cemfi

working paper

2502

# The information matrix test for Markov switching autoregressive models with covariate-dependent transition probabilities

Dante Amengual
Gabriele Fiorentini
Enrique Sentana

January 2025

CENTRO DE ESTUDIOS
MONETARIOS Y FINANCIEROS
Casado del Alisal 5, 28014 Madrid, Spain
www.cemfi.es

CEMFI Working Paper No. 2502
January 2025

# The information matrix test for Markov switching autoregressive models with covariate-dependent transition probabilities


## Abstract

The EM principle implies the moments underlying the information matrix test for multivariate Markov switching autoregressive models with covariate-dependent transition probabilities are the smoothed values of the moments we would test were the latent Markov chain observed. Thus, we identify components related to the heteroskedasticity, skewness and kurtosis of the multivariate regression residuals for each of the regimes, the neglected multivariate heteroskedasticity of the generalised residuals for each of the columns of the transition matrix, and a final component that assesses the conditional independence of these generalised residuals and the regression residuals, their squares and cross-products given the observed variables.


JEL Codes: C22, C32, C46, C52.

Keywords: Expectation - maximisation principle, incomplete data, Hessian matrix, outer product of the score, Term spread recession forecasting.

Dante Amengual
CEMFI
amengual@cemfi.es

Gabriele Fiorentini
Università di Firenze and RCEA
gabriele.fiorentini@unifi.it

Enrique Sentana
CEMFI
sentana@cemfi.es

**Acknowledgement**

We would like to thank audiences at Montreal, EC JRC (Ispra), Maryland, the Federal Reserve Board, Torcuato DiTella, the 2024 NBER-NSF Time Series conference (Penn), the Villa Mondragone Time Series Symposium in honour of Marco Lippi (Rome) and the 2024 annual LAMES conference (Montevideo) for useful comments, discussions and suggestions. We are particularly grateful to Martín Solá for both his insightful comments and his help with the empirical application. Of course, the usual caveat applies. The first and third authors gratefully acknowledge financial support from grant PID2021-128963NB-I00 funded by MICIU/AEI/1013039/501100011033 and ERDF/EU, while the second one is thankful to MIUR through the PRIN project "High-dimensional time series for structural macroeconomic analysis in times of pandemic".

# 1 Introduction

The Markov switching regression models originally introduced in econometrics by Goldfeld and Quandt (1973) and extended to autoregressive specifications by Hamilton (1989a,b) have been extensively employed by researchers and practitioners for the last few decades because of their flexibility to capture occasional but recurrent shifts in the dynamics of macroeconomic and financial time series. Early examples included business cycles (Hamilton (1989b) and Lam (1990)), interest rates (Hamilton (1989a), Evans and Lewis (1995) and Garcia and Perron (1996)), stock returns (Cecchetti, Lam, and Mark (1990)) and their volatility (Schwert (1989)), exchange rates (Engel and Hamilton (1996)) or risk aversion in financial markets (Turner, Startz, and Nelson (1989)). These papers assumed an autonomous underlying Markov chain process with constant transition probabilities but a second generation of models with probabilities that depend on observable variables also suggested by Goldfeld and Quandt (1973) soon came to the fore in order to improve the prediction of turning points. Initial contributions include Filardo (1994), Filardo and Gordon (1996), Diebold, Lee and Weinbach (1996), Gray (1996), Ravn and Solá (1999), Bekaert, Hodrick and Marshall (2001), Ang and Bekaert (2002a,b), Martínez Peria (2002), and Simpson, Osborn and Sensier (2001).

The popularity of Markov switching autoregressive models is confirmed by the enormity of the literature that studies them. Textbook treatments include Hamilton (1994), Krolzig (1997), Kim and Nelson (1999) and Frühwirth-Schnatter (2007), while Hamilton (2008, 2016), Guidolin (2011) and Ang and Timmermann (2012) provide useful surveys of applications in macroeconomics and finance. Despite their popularity, though, estimated Markov switching models are hardly ever subject to specification tests. Part of the reason could be that specification testing does not fit well with the Bayesian approach that has become common among practitioners. And even though Hamilton (1996) proposed several moment tests based on the log-likelihood scores to assess correct model specification, they are seldom used, possibly due to the difficulty in intuitively interpreting the influence functions involved, which is important for suggesting the directions on which model revisions should focus. And yet, in an influential recent paper, Pouzo, Psaradakis and Solá (2022) (PPS henceforth) highlight the inconsistencies that arise from misspecifying the data generating process in models with covariate-dependent transition probabilities.

For that reason, we discuss in detail the application of the information matrix (IM) test to the Markov switching autoregressions with transition probabilities that may be functions of lagged observed variables studied by PPS. As is well known, the original IM test introduced by White (1982) directly assesses the IM equality, which states that the sum of the Hessian matrix and the outer product of the score vector should be zero in expectation when the estimated model is correctly specified. We use recent results in Amengual, Fiorentini and Sentana (2024a), which say that in models in which the observations can be viewed as incomplete data, in the sense of Dempster, Laird and Rubin (1977), the law of iterated expectations implies that the influence functions involved in the IM test coincide with the expected value given the full sample of observed data of the influence functions that would be tested were the latent variables observed.

Exploiting the Expectation - Maximisation (EM) principle in the context of these models is hardly new. In fact, Hamilton (1990) suggested the EM algorithm for the purposes of computing the scores and estimating the model parameters, whose fixed-point equations provide rather intuitive expressions for the maximum likelihood estimators. In addition, Hamilton (1996) relied on those expressions to derive score-based specification tests. In turn, Diebold, Lee and Weinbach (1996) explained how to adapt the same algorithm so that it can be used to tackle models with dynamic transition matrices (see also Krolzig (1997)). However, none of these authors exploited in full the potential of the EM principle for testing purposes.

In this respect, we show that the use of the EM principle provides a very intuitive interpretation of the influence functions underlying the IM test. Specifically, we prove that the influence functions associated to the mean and variance parameters within each regime coincide with the multivariate version of the ones derived by Hall (1987) for linear regression models in Amengual, Fiorentini and Sentana (2022), but written in terms of residuals computed as if all observations came from the $k^{th}$ regime and weighted by the smoothed probability that each observation belongs to that regime. Thus, the IM test is effectively testing the unconditional skewness and kurtosis of those conditionally standardised residuals, as in Jarque and Bera (1980) and Amengual, Fiorentini and Sentana (2024b), as well as their conditional heteroskedasticity given the lagged dependent variables and their squares and cross-products, as in Sentana (1994), and their conditional skewness given past observations, as in Bera and Lee (1993). In turn, we show that the influence functions of the IM test associated to the parameters of the transition matrix correspond to the smoothed value given the observed data of the outer-product of what Gouriéroux, Monfort, Trognon and Renault (1987) called the generalised multinomial logit residuals minus their conditional covariance matrix times the levels and cross-products of the conditioning variables that determine the transition probabilities, as in Amengual, Fiorentini and Sentana (2025). Finally, we obtain a third group of influence functions that assess the conditional independence of these generalised residuals and the regression residuals, their squares and cross-products given the observed variables. An additional advantage of explicitly relating the influence functions of the IM test of the complete and incomplete data models is that it is easy to determine which of those functions is either redundant or spanned by the score vector, which is crucial for determining the right number of degrees of freedom, as illustrated in Amengual, Fiorentini and Sentana (2024a).

Classical tests (i.e. Likelihood ratio, Wald and score or Lagrange Multiplier (LM)) for the number of regimes in a switching regime model are a devilish problem even if one assumes that the distribution of the shocks is Gaussian (see e.g. Carrasco, Hu and Ploberger (2014), Kasahara and Shimotsu (2018), Qu and Zhuo (2021) and Amengual, Bei, Carrasco and Sentana (2025)). By comparison, testing Gaussianity of the underlying components against a more flexible family of parametric distributions while maintaining that the number of regimes and their transition matrix is correctly specified would be relatively straightforward if one also relied on the EM principle to obtain expressions for the scores and information matrix of the model under the alternative evaluated under the null along the lines of Almuzara, Amengual and Sentana (2019). In this respect, one advantage of the IM test is that while it is not designed to focus on any particular aspect of the model specification, it has power to detect misspecification in each of its ingredients, as we will see below.

The rest of the paper is organised as follows. We start by formally introducing the model in section 2.1. After a quick review of White's (1982) original IM test in section 2.2, we derive its expressions for the complete and incomplete data cases in sections 2.3 and 2.4, respectively. Next, we present the results of extensive Monte Carlo simulations in section 3, followed by an application of our tests to the empirical illustration in PPS in section 4. Finally, we conclude by discussing some interesting extensions in section 5.

# 2 Theoretical analysis

## 2.1 The model

We are interested in studying the correct specification of the following model for an observable, $M$-variate time series process $\mathbf{y}_t$:

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\nu}(\boldsymbol{\xi}_t) + \mathbf{A}(\boldsymbol{\xi}_t)\mathbf{y}_{t-1} + \boldsymbol{\Gamma}^{1/2}(\boldsymbol{\xi}_t)\boldsymbol{\varepsilon}_t^*, \\ \boldsymbol{\varepsilon}_t^*|I_{t-1}, \Xi_t \ldots &\sim N(\mathbf{0}, \mathbf{I}_M), \\ \boldsymbol{\xi}_t|I_{t-1}, \Xi_{t-1} \ldots &\sim MC[\mathbf{P}(I_{t-1})], \end{aligned} \quad (1)$$

where $I_{t-1}$ is the information set generated by lagged values of $\mathbf{y}_t$ and other $L$ strongly exogenous observed variables $\mathbf{z}_t$, $\Xi_{t-1}$ the information set generated by lagged values of the latent Markov chain (MC) process with $K$ states $\boldsymbol{\xi}_t' = (\xi_1, \ldots, \xi_k, \ldots, \xi_K)$, which is nothing other than an exhaustive collection of $K$ mutually exclusive (0-1) random variables, and $\mathbf{P}(I_{t-1})$ its transition matrix, which may depend on both $\mathbf{z}_{t-1}$ and $\mathbf{y}_{t-1}$.[1]

[1] Hamilton (1989b) considered a slightly different version of (1) in which there is no explicit drift $\boldsymbol{\nu}(\boldsymbol{\xi}_t)$ but $\mathbf{y}_t$ and $\mathbf{y}_{t-1}$ are expressed in terms of deviations of the unconditional means of the regime prevailing at times $t$ and $t-1$, respectively. Given that Hamilton (1994) explains how such a model can be equivalently written as a model with a higher number of regimes, we shall not consider it separely.

To ensure that $\mathbf{P}(I_{t-1})$ is indeed a transition matrix, we follow the previous literature and model the $j^{th}$ column of the transition matrix using the following multinomial logit parametrisation:

$$P(\xi_{kt} = 1|\xi_{jt-1} = 1, I_{t-1}; \boldsymbol{\beta}_j) = p_{kjt}(\boldsymbol{\beta}_j) = \frac{e^{\boldsymbol{\beta}'_{kj}\mathbf{z}^a_{t-1}}}{\sum_{\ell=1}^{K} e^{\boldsymbol{\beta}'_{\ell j}\mathbf{z}^a_{t-1}}}, \quad k, j = 1, \ldots, K, \tag{2}$$

which for simplicity of notation we write as a function of $\mathbf{z}^a_{t-1} = (1, \mathbf{z}'_{t-1})'$ only even though in practice it can depend on $\mathbf{y}_{t-1}$ too, where $\boldsymbol{\beta}_j = (\boldsymbol{\beta}'_{1j}, \ldots, \boldsymbol{\beta}'_{Kj})'$ collects the intercepts and slope coefficient vectors associated to the probabilities of transitioning from state $j$ at time $t-1$ to any other state at time $t$. This ensures that $\sum_{k=1}^{K} p_{kjt}(\boldsymbol{\beta}_j) = 1$ for all $\boldsymbol{\beta}_j$, $j$ and $t$, which in turn implies that we can skip one category per column. For identification purposes, we follow the usual practice of setting $\boldsymbol{\beta}_{1j} = \mathbf{0}$ for all $j$, so that the transition matrix is a function of the stacked parameter vector $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_K)'$, with $\boldsymbol{\beta}_j$ redefined henceforth as $(\boldsymbol{\beta}'_{2j}, \ldots, \boldsymbol{\beta}'_{Kj})'$ in a slight abuse of notation. Nevertheless, this arbitrary normalisation is without loss of generality because Lemma 1 in Amengual, Fiorentini and Sentana (2024a) implies that IM tests are numerically invariant to bijective reparametrisations.

As is well know, when $\mathbf{P}(I_{t-1}) = \boldsymbol{\pi}(I_{t-1})\boldsymbol{\imath}'_K$, where $\boldsymbol{\imath}_K$ is a vector of $K$ ones, the MC becomes an independent sequence of categorical random variables whose conditional probabilities are $\boldsymbol{\pi}(I_{t-1})$, which we will denote by $MN[\boldsymbol{\pi}(I_{t-1})]$, where $MN$ stands for multinomial.

