

working paper

2601

Information matrix tests for switching regressions

Dante Amengual
Gabriele Fiorentini
Enrique Sentana

February 2026

cemfi

Information matrix tests for switching regressions

Abstract

The EM principle implies the moments underlying the information matrix test for switching regressions are the expectation given the data of the moments one would test if one knew the subpopulation each observation originated from. Thus, we identify components related to conditional heteroskedasticity, conditional and unconditional skewness, and unconditional kurtosis of regression residuals within each regime. Simulations indicate analytical expressions for the asymptotic covariance matrix of those moments adjusted for sampling variability in parameter estimators provide reliable finite sample sizes and good power against various alternatives, especially combined with the parametric bootstrap. We apply the test to cross-country convergence regressions.

JEL Codes: C24, C34, C52, O47.

Keywords: Asymmetry, convergence regressions, expectation - maximization principle, heteroskedasticity, incomplete data, kurtosis.

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, the XIV Zaragoza Time Series Workshop and the CFE 2025 conference for useful comments, discussions and suggestions. Jiaxuan Ren provided valuable research assistance for 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.

1 Introduction

Models in which the regression coefficients may take a discrete number of values have been used for decades in a variety of empirical applications. Inspired by the finite mixture models studied at least as far back as Pearson (1884), Quandt (1972) introduced a switching regression model in which ‘nature’ randomly chooses the values of those regression coefficients and residual variances according to some probabilistic rule. The fact that there are explanatory variables affecting the conditional mean of the dependent variables within each subpopulation or regime in these “mixture of regressions” models have made them more popular than traditional mixture models in economic applications. For example, they have been employed for identifying “convergence clubs” of countries based on per capita GDP, as well as within-country clustering in household income and wealth distributions (see Johnson and Papageorgiou (2020) and Cowell and Flachaire (2015) for some recent surveys describing the use of switching regressions in those areas).

The theoretical econometric literature has focused on numerical algorithms for estimating the model parameters by maximum likelihood (see e.g. Hartley (1978)), the consistency and asymptotic normality of the maximum likelihood estimators of those parameters (see e.g. Kiefer (1978)), or testing procedures for determining the existence of more than one regime (see e.g. Cho and White (2007)). However, relatively little work has been done in testing the correct specification of the estimated model, with the exception of Hamilton (1996), who derived score-based specification tests using the Expectation Maximisation (EM) principle in a popular time-series generalisation of the model in which the observed series follow autoregressions within each regime while the regimes evolve according to a discrete Markov chain.

In this paper, we also exploit the EM principle but for the purposes of deriving the information matrix (IM) test of multivariate switching regression models. 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 regular 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 observed data of the influence functions that one would test if one knew the subpopulation each observation originated from. We also rely on the fact that Amengual, Fiorentini and Sentana (2024a) exploit the law of iterated variances to express the asymptotic covariance matrices required to compute the IM test statistic as the difference between the unconditional covariance matrix of the relevant influence functions in the complete data model and the mean of their covariance

matrix conditional on the observed variables. In this respect, an additional, non-trivial benefit of 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, so that researchers know the correct rank of the relevant covariance matrix, and consequently, the right number of degrees of freedom.

One advantage of the IM test is that it is not a priori targeted to either the number of regimes, the conditional normality of the disturbances in each of them or their homoskedasticity, while having power to detect misspecification in each of those ingredients of the model, as we will see below. In addition, the EM principle provides a very intuitive interpretation of the influence functions underlying the IM test. Specifically, we prove that the only influence functions that matter in multivariate switching regression models are those associated to the conditional mean and variance parameters within each regime, which coincide with the multivariate version of the ones derived by Hall (1987) for linear regression models, but written in terms of residuals computed as if all observations came from the k^{th} regime and weighted by the posterior probability that each observation belongs to that regime. Thus, the IM test is effectively testing the unconditional kurtosis of those conditionally standardised residuals, as in Jarque and Bera (1980), their conditional heteroskedasticity given the exogenous mean regressors, their squares and cross-products, as in White (1980), and their conditional skewness given those regressors, as in Bera and Lee (1993).

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, which is followed by an application of our tests to the empirical analysis of cross-country convergence in Di Vaio and Enflo (2011) in section 4. Finally, we conclude by discussing some extensions in section 5. Proofs and auxiliary results are relegated to appendices.

2 Theoretical analysis

2.1 The model

Consider the following model

$$\begin{aligned} \mathbf{y} &= \sum_{k=1}^K \xi_k (\mathbf{B}_k \mathbf{x} + \mathbf{\Gamma}_k^{1/2} \boldsymbol{\varepsilon}^*), \\ \boldsymbol{\varepsilon}^* | \boldsymbol{\xi}, \mathbf{x} &\sim N(\mathbf{0}, \mathbf{I}_M), \\ \boldsymbol{\xi} | \mathbf{x} &\sim MN(\boldsymbol{\pi}), \end{aligned} \tag{1}$$

where \mathbf{y} is a vector of M dependent variables, \mathbf{x} is a vector of P weakly exogenous regressors that usually includes a constant, \mathbf{B}_k is an $M \times P$ matrix of regression coefficients, $\mathbf{\Gamma}_k$ an $M \times M$

positive definite matrix of residual variances, $\boldsymbol{\xi}' = (\xi_1, \dots, \xi_k, \dots, \xi_K)$ is a latent multinomial (MN) variable that can take K different values and $\boldsymbol{\pi}$ is the vector that contains its K probabilities.

We find it convenient to use the ‘‘multinomial logit’’ specification

$$\pi_1 = \frac{1}{\sum_{l=2}^K e^{\beta_l} + 1}, \quad \pi_k = \frac{e^{\beta_k}}{\sum_{l=2}^K e^{\beta_l} + 1} \quad (k = 2, \dots, K), \quad (2)$$

to ensure that $\boldsymbol{\pi}$ remains within the unit simplex, but we could work with alternative specifications instead without any consequence because Amengual, Fiorentini and Sentana (2024a) show that the IM test statistic is numerically invariant to bijective reparametrisations.

2.2 White’s (1982) original IM test

Let $\boldsymbol{\phi}$ denote the vector of parameters of interest, which for model (1) include $\mathbf{b} = (\mathbf{b}'_1, \dots, \mathbf{b}'_K)'$, with $\mathbf{b}_k = \text{vec}(\mathbf{B}_k)$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}'_1, \dots, \boldsymbol{\gamma}'_K)'$, with $\boldsymbol{\gamma}_k = \text{vech}(\boldsymbol{\Gamma}_k)$, and $\boldsymbol{\beta} = (\beta_2, \dots, \beta_K)'$. We can write the log-likelihood function of a random sample of size N on \mathbf{y}_n conditional on \mathbf{x}_n as

$$L_N(\boldsymbol{\phi}) = \sum_{n=1}^N \ln f(\mathbf{y}_n | \mathbf{x}_n; \boldsymbol{\phi}) = \sum_{n=1}^N l_n(\boldsymbol{\phi}), \quad (3)$$

where $f(\mathbf{y}_n | \mathbf{x}_n; \boldsymbol{\phi})$ denotes the density of \mathbf{y}_n given \mathbf{x}_n . Consequently, the average score and Hessian will be given by

$$\bar{\mathbf{s}}_N(\boldsymbol{\phi}) = \frac{1}{N} \sum_{n=1}^N \mathbf{s}_n(\boldsymbol{\phi}) = \frac{1}{N} \sum_{n=1}^N \frac{\partial l_n(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}}, \quad (4)$$

$$\bar{\mathbf{h}}_N(\boldsymbol{\phi}) = \frac{1}{N} \sum_{n=1}^N \mathbf{h}_n(\boldsymbol{\phi}) = \frac{1}{N} \sum_{n=1}^N \frac{\partial^2 l_n(\boldsymbol{\phi})}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'}. \quad (5)$$

If we call $\hat{\boldsymbol{\phi}}_N$ the unrestricted maximum likelihood estimators (MLE) of the parameters of interest, we will have that $\bar{\mathbf{s}}_N(\hat{\boldsymbol{\phi}}_N) = \mathbf{0}$ and $\bar{\mathbf{h}}_N(\hat{\boldsymbol{\phi}}_N)$ negative definite. In what follows, we maintain the regularity conditions White (1982) assumed to prove his Theorem 4.1 on the IM test, which among other things guarantee the consistency of $\hat{\boldsymbol{\phi}}_N$, and its asymptotic normality with a full-rank information matrix when we centre this MLE around its true values, $\boldsymbol{\phi}_0$, and suitably scale it by \sqrt{N} .

