014	0	24.0606	0.0014
	AR(7)	0	23.1804	0.0015	0	25.8892	0.0015	0	21.9696	0.0014	0	24.6784	0.0014
	AR(8)	0	23.1827	0.0014	0	26.5686	0.0014	0	21.9848	0.0014	0	25.3707	0.0014
AR(9)	0	23.1861	0.0015	0	27.3244	0.0015	0	22.0034	0.0014	0	26.1417	0.0014	
1000	MA(1)	8.4	<i>23.1196</i>	0.0031	81.2	23.2597	0.0031	0	<i>21.9141</i>	0.0032	65.2	22.0543	0.0032
	ARMA(1,1)	50.8	23.1065	0.0345	18.8	23.3868	0.0345	68.2	21.8357	0.0349	34.8	<i>22.1160</i>	0.0349
	AR(1)	0	23.4616	0.0006	0	23.6017	0.0006	0	22.2755	0.0007	0	22.4157	0.0007
	AR(2)	0.6	23.1511	0.0011	0	23.4314	0.0011	0	21.9525	0.0010	0	22.2328	0.0010
	AR(3)	2	23.1371	0.0014	0	23.6043	0.0014	31.8	21.9238	0.0014	0	22.3910	0.0014
	AR(4)	0.6	23.1395	0.0014	0	23.8403	0.0014	0	21.9437	0.0014	0	22.6444	0.0014
	AR(5)	1.2	23.1381	0.0014	0	24.1192	0.0014	0	21.9545	0.0014	0	22.9356	0.0014
	AR(6)	36.4	23.1291	0.0014	0	24.4372	0.0014	0	21.9533	0.0014	0	23.2614	0.0014
	AR(7)	0	23.1719	0.0015	0	24.8537	0.0015	0	21.9611	0.0014	0	23.6429	0.0014
	AR(8)	0	23.1652	0.0014	0	25.2675	0.0014	0	21.9673	0.0014	0	24.0695	0.0014
AR(9)	0	23.1586	0.0015	0	25.7280	0.0015	0	21.9759	0.0014	0	24.5453	0.0014	

The model of interest is a $MA(1)$ with intercept 0.1, moving average parameter 0.2 and variance of the errors equal to 1. The total number of replications is $R = 500$ and the simulated time series contains $T = 100$ observations. The number of simulated paths H is 1 and 10, and N (first column) equals 100, 500, and 1000. The auxiliary model is either the model of interest itself or an $AR(r)$ or the $ARMA(1,1)$ (column Aux. Model). The table reports the percentage of choosing each model as the best (column Rate), the average value and the variance of the criteria (columns Mean and Var respectively) over the 500 replications. Numbers in bold denote the model chosen by the criteria. Numbers in italics denote the second best model chosen by the criteria.