The model in (1) nests a broad class of simpler models regularly used in empirical work, including:

1. $\mathbf{y}_t|I_{t-1}, \Xi_t, \ldots \sim N(\boldsymbol{\nu} + \mathbf{A}\mathbf{y}_{t-1}, \boldsymbol{\Gamma})$ (vector autoregressions or VARs)

2. $\boldsymbol{\xi}_t|I_{t-1}, \Xi_{t-1}, \ldots \sim MC[\mathbf{P}(I_{t-1})]$ (multinomial logit models and Markov chains with covariate-dependent transitions)

3. $\mathbf{y}_t|I_{t-1}, \Xi_t, \ldots \sim N[\boldsymbol{\nu}(\boldsymbol{\xi}_t), \boldsymbol{\Gamma}(\boldsymbol{\xi}_t)], \boldsymbol{\xi}_t|I_{t-1}, \Xi_{t-1}, \ldots \sim MN(\boldsymbol{\pi})$ (finite Gaussian mixtures)

4. $\mathbf{y}_t|I_{t-1}, \Xi_t, \ldots \sim N[\boldsymbol{\nu}(\boldsymbol{\xi}_t) + \mathbf{A}(\boldsymbol{\xi}_t)\mathbf{y}_{t-1}, \boldsymbol{\Gamma}(\boldsymbol{\xi}_t)], \boldsymbol{\xi}_t|I_{t-1}, \Xi_{t-1}, \ldots \sim MN[\boldsymbol{\pi}(I_{t-1})]$ (dynamic switching regressions with covariate-dependent probabilities)

but it differs from the first two because of the hidden regimes and from the last two because inferences about those regimes require smoothing rather than just filtering.

Although we will not discuss them in detail in the next subsections, extensions of (1) to more complex conditional mean and variance specifications for the distribution of $\mathbf{y}_t|I_{t-1}, \Xi_t, \ldots$, including higher-order autoregressions, or higher-order dependence for $\boldsymbol{\xi}_t|I_{t-1}, \Xi_{t-1} \ldots$ are tedious but otherwise straightforward.

In principle, a researcher should explicitly specify a joint data generating process (DGP) for $\mathbf{y}_t$ and $\mathbf{z}_t$, as PPS do. However, we follow the empirical and theoretical literatures in treating $\mathbf{z}_t$ as a strongly exogenous process that depends on an additional variation-free vector of parameters (see Engle, Hendry and Richard (1983)), so that a joint model is not necessary either for obtaining efficient estimators of the parameters of (1) because of the weak exogeneity of $\mathbf{z}_t$ or for filtering and smoothing purposes because $\mathbf{z}_t$ is not Granger caused by either $\mathbf{y}_t$ or $\boldsymbol{\xi}_t$. Nevertheless, we explicitly study the power of our proposed tests to detect departures from this assumption in section 3.

## 2.2 White's (1982) original IM test

Let $\boldsymbol{\phi}$, with $dim(\boldsymbol{\phi}) < \infty$, denote the vector of model parameters. On the basis of the usual prediction error decomposition, and ignoring initial conditions, we can write the log-likelihood function of a sample of size $T$ on $\mathbf{y}_t$ as

$$L_T(\boldsymbol{\phi}) = \sum\nolimits_{t=1}^{T} \ln f(\mathbf{y}_t | I_{t-1}; \boldsymbol{\phi}) = \sum\nolimits_{t=1}^{T} l_t(\boldsymbol{\phi}), \tag{3}$$

where $f(\mathbf{y}_t | I_{t-1}; \boldsymbol{\phi})$ denotes the conditional density of $\mathbf{y}_t$ given its past values and the past values of $\mathbf{z}_t$ contained in $I_{t-1}$. Hence, the average score and Hessian will be given by

$$\bar{\mathbf{s}}_T(\boldsymbol{\phi}) = \frac{1}{T} \sum\nolimits_{t=1}^{T} \frac{\partial l_t(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}} = \frac{1}{T} \sum\nolimits_{t=1}^{T} \mathbf{s}_t(\boldsymbol{\phi}), \tag{4}$$

$$\bar{\mathbf{h}}_T(\boldsymbol{\phi}) = \frac{1}{T} \sum\nolimits_{t=1}^{T} \frac{\partial^2 l_t(\boldsymbol{\phi})}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'} = \frac{1}{T} \sum\nolimits_{t=1}^{T} \mathbf{h}_t(\boldsymbol{\phi}). \tag{5}$$

In what follows, we maintain the regularity conditions that PPS impose on the DGP of $\mathbf{y}_t$, its conditional log density in (3) and its first-two derivatives in (4) and (5), which guarantee the consistency of the maximum likely estimators (MLE) of the model parameters, $\hat{\boldsymbol{\phi}}_T$, and their asymptotic normality when we centre them around their true values, $\boldsymbol{\phi}_0$, and suitably scale them by $\sqrt{T}$.

In this context, the IM test directly assesses the IM equality, which states that the sum of the Hessian matrix and the outer product of the score (OPS) vector should be zero in expected value when the estimated model is correctly specified.

As Newey (1985) and Tauchen (1985) showed, the IM test can be regarded as a moment test based on the influence functions:

$$vech[\mathbf{h}_t(\boldsymbol{\phi}) + \mathbf{s}_t(\boldsymbol{\phi})\mathbf{s}_t'(\boldsymbol{\phi})]. \tag{6}$$

In practice, we evaluate these influence functions at the MLE of $\boldsymbol{\phi}$, $\hat{\boldsymbol{\phi}}_T$, so we need the asymptotic covariance matrix of

$$\frac{\sqrt{T}}{T} \sum\nolimits_{t=1}^{T} vech[\mathbf{h}_t(\hat{\boldsymbol{\phi}}_T) + \mathbf{s}_t(\hat{\boldsymbol{\phi}}_T)\mathbf{s}_t'(\hat{\boldsymbol{\phi}}_T)].$$

Chesher (1983) and Lancaster (1984) realised that one can use the generalised information matrix equality to obtain the expected value of the Jacobian of the influence functions with respect to $\boldsymbol{\phi}$ from the covariance matrix between them and the score evaluated at the true values of the parameters. Thus, we simply need the residual covariance matrix from their least squares projection onto the linear span of $\mathbf{s}_t(\boldsymbol{\phi}_0)$:

$$\mathcal{R}(\boldsymbol{\phi}_0) - \mathcal{U}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}(\boldsymbol{\phi}_0)\mathcal{U}'(\boldsymbol{\phi}_0),$$

$$\begin{bmatrix} \mathcal{R}(\boldsymbol{\phi}_0) & \mathcal{U}(\boldsymbol{\phi}_0) \\ \mathcal{U}'(\boldsymbol{\phi}_0) & \mathcal{I}(\boldsymbol{\phi}_0) \end{bmatrix} = V \left\{ \begin{array}{c} vech[\mathbf{h}_t(\boldsymbol{\phi}_0) + \mathbf{s}_t(\boldsymbol{\phi}_0)\mathbf{s}_t'(\boldsymbol{\phi}_0)] \\ \mathbf{s}_t(\boldsymbol{\phi}_0) \end{array} \right\}. \tag{7}$$

Therefore, the infeasible IM test statistic is the quadratic form

$$T\left\{\frac{1}{T}\sum\nolimits_{t=1}^{T} vech'[\mathbf{h}_t(\hat{\boldsymbol{\phi}}_T) + \mathbf{s}_t(\hat{\boldsymbol{\phi}}_T)\mathbf{s}_t'(\hat{\boldsymbol{\phi}}_T)]\right\}$$
$$\times[\mathcal{R}(\boldsymbol{\phi}_0) - \mathcal{U}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}(\boldsymbol{\phi}_0)\mathcal{U}(\boldsymbol{\phi}_0)]^{-}$$
$$\times\left\{\frac{1}{T}\sum\nolimits_{t=1}^{T} vech[\mathbf{h}_t(\hat{\boldsymbol{\phi}}_T) + \mathbf{s}_t(\hat{\boldsymbol{\phi}}_T)\mathbf{s}_t'(\hat{\boldsymbol{\phi}}_T)]\right\}.$$

A generalised inverse is often necessary because some of the influence functions underlying the IM test may be an exact linear combination of $\mathbf{s}_t(\boldsymbol{\phi}_0)$ or appear multiple times. As a result, the number of degrees of freedom of the asymptotic $\chi^2$ distribution under the null of correct specification is $rank[\mathcal{R}(\boldsymbol{\phi}_0) - \mathcal{U}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}(\boldsymbol{\phi}_0)\mathcal{U}(\boldsymbol{\phi}_0)]$, which requires careful derivation, something that our EM-based procedure helps with, as we illustrate in sections 2.3 and 2.4 below.

Chesher (1983) and Lancaster (1984) suggested a feasible version as $T$ times the $R^2$ in the regression of a vector of $T$ ones onto $\mathbf{s}_t(\hat{\boldsymbol{\phi}}_T)$ and $vech[\mathbf{h}_t(\hat{\boldsymbol{\phi}}_T) + \mathbf{s}_t(\hat{\boldsymbol{\phi}}_T)\mathbf{s}_t'(\hat{\boldsymbol{\phi}}_T)]$ using an ordinary least squares routine robust to multicollinearity. The inclusion of $\mathbf{s}_t(\hat{\boldsymbol{\phi}}_T)$ as additional regressors makes the statistic robust to the fact that the influence functions are evaluated at $\hat{\boldsymbol{\phi}}_T$.

Nevertheless, this OPS regression has poor finite sample properties, as stressed by Horowitz (1994) among many others. Fortunately, the parametric bootstrap offers a simple way of improving the reliability of the IM test in empirical applications.

## 2.3 The IM test in the complete data model

The influence functions (6) are often difficult to interpret, so empirical researchers typically regard the IM test as a black-box that offers little guidance on how to improve the model specification when it is rejected. In the context of microeconometric applications, Chesher (1984) provided a reinterpretation of the IM test as a score test of unobserved heterogeneity in the model parameters, which is a first-order concern when dealing with data from individual agents, although the *i.i.d.* nature of his alternative reduces its appeal somewhat in time-series contexts (but see Amengual, Fiorentini and Sentana (2022), who use it to detect random coefficient variation in vector autoregressive processes).

Nevertheless, we can provide a very intuitive interpretation of (6) for the Markov switching model (1) by first considering what the IM test would be if the econometrician could observe not only $I_T$ but also $\Xi_T$. In that case, the time-series model would be the combination of a discrete Markov chain process for $\boldsymbol{\xi}_t$ driven by the dynamic transition matrix $\mathbf{P}(I_{t-1})$ and a conditionally Gaussian VAR model for each regime.

In view of the usual prediction error decomposition for time series models under the maintained assumption that $\mathbf{z}_t$ is a strongly exogenous process, the contribution from observation $t$ to the log-likelihood function from $\mathbf{y}_t$ and $\boldsymbol{\xi}_t$ conditional on the exogenous regressors can be written as the sum of two components, the conditional log-likelihood function of the Gaussian models for $\mathbf{y}_t$ given $\boldsymbol{\xi}_t$, $I_{t-1}$ and $\Xi_{t-1}$:

$$\begin{aligned}\ln f(\mathbf{y}_t|\boldsymbol{\xi}_t, I_{t-1}, \Xi_{t-1}; \mathbf{b}, \boldsymbol{\gamma}) &= \sum_{k=1}^{K} \xi_{kt} \ln f(\mathbf{y}_t|\xi_{kt}=1, I_{t-1}, \Xi_{t-1}; \mathbf{b}_k, \boldsymbol{\gamma}_k) \\ &= \sum_{k=1}^{K} \xi_{kt} \left[ -\frac{M}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Gamma}_k| - \frac{1}{2}\boldsymbol{\varepsilon}_t^{*\prime}(\mathbf{b}_k, \boldsymbol{\gamma}_k)\boldsymbol{\varepsilon}_t^{*}(\mathbf{b}_k, \boldsymbol{\gamma}_k) \right], \quad (8)\end{aligned}$$

where $\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k) = \boldsymbol{\Gamma}_k^{-1/2}(\mathbf{y}_t - \mathbf{B}_k \mathbf{y}_{t-1}^a)$, $\mathbf{b} = (\mathbf{b}_1', \ldots, \mathbf{b}_K')'$, $\mathbf{b}_k = vec(\mathbf{B}_k) = vec(\boldsymbol{\nu}_k, \mathbf{A}_k)$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_K')'$ and $\boldsymbol{\gamma}_k = vech(\boldsymbol{\Gamma}_k)$, and the marginal log-likelihood function of the first-order Markov chain $\boldsymbol{\xi}_t$ given $\Xi_{t-1}$ and $I_{t-1}$:

$$\ln p(\boldsymbol{\xi}_t|I_{t-1}, \Xi_{t-1}; \boldsymbol{\beta}) = \sum_{j=1}^{K} \xi_{jt-1} \sum_{k=1}^{K} \xi_{kt} \ln p_{kjt}(\boldsymbol{\beta}_j). \quad (9)$$

The multiple sequential cuts on the vector of $K[3M(M+1)/2 + (K-1)(L+1)]$ parameters performed by the additive log-likelihood decompositions above allows us to obtain very easily the score vectors, Hessian matrices, and influence functions for the IM test. On that basis, we can show that:

**Proposition 1** *1. The IM matrix test of model (1) with transition matrix (2) when $\boldsymbol{\xi}_t$ is observed coincides with a moment test based on the following groups of influence functions:*

$$\xi_{kt}\mathbf{H}_2[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \otimes vech(\mathbf{y}_{t-1}^a \mathbf{y}_{t-1}^{a\prime}), k = 1, \ldots, K \quad (10)$$

$$\xi_{kt}\mathbf{H}_3[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \otimes \mathbf{y}_{t-1}^a, \quad (11)$$

$$\xi_{kt}\mathbf{H}_4[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)], \quad (12)$$

$$\xi_{jt-1}(vech[\mathbf{u}_{jt}(\boldsymbol{\beta}_j)\mathbf{u}_{jt}'(\boldsymbol{\beta}_j) - \{diag[\mathbf{p}_{tj}(\boldsymbol{\beta}_j)] - \mathbf{p}_{tj}(\boldsymbol{\beta}_j)\mathbf{p}_{tj}'(\boldsymbol{\beta}_j)\}]) \otimes vech(\mathbf{z}_{t-1}^a \mathbf{z}_{t-1}^{a\prime}) \quad (13)$$

$$\xi_{kt}\xi_{jt-1} \left\{ \begin{array}{c} \mathbf{H}_1[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \otimes \mathbf{y}_{t-1}^a \\ \mathbf{H}_2[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \end{array} \right\} \otimes [\mathbf{u}_{jt}(\boldsymbol{\beta}_j) \otimes \mathbf{z}_{t-1}^a] \quad (14)$$