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:

$$\text{vech}[\mathbf{h}_n(\boldsymbol{\phi}) + \mathbf{s}_n(\boldsymbol{\phi})\mathbf{s}'_n(\boldsymbol{\phi})]. \quad (6)$$

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

$$\frac{\sqrt{N}}{N} \sum_{n=1}^N \text{vech}[\mathbf{h}_n(\hat{\boldsymbol{\phi}}_N) + \mathbf{s}_n(\hat{\boldsymbol{\phi}}_N)\mathbf{s}'_n(\hat{\boldsymbol{\phi}}_N)].$$

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 ϕ from the covariance matrix between them and the score evaluated at the true values of the parameters. Specifically, we simply need the residual covariance matrix from their least squares projection onto the linear span of $\mathbf{s}_n(\phi_0)$, which is given by

$$\mathcal{R}(\phi_0) - \mathcal{U}(\phi_0)\mathcal{I}^{-1}(\phi_0)\mathcal{U}'(\phi_0), \quad (7)$$

$$\begin{bmatrix} \mathcal{R}(\phi_0) & \mathcal{U}(\phi_0) \\ \mathcal{U}'(\phi_0) & \mathcal{I}(\phi_0) \end{bmatrix} = V \left\{ \begin{array}{c} \text{vech}[\mathbf{h}_n(\phi_0) + \mathbf{s}_n(\phi_0)\mathbf{s}'_n(\phi_0)] \\ \mathbf{s}_n(\phi_0) \end{array} \right\}. \quad (8)$$

Therefore, the infeasible IM test statistic is the quadratic form

$$N \left\{ \frac{1}{N} \sum_{n=1}^N \text{vech}'[\mathbf{h}_n(\hat{\phi}_N) + \mathbf{s}_n(\hat{\phi}_N)\mathbf{s}'_n(\hat{\phi}_N)] \right\} [\mathcal{R}(\phi_0) - \mathcal{U}(\phi_0)\mathcal{I}^{-1}(\phi_0)\mathcal{U}(\phi_0)]^{-} \\ \times \left\{ \frac{1}{N} \sum_{n=1}^N \text{vech}[\mathbf{h}_n(\hat{\phi}_N) + \mathbf{s}_n(\hat{\phi}_N)\mathbf{s}'_n(\hat{\phi}_N)] \right\}, \quad (9)$$

where the superscript “ $-$ ” denotes a generalised inverse, which is often necessary because some of the influence functions in (6) may be an exact linear combination of $\mathbf{s}_n(\phi_0)$ or appear multiple times, as we will see in section 2.3 below. As a result, the number of degrees of freedom of the asymptotic χ^2 distribution under the null of correct specification is $\text{rank}[\mathcal{R}(\phi_0) - \mathcal{U}(\phi_0)\mathcal{I}^{-1}(\phi_0)\mathcal{U}(\phi_0)]$, which requires careful derivation, something that our EM-based procedure helps with, as we illustrate in section 2.4 below.

Chesher (1983) and Lancaster (1984) also suggested a feasible version of (9) as N times the R^2 in the regression of a vector of N ones onto $\mathbf{s}_n(\hat{\phi}_N)$ and $\text{vech}[\mathbf{h}_n(\hat{\phi}_N) + \mathbf{s}_n(\hat{\phi}_N)\mathbf{s}'_n(\hat{\phi}_N)]$ using an ordinary least squares (OLS) routine robust to multicollinearity. The inclusion of $\mathbf{s}_n(\hat{\phi}_N)$ as additional regressors makes the statistic robust to the fact that the influence functions are evaluated at $\hat{\phi}_N$. Nevertheless, as explained by Horowitz (1994) and the references therein, this “outer product of the score” (OPS) regression has very poor finite sample properties, so in our work below we will also rely on the parametric bootstrap applied to a feasible version of (9) that evaluates the theoretical expression (7) at the MLE $\hat{\phi}_N$. The theoretical results in Beran (1988) imply that given that the asymptotic distribution of the test statistic (9) is chi-square, and therefore pivotal, the bootstrapped critical values should not only be valid, but also their errors should be of a lower order of magnitude under additional regularity conditions that guarantee the validity of a higher-order Edgeworth expansion.

2.3 The IM test in the complete data model

2.3.1 The influence functions

Typically, the influence functions (6) are difficult to interpret, so empirical researchers often regard the IM test as a black-box that offers little guidance on how to improve the model specification when it is rejected, even though Chesher (1984) provided a useful reinterpretation as a score test of unobserved heterogeneity in the model parameters, which is a first-order concern in empirical microeconometrics.

We can provide an alternative intuitive interpretation of (6) for the switching regression model (1) by first considering what the IM test would be if the econometrician could observe not only \mathbf{y}_n but also $\boldsymbol{\xi}_n$, so that she will know the subpopulation of origin of all observations. Under the maintained assumption that the elements in \mathbf{x}_n are weakly exogenous variables, their marginal log-likelihood is irrelevant, and we can focus on the contribution from observation n to the joint log-likelihood function of \mathbf{y}_n and $\boldsymbol{\xi}_n$ conditional on the exogenous regressors (see Engle, Hendry and Richard (1983)). In turn, this can be written as the sum of two components. Specifically,

$$\ln f(\mathbf{y}_n, \boldsymbol{\xi}_n | \mathbf{x}_n; \phi) = \ln f(\mathbf{y}_n | \boldsymbol{\xi}_n, \mathbf{x}_n; \mathbf{b}, \boldsymbol{\gamma}) + \ln p(\boldsymbol{\xi}_n | \mathbf{x}_n; \boldsymbol{\beta}), \quad (10)$$

where the conditional log-likelihood function of the regression models for \mathbf{y}_n given $\boldsymbol{\xi}_n$ and \mathbf{x}_n is

$$\begin{aligned} \ln f(\mathbf{y}_n | \boldsymbol{\xi}_n, \mathbf{x}_n; \mathbf{b}, \boldsymbol{\gamma}) &= \sum_{k=1}^K \xi_{kt} \ln f(\mathbf{y}_n | \xi_{kt} = 1, \mathbf{x}_n; \mathbf{b}_k, \boldsymbol{\gamma}_k) \\ &= \sum_{k=1}^K \xi_{kt} \left[-\frac{M}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Gamma}_k| - \frac{1}{2} \boldsymbol{\varepsilon}_n^{*'}(\mathbf{b}_k, \boldsymbol{\gamma}_k) \boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \boldsymbol{\gamma}_k) \right], \quad (11) \end{aligned}$$

with $\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \boldsymbol{\gamma}_k) = \boldsymbol{\Gamma}_k^{-1/2}(\mathbf{y}_n - \mathbf{B}_k \mathbf{x}_n)$ denoting the value of \mathbf{y}_n conditionally standardised according to the k^{th} regime, and the marginal log-likelihood function of the multinomial variable $\boldsymbol{\xi}_n$ given \mathbf{x}_n is

$$\ln p(\boldsymbol{\xi}_n | \mathbf{x}_n; \boldsymbol{\beta}) = \sum_{k=1}^K \xi_{kn} \ln \pi_k(\boldsymbol{\beta}), \quad (12)$$

with $\pi_k(\boldsymbol{\beta}) = \Pr(\boldsymbol{\xi}_n = 1 | \mathbf{x}_n; \phi) \forall n$ specified in (2). The multiple sequential cuts on the vector of $K[Mp + M(M+1)/2 + 1] - 1$ parameters ϕ performed by this additive log-likelihood decomposition allows us to obtain very easily the score vector and Hessian matrix that appear in the influence functions for the IM test (6). On this basis, we can show that:

Proposition 1 1. *The IM test of model (1) when $\boldsymbol{\xi}_n$ is observed coincides with a moment test*

based on the following groups of influence functions:

$$\mathbf{m}_{hn}^k(\mathbf{b}_k, \gamma_k) = \xi_{kn} \mathbf{H}_2[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k)] \otimes \text{vech}(\mathbf{x}_n \mathbf{x}_n'), \quad (13)$$

$$\mathbf{m}_{an}^k(\mathbf{b}_k, \gamma_k) = \xi_{kn} \mathbf{H}_3[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k)] \otimes \mathbf{x}_n, \text{ and} \quad (14)$$

$$\mathbf{m}_{fn}^k(\mathbf{b}_k, \gamma_k) = \xi_{kn} \mathbf{H}_4[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k)] \quad (15)$$

for $k = 1, \dots, K$, where

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

is the $\binom{M+r-1}{r}$ vector containing the distinct multivariate Hermite polynomials of order r of a standardised random vector $\boldsymbol{\varepsilon}^*$, which coincide with the product of the relevant univariate Hermite polynomials.

2. The asymptotic covariance matrix 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 (13)-(15) onto the following influence functions:

$$\left. \begin{aligned} \mathbf{m}_{ln}^k(\mathbf{b}_k, \gamma_k) &= \xi_{kn} \mathbf{H}_1[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k)] \otimes \mathbf{x}_n \\ \mathbf{m}_{sn}^k(\mathbf{b}_k, \gamma_k) &= \xi_{kn} \mathbf{H}_2[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k)] \end{aligned} \right\} k = 1, \dots, K, \quad (16)$$

$$\mathbf{u}_n(\boldsymbol{\beta}), \quad (17)$$

where

$$\mathbf{u}_n(\boldsymbol{\beta}) = [u_{2n}(\boldsymbol{\beta}), \dots, u_{Kn}(\boldsymbol{\beta})]' = [\xi_{2n} - \pi_2(\boldsymbol{\beta}), \dots, \xi_{Kn} - \pi_K(\boldsymbol{\beta})]' \quad (18)$$

are the generalised residuals of the multinomial logit (2), which span the score vectors corresponding to \mathbf{b}_k , γ_k and $\boldsymbol{\beta}$, respectively.

3. The asymptotic distribution of the IM test under correct specification when \mathbf{x}_n has finite fourth moments will be a χ^2 random variable with degrees of freedom equal to

$$K \left\{ \left[\frac{P(P+1)}{2} - \mathbb{I}(1 \in \langle \mathbf{x} \rangle) \right] \frac{(M+1)M}{2} + \frac{P(M+2)(M+1)M}{6} + \frac{(M+3)(M+2)(M+1)M}{24} \right\}, \quad (19)$$

where $\mathbb{I}(\cdot)$ denotes the indicator function and $\langle \mathbf{x} \rangle$ the linear span of \mathbf{x} .