Table 10: Monte Carlo results for $\theta = 0.5$ and $T = 10000$

N	Aux. Model	$H = 1$						$H = 10$					
		AIC_{TM}			IC_{TM}			AIC_{TM}			IC_{TM}		
		Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var
100	MA(1)	68.8	21.8472	0.0050	100	22.4534	0.0050	51.8	20.6481	0.0050	100	21.2543	0.0050
	ARMA(1,1)	31.2	<i>21.8866</i>	0.0177	0	<i>23.0990</i>	0.0177	48.2	<i>20.6561</i>	0.0169	0	<i>21.8686</i>	0.0169
	AR(1)	0	24.2082	0.0071	0	24.8144	0.0071	0	23.0504	0.0077	0	23.6567	0.0077
	AR(2)	0	22.8600	0.0010	0	24.0725	0.0010	0	21.6772	0.0010	0	22.8897	0.0010
	AR(3)	0	22.3474	0.0008	0	24.3681	0.0008	0	21.1334	0.0008	0	23.1542	0.0008
	AR(4)	0	22.1588	0.0015	0	25.1899	0.0015	0	20.9484	0.0015	0	23.9796	0.0015
	AR(5)	0	22.1141	0.0020	0	26.3577	0.0020	0	20.9171	0.0021	0	25.1607	0.0021
	AR(6)	0	22.1107	0.0025	0	27.7689	0.0025	0	20.9389	0.0024	0	26.5970	0.0024
	AR(7)	0	22.2062	0.0027	0	29.4809	0.0027	0	21.0005	0.0027	0	28.2752	0.0027
	AR(8)	0	22.2854	0.0028	0	31.3788	0.0028	0	21.0796	0.0028	0	30.1730	0.0028
AR(9)	0	22.3580	0.0029	0	33.4722	0.0029	0	21.1739	0.0029	0	32.2881	0.0029	
500	MA(1)	49	21.9105	0.0050	100	22.1362	0.0050	35.6	<i>20.7114</i>	0.0050	100	20.9371	0.0050
	ARMA(1,1)	42.4	<i>21.9259</i>	0.0177	0	<i>22.3774</i>	0.0177	59.6	20.6955	0.0169	0	<i>21.1469</i>	0.0169
	AR(1)	0	24.2716	0.0071	0	24.4973	0.0071	0	23.1138	0.0077	0	23.3395	0.0077
	AR(2)	0	22.8994	0.0010	0	23.3508	0.0010	0	21.7166	0.0010	0	22.1681	0.0010
	AR(3)	0	22.3547	0.0008	0	23.1072	0.0008	0	21.1408	0.0008	0	21.8932	0.0008
	AR(4)	0	22.1262	0.0015	0	23.2548	0.0015	0	20.9158	0.0015	0	22.0444	0.0015
	AR(5)	0	22.0335	0.0020	0	23.6136	0.0020	0	20.8365	0.0021	0	22.4166	0.0021
	AR(6)	8.6	21.9741	0.0025	0	24.0809	0.0025	1.6	20.8022	0.0024	0	22.9090	0.0024
	AR(7)	0	22.0055	0.0027	0	24.7143	0.0027	2.6	20.7999	0.0027	0	23.5086	0.0027
	AR(8)	0	22.0128	0.0028	0	25.3987	0.0028	0.6	20.8069	0.0028	0	24.1929	0.0028
AR(9)	0	22.0054	0.0029	0	26.1438	0.0029	0	20.8213	0.0029	0	24.9596	0.0029	
1000	MA(1)	35.2	21.9350	0.0050	97.8	22.0752	0.0050	26.8	<i>20.7359</i>	0.0050	94	20.8761	0.0050
	ARMA(1,1)	44.4	<i>21.9474</i>	0.0177	2.2	<i>22.2277</i>	0.0177	60.4	20.7170	0.0169	6	<i>20.9973</i>	0.0349
	AR(1)	0	24.2961	0.0071	0	24.4362	0.0071	0	23.1383	0.0077	0	23.2784	0.0077
	AR(2)	0	22.9209	0.0010	0	23.2012	0.0010	0	21.7381	0.0010	0	22.0184	0.0010
	AR(3)	0	22.3722	0.0008	0	22.8394	0.0008	0	21.1583	0.0008	0	21.6255	0.0014
	AR(4)	0	22.1387	0.0015	0	22.8394	0.0015	0	20.9283	0.0015	0	21.6291	0.0014
	AR(5)	0	22.0400	0.0020	0	23.0210	0.0020	0	20.8430	0.0021	0	21.8241	0.0014
	AR(6)	14.8	21.9736	0.0025	0	23.2817	0.0025	0.6	20.8017	0.0024	0	22.1098	0.0014
	AR(7)	0	21.9970	0.0027	0	23.6789	0.0027	3.2	20.7914	0.0027	0	22.4732	0.0014
	AR(8)	0	21.9953	0.0028	0	24.0975	0.0028	8	20.7894	0.0028	0	22.8917	0.0014
AR(9)	5.6	21.9779	0.0029	0	24.5473	0.0029	1	20.7938	0.0029	0	23.3632	0.0014	

The model of interest is a $MA(1)$ with intercept 0.1, moving average parameter 0.5 and variance of the errors equal to 1. The total number of replications is $R = 500$ and the simulated time series contains $T = 100$ observations. The number of simulated paths H is 1 and 10, and N (first column) equals 100, 500, and 1000. The auxiliary model is either the model of interest itself or an $AR(r)$ or the $ARMA(1,1)$ (column Aux. Model). The table reports the percentage of choosing each model as the best (column Rate), the average value and the variance of the criteria (columns Mean and Var respectively) over the 500 replications. Numbers in bold denote the model chosen by the criteria. Numbers in italics denote the second best model chosen by the criteria.