*for $k, j = 1, \ldots, K$, where $\mathbf{y}_{t-1}^a = (1, \mathbf{y}_{t-1}')'$,*

$$\mathbf{H}_r(\boldsymbol{\varepsilon}^*) = \begin{bmatrix} H_{r,0,\cdots,0}(\boldsymbol{\varepsilon}^*) \\ H_{r-1,1,\cdots,0}(\boldsymbol{\varepsilon}^*) \\ \vdots \\ H_{0,\cdots,0,r}(\boldsymbol{\varepsilon}^*) \end{bmatrix} = \begin{bmatrix} H_r(\varepsilon_1^*) \\ H_{r-1}(\varepsilon_1^*)H_1(\varepsilon_2^*) \\ \vdots \\ H_r(\varepsilon_M^*) \end{bmatrix}$$

*is the* $\binom{M+r-1}{r}$ *vector containing the distinct multivariate Hermite polynomials of order* $r$ *of a standardised random vector* $\boldsymbol{\varepsilon}^*$*, and*

$$\mathbf{u}_{jt}(\boldsymbol{\beta}_j) = [u_{2jt}(\boldsymbol{\beta}_j), \ldots, u_{Kjt}(\boldsymbol{\beta}_j)]' = (\xi_{2t} - p_{2jt}(\boldsymbol{\beta}_j), \ldots, \xi_{Kt} - p_{Kjt}(\boldsymbol{\beta}_j)]', \quad (15)$$

*is the vector of generalised residuals associated to the non-normalised elements of the* $j^{th}$ *column of the transition matrix*

$$\mathbf{p}_{jt}(\boldsymbol{\beta}_j) = [p_{2jt}(\boldsymbol{\beta}_j), \ldots, p_{Kjt}(\boldsymbol{\beta}_j)]'.$$

2. *The asymptotic covariance matrices of those influence functions evaluated at the MLE corrected for the sampling uncertainty in estimating the model parameters is the residual covariance matrix in the multivariate regression of (10)-(14) onto the following influence functions*

$$\xi_{kt}\mathbf{H}_1[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \otimes \mathbf{y}_{t-1}^a, \quad (16)$$
$$\xi_{kt}\mathbf{H}_2[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)], \quad (17)$$
$$\xi_{jt-1}\mathbf{u}_{jt}(\boldsymbol{\beta}_j) \otimes \mathbf{z}_{t-1}^a. \quad (18)$$

*for* $k, j = 1, \ldots, K$.

3. *The asymptotic distribution of the IM test under correct specification will be a* $\chi^2$ *random variable with degrees of freedom equal to*

$$K\left\{\frac{M(M+1)\left(11M^2+35M+14\right)}{24} + \frac{K(K-1)(L+1)(L+2)}{4} + \frac{3M(M+1)(K-1)(L+1)}{2}\right\}. \quad (19)$$

Proposition 1 allows us to provide a very intuitive interpretation to the influence functions underlying the IM test. In particular, (10) can be regarded as the multivariate counterpart to White's (1980) heteroskedasticity test, while (11) is a multivariate version of what Bera and Lee (1993) called a test for conditional "heterocliticity" in linear regression models, and (12) the multivariate analogue to the Kiefer and Salmon (1983) version of the kurtosis component of the Jarque and Bera (1980) test.

In turn, (13) is effectively testing the conditional mean independence of the conditionally demeaned outer product of the generalised residuals associated to the $j^{th}$ column of the transition matrix with respect to the elements of $\mathbf{z}_{t-1}$. Thus, it also resembles a multivariate version of White's (1980) test for residual conditional heteroskedasticity, which in turn confirms Chesher's (1984) reinterpretation of the IM test as a score test for neglected unobserved heterogeneity.

Finally, (14) assesses the conditional independence – given $\mathbf{y}_{t-1}$ and $\mathbf{z}_{t-1}$ – between the standardised innovations of the regimes, their squares and cross-products and the generalised residuals of the multinomial logit models determining the transition matrix. These influence functions arise because the model implicitly assumes that the parameters of the conditional mean and covariance matrices of each regime at time $t$ do not depend on the regime that prevailed at time $t-1$.

In contrast, there are no IM influence functions associated to the cross-products of the mean and variance parameters across regimes because the Hessian is block diagonal and the cross-products of the score vectors (16) and (17) for different $k's$ are identically 0. For analogous reasons, there are no IM influence functions for the cross-products of the parameters of the multinomial logit models for different columns of the transition matrix either. Consequently, $rank[\mathcal{R}(\boldsymbol{\phi}_0) - \mathcal{U}(\boldsymbol{\phi}_0)\mathcal{I}^{-1}(\boldsymbol{\phi}_0)\mathcal{U}(\boldsymbol{\phi}_0)]$ is substantially lower than the dimension of (6).

In summary, the IM test for the entire model coincides with the combination of the IM matrix test for linear VAR(1) models in Amengual, Fiorentini and Sentana (2022) for each of the $K$ regimes, the IM test for multinomial logit models in Amengual, Fiorentini and Sentana (2025) for each of the $K$ columns of the transition probability matrix, and some additional cross-model terms related to the independence of the residuals of those two types of models.

Importantly, when the transition matrix is constant so that $\mathbf{z}_{t-1}^a = 1$ for all $t$, a linear combination of some of the elements of (14) can be spanned by the scores (16) and (17) because

$$
\begin{aligned}
&\sum_{j=1}^{K} \xi_{kt}\xi_{jt-1} \left\{ \begin{array}{c} \mathbf{H}_1[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \otimes \mathbf{y}_{t-1}^a \\ \mathbf{H}_2[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \end{array} \right\} \otimes \left\{ \left[ \begin{array}{c} \xi_{2t} - p_{2j}(\boldsymbol{\beta}_j) \\ \vdots \\ \xi_{kt} - p_{Kj}(\boldsymbol{\beta}_j) \\ \vdots \\ \xi_{Kt} - p_{Kj}(\boldsymbol{\beta}_j) \end{array} \right] \right\} \\
= \; & \xi_{kt} \left\{ \begin{array}{c} \mathbf{H}_1[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \otimes \mathbf{y}_{t-1}^a \\ \mathbf{H}_2[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \end{array} \right\} \otimes [\mathbf{e}_k - \mathbf{p}_j(\boldsymbol{\beta}_j)],
\end{aligned}
$$

where $\mathbf{e}_k$ is a $(K-1)$ vector of 0's except for a 1 in position $k-1$, since $\xi_{kt}\xi_{lt} = I(k=l)$ and $\sum_{j=1}^{K} \xi_{jt-1} = 1$. In that case, therefore, we should eliminate all the elements of (14) corresponding to one column of the transition matrix, say the first one.

## 2.4 The IM test in the incomplete data model

Let

$$
w_{kt|T}(\boldsymbol{\phi}) = P(\xi_{kt} = 1 | I_T; \boldsymbol{\phi}) \tag{20}
$$

denote the smoothed probability of the marginal event $\xi_{kt} = 1$, and

$$
w_{kjt|T}(\boldsymbol{\phi}) = P(\xi_{kt} = 1, \xi_{jt-1} = 1 | I_T; \boldsymbol{\phi}) \tag{21}
$$

the corresponding smoothed probability for the joint event $\xi_{kt} = 1, \xi_{jt-1} = 1$. Importantly, Kim's (1994) smoother, which computes (20) from (21), is exact in this context because $\mathbf{z}_t$ is strongly exogenous.[2]

We can then show that

[2] Chapter 16 of Hamilton (1994) provides recursive expressions for these quantities, while Friedmann (1994) and Yang (2001) provide non-recursive ones.

**Proposition 2** *1. The IM matrix test of model (1) with transition matrix (2) when $\boldsymbol{\xi}_t$ is unobserved coincides with a moment test based on the following groups of influence functions:*

$$w_{kt|T}(\boldsymbol{\phi}) \cdot \mathbf{H}_2[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \otimes vech(\mathbf{y}_{t-1}^a \mathbf{y}_{t-1}^{a\prime}), \tag{22}$$
$$w_{kt|T}(\boldsymbol{\phi}) \cdot \mathbf{H}_3[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \otimes \mathbf{y}_{t-1}^a, \tag{23}$$
$$w_{kt|T}(\boldsymbol{\phi}) \cdot \mathbf{H}_4[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)], \tag{24}$$

*for $k = 1, \ldots, K$,*

$$\{[w_{kjt|T}(\boldsymbol{\phi}) - p_{kjt}(\boldsymbol{\beta}_j) w_{jt-1|T}(\boldsymbol{\phi})][1-2p_{kjt}(\boldsymbol{\beta}_j)]\} \otimes vech(\mathbf{z}_{t-1}^a \mathbf{z}_{t-1}^{a\prime}) \tag{25}$$

$$\{[w_{kjt|T}(\boldsymbol{\phi}) - p_{kjt}(\boldsymbol{\beta}_j) w_{jt-1|T}(\boldsymbol{\phi})] p_{\ell jt}(\boldsymbol{\beta}_j) + [w_{\ell jt|T}(\boldsymbol{\phi}) - p_{\ell jt}(\boldsymbol{\beta}_j) w_{jt-1|T}(\boldsymbol{\phi})] p_{kjt}(\boldsymbol{\beta}_j)\} \otimes vech(\mathbf{z}_{t-1}^a \mathbf{z}_{t-1}^{a\prime}) \tag{26}$$

*for $j = 1, \ldots, K$, $k = 2, \ldots, K$ and $\ell = 3, \ldots, K$, with $k > \ell$, and*

$$\left\{ \begin{array}{c} \mathbf{H}_1[\varepsilon_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \otimes \mathbf{y}_{t-1}^a \\ \mathbf{H}_2[\varepsilon_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \end{array} \right\} \otimes \left\{ w_{kjt|T} \left[ \begin{array}{c} -p_{2jt}(\boldsymbol{\beta}_j) \\ \vdots \\ 1 - p_{kjt}(\boldsymbol{\beta}_j) \\ \vdots \\ -p_{Kjt}(\boldsymbol{\beta}_j) \end{array} \right] \otimes \mathbf{z}_{t-1}^a \right\} \tag{27}$$

*for $k, j = 1, \ldots, K$.*

*2. The asymptotic covariance matrices of the (scaled) sampling averages of (22)-(27) evaluated at the MLE corrected for the sampling uncertainty in estimating the model parameters is the covariance matrix of the residuals in the limiting least squares projections of those (scaled) sample averages onto the linear span of the (scaled) sample averages of the following influence functions*

$$w_{kt|T}(\boldsymbol{\phi}) \mathbf{H}_1[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \otimes \mathbf{y}_{t-1}^a, \tag{28}$$
$$w_{kt|T}(\boldsymbol{\phi}) \mathbf{H}_2[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)], \tag{29}$$

*for $k = 1, \ldots, K$ and*

$$[w_{kjt|T}(\boldsymbol{\phi}) - p_{kjt}(\boldsymbol{\beta}_j) w_{jt-1|T}(\boldsymbol{\phi})] \otimes \mathbf{z}_{t-1}^a \tag{30}$$

*for $j = 1, \ldots, K$ and $k = 2, \ldots, K$.*

*3. The asymptotic distribution of the IM test will be a $\chi^2$ random variable with the same number of degrees of freedom (19) as in Proposition 1.*

It is straightforward to see that the sum of the influence functions (28) over $t$ from 1 to $T$ constitutes a basis for the scores of $\mathbf{b}_k$ obtained by means of the EM principle in Hamilton (1990), while the corresponding sum of (29) is the basis for the scores for $\boldsymbol{\gamma}_k$ obtained in the same manner. In turn, the analogous sum of the influence functions (30) effectively coincide with the scores derived by Diebold, Lee and Weinbach (1996) for the parameters characterising the transition probability matrix using again the EM principle. Intuitively, the sum of (28) coincides with the average of the usual orthogonality conditions for the regression parameters of the $k^{th}$

regime weighted by the smoothed probability that $\mathbf{y}_t$ belongs to that regime. Similarly, the sum of (29) forces the covariance matrix of the standardised residuals of the $k^{th}$ multivariate regression to be the identity matrix once we weight them again by the smoothed probability that $\mathbf{y}_t$ belongs to that regime. Finally, the sum of (30) imposes that the differences between the smoothed probability of the joint event $\xi_{kt} = 1, \xi_{jt-1} = 1$ and the product of the smoothed probability of $\xi_{jt-1} = 1$ multiplied by the probability from transitioning from state $j$ to state $k$ should be orthogonal to the lagged values of the observed variables that determine the probability of this transition.

In addition, we can interpret (22) as the expected value of the influence functions that underlie the multivariate test for heteroskedasticity in the regression residuals of the $k^{th}$ regime weighted again by the smoothed probability that $\mathbf{y}_t$ belongs to that regime. Entirely analogous comments apply to (23) in relation to testing for conditional heterocliticity in those residuals, and (24) as far as their unconditional kurtosis in concerned. In turn, we can interpret (25) as the expected value given the observed data of the IM influence functions related to the multinomial logit probability that the Markov chain will sojourn in regime $k$ for one additional period, while (26) corresponds to analogous conditional expected value of the IM influence functions associated to the probability that it will transit from regime $j$ at time $t-1$ to either regime $k$ or regime $\ell$ at $t$. Finally, we can interpret the cross-model influence functions (27) as checking the conditional independence – given $\mathbf{y}_{t-1}$ and $\mathbf{z}_{t-1}$ – between the standardised innovations of the regimes, their squares and cross-products and the generalised residuals of the multinomial logit models.