Importantly, there are no IM influence functions associated to the cross-products of the conditional mean and variance parameters across regimes because the Hessian is block diagonal and the cross-products of the mean and variance score vectors for different k 's are identically 0 in view of (16) since $\xi_k \xi_\ell = 0$ for $\forall k \neq \ell$.

In addition, there are no IM influence functions for the multinomial model because they are spanned by the score (17) (see Amengual, Fiorentini and Sentana (2025b)). Intuitively, this simply reflects the fact that (2) is effectively a convenient parametrisation of an unrestricted, non-parametric model.

Finally, there are no IM influence functions for the cross-products of the conditional mean and variance parameters for each regime with the parameters of the multinomial probabilities because the Hessian is also block diagonal between $(\mathbf{b}_k, \boldsymbol{\gamma}_k)$ and $\boldsymbol{\beta}$, $\xi_k^2 = \xi_k \forall k$ and $\xi_k \xi_\ell = 0$ for $\forall k \neq \ell$, so that one can span the Kronecker product of (16) with (17) using the influence functions (16).

Proposition 1 allows us to provide a very intuitive interpretation to the influence functions underlying the IM test. In particular, (13) can be regarded as the multivariate counterpart to White’s (1980) heteroskedasticity test. In turn, (14) is a multivariate version of what Bera and Lee (1993) called a test for “heterocliticity” (i.e., conditional asymmetry) in linear regression models, which will include unconditional skewness of the standardised residuals if the regressors include a constant term. Finally, (15) is the multivariate analogue to the Kiefer and Salmon (1983) version of the kurtosis component of the Jarque and Bera (1980) test. Consequently, the IM test for the entire model coincides with the combination of the IM matrix test for multivariate linear regression models in Amengual, Fiorentini and Sentana (2022) for each of the K regimes. Therefore, the number of degrees of freedom equals K times the degrees of freedom of the IM test for a multivariate regression model, which is substantially lower than a simple count of the number of influence functions in (6).

2.3.2 The asymptotic covariance matrix

Given that $\xi_k \xi_\ell = 0$ for $\forall k \neq \ell$, the IM influence functions and scores are uncorrelated across components. Therefore, one would simply need to compute the joint conditional covariances of (13)-(17) within each regime. Let us start with the conditional covariances of (13)-(16) with (17). To compute these conditional covariances, we simply need to compute $E\{\xi_{kn} \mathbf{H}_r[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \mathbf{u}'_n(\boldsymbol{\beta})\}$ for $r = 1, \dots, 4$ thanks to the law of iterated expectations. At the true values of the parameters, though, $\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \boldsymbol{\gamma}_k) = \boldsymbol{\varepsilon}_n^*$ when $\xi_{kn} = 1$, which is independent of $\boldsymbol{\xi}_n$ in view of (1). Consequently, all these conditional covariances will be 0 too.

Let us now turn to the joint conditional covariances of (13)-(16) within each regime. Amengual, Fiorentini and Sentana (2022) provide computationally simple closed-form expressions for the joint conditional covariance matrix of the IM influence functions for a multivariate regression model with a single regime, which scaled by π_k^{-1} provide the corresponding expressions for the conditional covariance matrix of (13), (14), (15) and (16) for regime k because in large samples a fraction π_k of the observations will come from that regime. Each of these K conditional covariance matrices will

be almost block-diagonal because all multivariate Hermite polynomials have a zero conditional mean and a known constant conditional variance given \mathbf{x}_n ,¹ and they are uncorrelated for different orders (see e.g. Holmquist (1996) or Rahman (2017)). The only exception affects the covariance of (13) with $m_{sn}^k(\mathbf{b}_k, \gamma_k)$, which is a full-rank linear transformation of the score of the residual covariance matrix γ_k . As a result, one would only need to regress (13) on $\mathbf{H}_2[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k)]$ for those observations for which $\xi_{kn} = 1$ to purge the IM influence functions from sampling uncertainty resulting from the estimation of the mean and variance parameters of the k^{th} regime. But when the regressors include either a constant term or an exhaustive set of dummy variables that add up to a constant term, such a regression will be effectively eliminating the $\binom{M+1}{2}$ influence functions corresponding to $\mathbf{H}_2[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k)] \otimes \mathbf{1}$ from (13) because the sample average of $\xi_{kn} \mathbf{H}_2[\boldsymbol{\varepsilon}_n^*(\hat{\mathbf{b}}_k, \hat{\gamma}_k)]$ is equal to $\mathbf{0}$ from the first order conditions for γ_k , which explains the presence of the term

$$\frac{P(P+1)}{2} - \mathbb{I}(1 \in \langle \mathbf{x} \rangle) = \text{rank}\{V[\text{vech}(\mathbf{x}_n \mathbf{x}_n')]\}$$

in (19).

Let

$$\bar{\mathbf{m}}_{lN}^k(\boldsymbol{\theta}) = \frac{1}{\sum_{n=1}^N \xi_{kn}} \sum_{n=1}^N \mathbf{m}_{ln}^k(\boldsymbol{\theta}) \text{ for } l = h, a, f.$$

and assume that \mathbf{x}_n has finite fourth moments. Then, the almost perfect block-diagonality of the conditional covariance matrix of the influence functions that appear in Proposition 1 that we have just seen imply that the IM test of a switching regression model will be asymptotically equivalent under the null hypothesis of correct specification to the sum over k of the sum of the following three moment tests:

$$h_{hN}^k = N \cdot \bar{\mathbf{m}}_{hn}^{k'}(\hat{\boldsymbol{\theta}}_N) \hat{V}^{-1}[\mathbf{m}_{hN}^k(\hat{\boldsymbol{\theta}}_N)] \bar{\mathbf{m}}_{hN}^k(\hat{\boldsymbol{\theta}}_N), \quad (20)$$

$$h_{aN}^k = N \cdot \bar{\mathbf{m}}_{an}^{k'}(\hat{\boldsymbol{\theta}}_N) \hat{V}^{-1}[\mathbf{m}_{aN}^k(\hat{\boldsymbol{\theta}}_N)] \bar{\mathbf{m}}_{aN}^k(\hat{\boldsymbol{\theta}}_N), \quad (21)$$

and

$$h_{fN}^k = N \cdot \bar{\mathbf{m}}_{fN}^{k'}(\hat{\boldsymbol{\theta}}_N) \hat{V}^{-1}[\mathbf{m}_{fN}^k(\hat{\boldsymbol{\theta}}_N)] \bar{\mathbf{m}}_{fN}^k(\hat{\boldsymbol{\theta}}_N), \quad (22)$$

where

$$\lim_{N \rightarrow \infty} V[\sqrt{N} \bar{\mathbf{m}}_{hT}(\hat{\boldsymbol{\theta}}_N)] = \pi_k^{-1} V\{\mathbf{H}_2(\boldsymbol{\varepsilon}_{kt}^*)\} \otimes V[\text{vech}(\mathbf{x}_n \mathbf{x}_n')] \quad (23)$$

$$\lim_{N \rightarrow \infty} V[\sqrt{N} \bar{\mathbf{m}}_{aT}(\hat{\boldsymbol{\theta}}_N)] = \pi_k^{-1} V\{\mathbf{H}_3[\boldsymbol{\varepsilon}_n^*(\boldsymbol{\theta})]\} \otimes E(\mathbf{x}_n \mathbf{x}_n'), \text{ and} \quad (24)$$

$$\lim_{N \rightarrow \infty} V[\sqrt{N} \bar{\mathbf{m}}_{kT}(\hat{\boldsymbol{\theta}}_N)] = \pi_k^{-1} V\{\mathbf{H}_4[\boldsymbol{\varepsilon}_n^*(\boldsymbol{\theta})]\},$$

¹Specifically, the diagonal elements of $V[\mathbf{H}_2(\boldsymbol{\varepsilon}^*)]$ are $V(\varepsilon_i^{*2}) = 2$ and $V(\varepsilon_i^* \varepsilon_{i'}^*) = 1$, for $i' \neq i$, while those of $V[\mathbf{H}_3(\boldsymbol{\varepsilon}^*)]$ are $V(\varepsilon_i^{*3} - 3\varepsilon_i^*) = 6$, $V(\varepsilon_i^{*2} \varepsilon_{i'}^* - \varepsilon_{i'}^*) = 2$ for $i' \neq i$ and $V(\varepsilon_i^* \varepsilon_{i'}^* \varepsilon_{i''}^*) = 1$ for $i'' \neq i' \neq i$. Finally, the diagonal elements of $V[\mathbf{H}_4(\boldsymbol{\varepsilon}^*)]$ are $V[(\varepsilon_i^{*2} - 3\varepsilon_i^*)^2 - 6] = 24$, $V(\varepsilon_i^{*2} \varepsilon_{i'}^{*2} - \varepsilon_i^{*2} - \varepsilon_{i'}^{*2} + 1) = 4$ for $i' \neq i$, $V(\varepsilon_i^{*3} \varepsilon_{i'}^* - 3\varepsilon_i^* \varepsilon_{i'}^*) = 6$ for $i' \neq i$, $V(\varepsilon_i^{*2} \varepsilon_{i'}^* \varepsilon_{i''}^* - \varepsilon_{i'}^* \varepsilon_{i''}^*) = 2$ for $i'' \neq i' \neq i$, and $V(\varepsilon_i^* \varepsilon_{i'}^* \varepsilon_{i''}^* \varepsilon_{i'''}^*) = 1$ for $i''' \neq i'' \neq i' \neq i$ (see Amengual, Fiorentini and Sentana (2024b) for further details).

which converge in distribution to three independent chi-square random variables whose degrees of freedom are $\binom{M+1}{2} \text{rank}\{V[\text{vech}(\mathbf{x}_n \mathbf{x}'_n)]\}$, $\binom{M+2}{3}$ and $\binom{M+3}{4}$, respectively.