Table 11: Monte Carlo results for $\theta = 0.8$ and $T = 10000$

N	Aux. Model	$H = 1$						$H = 10$					
		AIC_{IM}			IC_{IM}			AIC_{IM}			IC_{IM}		
		Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var	Rate	Mean	Var
100	MA(1)	69.58	19.2698	0.0130	100	19.8760	0.0130	62.5	18.0532	0.0135	100	18.6594	0.0135
	ARMA(1,1)	30.42	<i>19.3008</i>	0.0216	0	<i>20.5132</i>	0.0216	37.5	<i>18.0737</i>	0.0219	0	<i>19.2862</i>	0.0219
	AR(1)	0	26.3585	0.0875	0	26.9647	0.0875	0	25.5135	0.1795	0	26.1198	0.1795
	AR(2)	0	24.5225	0.0822	0	25.7350	0.0822	0	23.4126	0.1020	0	24.6251	0.1020
	AR(3)	0	23.3011	0.0330	0	25.3219	0.0330	0	22.0186	0.0263	0	24.0394	0.0263
	AR(4)	0	22.4545	0.0143	0	25.4857	0.0143	0	21.1685	0.0112	0	24.1997	0.0112
	AR(5)	0	21.8378	0.0061	0	26.0814	0.0061	0	20.5838	0.0050	0	24.8274	0.0050
	AR(6)	0	21.3041	0.0021	0	26.9622	0.0021	0	20.1301	0.0021	0	25.7883	0.0021
	AR(7)	0	21.0204	0.0010	0	28.2952	0.0010	0	19.8091	0.0009	0	27.0839	0.0009
	AR(8)	0	20.8081	0.0006	0	29.9015	0.0006	0	19.5852	0.0006	0	28.6786	0.0006
AR(9)	0	20.6325	0.0008	0	31.7467	0.0008	0	19.4407	0.0007	0	30.5549	0.0007	
500	MA(1)	52.5	19.3331	0.0130	100	19.5589	0.0130	42.86	<i>18.1165</i>	0.0135	100	18.3423	0.0135
	ARMA(1,1)	47.5	<i>19.3401</i>	0.0216	0	<i>19.7916</i>	0.0216	57.14	18.1131	0.0219	0	<i>18.5645</i>	0.0219
	AR(1)	0	26.4219	0.0875	0	26.6476	0.0875	0	25.5769	0.1795	0	25.8026	0.1795
	AR(2)	0	24.5619	0.0822	0	25.0133	0.0822	0	23.4520	0.1020	0	23.9034	0.1020
	AR(3)	0	23.3085	0.0330	0	24.0609	0.0330	0	22.0260	0.0263	0	22.7784	0.0263
	AR(4)	0	22.4219	0.0143	0	23.5505	0.0143	0	21.1359	0.0112	0	22.2645	0.0112
	AR(5)	0	21.7572	0.0061	0	23.3373	0.0061	0	20.5032	0.0050	0	22.0833	0.0050
	AR(6)	0	21.1674	0.0021	0	23.2742	0.0021	0	19.9935	0.0021	0	22.1003	0.0021
	AR(7)	0	20.8198	0.0010	0	23.5285	0.0010	0	19.6085	0.0009	0	22.3172	0.0009
	AR(8)	0	20.5355	0.0006	0	23.9214	0.0006	0	19.3125	0.0006	0	22.6985	0.0006
AR(9)	0	20.2799	0.0008	0	24.4182	0.0008	0	19.0881	0.0007	0	23.2265	0.0007	
1000	MA(1)	48.75	19.3576	0.0130	99.17	19.4978	0.0130	41.43	<i>18.1470</i>	0.0135	98.93	18.2812	0.0135
	ARMA(1,1)	51.25	<i>19.3616</i>	0.0216	0.83	<i>19.6419</i>	0.0216	58.57	18.1346	0.0219	1.07	<i>18.4149</i>	0.0219
	AR(1)	0	26.4464	0.0875	0	26.5865	0.0875	0	25.6014	0.1795	0	25.7415	0.1795
	AR(2)	0	24.5834	0.0822	0	24.8637	0.0822	0	23.4735	0.1020	0	23.7538	0.1020
	AR(3)	0	23.3260	0.0330	0	23.7932	0.0330	0	22.0435	0.0263	0	22.5107	0.0263
	AR(4)	0	22.4344	0.0143	0	23.1352	0.0143	0	21.1484	0.0112	0	21.8492	0.0112
	AR(5)	0	21.7637	0.0061	0	22.7447	0.0061	0	20.5097	0.0050	0	21.4908	0.0050
	AR(6)	0	21.1670	0.0021	0	22.4750	0.0021	0	19.9930	0.0021	0	21.3011	0.0021
	AR(7)	0	20.8113	0.0010	0	22.4931	0.0010	0	19.6000	0.0009	0	21.2818	0.0009
	AR(8)	0	20.5180	0.0006	0	22.6202	0.0006	0	19.2950	0.0006	0	21.3973	0.0006
AR(9)	0	20.2524	0.0008	0	22.8218	0.0008	0	19.0606	0.0007	0	21.6301	0.0007	