Naturally, the results in section 2.3 imply that in the incomplete data case there are no IM influence functions either associated to the cross-products of the mean and variance parameters across regimes or the cross-products of the parameters of the multinomial logit models associated to different columns of the transition matrix. Those results also imply that if the transition matrix is constant, then we should once again eliminate one column of (27) to avoid that a linear combination of these influence functions be spanned by the score vector. In turn, if there are only two regimes, we should use (25) for the second regime and (26) for the first one given our normalisation.

The calculation of the asymptotic covariance matrices is more involved than in the observed case. The source of the problem is that the smoothed values of the state variables are serially correlated, which in turn implies that both (22)-(27) and (28)-(30) will be serially correlated too.

From a practical point of view, there are three possible solutions. The first one draws inspiration from Hamilton (1996), who in a univariate model with constant transition matrix effectively used the first differences of the cumulative sums of (28), (29) and (30) to obtain the contributions to the scores from the conditional distribution of $\mathbf{y}_t$ given $I_{t-1}$. Given that such contributions evaluated at the true parameter values constitute a martingale difference sequence under correct specification, their sample covariance matrix provides a consistent OPS estimator of the information matrix.

For the purposes of the IM test, though, we need to prove that the martingale difference property holds not only for the scores, but also for the influence functions underlying it. To understand why this is indeed the case, let us consider the mean parameters for the sake of brevity. The first difference of the cumulative sum of (28) up to observation $t$ is

$$\begin{aligned}
& \sum_{s=1}^{t} w_{ks|t}(\boldsymbol{\phi})\mathbf{H}_1[\boldsymbol{\varepsilon}_s^*(\mathbf{b}_k,\boldsymbol{\gamma}_k)]\otimes\mathbf{y}_{s-1}^a - \sum_{s=1}^{t-1} w_{ks|t-1}(\boldsymbol{\phi})\mathbf{H}_1[\boldsymbol{\varepsilon}_s^*(\mathbf{b}_k,\boldsymbol{\gamma}_k)]\otimes\mathbf{y}_{s-1}^a \\
= \ & w_{kt|t}(\boldsymbol{\phi})\mathbf{H}_1[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k,\boldsymbol{\gamma}_k)]\otimes\mathbf{y}_{t-1}^a + \sum_{s=1}^{t-1}[w_{ks|t}(\boldsymbol{\phi}) - w_{ks|t-1}(\boldsymbol{\phi})]\mathbf{H}_1[\boldsymbol{\varepsilon}_s^*(\mathbf{b}_k,\boldsymbol{\gamma}_k)]\otimes\mathbf{y}_{s-1}^a.
\end{aligned}$$

Given the orthogonality properties of conditional expectations, updates in the smoothed probabilities

$$[w_{ks|t}(\boldsymbol{\phi}) - w_{ks|t-1}(\boldsymbol{\phi})]$$

are necessarily mean independent of $I_{t-1}$, which effectively makes the second term non-linearly unpredictable from the point of view of $t-1$. In addition,

$$\begin{aligned}
w_{kt|t}(\boldsymbol{\phi})\mathbf{H}_1[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k,\boldsymbol{\gamma}_k)] &= E\{\xi_{kt}\mathbf{H}_1[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k,\boldsymbol{\gamma}_k)]|I_{t-1}\} \\
&= E[E\{\xi_{kt}\mathbf{H}_1[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k,\boldsymbol{\gamma}_k)]|\xi_{kt},I_{t-1}\}|I_{t-1}] = 0
\end{aligned}$$

because the inner expectation is either 0 when $\xi_{kt}=0$ with probability $1-P(\xi_{kt}=1|I_{t-1})$ or $E\{\mathbf{H}_1[\boldsymbol{\varepsilon}_t^*(\mathbf{b}_k,\boldsymbol{\gamma}_k)]|\xi_{kt}=1,I_{t-1}\}=0$ when $\xi_{kt}=1$ with probability $P(\xi_{kt}=1|I_{t-1})$, which means that the first term is also non-linearly unpredictable from the same point of view. Importantly, exactly the same arguments apply to (22)-(27), so the first differences of their cumulative sums will also be martingale difference sequences under correct specification. Intuitively, the conditional version of the information matrix equality ensures the martingale difference nature of those influence functions. Consequently, we can follow Chesher (1983) and Lancaster (1984) in regressing a vector of $T$ ones onto the first differences of the cumulative sums of all the influence functions that appear in Proposition 2.

The second practical approach exploits Corollary 1 in Amengual, Fiorentini and Sentana (2024a), which expresses the theoretical covariance matrix of the IM influence functions and the scores as the difference between the unconditional covariance matrix of the analogous influence functions in the complete data model and the mean of their covariance matrix conditional on the observed variables. The first element of this expression coincides with the covariance matrix of the influence functions that appear in Proposition 1. The conditional covariance matrix of their sum given $I_T$ is trickier because $\boldsymbol{\xi}_t$ is serially correlated. Nevertheless, we can exploit the following expressions in Krolzig (1994) to compute their conditional autocovariances:

$$\begin{aligned} V(\boldsymbol{\xi}_t|I_T;\boldsymbol{\phi}) &= diag[\mathbf{w}_{t|T}(\boldsymbol{\phi})] - \mathbf{w}_{t|T}(\boldsymbol{\phi})\mathbf{w}'_{t|T}(\boldsymbol{\phi}), \\ Cov(\boldsymbol{\xi}_{t+1},\boldsymbol{\xi}_t|I_T;\boldsymbol{\phi}) &= \mathbf{P}(I_t)V(\boldsymbol{\xi}_t|I_T;\boldsymbol{\phi}) - \mathbf{w}_{t+1|T}(\boldsymbol{\phi})\mathbf{w}'_{t|T}(\boldsymbol{\phi}), \\ Cov(\boldsymbol{\xi}_{t+2},\boldsymbol{\xi}_t|I_T;\boldsymbol{\phi}) &= \mathbf{P}^2(I_t)V(\boldsymbol{\xi}_t|I_T;\boldsymbol{\phi}) - \mathbf{w}_{t+2|T}(\boldsymbol{\phi})\mathbf{w}'_{t|T}(\boldsymbol{\phi}), \end{aligned}$$

etc., where $\mathbf{w}_{t|T} = [w_{1t|T}(\boldsymbol{\phi}), \ldots, w_{Kt|T}(\boldsymbol{\phi})\}'$. On this basis, we can tediously obtain the conditional autocovariances of all the influence functions that appear in Proposition 2, whose doubly infinite sum will give us the required long-run covariance matrix of the sample average of those influence functions. Although we cannot obtain closed-form expressions for the matrix of autocovariance generating functions evaluated at 1, we can follow a procedure analogous to the one in Almuzara, Amengual and Sentana (2019), which effectively truncates the sum once the effect of including additional autocovariances is numerically negligible.

Finally, there is a third solution that combines aspects of the previous two. Given that the parametric model is fully specified, one can compute the elements of (7) evaluated at the maximum likelihood estimates to any desired degree of accuracy using the first method but in a single simulated sample of size $T_s$, where $T_s$ is much larger than $T$, with the true parameter values set to the estimated ones, as explained in Mencía and Sentana (2012).

# 3 Monte Carlo simulations

As stated in Proposition 2, the asymptotic distribution of our proposed IM test is $\chi^2$ with degrees of freedom equal to (19). Similarly, each group of influence functions for the IM test that appear in Proposition 2 also gives rise to moments tests whose limiting distributions are $\chi^2$, with degrees of freedom equal to $\frac{1}{4}KM^2(M+1)(M+3)$ (conditional heteroskedasticity), $\frac{1}{6}K(M+2)(M+1)M^2$ (conditional heterocliticity given $\mathbf{y}_{t-1}$ only), $\frac{1}{6}K(M+2)(M+1)M$ (unconditional skewness), $\frac{1}{24}K(M+3)(M+2)(M+1)M$ (unconditional kurtosis), $\frac{1}{4}K^2(K-1)(M+1)(M+2)$ (multinomial logit conditional heteroskedasticity) and $\frac{3}{2}KM(M+1)(K-1)(L+1)$ (independence between regression residuals and generalised ones). However, these asymptotic approximations might not be very reliable in finite samples. For that reason, we conduct some Monte Carlo experiments to study the rejection rates under the null of correct specification in sample sizes of $T = 250$ and $T = 1,000$. To do so, we simulate data from the following conditional model:

$$
\begin{aligned}
y_t|I_{t-1}, \Xi_t &\sim N(\nu_{\xi_t} + a_{\xi_t,1} y_{t-1}, \gamma^2_{\xi_t}) \\
P(\xi_t &= 1|\xi_{t-1} = 1, I_{t-1}, \Xi_{t-1}) = \frac{1}{1 + e^{\boldsymbol{\beta}'_{21}\mathbf{z}^a_{t-1}}} \quad (31)\\
P(\xi_t &= 2|\xi_{t-1} = 2, I_{t-1}, \Xi_{t-1}) = \frac{e^{\boldsymbol{\beta}'_{22}\mathbf{z}^a_{t-1}}}{1 + e^{\boldsymbol{\beta}'_{22}\mathbf{z}^a_{t-1}}}, \quad (32)
\end{aligned}
$$

where $\mathbf{z}^a_{t-1} = (1, z_{t-1})'$, and $z_{t-1}$ is the term spread variable in the empirical application in section 4, which we treat as fixed in repeated samples given its assumed strongly exogenous nature under the null.[3] We borrow some of the true parameter configurations from the related Monte Carlo experiments in PPS for the design in which the correlations between the standardised shocks to $y_t$ within each regime and the regressor $z_t$ are 0.

We consider three different versions of this model to assess the different components of our proposed IM test in practice:

1. A model with a constant transition matrix, so that $\beta_{21,2} = \beta_{22,2} = 0$.

   As for the 8 free parameters, we set $\nu_1 = -1$, $\nu_2 = 1$, $a_{11} = a_{21} = 0.9$, $\gamma^2_1 = \gamma^2_2 = 1$, and $\beta_{21,1}$ and $\beta_{22,1}$ such that $P(\xi_t = 1|\xi_{t-1} = 1, I_{t-1}, \Xi_{t-1}) = p = 0.76$ and $P(\xi_t = 2|\xi_{t-1} = 2, I_{t-1}, \Xi_{t-1}) = q = 0.93$. In this case, the IM test is exclusively based on (22), (23), (24) and (27) for each of the regimes, as the sample means of (25) and (26) are identically equal to 0 at the maximum likelihood estimators. Consequently, the IM test follows a $\chi^2$ with 16 degrees of freedom in large samples despite (6) having 36 different influence functions.

2. A model without dynamics within regimes, so that $a_{1,1} = a_{2,1} = 0$, in which the transitions between regimes are governed by the logit models (31) and (32). As for the remaining 8 free parameters, we set $\nu_1 = -10$, $\nu_2 = 10$, $\gamma^2_1 = \gamma^2_2 = (1 - .9^2)^{-1}$, and $\beta_{21,1} = -2$, $\beta_{21,2} = -0.5$, $\beta_{22,1} = 2$ and $\beta_{22,2} = 0.5$. Relative to the previous case, one needs to add the influence function (25) and (26) for the second and first regime, respectively, but eliminate the influence functions (22) for each of the regimes and only preserve the component of (23) related to the constant, which effectively checks unconditional skewness. Consequently, the joint IM test has 26 degrees of freedom rather than 36, which are the number of different elements of (6) in this case.

3. The combination of the previous two, which requires a total of 10 parameters. The IM test now includes all the influence functions and therefore has 40 degrees of freedom rather than

[3] For $T = 250$, we downloaded updated data for the period 1959Q4 and 2024Q2 from the New York Fed web page https://www.newyorkfed.org/research/capital_markets/ycfaq#/interactive, while for $T = 1000$, we use four suitably concatenated copies of this series.

$10 \times 11/2 = 55$. The values of the conditional mean and variance parameters are set as in the first model while the transition probability parameters as in the second one.

For each of these specification, we look at the OPS version of the joint IM test and its relevant components. Following Hamilton (1996), we also compute White's (1987) dynamic information matrix test based on the autocorrelation matrices of the scores, which is effectively testing the conditional version of the first Barlett identity rather than the unconditional version of the second one (see chapter 11 of White (1994) for further details). To keep the number of degrees of freedom of this last test manageable, we only look at the scores' marginal first-order autocorrelations, so that the number of influence functions coincides with the number of model parameters as opposed to its square.[4]

In all the null designs, we generate $10,000$ samples and compare asymptotic critical values to those based on a parametric bootstrap procedure in which we simulate $B = 99$ samples from the model estimated under the null.[5] A substantially larger value of $B$ is feasible in an empirical application such as the one in section 4, but prohibitively costly for $10,000$ Monte Carlo replications.

To ensure that we have indeed maximised the log-likelihood function, in each Monte Carlo replication we check that the Euclidean norm of the sum of the scores evaluated at the maximum likelihood estimators is less than $10^{-5}$. We also check that the $TR^2$ version of the IM test does not decrease when we add influence functions as regressors in the regression of the vector of ones on those influence functions and the scores. When these checks fail, we discard the replication. We apply exactly the same checks to the parametric bootstrap procedure until we get 99 valid bootstrap samples, although if the number of samples required exceeds 109, we also discard the entire Monte Carlo replication. In practice, we find that these checks are almost always satisfied, with the proportion of discarded replications often zero or at most 0.1% in all designs except in models with asymmetric Student $t$ innovations when it is slightly larger, possibly because an occasional unusually large outlier makes the likelihood maximisation algorithm numerically unstable.