Given that the diagonal covariance matrices of $\mathbf{H}_k(\boldsymbol{\varepsilon}_n^*)$ for $k = 2, 3, 4$ do not depend on any unknown quantities under the null of correct specification as we explain in footnote 1, under standard regularity conditions one can consistently estimate (23) and (24) by simply replacing $V[\text{vech}(\mathbf{x}_n \mathbf{x}'_n)]$ and $E(\mathbf{x}_n \mathbf{x}'_n)$ by their sample counterparts and the π'_k s by the corresponding sample frequencies.

2.4 The IM test in the incomplete data model

2.4.1 The influence functions

In many relevant empirical situations, the subpopulation of origin of the observations is unknown, which renders $\boldsymbol{\xi}$ unobserved. In that case, the relevant log-likelihood function for inference purposes is given by

$$\sum_{n=1}^N \ln f(\mathbf{y}_n | \mathbf{x}_n; \boldsymbol{\phi}),$$

where

$$f(\mathbf{y}_n | \mathbf{x}_n; \boldsymbol{\phi}) = \sum_{k=1}^K \pi_k f(\mathbf{y}_n | \xi_{kt} = 1, \mathbf{x}_n; \mathbf{b}_k, \boldsymbol{\gamma}_k)$$

because, once again, we can ignore the marginal component of the regressors due to their weak exogeneity.

Unfortunately, the additive decomposition of the complete data log-likelihood (10) and its associated sequential cuts that simplified the expressions for the score vector and Hessian matrix no longer apply. For that reason, many authors have proposed the EM algorithm popularised by Dempster, Laird and Rubin (1977) for estimating such highly non-linear models, whose recursions consist of a sequence of reweighted linear regression estimators (see e.g. Hartley (1978) for the univariate case and Appendix B for the multivariate one).

As we argued in the introduction, we can also successfully apply the EM principle to simplify the calculation of the IM test in this model. Let $\boldsymbol{\zeta} = (\mathbf{y}', \boldsymbol{\xi}')'$ denote the complete data and $\mathbf{y} = (\mathbf{I}_M | \mathbf{0}_{M \times K}) \boldsymbol{\zeta}$ its mapping to the observed data \mathbf{y} , which implicitly defines a subset R of $\boldsymbol{\zeta}$'s compatible with the observations. Given that the boundary of R does not depend on $\boldsymbol{\phi}$ in model (1), we can use the law of iterated expectations to prove a conditional version of Proposition 1 in Amengual, Fiorentini and Sentana (2024a), which says that for each value of the exogenous regressors \mathbf{x} , the influence functions of the IM test for the incomplete data based on the conditional log-likelihood function $\ln f(\mathbf{y}_n | \mathbf{x}_n; \boldsymbol{\phi})$ are the expected value conditional of the observed variables of the influence functions underlying the IM test for the complete data based on the joint log-likelihood function (10).

Let

$$w_{kn}(\boldsymbol{\phi}) = E(\xi_{kn} | \mathbf{y}_n, \mathbf{x}_n; \boldsymbol{\phi}) = \Pr(\xi_{kn} = 1 | \mathbf{y}_n, \mathbf{x}_n; \boldsymbol{\phi}) = \frac{\pi_{kn}(\boldsymbol{\beta}) |\boldsymbol{\Gamma}_k|^{-1/2} \phi_M[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)]}{\sum_{j=1}^K \pi_{jn}(\boldsymbol{\beta}) |\boldsymbol{\Gamma}_j|^{-1/2} \phi_M[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_j, \boldsymbol{\gamma}_j)]} \quad (25)$$

denote the posterior probability that observation n came from the k^{th} regime, where $\phi_M(\cdot)$ is the M -variate spherical normal density. We can prove the following result:

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

$$w_{kn}(\boldsymbol{\phi}) \mathbf{H}_2[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \otimes \text{vech}(\mathbf{x}_n \mathbf{x}_n'), \quad (26)$$

$$w_{kn}(\boldsymbol{\phi}) \mathbf{H}_3[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \otimes \mathbf{x}_n, \quad \text{and} \quad (27)$$

$$w_{kn}(\boldsymbol{\phi}) \mathbf{H}_4[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)], \quad (28)$$

for $k = 1, \dots, K$.

2. *The asymptotic covariance matrix 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 (26)-(28) onto the following influence functions:*

$$w_{kn}(\boldsymbol{\phi}) \mathbf{H}_1[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \otimes \mathbf{x}_n, \quad (29)$$

$$w_{kn}(\boldsymbol{\phi}) \mathbf{H}_2[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)], \quad (30)$$

for $k = 1, \dots, K$ and

$$w_{kn}(\boldsymbol{\phi}) - \pi_k(\boldsymbol{\beta}) \quad (31)$$

for $k = 2, \dots, K$.

3. *The asymptotic distribution of the IM test will be a χ^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 (29) constitutes a basis for the scores of \mathbf{b}_k obtained by means of the EM principle by Hartley (1978) for the univariate case and Appendix B for the multivariate one, while the sum of (30) is the basis for the scores of $\boldsymbol{\gamma}_k$ obtained in the same manner. Intuitively, the sum of (29) coincides with the average of the usual orthogonality conditions for the regression parameters of the k^{th} regime weighted by the posterior probability that \mathbf{y}_n belongs to that regime. Similarly, the sum of (30) 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 posterior probability that \mathbf{y}_n belongs to that regime. Finally, (31) imposes that the differences between the posterior and prior probabilities of the event $\xi_{kn} = 1$ have a zero sample mean.

In addition, we can interpret (26) 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 posterior probability that \mathbf{y}_n belongs to that regime. Entirely analogous comments

apply to (27) in relation to testing for conditional heterocliticity in those residuals given the regressors, and (28) as far as their unconditional kurtosis is concerned.

Once more, there are no influence functions associated to the cross-products of the mean and variance parameters across regimes. Similarly, there are no influence functions for the multinomial probabilities or their cross-products with the conditional mean and variance parameters. In this respect, the use of the EM principle proves invaluable for figuring out the right number of degrees of freedom for the IM test is (19) rather than the number of influence functions in (6).

2.4.2 The asymptotic covariance matrix

As we explained at the end of section 2.2, the feasible version of IM test statistic suggested by Chesher (1983) and Lancaster (1984), which is N times the R^2 in the regression of a vector of N ones onto $\mathbf{s}_n(\hat{\phi}_N)$ and $\text{vech}[\mathbf{h}_n(\hat{\phi}_N) + \mathbf{s}_n(\hat{\phi}_N)\mathbf{s}'_n(\hat{\phi}_N)]$, effectively replaces the elements of $\mathcal{R}(\phi_0)$, $\mathcal{U}(\phi_0)$ and $\mathcal{I}(\phi_0)$ by their sample counterparts evaluated at $\hat{\phi}_N$. However, given that this yields very noisy estimators that induce substantial size distortions under the null hypothesis of correct specification, we propose another feasible version of the IM test that replaces ϕ_0 with $\hat{\phi}_N$ in the theoretical expressions for $\mathcal{R}(\phi_0)$, $\mathcal{U}(\phi_0)$ and $\mathcal{I}(\phi_0)$ and unconditional expectations by sample averages.

Although the calculation of those asymptotic covariance matrices is more involved than in the observed case, we can still obtain “closed-form” expressions for them conditional on \mathbf{x} as follows. Let $\mathbf{n}(\zeta; \phi, \mathbf{x})$ denote a vector influence functions of the complete data ζ such that

$$E_{\zeta}[\mathbf{n}(\zeta, \mathbf{x}; \phi)|\mathbf{x}] = \mathbf{0}$$

when both the expectation and the influence function are evaluated at the same value of the model parameters, ϕ . In addition, let

$$\mathbf{m}(\mathbf{y}, \mathbf{x}; \phi) = E_{\zeta|\mathbf{y}, \mathbf{x}}[\mathbf{n}(\zeta, \mathbf{x}; \phi)|\mathbf{x}].$$

We can then prove the following conditional version of Proposition 2 in Amengual, Fiorentini and Sentana (2024a), which exploits again the fact that the complete model is much simpler to work with than the incomplete one:

$$V_{\mathbf{y}|\mathbf{x}}[\mathbf{m}(\mathbf{y}, \mathbf{x}; \phi)] = V_{\zeta|\mathbf{x}}[\mathbf{n}(\zeta, \mathbf{x}; \phi)] - E_{\mathbf{y}|\mathbf{x}}\{V_{\zeta|\mathbf{y}, \mathbf{x}}[\mathbf{n}(\zeta, \mathbf{x}; \phi)]\}$$

and

$$\begin{aligned} E_{\mathbf{y}|\mathbf{x}} \left[\mathbf{m}(\mathbf{y}, \mathbf{x}; \phi) \frac{\partial \ln f(\mathbf{y}|\mathbf{x}; \phi)}{\partial \phi'} \right] &= -E_{\mathbf{y}|\mathbf{x}} \left[\frac{\partial \mathbf{m}(\mathbf{y}, \mathbf{x}; \phi)}{\partial \phi'} \right] \\ &= E_{\zeta|\mathbf{x}} \left[\mathbf{n}(\zeta, \mathbf{x}; \phi) \frac{\partial \ln f(\zeta|\mathbf{x}; \phi)}{\partial \phi} \right] - E_{\mathbf{y}|\mathbf{x}} \left\{ \text{cov}_{\zeta|\mathbf{y}, \mathbf{x}} \left[\mathbf{n}(\zeta, \mathbf{x}; \phi), \frac{\partial \ln f(\zeta|\mathbf{x}; \phi)}{\partial \phi} \right] \right\}. \end{aligned}$$

The first elements of the above expressions coincide with the covariance matrix of the influence functions and scores of the complete model obtained in section 2.3. As for the second elements, for a given value of \mathbf{x}_n , we can easily obtain analytical expressions for the relevant expected values conditional on the observed dependent variables \mathbf{y}_n and the regressors using the results in section 3.3 of Amengual, Fiorentini and Sentana (2024a). Unlike what happens in the case of finite mixtures, though, we must now use the numerical quadrature procedure employed in that paper N times because the posterior weights (25) depend on the values of the regressors, which change from observation to observation. Once this is done, we can consistently estimate the covariance matrices without conditioning on \mathbf{x} by simply taking sample averages of the theoretical conditional expressions.