The model of interest is a $MA(1)$ with intercept 0.1, moving average parameter 0.8 and variance of the errors equal to 1. The total number of replications is $R = 500$ and the simulated time series contains $T = 100$ observations. The number of simulated paths H is 1 and 10, and N (first column) equals 100, 500, and 1000. The auxiliary model is either the model of interest itself or an $AR(r)$ or the $ARMA(1,1)$ (column Aux. Model). The table reports the percentage of choosing each model as the best (column Rate), the average value and the variance of the criteria (columns Mean and Var respectively) over the 500 replications. Numbers in bold denote the model chosen by the criteria. Numbers in italics denote the second best model chosen by the criteria.

Table 12: Empirical application with sample 2003–2007

N	Aux.	$H = 1$		$H = 10$	
	Model	AIC_{IM}	IC_{IM}	AIC_{IM}	IC_{IM}
100	ARCH(1)	13.5391	14.1453	-10.0531	-9.4469
	ARCH(2)	8.0593	<i>9.2717</i>	<i>-10.5561</i>	-9.3436
	ARCH(3)	21.7615	23.7822	-8.8213	-6.8006
	ARCH(4)	7.9145	10.9456	-9.9379	-6.9068
	ARCH(5)	7.9161	12.1596	-9.8875	-5.6439
	ARCH(6)	6.9432	12.6013	-8.0767	-2.4186
	ARCH(7)	6.8019	14.0766	-8.3622	-1.0874
	ARCH(8)	6.7399	15.8333	-8.2522	0.8413
	ARCH(9)	6.8039	17.9181	-8.3079	2.8062
	GARCH(1,1)	-8.7668	-7.5543	-11.0157	-9.8033
500	ARCH(1)	13.6025	13.8282	-9.9898	-9.7641
	ARCH(2)	8.0987	<i>8.5501</i>	<i>-10.5167</i>	<i>-10.0653</i>
	ARCH(3)	21.7688	22.5213	-8.8140	-8.0615
	ARCH(4)	7.8818	9.0105	-9.9706	-8.8419
	ARCH(5)	7.8354	9.4155	-9.9681	-8.3880
	ARCH(6)	6.8066	8.9134	-8.2134	-6.1066
	ARCH(7)	6.6012	9.3100	-8.5628	-5.8541
	ARCH(8)	6.4672	9.8531	-8.5248	-5.1389
	ARCH(9)	<i>6.4513</i>	10.5897	-8.6606	-4.5222
	GARCH(1,1)	-8.7274	-8.2760	-10.9764	-10.5249
1000	ARCH(1)	13.6270	13.7671	-9.9653	-9.8251
	ARCH(2)	8.1202	8.4005	<i>-10.4952</i>	<i>-10.2149</i>
	ARCH(3)	21.7863	22.2535	-8.7965	-8.3293
	ARCH(4)	7.8943	8.5951	-9.9581	-9.2573
	ARCH(5)	7.8419	8.8230	-9.9616	-8.9806
	ARCH(6)	6.8061	<i>8.1141</i>	-8.2139	-6.9058
	ARCH(7)	6.5927	8.2746	-8.5713	-6.8895
	ARCH(8)	6.4497	8.5520	-8.5423	-6.4400
	ARCH(9)	6.4238	8.9933	-8.6881	-6.1186
	GARCH(1,1)	-8.7059	-8.4256	-10.9549	-10.6746
100	AR(1)	-10.4436	<i>-9.2371</i>	-6.6166	-5.4041
	AR(2)	-10.0119	-7.9912	-5.1512	-3.1305
	AR(3)	-10.2961	-7.2650	-6.0727	-3.0416
	AR(4)	-12.3629	-8.1193	-11.0190	-6.7754
	AR(5)	-11.7664	-6.1082	-7.6113	-1.9532
	AR(6)	-12.5222	-5.2475	-10.0463	-2.7715
	AR(7)	-11.9616	-2.8682	-8.8222	0.2713
	AR(8)	-13.0263	-1.9122	-10.4620	0.6522
	AR(9)	<i>-21.8050</i>	-8.4680	<i>-15.3171</i>	-1.9801
	ARMA(1,1)	-22.8581	-20.8373	-18.9235	-16.9028
500	AR(1)	-10.4042	-9.9528	-6.5772	-6.1257
	AR(2)	-10.0046	-9.2522	-5.1439	-4.3914
	AR(3)	-10.3288	-9.2001	-6.1054	-4.9767
	AR(4)	-12.4436	-10.8635	-11.0996	-9.5195
	AR(5)	-11.9030	-9.7962	-7.7479	-5.6411
	AR(6)	-12.7229	-10.0141	-10.2469	-7.5382
	AR(7)	-12.2343	-8.8484	-9.0948	-5.7089
	AR(8)	-13.3790	-9.2406	-10.8146	-6.6763
	AR(9)	<i>-22.2457</i>	<i>-17.2796</i>	<i>-15.7577</i>	<i>-10.7917</i>
	ARMA(1,1)	-22.8507	-22.0983	-18.9162	-18.1637
1000	AR(1)	-10.3827	-10.1024	-6.5557	-6.2754
	AR(2)	-9.9871	-9.5199	-5.1264	-4.6592
	AR(3)	-10.3163	-9.6155	-6.0929	-5.3921
	AR(4)	-12.4371	-11.4560	-11.0931	-10.1121
	AR(5)	-11.9035	-10.5954	-7.7484	-6.4404
	AR(6)	-12.7314	-11.0495	-10.2554	-8.5736
	AR(7)	-12.2518	-10.1495	-9.1123	-7.0100
	AR(8)	-13.4065	-10.8370	-10.8421	-8.2727
	AR(9)	<i>-22.2842</i>	<i>-19.2008</i>	<i>-15.7962</i>	<i>-12.7129</i>
	ARMA(1,1)	-22.8332	-22.3661	-18.8987	-18.4315