In Table 1, we present the results for the three different versions of the model grouping together the tests for unconditional asymmetry and kurtosis in the interest of space. We find substantial over-rejections using the asymptotic-based critical values, which confirms the need for

[4] Including all possible scores' first-order autocorrelations would result in a test with 64 degrees of freedom in the first and second models, and 100 degrees of freedom in the third one. Higher-order autocorrelations would multiply these numbers by the number of lags.

[5] Given the number of replications, the 95% asymptotic confidence intervals for the Monte Carlo rejection probabilities under the null are (.80,1.20), (4.57,5.43) and (9.41,10.59) at the 1%, 5% and 10% levels, respectively.

finite sample size adjustments. As can be seen, our parametric bootstrap procedure substantially improves the reliability of the tests in finite samples. Still, the rejection rates of the joint IM test, the MN logit component and the cross-model test are lower than their nominal values in samples of size $T = 250$, especially so in the versions of the model with covariate-dependent transitions, with no other noticeable differences across specifications. As expected, though, all bootstrap-based p-values become reliable for $T = 1000$.

Turning now to the ability of the IM test to detect misspecification, for each version of the model considered in the size experiments we specify several alternatives to assess finite sample power using this time $2,500$ replications. For the model with AR(1) processes within regimes but a constant transition matrix in particular, we consider:

1a conditionally homoskedastic Gaussian AR(2) processes within regimes but the same constant transition matrix as under the null. The AR(2) parameters are $a_{1,1} = a_{2,1} = 1.1$ and $a_{1,2} = a_{2,2} = -0.18$. This amounts to adding an extra autoregressive root equal to 0.2 to the specification of the conditional means of the regimes under the null.

1b conditionally Gaussian AR(1)-GQARCH(1,1) processes within regimes but the same autoregressive specification and constant transition matrix as under the null. Specifically, we generate the conditional variance of the (standardised) autoregressive innovations $\varepsilon_t^* = (y_t - \nu_{\xi_t} - a_{\xi_t,1}y_{t-1})\gamma_{\xi_t}^{-1}$ as $\psi_t^2 = 0.18 + 0.2(\varepsilon_{t-1}^* - 0.9)^2 + 0.6\psi_{t-1}^2$.

1c conditionally homoskedastic AR(1) processes within regimes with the same means, variances and constant transition matrix as under the null but with asymmetric Student $t$ innovations within regimes whose shape parameters are $\eta_0 = \eta_1 = 1/12$, $b_0 = -5$ and $b_1 = 5$ (see Mencía and Sentana (2012) for details).

1d conditionally homoskedastic Gaussian AR(1) processes within regimes in which two univariate logit models that depend on the single strictly exogenous regressor $z_t$ determine the transitions between regimes, with $\beta_{21,1} = -2$, $\beta_{22.1} = 2$, and $\beta_{21,2} = \beta_{22,2} = 0.5$.

We report the Monte Carlo rejection rates in Table 2. Under incorrect specification, we find non-negligible power against alternatives 1a to 1c. Still, some components of the IM test are better suited than others to detect specific alternatives. For example, the heteroskedasticity component of the IM test is good at detecting GQARCH effects (see Panel B) while the asymmetry and kurtosis ones capture non-normality (see Panel C). As expected, the test based on the autocorrelation of the scores is the most powerful under misspecified conditional mean dynamics, but the cross-model component of the IM test is almost on par with it despite the fact that none of its influence functions targets neglected serial correlation. Unfortunately, all tests struggle to detect time-variation in the transition probabilities when the estimated model assumes that they are constant, which is not surprising given that none of them involve $z_t$.

We provide some additional insights by looking at the median of the maximum likelihood estimators and their Monte Carlo interquartile ranges (IQRs) under incorrect specification. As can be seen in Table 3, power is intimately related to the extent of the inconsistencies in the parameter estimators. For example, the results in Panel D show that the low power of the tests to detect covariate-dependent transitions may be partly the result of the fact that the conditional mean and variance parameters of both regimes appear to be consistently estimated despite incorrectly assuming a constant transition probability (see Pouzo, Psaradakis and Solá (2024)).

For the model without dynamics within regimes in which two univariate logit models that depend on the single strictly exogenous regressor $z_t$ determine the transitions between regimes, we consider the following misspecified alternatives:

2a conditionally homoskedastic Gaussian AR(1) processes within regimes with $a_{1,1} = a_{2,1} = 0.2$, but the same covariate-dependent transition matrix as under the null

2b conditionally Gaussian serially uncorrelated GQARCH(1,1) processes as in 1b and the same parameters and covariate-dependent transition matrix as under the null

2c asymmetric Student $t$ innovations as in 1c with the same parameters and covariate-dependent transition matrix as under the null

2d no dynamics within regimes with a transition matrix that depends on $z_t$ when this variable is predetermined instead of strongly exogenous. Specifically, we allow for contemporaneous correlation between the shocks driving $y_t$ within each regime and the shocks driving $z_t$ by simulating $y_t = \nu_{\xi_t} + \gamma_{\xi_t}(\epsilon_t + \rho\eta_t)$, with $\epsilon_t \sim i.i.d.\ N[0, (1+\rho^2)^{-1}]$, $\rho = .8$ as in PPS, and $\eta_t$ the standardised residuals from the autoregression of the slope of the term structure on a constant and its first four lags.

We report the results of this second set of power experiments in Table 4. In this case, the cross-model component of the IM test is the best at detecting misspecified conditional mean dynamics (Panel A), while the (conditional and unconditional) skewness and kurtosis components detect non-normality of the innovations (Panel C). Somewhat surprisingly, the test based on the autocorrelation of the scores is better than the conditional heteroskedasticity test for detecting misspecified conditional variance dynamics, even though they both exploit the non-zero autocorrelation of the square observations (Panel B). All tests have a hard time in detecting failure of regressor strict exogeneity, with the exception of the cross-model component, which has nonnegligible power in the larger samples (Panel D). In turn, the MN logit heteroskedasticity component of the IM test has very little power for all these alternatives. Finally, the median parameter estimates and IQRs in Table 5 confirm that power is intimately related to the extent of the inconsistencies in the parameter estimators, as in the first set of power experiments.

The last set of power experiments correspond to the general model with conditionally homoskedastic Gaussian AR(1) processes within regimes in which two univariate logit models that depend on the single strictly exogenous regressor $z_t$ determine the transitions between regimes. We consider the following alternatives:

3a conditionally homoskedastic Gaussian AR(2) processes as in 1a and the same parameters and covariate-dependent transition matrix as under the null

3b conditionally Gaussian AR(1)-GQARCH(1,1) processes as in 1b and 2b and the same parameters and covariate-dependent transition matrix as under the null

3c conditionally homoskedastic AR(1) processes with asymmetric $t$ innovations as in 1c and 2c and the same parameters and covariate-dependent transition matrix as under the null

3d conditionally homoskedastic Gaussian AR(1) processes within regimes with a transition matrix that depends on $z_t$ when this variable is predetermined instead of strongly exogenous. Once again, we allow for contemporaneous correlation between the shocks driving $y_t$ within each regime with the shocks driving $z_t$ by simulating $y_t = \nu_{\xi_t} + a_{\xi_t,1}y_{t-1} + \gamma_{\xi_t}(\epsilon_t + \rho\eta_t)$, with the remaining details as in 2d.

We report the Monte Carlo rejection rates in Table 6. In this case, all tests have some power against predetermined regressors, especially for the larger sample size. Misspecified conditional mean dynamics is mostly detected by the cross-model component of the IM test, and especially the serial correlation test for the scores. As in Table 2, the heteroskedasticity test is again the best at detecting GQARCH dynamics in the conditional variance, while the skewness and kurtosis components of the IM test are unsurprisingly best for capturing departures from normality. Once more, the results in Table 7 confirm the close connection between parameter inconsistencies and test power.

In summary, our simulation exercises confirm the need for finite sample size adjustments under correct specification, with the simple parametric bootstrap procedure achieving rather accurate sizes. We also find that the IM test has nonnegligible power against several empirically relevant alternatives, with some components better suited than others to detect specific forms of specification failure. It is precisely this diversity in the responses of the IM test components what provides very useful guidance to identify the possible causes of misspecification.

# 4 Empirical illustration

As an illustration of their theoretical results, PPS investigate the potential contribution of the spread between the rates of the 10-year Treasury note and the 3-month Tbill in predicting regime changes in US real output growth. To do so, they obtained quarterly data for the period 1954:3-2009:2 from the FRED database for interest rates. Then, they estimated a two-regime Markov switching model for GDP growth that postulates univariate conditionally homoskedastic Gaussian AR(4) processes within each regime with different intercepts but common autoregressive coefficients and residual variances, together with two univariate logit models for the transition probabilities that depend on the term spread as the single regime-change predictor. Although PPS also considered joint maximum likelihood estimators of an augmented model that postulates another univariate AR(4) process for the term spread regressor, we focus on what they call partial maximum likelihood estimator, which treats the predictor variable as strongly exogenous by implicitly assuming no correlation between its innovations and the innovations in GDP growth.

We are able to practically replicate the parameter estimates in PPS, with some very minor differences due to slightly different initialisation conditions for the filter. However, we need to modify our tests to take into account the restriction that the autoregressive coefficients and residual variances are common across regimes. Although the modification is relatively straightforward thanks to the chain rule for first- and second-order derivatives, the interpretation of the components of the tests related to the autoregressive coefficients and their degrees of freedom are different. To improve the reliability of the parametric boostrap-based p-values, we consider $B = 9999$ simulations.

Looking at the various components of the IM test, we are unable to reject the null hypothesis of correct specification of the two logit models, or the conditional and unconditional symmetry and platykurtosis of the innovations in the autoregressive process. In contrast, we find some weak evidence against the lack of correlation between the levels and squares of those innovations and the generalised residuals of the two logit models (*p-value* = 7.68%), and much stronger evidence against the null of conditional homoskedasticity of the AR(4) innovations (*p-value* = 0.91%). When we consider all those components together, we find that the *p-value* of the joint IM test is 3.82%. We interpret our results as suggestive evidence that the autoregressive coefficients may in fact be regime-specific rather than common.

## 5 Directions for further research

The model we have considered in this paper may be excessively general for some purposes. For example, an empirical researcher might have good reasons to restrict some elements of $\mathbf{b}_k$ or $\boldsymbol{\gamma}_k$ to be common across regimes. As we explained in the empirical application in section 4, the usual chain rules for first and second derivatives would immediately yield the relevant influence functions for the IM test and their asymptotic covariance matrices in those restricted models. Similarly, in models with three or more regimes, it may be sensible to assume that the transition matrix is tridiagonal so that the state variable can only move to adjacent regimes, in which case the effective dimension of the multinomial models for each of the columns will be reduced. Alternatively, one could consider a dynamic ordered logit model that determines the current regime as a function of a dynamic latent variable and some strongly exogenous regressors, as in Chang, Choi and Park (2017) (see Duekker, Solá and Spagnolo (2007) for a closely related proposal).

Another straightforward extension of our test would be to consider the joint model for $\mathbf{y}_t$ and $\mathbf{z}_t$ proposed by PPS and used in some of our Monte Carlo simulations, in which the determinants of the transition matrix probabilities follow an autoregressive process whose shocks are correlated to the shocks to the observed variables within each regime. Such a test would add the influence functions for the observed VAR(1) process for $\mathbf{z}_t$ discussed in section 2.3 together with some additional influence functions related to the correlation between the shocks. Similarly, we could consider models in which in addition to the recurrent regimes, there is a one-off structural break, as in Ravn and Solá (1999) or Psaradakis and Solá (2024).

In fact, the IM tests considered in this paper can be extended to a much wider class of dynamic state space models with discrete and continuous latent variables that are routinely used in macroeconometric and empirical finance applications, including linear and non-linear models with stochastic volatility and non-Gaussian shocks, as long as they can be written in the incomplete data framework of Dempster, Laird and Rubin (1978) after a suitable data augmentation. The main difference would be that numerical techniques, such as Markov chain Monte Carlo or particle filters, would often be required for smoothing purposes.

# References

Almuzara, T., D. Amengual and E. Sentana (2019): "Normality tests for latent variables", *Quantitative Economics* 10, 981–1017.

Amengual, D., Bei, X., Carrasco, M. and Sentana, E. (2025): "Score-type tests for Markov switching models", mimeo, CEMFI.

Amengual, D., Fiorentini, G. and Sentana, E. (2022): "Tests for random coefficient variation in vector autoregressive models", in J.J. Dolado, L. Gambetti and C. Matthes (eds.) *Essays in honour of Fabio Canova*, *Advances in Econometrics* 44B, 1-35, Emerald.

Amengual, D., Fiorentini, G. and Sentana, E. (2024a): "Information matrix tests for Gaussian mixtures", CEMFI Working Paper 2401.

Amengual, D., Fiorentini, G. and Sentana, E. (2024b): "Multivariate Hermite polynomials and information matrix tests", forthcoming in *Econometrics and Statistics.*

Amengual, D., Fiorentini, G. and Sentana, E. (2025): "Information matrix tests for multinomial logit models", *Economics Letters* 247, 112180.

Ang, A. and Bekaert, G. (2002a): "International asset allocation with regime shifts", *Review of Financial Studies* 15, 817-858.

Ang, A. and Bekaert, G. (2002b): "Regime switches in interest rates", *Journal of Business and Economic Statistics* 20, 163-182.

Ang, A. and Timmermann, A. (2012): "Regime changes and financial markets", *Annual Review of Financial Economics* 4, 313-337.

Bekaert, G., Hodrick, R.J. and Marshall, D.A. (2001): "Peso problem explanations for term structure anomalies", *Journal of Monetary Economics* 48, 241-270.

Bera, A. and Lee, S (1993): "Information matrix test, parameter heterogeneity and ARCH: a synthesis", *Review of Economic Studies* 60, 229–240.