3 Monte Carlo simulations

3.1 Size properties

The asymptotic χ^2 distribution of the IM test statistic may offer a poor approximation in finite samples. For that reason, we assess test sizes by simulating 10,000 samples of length $N = 75$ and $N = 300$ for univariate switching regressions with Gaussian errors within each of two possible regimes, a design motivated by the empirical application in section 4. We construct the regressor matrix by combining a constant with a standard normal random variable, which we keep fixed in repeated samples.² In addition, we set $\boldsymbol{\pi} = (1/4, 3/4)'$, $\mathbf{b}_1 = (5, -1.5)'$, $\mathbf{b}_2 = (0, 0.2)'$, and $\boldsymbol{\gamma} = (0.5, 1)'$. We also consider a parametric bootstrap procedure in which we simulate $B = 99$ samples from model (1) with $\boldsymbol{\phi}$ replaced with $\hat{\boldsymbol{\phi}}_N$.

As stated in Proposition 2, the asymptotic distribution of our proposed IM test is χ^2 with degrees of freedom equal to (19). Similarly, each group of the IM influence functions that appear in Proposition 2.1 also give rise to moments tests whose limiting distributions are χ^2 , with degrees of freedom equal to $K \binom{M+1}{2} [\binom{P+1}{2} - 1]$ (conditional heteroskedasticity), $KP \binom{M+2}{3}$ (conditional and unconditional skewness), and $K \binom{M+3}{4}$ (unconditional kurtosis).

As expected, we find huge over-rejections using the asymptotic-based critical values when we rely on sample second moments. In contrast, we typically find much smaller under-rejections using our theoretical expressions for the covariance matrices of the influence functions. Finally, combining those expressions with the parametric bootstrap practically eliminates size distortions.

²To reduce the effects of Monte Carlo variability even further, one could draw the Gaussian variable by applying the standard normal quantile function to N points equally spaced between $.5/N$ and $1 - .5/N$.

3.2 Power properties

It is also of interest to investigate the power properties of our test. To do so, we simulate 2,500 samples of the same length as in the previous subsection from the following four alternatives:

- DGP A: *Switching regression with two groups but asymmetric t errors*: We again set $\boldsymbol{\pi} = (1/4, 3/4)'$, $\mathbf{b}_1 = (5, -1.5)'$, $\mathbf{b}_2 = (0, 0.2)'$, $\boldsymbol{\gamma} = (0.5, 1)'$ but the errors ε_{kn} follow a zero mean, unit variance asymmetric t distribution with 12 degrees of freedom and skewness parameters -5 and 5 , respectively (see Mencía and Sentana (2012) for details).
- DGP B: *Switching regression with two groups and Gaussian errors but heterogenous intercepts, slopes and residual variances*: Once more, we choose $\boldsymbol{\pi} = (1/4, 3/4)'$, but within each regime $k = 1, 2$, the augmented coefficient vector $(b_{k0n}, b_{k1n}, \log \gamma_{kn})$ follows a trivariate normal distribution with mean vectors $\boldsymbol{\mu}_1 = (5, -1.5, \log 0.1)'$ and $\boldsymbol{\mu}_2 = (0, 0.2, \log 0.2)'$, and equicorrelated covariance matrices that combine a common pairwise correlation equal to $2/3$ with $(1.5, 1.0, 0.5)'$ as standard deviations.³
- DGP C: *Switching regression with three groups and Gaussian errors*: We set $\boldsymbol{\pi} = (1/3, 1/3, 1/3)'$, the regime-specific regression coefficients to $\mathbf{b}_1 = (2.0, -0.5)'$, $\mathbf{b}_2 = (1.0, -0.2)'$, and $\mathbf{b}_3 = (0, 0.5)'$, and common variances $\gamma_k = 0.1$ for $k = 1, 2, 3$.
- DGP D: *Non-switching regression with errors that follow a skewed generalised error distribution (SGED)*: We consider the single-regime linear regression $y_n = \mathbf{b}'\mathbf{x}_n + \gamma^{1/2}\varepsilon_n^*$ with $\mathbf{b} = (0, 0.2)'$, $\gamma = 1$, and $\varepsilon_n^*|x_n \sim i.i.d.$ SGED with skewness parameter $\lambda = 0.70136$ and tail parameter $\nu = 1.19668$ standardised to have zero mean and unit variance, for which we have used the expressions in Theodossiou (1998) and Fernández and Steel (1998).

As can be seen in Table 2, we find non-negligible power against all these four DGPs. However, some components of the IM test are better suited than others to detect specific departures from the null. In particular, the heteroskedasticity component captures continuous (Panel B) and especially discrete (Panel C) unobserved heterogeneity. In turn, the kurtosis component and especially the

³To sample from the trivariate spherical normal underlying the random coefficients in both regimes, we can use spherical coordinates and combine the square root of a χ_3^2 that determines its length with an independent uniform on the surface of the unit sphere in \mathbb{R}^3 that determines its direction. To obtain the latter, one can draw two independent random variables v_1 and v_2 that are uniform on $(0, 1)$ each, and then transform them into two angles that define the spherical coordinates by setting $\theta = 2\pi v_1$ and $\phi = \arccos(2v_2 - 1)$. Finally, one transforms those spherical coordinates into Cartesian ones as $u_1 = \sin \phi \cos \theta$, $u_2 = \sin \phi \sin \theta$ and $u_3 = \cos \phi$. This procedure generalises the use of polar coordinates to draw a standard bivariate normal by combining a uniform angle between 0 and 2π that determines its direction with an independent χ_2^2 whose square root determines its Euclidean norm. To minimise Monte Carlo uncertainty, we force v_1 and v_2 to lie on an equally spaced grid with five points each over $[0, 1]$ and fix a third independent discrete uniform variable v_3 over an equally spaced grid between 0 and 1 with either 3 points when $N = 75$ or 12 points when $N = 300$, finally using the gamma quantile function for transforming v_3 into a χ_3^2 .

skewness one detect non-Gaussian residuals (Panel A). Finally, the kurtosis component is the most powerful when there is a single regime but the errors follow an SGED distribution (Panel D).⁴

In summary, our decomposition of the joint IM test into its components provides flexible diagnostics adaptable to diverse empirical settings and sample sizes while, at the same time, delivers directions to refine the estimated model when the specification of some of its aspects is rejected in the sample at hand.

4 Empirical illustration

Di Vaio and Enflo (2011) analyse the β -convergence of real per capita income using data for 64 countries over the period 1870–2003, which they obtain from the 2007 Maddison database. The main novelty of their approach is that when they regress the average annual rate of growth on initial (log) income, they allow the intercept, slope and residual variance to vary across two different but unknown groups of countries by means of a univariate version of the switching regression model (1).

They find that real per capita income converges for one group of countries but diverges for the other. They also find that the residual variance of the second group is much larger than the residual variance of the first one. Finally, they assign countries to those two groups on the basis of the estimated posterior probabilities $w_{kn}(\hat{\phi}_N)$.

In Figure 1a, we depict the estimated regression lines and classification of countries that we obtain using the 2010 Maddison database instead of the 2007 version they use.⁵ Still, the parameter estimates we report in the first block of columns of Panel A in Table 3 are (almost) identical to the ones in Table 3 of Di Vaio and Enflo (2011).

We then take their specification as given and apply the IM test and its components to this switching regression model. As can be seen from the second and third block of columns of Panel A in Table 3, we are unable to reject the null hypotheses of conditional homoskedasticity and homocliticity and unconditional mesokurtosis for the two regimes, so their chosen specification seems compatible with the data.

Di Vaio and Enflo (2011) also find two unknown groups of countries during the subperiod 1950–2003, both of which converged, but the speed of convergence was noticeably lower for one of them and the residual variance larger, as can be seen in Figure 1b. We include the parameter estimates for this more recent subsample in Panel B of Table 3. Once again, the IM test and

⁴We obtain a similar pattern when we consider bootstrap versions of the OPS version of the IM tests but the rejection rates are significantly lower compared to those in Table 2. These results are available on request.

⁵For computing rates of growth, we identify Russia with the USSR.

its components we report in the same panel are unable to reject the correct specification of their model over this subperiod.

5 Extensions and 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 of the regressions coefficients or the residual variances to be common across regimes. Fortunately, 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 as a function of the corresponding elements in the unrestricted model that we have considered in Proposition 2.