The model of interest is the stochastic volatility model. The underlying data are daily log-returns of IBM from 01.01.2003 up to 31.12.2007 (1258 days). The number of simulated paths H is 1 and 10, and N (first column) equals 100, 500, and 1000. The auxiliary model is either $AR(r)$ or $ARMA(1,1)$ or $ARCH(r)$ or $GARCH(1,1)$ (column Aux. Model). The entries represent the values of AIC_{IM} and IC_{IM} . Numbers in bold denote the model chosen by the criteria. Numbers in italics denote the second best model chosen by the criteria.

Table 13: Empirical application with sample 2004–2008

N	Aux.	$H = 1$		$H = 10$	
	Model	AIC_{IM}	IC_{IM}	AIC_{IM}	IC_{IM}
100	ARCH(1)	24.6159	25.2222	27.0852	27.6914
	ARCH(2)	7.1603	8.3728	6.8566	8.0690
	ARCH(3)	6.7257	8.7464	-8.4580	-6.4373
	ARCH(4)	7.1083	10.1394	-8.7224	-5.6912
	ARCH(5)	18.2875	22.5311	-8.7705	-4.5269
	ARCH(6)	14.9909	20.6490	-9.4171	-3.7590
	ARCH(7)	15.1704	22.4451	-9.9766	-2.7018
	ARCH(8)	15.5943	24.6877	-10.0217	-0.9283
	ARCH(9)	15.4400	26.5541	-9.8151	1.2990
	GARCH(1,1)	-8.4924	-7.2800	-10.4355	-9.2230
500	ARCH(1)	24.6793	24.9050	27.1485	27.3743
	ARCH(2)	7.1997	7.6512	6.8959	7.3474
	ARCH(3)	6.7330	7.4854	-8.4507	-7.6982
	ARCH(4)	7.0756	8.2043	-8.7550	-7.6264
	ARCH(5)	18.2069	19.7870	-8.8512	-7.2711
	ARCH(6)	14.8543	16.9611	-9.5538	-7.4470
	ARCH(7)	14.9697	17.6785	-10.1772	-7.4685
	ARCH(8)	15.3217	18.7076	-10.2944	-6.9084
	ARCH(9)	15.0873	19.2257	-10.1678	-6.0294
	GARCH(1,1)	-8.4531	-8.0016	-10.3961	-9.9447
1000	ARCH(1)	24.7038	24.8439	27.1730	27.3132
	ARCH(2)	7.2212	7.5015	6.9174	7.1977
	ARCH(3)	6.7505	7.2177	-8.4332	-7.9660
	ARCH(4)	7.0881	7.7889	-8.7425	-8.0417
	ARCH(5)	18.2134	19.1944	-8.8447	-7.8636
	ARCH(6)	14.8538	16.1618	-9.5543	-8.2462
	ARCH(7)	14.9612	16.6430	-10.1857	-8.5039
	ARCH(8)	15.3042	17.4064	-10.3119	-8.2096
	ARCH(9)	15.0598	17.6293	-10.1953	-7.6258