Carrasco, M., Hu, L. and Ploberger, W. (2014): "Optimal test for Markov switching parameters", *Econometrica* 82, 765–784.

Cecchetti, S. G., Lam, P. and Mark, N.C. (1990): "Mean reversion in equilibrium asset prices", *American Economic Review* 80, 398-418.

Chang, Y., Choi, Y. and Park, J.Y. (2017): "A new approach to model regime switching", *Journal of Econometrics* 196, 127–143.

Chesher, A. (1983): "The information matrix test: simplified calculation via a score test interpretation", *Economics Letters* 13, 45–48.

Chesher, A. (1984): "Testing for neglected heterogeneity", *Econometrica* 52, 865–872.

Dempster, A., N. Laird, and D. Rubin (1977): "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society Series B* 39, 1–38.

Diebold, F.X., Lee, J. H., and Weinbach, G. C. (1996): "Regime switching and endogenous transition probabilities", in C. Hargreaves (ed.) *Non-stationary time series analysis and cointegration*, 284-302, Oxford University Press.

Dueker, M.J., Solá, M. and Spagnolo, F. (2007): "Contemporaneous threshold autoregressive models: estimation, testing and forecasting", *Journal of Econometrics* 141, 517-547.

Engel, C. and Hakkio, C.S. (1996): "The distribution of the exchange rate in the EMS", *International Journal of Finance and Economics* 1, 55–67.

Engel, C. and Hamilton, J. (1990): "Long swings in the dollar: are they in the data and do markets know it?", *American Economic Review* 80, 689-713.

Engle, R.F., Hendry, D.F. and Richard, J.-F. (1983): "Exogeneity", *Econometrica* 51, 277–304.

Evans, M.D. and Lewis, K.K. (1995): "Do expected shifts in inflation affect estimates of the long-run Fisher relation?", *Journal of Finance* 50, 225-253.

Filardo, A.J. (1994): "Business-cycle phases and their transitional dynamics", *Journal of Business and Economic Statistics* 12, 299-308.

Filardo, A.J. and Gordon, S.F. (1998): "Business cycle durations", *Journal of Econometrics* 85, 99–123.

Friedmann, R. (1994): "Untersuchung der zinsentwicklung auf regimewechsel", mimeo, Universitat des Saarlandes.

Frühwirth-Schnatter, S. (2007): *Finite mixture and Markov switching models*, Springer.

Gable, J., Norden, S.V. and Vigfusson, R. (1997): "Analytical derivatives for Markov switching models", *Computational Economics* 10, 187–194.

Garcia, R. and Perron, P. (1996): "An analysis of the real interest rate under regime shifts", *Review of Economics and Statistics* 78, 111-125.

Goldfeld, S.M. and Quandt, R.E. (1973): "A Markov model for switching regressions", *Journal of Econometrics* 1, 3-15.

Gouriéroux, C., A. Monfort, E. Renault and A. Trognon (1987): "Generalized residuals", *Journal of Econometrics* 34, 5–32.

Gray, S.F. (1996): "Modeling the conditional distribution of interest rates as a regime-switching process", *Journal of Financial Economics* 42, 27-62.

Guidolin, M. (2011): "Markov switching models in empirical finance", in Drukker, D.M. (ed.) *Missing data methods: time-series methods and applications*, *Advances in Econometrics* 27, 1-86, Emerald.

Hall, A. (1987): "The information matrix test for the linear model", *Review of Economic Studies*

54, 257–263.

Hamilton, J.D. (1989a): “Rational expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates”, *Journal of Economic Dynamics and Control* 12, 385-423.

Hamilton, J.D. (1989b): “A new approach to the economic analysis of nonstationary time series and the business cycle”, *Econometrica* 57, 357–384.

Hamilton, J.D. (1990): “Analysis of time series subject to changes in regime”, *Journal of Econometrics* 45, 39-70.

Hamilton, J.D. (1993): “Estimation, inference, and forecasting of time series subject to changes in regime”, in G. S. Maddala, C. R. Rao, and H. D. Vinod (eds.) *Handbook of Statistics* 11, North-Holland.

Hamilton, J.D. (1994): *Time series models*, Princeton.

Hamilton, J.D. (1996): “Specification testing in Markov-switching time-series models”, *Journal of Econometrics* 70, 127-157.

Hamilton, J.D. (2008): “Regime-switching models”, in Durlauf, S. and Blume, L. (eds) *New Palgrave Dictionary of Economics*, Palgrave Macmillan.

Hamilton, J.D. (2016): “Macroeconomic regimes and regime shifts”, in J.B. Taylor and H. Uhlig (eds.) *Handbook of Macroeconomics* 2, 163-201, Elsevier.

Holmquist, B. (1996): “The d-variate vector Hermite polynomial of order k”, *Linear Algebra and its Applications* 237/238, 155–190.

Horowitz, J. (1994): “Bootstrap-based critical values for the information matrix test”, *Journal of Econometrics* 61, 395–411.

Jarque, C.M. and Bera, A.K. (1980): “Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals”, *Economic Letters* 6, 255–259.

Kasahara, H. and Shimotsu, K. (2018): “Testing the number of regimes in Markov regime switching models”, mimeo, University of British Columbia.

Kiefer, N.M. and Salmon, M. (1983): “Testing normality in econometric models”, *Economic Letters* 11, 123–127.

Kim, C.-J. (1994): “Dynamic linear models with Markov-switching”, *Journal of Econometrics* 60, 1-22.

Kim, C.J. and Nelson, C.R. (1999): *State-space models with regime switching: classical and Gibbs-sampling approaches with applications*, MIT Press.

Krolzig, H.M. (1997): *Markov-switching vector autoregressions modelling, statistical inference, and application to business cycle analysis*, Springer.

Lam, P.S. (1990): “The Hamilton model with a general autoregressive component: estimation and comparison with other models of economic time series”, *Journal of Monetary Economics* 26, 409–432.

Lancaster, A. (1984): “The covariance matrix of the information matrix test”, *Econometrica* 52, 1051–1053.

Magnus, J.R. and Neudecker, H. (2019): *Matrix differential calculus with applications in Statistics and Econometrics*, 3rd edition, Wiley.

Martínez Peira, M.S. (2002): “A regime-switching approach to the study of speculative attacks: a focus on EMS crises”, *Empirical Economics* 27, 299-334.

Mencía, J. and Sentana, E. (2012): “Distributional tests in multivariate dynamic models with Normal and Student $t$ innovations”, *Review of Economics and Statistics* 94, 133-152.

Newey, W.K. (1985): “Maximum likelihood specification testing and conditional moment tests”, *Econometrica* 53, 1047–70.

Pouzo, D., Psaradakis, Z. and Solá, M. (2022): “Maximum likelihood estimation in Markov regime-switching model with covariate-dependent transition probabilities”, *Econometrica* 90, 1681–1710.

Pouzo, D., Psaradakis, Z. and Solá, M. (2024): “On the robustness of mixture models in the presence of hidden Markov regimes with covariate-dependent transition probabilities”, Department of Economics Working Paper 2024_04, Universidad Torcuato Di Tella.

Psaradakis, Z. and Solá, M. (2024): “Markov-switching models with state-dependent time-varying transition probabilities”, *Econometrics and Statistics* 29, 49–63.

Qu, Z. and Zhuo, F. (2021): “Likelihood ratio based tests for Markov regime switching”, *Review of Economic Studies* 88, 937–968.

Rahman, S. (2017): “Wiener–Hermite polynomial expansion for multivariate Gaussian probability measures”, *Journal of Mathematical Analysis and Applications* 454, 303–334.

Ravn, M.O. and Solá, M. (1999): “Business cycle dynamics: predicting transitions with macrovariables”, in Rothman, P. (ed.) *Nonlinear time series analysis of economic and financial data*, 231–265, Springer.

Schwert, G.W. (1989): “Business cycles, financial crises, and stock volatility”, in: K. Brunner and A.H. Meltzer (eds.), *IMF policy advice, market volatility, commodity price rules, and other essays*, *Carnegie-Rochester Conference Series in Public Policy* 31, 83-126.

Sentana, E. (1995): “Quadratic ARCH models”, *Review of Economic Studies* 62, 639–661.

Simpson, P.W., Osborn, D.R. and Sensier, M. (2001): “Modelling business cycle movements in the UK economy”, *Economica* 68, 243–267.

Tauchen, G. (1985): “Diagnostic testing and evaluation of maximum likelihood models”, *Journal of Econometrics* 30, 415–443.

Turner, C., Startz, R. and Nelson, C. (1989): “A Markov model of heteroskedasticity, risk, and learning in the stock market”, *Journal of Financial Economics* 25, 3-22.

White, H. (1980): “A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity”, *Econometrica* 48, 817–838.

White, H. (1982): “Maximum likelihood estimation of misspecified models”, *Econometrica* 50, 1–25.

White, H. (1987): “Specification testing in dynamic models”, in T.B. Bewley (ed.), *Advances in Econometrics. Fifth World Congress*, vol I, 1-58, Cambridge University Press.

White, H. (1994): *Estimation, inference and specification analysis*, Cambridge University Press.

Yang, M. (2001): “Closed-form likelihood function of Markov-switching models”, *Economics Letters* 70, 319-326.

# Appendix

## Proofs

### Proof or Proposition 1

The multiple sequential cuts on the vector of parameters performed by the additive log-likelihood decompositions in (8) and (9) allows us to obtain very easily the score vectors, Hessian matrices, and influence functions for the IM test for the each of the $K$ conditional models for $\mathbf{y}_t$ given $\xi_{kt}$, and each of the $K$ marginal models for $\xi_{kt}$. Specifically, we can use the expressions for multivariate regression models in Amengual, Fiorentini and Sentana (2022) to write

$$\begin{aligned}\mathbf{s}_{\mathbf{b}_k t}(\boldsymbol{\phi}) &= \xi_{kt}(\mathbf{I}_M \otimes \boldsymbol{\Gamma}_k^{-1/2\prime})\{\mathbf{y}_{t-1}^a \otimes \mathbf{H}_{1t}[\boldsymbol{\varepsilon}_{kt}^{*\prime}(\mathbf{b}_k,\boldsymbol{\gamma}_k)]\},\\ \mathbf{s}_{\boldsymbol{\gamma}_k t}(\boldsymbol{\phi}) &= \xi_{kt}\frac{1}{2}\mathbf{D}_M'(\boldsymbol{\Gamma}_k^{-1/2\prime} \otimes \boldsymbol{\Gamma}^{-1/2\prime})\mathbf{D}_M\mathbf{H}_{2t}[\boldsymbol{\varepsilon}_{kt}^{*\prime}(\mathbf{b}_k,\boldsymbol{\gamma}_k)],\end{aligned}$$

where $\mathbf{D}_M$ is the duplication matrix of order $M$ (see Magnus and Neudecker (2019)), which confirms that the scores with respect to $\mathbf{b}_k$ and $\boldsymbol{\gamma}_k$ are spanned by $\{\mathbf{y}_t^a \otimes \mathbf{H}_{1t}[\boldsymbol{\varepsilon}_{kt}^{*\prime}(\mathbf{b}_k,\boldsymbol{\gamma}_k)]\}$ and $\mathbf{H}_{2t}[\boldsymbol{\varepsilon}_{kt}^{*\prime}(\mathbf{b}_k,\boldsymbol{\gamma}_k)]\}$, respectively. Amengual, Fiorentini and Sentana (2022) also show that the sum of the outer product of the score and Hessian corresponding to these parameters is spanned by

$$\mathbf{m}_{\mathbf{b}_k\mathbf{b}_{kr}}(\boldsymbol{\xi}_t;\mathbf{y}_{t-1};\boldsymbol{\phi}) = \xi_{kt}\mathbf{H}_2[\boldsymbol{\varepsilon}_{kt}^{*\prime}(\mathbf{b}_k,\boldsymbol{\gamma}_k)] \otimes vech(\mathbf{y}_{t-1}^a\mathbf{y}_{t-1}^{a\prime}), \tag{A1}$$

$$\mathbf{m}_{\mathbf{b}_k\boldsymbol{\gamma}_{kr}}(\boldsymbol{\xi}_t;\mathbf{y}_{t-1};\boldsymbol{\phi}) = \xi_{kt}\mathbf{H}_3[\boldsymbol{\varepsilon}_{kt}^{*\prime}(\mathbf{b}_k,\boldsymbol{\gamma}_k)] \otimes \mathbf{y}_{t-1}^a, \tag{A2}$$

$$\mathbf{m}_{\boldsymbol{\gamma}_k\boldsymbol{\gamma}_{kr}}(\boldsymbol{\xi}_t;\mathbf{y}_{t-1};\boldsymbol{\phi}) = \mathbf{H}_4[\boldsymbol{\varepsilon}_{kt}^{*\prime}(\mathbf{b}_k,\boldsymbol{\gamma}_k)], \tag{A3}$$

Since all multivariate Hermite polynomials have a zero conditional mean and a constant conditional variance given $\mathbf{y}_{t-1}^a$, and they are uncorrelated for different orders (see e.g. Holmquist (1996) or Rahman (2017)), one only needs to regress (A1) on $\mathbf{H}_{2t}[\boldsymbol{\varepsilon}_{kt}^{*\prime}(\mathbf{b}_k,\boldsymbol{\gamma}_k)$ for those observations for which $\xi_{kt} = 1$ to purge the IM influence functions from sampling uncertainty resulting from the estimation of the mean and variance parameters of the $k^{th}$ regime. Importantly, this results in the loss of $\binom{M+1}{2}$ degrees of freedom. Intuitively, the sample average of $\xi_{kt}\mathbf{H}_2[\boldsymbol{\varepsilon}_{kt}^{*\prime}(\hat{\mathbf{b}}_k,\hat{\boldsymbol{\gamma}}_k)]$ is equal to $\mathbf{0}$ from the first order conditions for $\boldsymbol{\gamma}_k$. Consequently, the information matrix test that compares the outer product of the score with the Hessian of the VAR(1) model for regime $k$ evaluated at the Gaussian MLE $\hat{\boldsymbol{\phi}}_T$ is asymptotically equivalent under the null hypothesis of correct specification to the sum of three quadratic forms in the sample averages of (A2), (A3) and a version of (A1) with $vech(\mathbf{y}_{t-1}^a\mathbf{y}_{t-1}^{a\prime})$ replaced by $vech(\mathbf{y}_{t-1}^a\mathbf{y}_{t-1}^{a\prime}) - E[vech(\mathbf{y}_{t-1}^a\mathbf{y}_{t-1}^{a\prime})]$ which converge in distribution to three independent chi-square random variables whose degrees of freedom are $\binom{M+2}{3}(M+1)$, $\binom{M+3}{4}$ and $\binom{M+1}{2}\left[\frac{(M+1)(M+2)}{2} - 1\right]$, respectively.