On the other hand, model (1) may be excessively restrictive in other contexts. For example, empirical researchers sometimes model the elements of $\boldsymbol{\pi}$ as functions of some potentially different exogenous regressors using a multinomial logit model. In Amengual, Fiorentini and Sentana (2025c) we develop the IM test for a version of that model without mean regressors. Given that the assumption of *i.i.d.* observations is also problematic in dynamic contexts, in Amengual, Fiorentini and Sentana (2025b) we derive the IM test for Markov switching processes in which each column of the transition matrix is given by a multinomial logit model. We are currently considering the application of the EM principle to obtain the IM test in several other empirically relevant models with latent variables.

References

- 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. (2025a): “Information matrix tests for multinomial logit models”, *Economics Letters* 247, 112180.
- Amengual, D., Fiorentini, G. and Sentana, E. (2025b): “Information matrix tests for Markov switching autoregressive models with covariate-dependent transition probabilities”, CEMFI Working Paper 2502.
- Amengual, D., Fiorentini, G. and Sentana, E. (2025c): “Information matrix tests for finite mixtures with covariate-dependent probabilities”, mimeo, CEMFI.
- Bera, A. and Lee, S (1993): “Information matrix test, parameter heterogeneity and ARCH: a synthesis”, *Review of Economic Studies* 60, 229–240.
- Beran, R. (1988): “Prepivoting test statistics: a bootstrap view of asymptotic refinements”, *Journal of the American Statistical Association* 83, 687–697.
- 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.
- Cho, J.S. and White, H. (2007) “Testing for regime switching”, *Econometrica* 75, 1671–1720.
- Cowell, F.A. and Flachaire, E. (2015): “Statistical methods for distributional analysis”, in A.B. Atkinson and F. Bourguignon (eds), *Handbook of Income Distribution* vol. 2A, 359–465, Elsevier.
- Dempster, A., Laird N., and Rubin, D. (1977): “Maximum likelihood from incomplete data via the EM algorithm”, *Journal of the Royal Statistical Society Series B* 39, 1–38.
- Di Vaio, G. and Enflo, K. (2011): “Did globalization drive convergence? Identifying cross-country growth regimes in the long run”, *European Economic Review* 55, 832–844.
- Engle, R.F., Hendry, D.F. and Richard, J.-F. (1983): “Exogeneity”, *Econometrica* 51, 277–304.
- Fernández, C. and Steel, M. F. J. (1998): “On Bayesian modelling of fat tails and skewness”, *Journal of the American Statistical Association*, 93, 359–371.
- Gouriéroux, C., Monfort, A., Renault, E. Trognon, A. (1987): “Generalized residuals”, *Journal of Econometrics* 34, 5–32.

- Hall, A. (1987): “The information matrix test for the linear model”, *Review of Economic Studies* 54, 257–263.
- Hamilton, J.D. (1996): “Specification testing in Markov-switching time-series models”, *Journal of Econometrics* 70, 127–157.
- Hartley, M.J. (1978): “Estimating mixtures of normal distributions and switching regressions: comment”, *Journal of the American Statistical Association* 73, 738–741.
- Holly, A. and Gardiol, L. (1995): An asymptotic expansion for the distribution of test criteria which are asymptotically distributed as chi-squared under contiguous alternatives”, chap 5 in G. S. Maddala, P.C.B. Phillips and T. N. Srinivasan (eds) *Advances in Econometrics and Quantitative Economics: Essays in honor of Professor C. R. Rao*, Blackwells.
- 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.
- Kiefer, N.M. (1978): “Discrete parameter variation: efficient estimation of a switching regression model”, *Econometrica* 46, 427–434.
- Lancaster, A. (1984): “The covariance matrix of the information matrix test”, *Econometrica* 52, 1051–1053.
- Louis, T.A. (1982): “Finding the observed information matrix when using the EM algorithm”, *Journal of the Royal Statistical Society Series B* 44, 226–233.
- Magnus, J.R. and Neudecker, H. (2019): *Matrix differential calculus with applications in Statistics and Econometrics*, 3rd edition, Wiley.
- 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.
- Pearson, K. (1894): “Contributions to the mathematical theory of evolution”, *Philosophical Transactions of the Royal Society London Series A* 185, 71–110.
- Quandt, R. E. (1972): “A new approach to estimating switching regressions”, *Journal of the American Statistical Association* 67, 306–310.
- Rahman, S. (2017): “Wiener–Hermite polynomial expansion for multivariate Gaussian probability measures”, *Journal of Mathematical Analysis and Applications* 454, 303–334.
- Tauchén, G. (1985): “Diagnostic testing and evaluation of maximum likelihood models”, *Journal of Econometrics* 30, 415–443.
- Theodossiou, P. (1998): “Financial data and the skewed generalized t distribution”, *Management*

Science, 44, 1650–1661.

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.

Appendices

A Proofs

A.1 Proof of Proposition 1

In the complete model, the multiple sequential cuts on the vector of parameters performed by the additive log-likelihood decompositions in (11) and (12) 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}_n given ξ_{kn} , as well as the marginal model for ξ_n . Specifically:

$$\begin{aligned} \frac{\partial \ln f(\mathbf{y}_n, \xi_n | \mathbf{x}_n; \mathbf{b}, \gamma, \beta)}{\partial \mathbf{b}_k} &= \frac{\partial \ln f(\mathbf{y}_n | \xi_n, \mathbf{x}_n; \mathbf{b}, \gamma)}{\partial \mathbf{b}_k} = \xi_{kn} \text{vec}[\Gamma_k^{-1/2} \boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k) \mathbf{x}'_n], \\ \frac{\partial \ln f(\mathbf{y}_n, \xi_n | \mathbf{x}_n; \mathbf{b}, \gamma, \beta)}{\partial \gamma_k} &= \frac{\partial \ln f(\mathbf{y}_n | \xi_n, \mathbf{x}_n; \mathbf{b}, \gamma)}{\partial \gamma_k} \\ &= -\xi_{kn} \frac{1}{2} \mathbf{D}'_M (\Gamma_k^{-1/2'} \otimes \Gamma_k^{-1/2'}) \text{vec}[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k) \boldsymbol{\varepsilon}_n^{*'}(\mathbf{b}_k, \gamma_k) - \mathbf{I}_M] \end{aligned}$$

and

$$\frac{\partial \ln f(\mathbf{y}_n, \xi_n | \mathbf{x}_n; \mathbf{b}, \gamma, \beta)}{\partial \pi_k} = \frac{\partial \ln p(\xi_n | \mathbf{x}_n; \beta)}{\partial \pi_k} = \xi_{kn} \frac{1}{\pi_k(\beta)},$$

for $k = 1, \dots, K$, where \mathbf{D}_M is the duplication matrix of order M (see Magnus and Neudecker (2019)). But given that

$$\frac{\partial \pi_k(\beta)}{\partial \beta_k} = \frac{1}{\left(\sum_{\ell=2}^K e^{\beta_\ell}\right)^2} \left[e^{\beta_k} \left(\sum_{\ell=2}^K e^{\beta_\ell}\right) - e^{2\beta_k} \right] = \pi_k(\beta) [1 - \pi_k(\beta)],$$

for $k = 2, \dots, K$,

$$\frac{\partial \pi_k(\beta)}{\partial \beta_j} = \frac{-e^{\beta_k} e^{\beta_j}}{\left(\sum_{\ell=2}^K e^{\beta_\ell}\right)^2} = -\pi_k(\beta) \pi_j(\beta)$$

for $j \neq k$, and

$$\frac{\partial \pi_1(\beta)}{\partial \beta_j} = -\frac{1}{\left(\sum_{\ell=2}^K e^{\beta_\ell}\right)^2} e^{\beta_j} = -\pi_1(\beta) \pi_j(\beta),$$

the score with respect to β_j ($j = 2, \dots, K$) will be given by

$$\frac{\partial \ln f(\mathbf{y}_n, \xi_n | \mathbf{x}_n; \mathbf{b}, \gamma, \beta)}{\partial \beta_j} = \frac{\partial \ln p(\xi_n | \mathbf{x}_n; \beta)}{\partial \beta_j} = \sum_{k=1}^K \xi_{kn} [I(j=k) - \pi_j(\beta)] = \xi_{jn} - \pi_j(\beta) = u_{jn}(\beta).$$

Consequently, we can write

$$\mathbf{s}_{\mathbf{b}_{kn}}(\phi) = \xi_{kn} (\mathbf{I}_M \otimes \Gamma_k^{-1/2'}) \{ \mathbf{x}_n \otimes \mathbf{H}_1[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k)] \}, \quad (\text{A1})$$

$$\mathbf{s}_{\gamma_{kn}}(\phi) = \xi_{kn} \frac{1}{2} \mathbf{D}'_M (\Gamma_k^{-1/2'} \otimes \Gamma_k^{-1/2'}) \mathbf{D}_M \mathbf{H}_2[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k)] \quad (\text{A2})$$

for $k = 1, \dots, K$, and

$$s_{\beta_j, n}(\phi) = u_{jn}(\beta) \quad (\text{A3})$$

for $j = 2, \dots, K$, which confirms that the scores of the conditional mean parameters will be spanned by $\{\mathbf{H}_1[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k) \otimes \mathbf{x}_n]\}$ because $(\mathbf{I}_M \otimes \boldsymbol{\Gamma}_k^{-1/2'})$ has full rank and $\{\mathbf{x}_n \otimes \mathbf{H}_1[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k)]\} = \mathbf{K}_{NN}\{\mathbf{H}_1[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k) \otimes \mathbf{x}_n]\}$, where \mathbf{K}_{NN} is the full rank commutation matrix of orders N, N (see Theorem 9.c in Magnus and Neudecker (2009)), the scores of the covariance parameters by $\mathbf{H}_2[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k)]$ because $(\boldsymbol{\Gamma}_k^{-1/2'} \otimes \boldsymbol{\Gamma}_k^{-1/2'})$ is also of full rank, and the score of the logit parameters by the generalised residuals, as stated in the second part of the proposition.