	GARCH(1,1)	-8.4316	-8.1513	-10.3746	-10.0943
100	AR(1)	-9.4081	-8.1957	-10.4344	-9.2220
	AR(2)	-13.2206	-11.1998	-9.0100	-6.9892
	AR(3)	-10.9748	-7.9437	-5.1172	-2.0861
	AR(4)	-11.5743	-7.3307	-9.4056	-5.1620
	AR(5)	-11.3426	-5.6844	-8.8515	-3.1934
	AR(6)	-13.8177	-6.5430	-9.4771	-2.2024
	AR(7)	-11.6719	-2.5785	-8.8038	0.2896
	AR(8)	-13.1929	-2.0787	-9.3069	1.8073
	AR(9)	-11.8119	1.5251	-9.4980	3.8390
	ARMA(1,1)	-16.5902	-14.5694	-16.1230	-14.1022
500	AR(1)	-9.3687	-8.9173	-10.3951	-9.9436
	AR(2)	-13.2132	-12.4608	-9.0026	-8.2502
	AR(3)	-11.0075	-9.8788	-5.1499	-4.0212
	AR(4)	-11.6549	-10.0748	-9.4863	-7.9062
	AR(5)	-11.4792	-9.3724	-8.9881	-6.8813
	AR(6)	-14.0184	-11.3096	-9.6778	-6.9690
	AR(7)	-11.9445	-8.5586	-9.0764	-5.6905
	AR(8)	-13.5456	-9.4072	-9.6595	-5.5212
	AR(9)	-12.2525	-7.2865	-9.9387	-4.9727
	ARMA(1,1)	-16.5828	-15.8304	-16.1156	-15.3632
1000	AR(1)	-9.3472	-9.0669	-10.3736	-10.0933
	AR(2)	-13.1957	-12.7286	-8.9851	-8.5180
	AR(3)	-10.9950	-10.2942	-5.1374	-4.4366
	AR(4)	-11.6484	-10.6674	-9.4798	-8.4987
	AR(5)	-11.4797	-10.1716	-8.9886	-7.6805
	AR(6)	-14.0269	-12.3451	-9.6863	-8.0044
	AR(7)	-11.9620	-9.8597	-9.0939	-6.9917
	AR(8)	-13.5731	-11.0036	-9.6870	-7.1176
	AR(9)	-12.2910	-9.2077	-9.9772	-6.8938
	ARMA(1,1)	-16.5653	-16.0982	-16.0981	-15.6309

The model of interest is the stochastic volatility model. The underlying data are daily log-returns of IBM from 01.01.2004 up to 31.12.2008 (1152 days). The number of simulated paths H is 1 and 10, and N (first column) equals 100, 500, and 1000. The auxiliary model is either $AR(r)$ or $ARMA(1,1)$ or $ARCH(r)$ or $GARCH(1,1)$ (column Aux. Model). The entries represent the values of AIC_{IM} and IC_{IM} . Numbers in bold denote the model chosen by the criteria. Numbers in italics denote the second best model chosen by the criteria.