Similarly, we can use the expressions for multinomial logit models in Amengual, Fiorentini and Sentana (2025) to show that

$$\begin{aligned}\mathbf{s}_{\boldsymbol{\beta}_{rj}}(\boldsymbol{\xi}_t, \mathbf{y}_{t-1}; \boldsymbol{\beta}_{rj}) &= \xi_{jt-1}\mathbf{u}_{rj}(\boldsymbol{\xi}_t, \mathbf{y}_{t-1}; \boldsymbol{\beta}_{rj}) \otimes \mathbf{z}^a_{t-1}, \\ \mathbf{h}_{\boldsymbol{\beta}_{rj}\boldsymbol{\beta}_{rj}}(\boldsymbol{\xi}_t, \mathbf{y}_{t-1}; \boldsymbol{\beta}_{rj}) &= -\xi_{jt-1}\{diag[\mathbf{p}_{rj}(\mathbf{y}_{t-1}; \boldsymbol{\beta}_{rj})] - \mathbf{p}_{rj}(\mathbf{y}_{t-1}; \boldsymbol{\beta}_{rj})\mathbf{p}'_{rj}(\mathbf{y}_{t-1}; \boldsymbol{\beta}_{rj})\} \\ &\quad \otimes \mathbf{z}^a_{t-1}\mathbf{z}^{a\prime}_{t-1},\end{aligned}$$

and

$$\begin{aligned}\mathbf{m}_{\boldsymbol{\beta}_{rj}\boldsymbol{\beta}_{rj}}(\boldsymbol{\xi}_t; \mathbf{y}_{t-1}; \boldsymbol{\beta}_{rj}) &= \xi_{jt-1}(vech[\mathbf{u}_{rj}(\boldsymbol{\xi}_t, \mathbf{y}_{t-1}; \boldsymbol{\beta}_{rj})\mathbf{u}'_{rj}(\boldsymbol{\xi}_t, \mathbf{y}_{t-1}; \boldsymbol{\beta}_{rj}) \\ &\quad -\{diag[\mathbf{p}_r(\mathbf{y}^a; \boldsymbol{\beta})] - \mathbf{p}_r(\mathbf{y}^a; \boldsymbol{\beta})\mathbf{p}'_r(\mathbf{y}^a; \boldsymbol{\beta})\}]) \\ &\quad \otimes vech(\mathbf{z}^a_{t-1}\mathbf{z}^{a\prime}_{t-1}),\end{aligned}$$

which gives rise to another $\frac{K(K-1)(L+1)(L+2)}{4}$ influence functions for each column of the transition matrix.

The sequential cuts in (8) and (9) also mean that both the Hessian matrix and the information matrix will be block-diagonal between the elements of the mean and variance parameters of each regime and the multinomial logit parameters characterising each column of the transition matrix. However, the outer-product of the scores corresponding to those elements are not necessarily 0, even though their expected value is. While the outer product of $[\mathbf{s}_{\mathbf{b}_k t}(\boldsymbol{\phi}), \mathbf{s}_{\boldsymbol{\gamma}_k t}(\boldsymbol{\phi})]$ and $[\mathbf{s}_{\mathbf{b}_\ell t}(\boldsymbol{\phi}), \mathbf{s}_{\boldsymbol{\gamma}_\ell t}(\boldsymbol{\phi})]$ will indeed be 0 for $\ell \neq k$ because $\xi_{kt}\xi_{\ell t} = 0$, and the same is true of the outer product of $\mathbf{s}_{\boldsymbol{\beta}_{rj}}(\boldsymbol{\xi}_t, \mathbf{y}_{t-1}; \boldsymbol{\beta}_{rj})$ and $\mathbf{s}_{\boldsymbol{\beta}_{r\ell}}(\boldsymbol{\xi}_t, \mathbf{y}_{t-1}; \boldsymbol{\beta}_{r\ell})$ for $\ell \neq j$ because $\xi_{jt-1}\xi_{\ell t-1} = 0$ too, the outer product of $\xi_{kt}[\mathbf{s}_{\mathbf{b}_k t}(\boldsymbol{\phi}), \mathbf{s}_{\boldsymbol{\gamma}_k t}(\boldsymbol{\phi})]$ with $\xi_{jt-1}\mathbf{s}_{\boldsymbol{\beta}_{rj}}(\boldsymbol{\xi}_t, \mathbf{y}_{t-1}; \boldsymbol{\beta}_{rj})$ will not generally be 0 on average in the sample when evaluated at the maximum likelihood estimators, which gives rise to the additional $\frac{3KM(M+1)(K-1)(L+1)}{2}$ influence functions (14).

## Proof or Proposition 2

This is a straightforward application of Proposition 1 in Amengual, Fiorentini and Sentana (2024a). The only complication arises because $\mathbf{u}_{jt}$ is unobserved, but if we express this vector as (15) and exploit the fact that $\xi_{kt}\xi_{\ell t} = I(k = \ell)$, we can tediously derive all the expressions in the statement of the Proposition.

Table 1: Finite sample rejection rates under correct specification.

| | | $T = 250$ | | | | | | $T = 1,000$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *Asymptotic* | | | *Bootstrap* | | | *Asymptotic* | | | *Bootstrap* | | |
| | *df* | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% | 10% | 5% | 1% |
| *Panel A*: AR(1) dynamics, constant transitions | | | | | | | | | | | | | |
| Joint | *16* | 76.77 | 68.72 | 52.30 | 7.67 | 3.07 | 0.27 | 43.42 | 33.49 | 19.07 | 9.48 | 4.59 | 0.81 |
| Heteroskedasticity | *4* | 24.78 | 15.91 | 5.86 | 9.29 | 4.46 | 0.69 | 15.86 | 9.63 | 3.07 | 10.09 | 4.98 | 0.98 |
| Heterocliticity | *2* | 19.89 | 12.26 | 4.01 | 10.10 | 4.92 | 0.89 | 14.36 | 8.38 | 2.21 | 10.14 | 5.02 | 0.90 |
| Skewness and kurtosis | *4* | 40.72 | 32.03 | 19.54 | 8.92 | 4.34 | 0.71 | 23.37 | 16.73 | 8.37 | 10.10 | 5.10 | 0.90 |
| Across models | *6* | 42.76 | 34.60 | 21.38 | 9.17 | 4.16 | 0.71 | 21.46 | 15.15 | 6.92 | 9.93 | 4.90 | 1.04 |
| Scores autocorrelation | *8* | 21.18 | 11.95 | 3.15 | 9.03 | 4.14 | 0.84 | 13.99 | 7.84 | 1.87 | 9.67 | 4.86 | 0.95 |
| *Panel B*: White noise, covariate-dependent transitions | | | | | | | | | | | | | |
| Joint | *26* | 99.19 | 98.61 | 96.45 | 3.28 | 0.86 | 0.09 | 76.81 | 69.46 | 54.80 | 8.30 | 3.94 | 0.61 |
| Skewness and kurtosis | *4* | 55.51 | 48.59 | 35.47 | 9.13 | 4.60 | 1.05 | 29.46 | 22.42 | 12.34 | 9.25 | 4.95 | 0.93 |
| MNL heteroskedasticity | *6* | 47.68 | 37.40 | 23.73 | 5.87 | 2.21 | 0.21 | 29.33 | 20.78 | 10.24 | 9.18 | 4.55 | 0.86 |
| Across models | *16* | 86.46 | 81.85 | 69.78 | 7.12 | 2.67 | 0.28 | 45.70 | 35.80 | 21.42 | 9.70 | 4.59 | 0.74 |
| Scores autocorrelation | *8* | 26.18 | 16.47 | 5.57 | 8.89 | 4.23 | 0.70 | 15.12 | 8.41 | 2.39 | 9.85 | 4.89 | 0.85 |
| *Panel C*: AR(1) dynamics, covariate-dependent transitions | | | | | | | | | | | | | |
| Joint | *40* | 98.72 | 97.44 | 92.12 | 5.37 | 1.83 | 0.13 | 78.00 | 69.44 | 51.70 | 9.46 | 4.78 | 0.87 |
| Heteroskedasticity | *4* | 27.68 | 18.47 | 6.64 | 9.28 | 4.43 | 0.75 | 16.30 | 9.81 | 2.87 | 10.06 | 4.78 | 0.83 |
| Heterocliticity | *2* | 19.41 | 11.54 | 3.81 | 9.07 | 4.31 | 0.87 | 13.58 | 7.36 | 1.82 | 9.86 | 4.97 | 1.05 |
| Skewness and kurtosis | *4* | 42.57 | 34.29 | 20.84 | 8.71 | 4.32 | 0.93 | 23.08 | 15.71 | 7.24 | 9.55 | 4.43 | 0.86 |
| MNL heteroskedasticity | *6* | 31.32 | 20.48 | 7.98 | 6.14 | 2.64 | 0.41 | 14.64 | 8.13 | 2.11 | 9.70 | 4.83 | 1.03 |
| Across models | *24* | 86.34 | 79.47 | 63.59 | 8.08 | 3.35 | 0.33 | 55.76 | 45.07 | 28.15 | 10.30 | 5.30 | 1.23 |
| Scores autocorrelation | *10* | 24.78 | 14.95 | 4.23 | 9.18 | 4.46 | 0.82 | 14.99 | 8.32 | 1.88 | 10.11 | 4.93 | 0.87 |

Notes: Monte Carlo empirical rejection rates based on 10,000 replications. *Asymptotic* contains p-values based on the asymptotic critical values while *Bootstrap* includes rejection rates based on a parametric bootstrap procedure in which we simulate $B = 99$ samples from the model estimated under the null. See section 3 for details about the DGPs.

Table 2: Finite sample rejections rates of the AR(1) dynamics and constant transitions model under the alternatives

| | | *Bootstrap* | |
|---|---|---|---|
| | *df* | $T = 250$ | $T = 1,000$ |
| *Panel A*: AR(2) dynamics | | | |
| Joint | *16* | 10.24 | 47.32 |
| Heteroskedasticity | *4* | 4.80 | 6.60 |
| Heteroclіticity | *2* | 4.08 | 6.40 |
| Skewness and kurtosis | *4* | 4.68 | 12.40 |
| Across models | *6* | 19.84 | 74.72 |
| Scores autocorrelation | *8* | 28.16 | 93.92 |
| *Panel B*: GQARCH | | | |
| Joint | *16* | 3.56 | 45.56 |
| Heteroskedasticity | *4* | 25.04 | 79.60 |
| Heteroclіticity | *2* | 11.80 | 50.56 |
| Skewness and kurtosis | *4* | 2.68 | 31.44 |
| Across models | *6* | 4.24 | 12.76 |
| Scores autocorrelation | *8* | 15.76 | 69.36 |
| *Panel C*: Asymmetric $t$ innovations | | | |
| Joint | *16* | 22.84 | 99.32 |
| Heteroskedasticity | *4* | 15.00 | 15.60 |
| Heteroclіticity | *2* | 62.48 | 95.52 |
| Skewness and kurtosis | *4* | 47.36 | 99.96 |
| Across models | *6* | 8.12 | 22.04 |
| Scores autocorrelation | *8* | 15.04 | 57.76 |
| *Panel D*: Covariate-dependent transitions | | | |
| Joint | *16* | 3.16 | 4.40 |
| Heteroskedasticity | *4* | 4.68 | 4.00 |
| Heteroclіticity | *2* | 4.20 | 4.04 |
| Skewness and kurtosis | *4* | 4.64 | 3.96 |
| Across models | *6* | 3.72 | 4.72 |
| Scores autocorrelation | *8* | 4.96 | 5.04 |

Notes: Monte Carlo empirical rejection rates at 5% significance level based on 2,500 replications. Rejection rates based on a parametric bootstrap procedure in which we simulate $B = 99$ samples from the model estimated under the null. See section 3 for details about the DGPs.