To obtain the influence functions of the information matrix test corresponding to the conditional log-likelihood function (11), it is convenient to apply to each of the components the expressions for the Hessian matrix of a Gaussian log-likelihood in a multivariate regression model in Amengual, Fiorentini and Sentana (2022). This yields:

$$\frac{\partial^2 \ln f(\mathbf{y}_n | \boldsymbol{\xi}_n, \mathbf{x}_n; \mathbf{b}, \boldsymbol{\gamma})}{\partial \mathbf{b}_k \partial \mathbf{b}'_k} = -\xi_{kn} \{\mathbf{x}_n \mathbf{x}'_n \otimes \boldsymbol{\Gamma}_k^{-1}\} \quad (\text{A4})$$

$$\frac{\partial^2 \ln f(\mathbf{y}_n | \boldsymbol{\xi}_n, \mathbf{x}_n; \mathbf{b}, \boldsymbol{\gamma})}{\partial \mathbf{b}_k \partial \boldsymbol{\gamma}'_k} = -\xi_{kn} [\mathbf{x}_n \boldsymbol{\varepsilon}'_n(\mathbf{b}_k, \gamma_k) \boldsymbol{\Gamma}_k^{-1/2} \otimes \boldsymbol{\Gamma}_k^{-1}] \mathbf{D}_M \quad (\text{A5})$$

and

$$\frac{\partial^2 \ln f(\mathbf{y}_n | \boldsymbol{\xi}_n, \mathbf{x}_n; \mathbf{b}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_k \partial \boldsymbol{\gamma}'_k} = -\xi_{kn} \mathbf{D}'_M \{ \boldsymbol{\Gamma}_k^{-1} \otimes [\boldsymbol{\Gamma}_k^{-1/2'} \boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k) \boldsymbol{\varepsilon}'_n(\mathbf{b}_k, \gamma_k) \boldsymbol{\Gamma}_k^{-1/2} - \frac{1}{2} \boldsymbol{\Gamma}_k^{-1}] \} \mathbf{D}_M. \quad (\text{A6})$$

Therefore, the sum of the outer product of the score and the Hessian of the conditional model yields the following three terms:

$$\mathbf{b}_k \mathbf{b}_k: \xi_{kn} [\mathbf{x}_n \mathbf{x}'_n \otimes \boldsymbol{\Gamma}_k^{-1/2'} \boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k) \boldsymbol{\varepsilon}'_n(\mathbf{b}_k, \gamma_k) \boldsymbol{\Gamma}_k^{-1/2}] - (\mathbf{x}_n \mathbf{x}'_n \otimes \boldsymbol{\Gamma}_k^{-1}), \quad (\text{A7})$$

$$\begin{aligned} \boldsymbol{\gamma}_k \mathbf{b}_k : & \xi_{kn} \frac{1}{2} \mathbf{D}'_M \text{vec} [\boldsymbol{\Gamma}_k^{-1/2'} \boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k) \boldsymbol{\varepsilon}'_n(\mathbf{b}_k, \gamma_k) \boldsymbol{\Gamma}_k^{-1/2} - \boldsymbol{\Gamma}_k^{-1}] [\mathbf{x}'_n \otimes \boldsymbol{\varepsilon}'_n(\mathbf{b}_k, \gamma_k) \boldsymbol{\Gamma}_k^{-1/2}] \\ & - \mathbf{D}'_M [\boldsymbol{\Gamma}_k^{-1/2'} \boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k) \mathbf{x}'_n \otimes \boldsymbol{\Gamma}_k^{-1}], \end{aligned} \quad (\text{A8})$$

and

$$\begin{aligned} \boldsymbol{\gamma}_k \boldsymbol{\gamma}_k : & \xi_{kn} \frac{1}{4} \mathbf{D}'_M \text{vec} [\boldsymbol{\Gamma}_k^{-1/2'} \boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k) \boldsymbol{\varepsilon}'_n(\mathbf{b}_k, \gamma_k) \boldsymbol{\Gamma}_k^{-1/2} - \boldsymbol{\Gamma}_k^{-1}] \\ & \times \text{vec}' [\boldsymbol{\Gamma}_k^{-1/2'} \boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k) \boldsymbol{\varepsilon}'_n(\mathbf{b}_k, \gamma_k) \boldsymbol{\Gamma}_k^{-1/2} - \boldsymbol{\Gamma}_k^{-1}] \mathbf{D}_M \\ & - \mathbf{D}'_M \{ \boldsymbol{\Gamma}_k^{-1} \otimes [\boldsymbol{\Gamma}_k^{-1/2'} \boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k) \boldsymbol{\varepsilon}'_n(\mathbf{b}_k, \gamma_k) \boldsymbol{\Gamma}_k^{-1/2} - \frac{1}{2} \boldsymbol{\Gamma}_k^{-1}] \} \mathbf{D}_M. \end{aligned} \quad (\text{A9})$$

A straightforward application of the arguments in the proof of Proposition 1 of Amengual, Fiorentini and Sentana (2024b) shows that the expressions for the *symmetrised* version of the sum of the Hessian and the outer product of the scores are spanned by (13), (14) and (15) (see Holly & Gardiol (1995) for the properties of the Holmquist (1996) symmetrisation operators in this context, as well as Theorem 9.c in Magnus and Neudecker (2009) for the permutation of Kronecker products using the commutation matrix).

All the remaining influence functions involving conditional mean and variance parameters from different components will be zero because $\xi_k \xi_\ell = 0 \forall k \neq \ell$.

Let us now look at the second derivatives of the marginal log-likelihood function (12). Given (A3), we will have

$$\frac{\partial^2 \ln p(\boldsymbol{\xi}_n | \mathbf{x}_n; \boldsymbol{\beta})}{\partial \beta_j \partial \beta'_j} = -\pi_j(\boldsymbol{\beta})[1 - \pi_j(\boldsymbol{\beta})]$$

and

$$\frac{\partial^2 \ln p(\boldsymbol{\xi}_n | \mathbf{x}_n; \boldsymbol{\beta})}{\partial \beta_j \partial \beta'_\ell} = \pi_j(\boldsymbol{\beta})\pi_\ell(\boldsymbol{\beta}).$$

Therefore,

$$\begin{aligned} \frac{\partial \ln p(\boldsymbol{\xi}_n | \mathbf{x}_n; \boldsymbol{\beta})}{\partial \beta_j} \frac{\partial \ln p(\boldsymbol{\xi}_n | \mathbf{x}_n; \boldsymbol{\beta})}{\partial \beta_j} + \frac{\partial^2 \ln p(\boldsymbol{\xi}_n | \mathbf{x}_n; \boldsymbol{\beta})}{\partial \beta_j \partial \beta_j} &= [\xi_{jn} - \pi_j(\boldsymbol{\beta})]^2 - \pi_j[1 - \pi_j(\boldsymbol{\beta})] \\ &= [1 - 2\pi_j(\boldsymbol{\beta})]u_{jn}(\boldsymbol{\beta}) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \ln p(\boldsymbol{\xi}_n | \mathbf{x}_n; \boldsymbol{\beta})}{\partial \beta_j} \frac{\partial \ln p(\boldsymbol{\xi}_n | \mathbf{x}_n; \boldsymbol{\beta})}{\partial \beta_\ell} + \frac{\partial^2 \ln p(\boldsymbol{\xi}_n | \mathbf{x}_n; \boldsymbol{\beta})}{\partial \beta_j \partial \beta_\ell} &= [\xi_{jn} - \pi_j(\boldsymbol{\beta})][\xi_{\ell n} - \pi_\ell(\boldsymbol{\beta})] + \pi_j(\boldsymbol{\beta})\pi_\ell(\boldsymbol{\beta}) \\ &= -\pi_\ell(\boldsymbol{\beta})[\xi_{jn} - \pi_j(\boldsymbol{\beta})] - \pi_j(\boldsymbol{\beta})[\xi_{\ell n} - \pi_\ell(\boldsymbol{\beta})] \end{aligned}$$

because $\xi_j^2 = \xi_j$ and $\xi_j \xi_\ell = 0 \forall k \neq \ell$. Consequently, these influence functions are spanned by (A3), so we can ignore them in computing the IM test.