Table 3: Parameter estimators of the AR(1) dynamics and constant transitions model

| Parameter | $\nu_1$ | $\nu_2$ | $a_{1,1}$ | $a_{2,1}$ | $\gamma_1$ | $\gamma_2$ | $p$ | $q$ |
|---|---|---|---|---|---|---|---|---|
| | | | | *Panel 0*: Null hypothesis | | | | |
| *True Values* | *1.00* | *-1.00* | *0.90* | *0.90* | *1.00* | *1.00* | *0.76* | *0.93* |
| Median ($T = 1,000$) | 1.00 | -1.01 | 0.90 | 0.90 | 0.99 | 1.00 | 0.76 | 0.93 |
| IQR ($T = 250$) | 0.38 | 0.26 | 0.06 | 0.03 | 0.19 | 0.09 | 0.12 | 0.04 |
| IQR ($T = 1,000$) | 0.17 | 0.12 | 0.03 | 0.02 | 0.10 | 0.05 | 0.06 | 0.02 |
| | | | | *Panel A*: AR(2) dynamics | | | | |
| Median ($T = 1,000$) | 1.18 | -1.13 | 0.93 | 0.92 | 1.00 | 0.99 | 0.78 | 0.93 |
| IQR ($T = 250$) | 0.37 | 0.27 | 0.05 | 0.03 | 0.16 | 0.09 | 0.08 | 0.04 |
| IQR ($T = 1,000$) | 0.17 | 0.13 | 0.02 | 0.01 | 0.08 | 0.05 | 0.04 | 0.02 |
| | | | | *Panel B*: GQARCH | | | | |
| Median ($T = 1,000$) | 1.06 | -0.93 | 0.88 | 0.91 | 0.92 | 0.96 | 0.74 | 0.93 |
| IQR ($T = 250$) | 0.33 | 0.22 | 0.06 | 0.04 | 0.27 | 0.17 | 0.13 | 0.04 |
| IQR ($T = 1,000$) | 0.14 | 0.11 | 0.03 | 0.02 | 0.14 | 0.09 | 0.06 | 0.02 |
| | | | | *Panel C*: Asymmetric $t$ innovations | | | | |
| Median ($T = 1,000$) | 1.27 | -0.86 | 0.89 | 0.91 | 0.66 | 1.07 | 0.75 | 0.95 |
| IQR ($T = 250$) | 0.35 | 0.23 | 0.05 | 0.03 | 0.22 | 0.16 | 0.11 | 0.03 |
| IQR ($T = 1,000$) | 0.15 | 0.12 | 0.02 | 0.02 | 0.11 | 0.09 | 0.06 | 0.01 |
| | | | | *Panel D*: Covariate-dependent transitions | | | | |
| Median ($T = 1,000$) | 1.00 | -1.02 | 0.90 | 0.90 | 0.99 | 1.00 | 0.80 | 0.93 |
| IQR ($T = 250$) | 0.33 | 0.24 | 0.06 | 0.03 | 0.18 | 0.09 | 0.10 | 0.04 |
| IQR ($T = 1,000$) | 0.14 | 0.12 | 0.02 | 0.01 | 0.08 | 0.04 | 0.05 | 0.02 |

Notes: Monte Carlo medians and interquantile ranges of parameter estimators. See section 3 for details about the DGPs.

Table 4: Finite sample power properties: White noise, covariate-dependent transitions

| | | *Bootstrap* | |
|---|---|---|---|
| | *df* | $T = 250$ | $T = 1,000$ |
| *Panel A*: AR(1) dynamics | | | |
| Joint | *26* | 64.56 | 100.00 |
| Skewness and kurtosis | *4* | 13.36 | 65.84 |
| MNL heteroskedasticity | *6* | 2.32 | 4.56 |
| Across models | *16* | 78.28 | 100.00 |
| Scores autocorrelation | *8* | 38.72 | 100.00 |
| *Panel B*: GQARCH | | | |
| Joint | *26* | 1.44 | 9.72 |
| Skewness and kurtosis | *4* | 2.32 | 37.60 |
| MNL heteroskedasticity | *6* | 2.48 | 4.16 |
| Across models | *16* | 4.68 | 12.24 |
| Scores autocorrelation | *8* | 14.24 | 76.88 |
| *Panel C*: Asymmetric $t$ innovations | | | |
| Joint | *26* | 14.68 | 99.64 |
| Skewness and kurtosis | *4* | 93.04 | 100.00 |
| MNL heteroskedasticity | *6* | 2.40 | 4.08 |
| Across models | *16* | 7.28 | 16.24 |
| Scores autocorrelation | *8* | 4.68 | 10.80 |
| *Panel D*: Predetermined regressors | | | |
| Joint | *26* | 4.20 | 20.08 |
| Skewness and kurtosis | *4* | 3.20 | 13.84 |
| MNL heteroskedasticity | *6* | 2.24 | 5.68 |
| Across models | *16* | 8.64 | 45.16 |
| Scores autocorrelation | *8* | 8.72 | 20.96 |

Notes: Monte Carlo empirical rejection rates at 5% significance level based on 2,500 replications. Rejection rates based on a parametric bootstrap procedure in which we simulate $B = 99$ samples from the model estimated under the null. See section 3 for details about the DGPs.

Table 5: Parameter estimators: White noise, covariate-dependent transitions

| Parameter | $\nu_1$ | $\nu_2$ | $\gamma_1$ | $\gamma_2$ | $\beta_{21,1}$ | $\beta_{21,2}$ | $\beta_{22,1}$ | $\beta_{22,2}$ |
|---|---|---|---|---|---|---|---|---|
| | | | *Panel 0*: Null hypothesis | | | | | |
| *True Value* | *10.00* | *-10.00* | *2.29* | *2.29* | *-2.00* | *0.50* | *2.00* | *0.50* |
| Median ($T = 1,000$) | 10.00 | -10.00 | 2.29 | 2.29 | -2.01 | 0.51 | 2.00 | 0.50 |
| IQR ($T = 250$) | 0.40 | 0.22 | 0.28 | 0.17 | 0.73 | 0.43 | 0.56 | 0.35 |
| IQR ($T = 1,000$) | 0.20 | 0.11 | 0.14 | 0.08 | 0.36 | 0.20 | 0.27 | 0.17 |
| | | | *Panel A*: AR(1) dynamics | | | | | |
| Median ($T = 1,000$) | 11.31 | -12.12 | 3.02 | 2.62 | -2.01 | 0.51 | 2.00 | 0.50 |
| IQR ($T = 250$) | 0.61 | 0.30 | 0.41 | 0.21 | 0.72 | 0.43 | 0.56 | 0.35 |
| IQR ($T = 1,000$) | 0.30 | 0.16 | 0.20 | 0.10 | 0.35 | 0.19 | 0.26 | 0.17 |
| | | | *Panel B*: GQARCH | | | | | |
| Median ($T = 1,000$) | 10.00 | -10.00 | 2.26 | 2.26 | -2.01 | 0.51 | 2.00 | 0.50 |
| IQR ($T = 250$) | 0.39 | 0.22 | 0.54 | 0.36 | 0.73 | 0.42 | 0.55 | 0.35 |
| IQR ($T = 1,000$) | 0.20 | 0.12 | 0.28 | 0.19 | 0.36 | 0.20 | 0.27 | 0.17 |
| | | | *Panel C*: Asymmetric $t$ innovations | | | | | |
| Median ($T = 1,000$) | 10.06 | -9.99 | 2.11 | 2.29 | -1.98 | 0.50 | 1.99 | 0.50 |
| IQR ($T = 250$) | 0.38 | 0.23 | 0.38 | 0.32 | 0.70 | 0.43 | 0.54 | 0.34 |
| IQR ($T = 1,000$) | 0.18 | 0.11 | 0.19 | 0.17 | 0.34 | 0.20 | 0.26 | 0.16 |
| | | | *Panel D*: Predetermined regressors | | | | | |
| Median ($T = 1,000$) | 10.02 | -10.02 | 2.57 | 2.18 | -2.02 | 0.53 | 1.95 | 0.53 |
| IQR ($T = 250$) | 0.38 | 0.19 | 0.41 | 0.17 | 0.72 | 0.43 | 0.54 | 0.35 |
| IQR ($T = 1,000$) | 0.19 | 0.10 | 0.20 | 0.09 | 0.35 | 0.19 | 0.26 | 0.17 |

Notes: Monte Carlo medians and interquantile ranges of parameter estimators. See section 3 for details about the DGPs.

Table 6: Finite sample power properties: AR(1) dynamics, covariate-dependent transitions

| | | *Bootstrap* | |
|---|---|---|---|
| | *df* | $T = 250$ | $T = 1,000$ |
| | | *Panel A*: AR(2) dynamics | |
| Joint | *40* | 6.24 | 36.44 |
| Heteroskedasticity | *4* | 5.04 | 5.32 |
| Heteroclititicity | *2* | 4.00 | 5.60 |
| Skewness and kurtosis | *4* | 5.08 | 12.12 |
| MNL heteroskedasticity | *6* | 2.72 | 5.36 |
| Across models | *24* | 12.48 | 54.16 |
| Scores autocorrelation | *10* | 27.56 | 93.44 |
| | | *Panel B*: GQARCH | |
| Joint | *40* | 2.68 | 21.44 |
| Heteroskedasticity | *4* | 26.28 | 77.08 |
| Heteroclititicity | *2* | 13.12 | 48.08 |
| Skewness and kurtosis | *4* | 2.04 | 34.48 |
| MNL heteroskedasticity | *6* | 3.48 | 4.96 |
| Across models | *24* | 4.52 | 11.40 |
| Scores autocorrelation | *10* | 14.80 | 65.84 |
| | | *Panel C*: Asymmetric $t$ innovations | |
| Joint | *40* | 15.00 | 95.92 |
| Heteroskedasticity | *4* | 13.88 | 16.52 |
| Heteroclititicity | *2* | 61.28 | 93.84 |
| Skewness and kurtosis | *4* | 55.72 | 100.00 |
| MNL heteroskedasticity | *6* | 8.60 | 9.00 |
| Across models | *24* | 11.28 | 18.56 |
| Scores autocorrelation | *10* | 13.76 | 46.24 |
| | | *Panel D*: Predetermined regressors | |
| Joint | *40* | 7.92 | 26.96 |
| Heteroskedasticity | *4* | 12.08 | 25.84 |
| Heteroclititicity | *2* | 19.28 | 52.00 |
| Skewness and kurtosis | *4* | 2.52 | 27.72 |
| MNL heteroskedasticity | *6* | 2.04 | 18.56 |
| Across models | *24* | 12.36 | 41.24 |
| Scores autocorrelation | *10* | 12.72 | 41.56 |

Notes: Monte Carlo empirical rejection rates at 5% significance level based on 2,500 replications. Rejection rates based on a parametric bootstrap procedure in which we simulate $B = 99$ samples from the model estimated under the null. See section 3 for details about the DGPs.

Table 7: Parameter estimators: AR(1) dynamics, covariate-dependent transitions

| Parameter | $\nu_1$ | $\nu_2$ | $a_{1,1}$ | $a_{2,1}$ | $\gamma_1$ | $\gamma_2$ | $\beta_{21,1}$ | $\beta_{21,2}$ | $\beta_{22,1}$ | $\beta_{22,2}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| *Panel 0*: Null hypothesis | | | | | | | | | | |
| *True Values* | *1.00* | *-1.00* | *0.90* | *0.90* | *1.00* | *1.00* | *-2.00* | *0.50* | *2.00* | *0.50* |
| Median ($T = 1,000$) | 1.00 | -1.01 | 0.90 | 0.90 | 0.99 | 1.00 | -2.02 | 0.52 | 2.00 | 0.51 |
| IQR ($T = 250$) | 0.32 | 0.24 | 0.06 | 0.03 | 0.18 | 0.09 | 1.20 | 0.66 | 0.80 | 0.57 |
| IQR ($T = 1,000$) | 0.14 | 0.11 | 0.02 | 0.01 | 0.09 | 0.04 | 0.52 | 0.28 | 0.36 | 0.24 |
| *Panel A*: AR(2) dynamics | | | | | | | | | | |
| Median ($T = 1,000$) | 1.19 | -1.14 | 0.92 | 0.92 | 0.99 | 0.99 | -2.06 | 0.49 | 1.99 | 0.48 |
| IQR ($T = 250$) | 0.33 | 0.25 | 0.05 | 0.03 | 0.16 | 0.09 | 1.00 | 0.54 | 0.69 | 0.49 |
| IQR ($T = 1,000$) | 0.15 | 0.12 | 0.02 | 0.01 | 0.08 | 0.04 | 0.43 | 0.24 | 0.33 | 0.21 |
| *Panel B*: GQARCH | | | | | | | | | | |
| Median ($T = 1,000$) | 1.07 | -0.95 | 0.88 | 0.91 | 0.93 | 0.96 | -1.90 | 0.53 | 2.01 | 0.45 |
| IQR ($T = 250$) | 0.28 | 0.21 | 0.05 | 0.03 | 0.26 | 0.17 | 1.09 | 0.62 | 0.74 | 0.51 |
| IQR ($T = 1,000$) | 0.12 | 0.11 | 0.02 | 0.02 | 0.13 | 0.09 | 0.53 | 0.29 | 0.35 | 0.22 |
| *Panel C*: Asymmetric $t$ innovations | | | | | | | | | | |
| Median ($T = 1,000$) | 1.23 | -0.89 | 0.89 | 0.90 | 0.69 | 1.07 | -1.76 | 0.40 | 2.19 | 0.64 |
| IQR ($T = 250$) | 0.29 | 0.22 | 0.04 | 0.03 | 0.21 | 0.16 | 1.12 | 0.63 | 0.75 | 0.54 |
| IQR ($T = 1,000$) | 0.14 | 0.11 | 0.02 | 0.01 | 0.10 | 0.08 | 0.50 | 0.27 | 0.35 | 0.23 |
| *Panel D*: Predetermined regressors | | | | | | | | | | |
| Median ($T = 1,000$) | 0.97 | -0.98 | 0.89 | 0.91 | 1.13 | 0.94 | -2.65 | 0.85 | 1.83 | 0.71 |
| IQR ($T = 250$) | 0.35 | 0.23 | 0.07 | 0.03 | 0.23 | 0.09 | 1.29 | 0.70 | 0.58 | 0.48 |
| IQR ($T = 1,000$) | 0.15 | 0.11 | 0.03 | 0.01 | 0.11 | 0.05 | 0.56 | 0.30 | 0.26 | 0.22 |

Notes: Monte Carlo medians and interquantile ranges of parameter estimators. See section 3 for details about the DGPs.