The only influence functions left are the ones corresponding to the combinations of the conditional mean and variance parameters with the logit parameters. Given that the Hessian is block-diagonal between those groups of parameters, we simply need to look at the products of the scores. Once again, the fact that $\xi_j \xi_\ell = 0 \forall k \neq \ell$ implies that the only terms we should focus on are

$$\begin{aligned} \mathbf{s}_{\mathbf{b}_k n}(\phi) \frac{\partial \ln p(\boldsymbol{\xi}_n | \mathbf{x}_n; \boldsymbol{\beta})}{\partial \beta_k} &= \xi_{kn} (\mathbf{I}_M \otimes \boldsymbol{\Gamma}_k^{-1/2'}) \{ \mathbf{x}_n \otimes \mathbf{H}_1[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \} u_{kn}(\boldsymbol{\beta}) \\ &= \xi_{kn} (\mathbf{I}_M \otimes \boldsymbol{\Gamma}_k^{-1/2'}) \{ \mathbf{x}_n \otimes \mathbf{H}_1[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] \} [1 - \pi_k(\boldsymbol{\beta})] \end{aligned}$$

and

$$\begin{aligned} \mathbf{s}_{\boldsymbol{\gamma}_k n}(\phi) \frac{\partial \ln p(\boldsymbol{\xi}_n | \mathbf{x}_n; \boldsymbol{\beta})}{\partial \beta_k} &= \xi_{kn} \frac{1}{2} \mathbf{D}'_M (\boldsymbol{\Gamma}_k^{-1/2'} \otimes \boldsymbol{\Gamma}_k^{-1/2'}) \mathbf{D}_M \mathbf{H}_2[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] u_{kn}(\boldsymbol{\beta}) \\ &= \xi_{kn} \frac{1}{2} \mathbf{D}'_M (\boldsymbol{\Gamma}_k^{-1/2'} \otimes \boldsymbol{\Gamma}_k^{-1/2'}) \mathbf{D}_M \mathbf{H}_2[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \boldsymbol{\gamma}_k)] [1 - \pi_k(\boldsymbol{\beta})] \end{aligned}$$

because $\xi_k^2 = \xi_k$. Consequently, these influence functions are spanned by (A1) and (A2), respectively, so we can also ignore them in computing the IM test.

Finally, the number of degrees of freedom reflects the number of elements in (13), (14) and (15), but with an adjustment for the possibility that the regressors either include a constant term or an exhaustive set of dummy variables that span a constant term. \square

A.2 Proof of Proposition 2

The proof is a straightforward application of a conditional version of Proposition 1 in Amengual, Fiorentini and Sentana (2024a) to the expressions that appear in the proof of Proposition 1 above. \square

B The EM algorithm

The EM principle implies that the scores of the log-likelihood function of \mathbf{y}_n given \mathbf{x}_n will be given by

$$\begin{aligned}
\frac{\partial \ln f(\mathbf{y}_n | \mathbf{x}_n; \mathbf{b}, \gamma, \boldsymbol{\pi})}{\partial \mathbf{b}_k} &= w_{kn}(\mathbf{b}, \gamma, \boldsymbol{\pi}) \text{vec}[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k) \mathbf{x}'_n] = w_{kn}(\mathbf{b}, \gamma, \boldsymbol{\pi}) \text{vec}[\boldsymbol{\Gamma}_k^{-1/2} \boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k) \mathbf{x}'_n], \\
\frac{\partial \ln f(\mathbf{y}_n | \mathbf{x}_n; \mathbf{b}, \gamma, \boldsymbol{\pi})}{\partial \gamma_k} &= \frac{\partial \ln f(\mathbf{y}_n | \boldsymbol{\xi}_n, \mathbf{x}_n; \mathbf{b}, \gamma)}{\partial \gamma_k} \\
&= w_{kn}(\mathbf{b}, \gamma, \boldsymbol{\pi}) \frac{1}{2} \mathbf{D}'_M (\boldsymbol{\Gamma}_k'^{-1} \otimes \boldsymbol{\Gamma}_k'^{-1}) \text{vec}[\boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k) \boldsymbol{\varepsilon}_n^{*'}(\mathbf{b}_k, \gamma_k) - \mathbf{I}_M] \\
&= w_{kn}(\mathbf{b}, \gamma, \boldsymbol{\pi}) \frac{1}{2} \mathbf{D}'_M \text{vec}[\boldsymbol{\Gamma}^{-1/2} \boldsymbol{\varepsilon}_n^*(\mathbf{b}_k, \gamma_k) \boldsymbol{\varepsilon}_n^{*'}(\mathbf{b}_k, \gamma_k) \boldsymbol{\Gamma}^{-1/2} - \boldsymbol{\Gamma}^{-1}], \\
\frac{\partial \ln f(\mathbf{y}_n | \mathbf{x}_n; \mathbf{b}, \gamma, \boldsymbol{\pi})}{\partial \pi_k} &= w_{kn}(\mathbf{b}, \gamma, \boldsymbol{\pi}) \frac{1}{\pi_k}.
\end{aligned}$$

These equations immediately imply that the recursions of the EM algorithm will be given by

$$\begin{aligned}
\hat{\mathbf{B}}_k^{(h)} &= \left[\frac{1}{N} \sum_{n=1}^N w_{kn}(\mathbf{b}^{(h-1)}, \boldsymbol{\gamma}^{(h-1)}, \boldsymbol{\pi}^{(h-1)}) \mathbf{y}_n \mathbf{x}'_n \right] \left[\frac{1}{N} \sum_{n=1}^N w_{kn}(\mathbf{b}^{(h-1)}, \boldsymbol{\gamma}^{(h-1)}, \boldsymbol{\pi}^{(h-1)}) \mathbf{x}_n \mathbf{x}'_n \right]^{-1}, \\
\hat{\boldsymbol{\Gamma}}_k^{(h)} &= \frac{1}{\hat{\pi}_k^{(h)}} \frac{1}{N} \sum_{n=1}^N w_{kn}(\mathbf{b}^{(h-1)}, \boldsymbol{\gamma}^{(h-1)}, \boldsymbol{\pi}^{(h-1)}) (\mathbf{y}_n - \hat{\mathbf{B}}_k^{(h)} \mathbf{x}_n) (\mathbf{y}_n - \hat{\mathbf{B}}_k^{(h)} \mathbf{x}_n)'
\end{aligned}$$

and

$$\hat{\pi}_k^{(h)} = \frac{1}{N} \sum_{n=1}^N w_{kn}(\mathbf{b}^{(h-1)}, \boldsymbol{\gamma}^{(h-1)}, \boldsymbol{\pi}^{(h-1)}). \tag{B11}$$

Given that (25) is homogeneous of degree zero in $\boldsymbol{\pi}$, in principle these posterior probabilities are compatible with values of $\boldsymbol{\pi}$ outside the unit simplex. Nevertheless, a useful property of the EM algorithm is that it automatically imposes the relevant inequality restrictions on the estimators of $\boldsymbol{\pi}$ because $\sum_{k=1}^K w_{kn}(\boldsymbol{\phi}) = 1$ regardless of the values of \mathbf{y}_n and $\boldsymbol{\phi}$. Therefore, there is no need to use parametrisation (2) at each iteration, only at the very last one, relying on (B11) for all previous ones.

Table 1: Finite sample properties: Null hypothesis

Panel A: Asymptotic size properties for $N = 75$						
Size	Test version					
	OPS			IM		
	10%	5%	1%	10%	5%	1%
Joint	95.29	92.74	84.73	5.52	3.84	1.91
White	39.11	26.78	10.40	6.56	3.50	1.10
Skewness	53.84	42.96	24.06	5.73	3.43	1.32
Kurtosis	60.05	53.01	39.95	3.25	2.13	0.90

Panel B: Bootstrap size properties for $N = 75$						
Size	Test version					
	OPS			IM		
	10%	5%	1%	10%	5%	1%
Joint	6.00	2.24	0.26	9.68	4.60	0.91
White	8.38	3.99	0.61	9.75	4.93	0.97
Skewness	8.51	3.83	0.55	9.98	4.82	0.91
Kurtosis	8.69	4.16	0.92	9.96	4.93	1.11

Panel C: Asymptotic size properties for $N = 300$						
Size	Test version					
	OPS			IM		
	10%	5%	1%	10%	5%	1%
Joint	66.52	58.08	42.01	9.11	5.78	2.57
White	24.07	15.68	5.18	8.10	4.44	1.32
Skewness	25.49	16.43	6.25	8.73	5.34	2.00
Kurtosis	31.78	24.77	15.27	6.28	3.72	1.53

Panel D: Bootstrap size properties for $N = 300$						
Size	Test version					
	OPS			IM		
	10%	5%	1%	10%	5%	1%
Joint	9.89	4.73	0.93	9.92	4.92	0.83
White	10.05	4.78	0.85	9.60	4.63	0.96
Skewness	9.98	4.79	1.19	9.93	5.18	1.17
Kurtosis	9.74	4.92	1.14	9.79	4.84	0.95

Notes: Monte Carlo empirical rejection rates based on 10,000 replications. OPS refers to the version of the statistic proposed by Chesher (1983) and Lancaster (1984), while IM to the feasible version that makes use of the theoretical expressions replacing the true parameter values ϕ_0 by their MLEs $\hat{\phi}_N$. Panels A and C contain rejection rates based on asymptotic critical values corresponding to a χ^2 distribution with 10, 4, 4 and 2 degrees of freedom (sorted by rows), while those in Panels B and D are based on a parametric bootstrap procedure in which we simulate $B = 99$ samples from the switching regression model estimated under the null.

Table 2: Finite sample properties: IM test (Bootstrap power)

Panel A: <i>Two groups, asymmetric t errors</i>						
Size	$N = 75$			$N = 300$		
	10%	5%	1%	10%	5%	1%
Joint	57.88	44.60	19.88	99.52	98.72	83.40
White	21.08	14.00	4.96	42.96	31.40	14.76
Skewness	67.04	53.56	23.92	99.88	99.64	91.56
Kurtosis	36.76	29.00	13.68	78.08	70.36	45.24

Panel B: <i>Two groups, heterogenous parameters</i>						
Size	$N = 75$			$N = 300$		
	10%	5%	1%	10%	5%	1%
Joint	51.32	28.76	4.40	99.80	98.36	72.44
White	22.36	12.40	3.24	59.92	48.64	26.36
Skewness	73.20	56.96	15.52	91.52	88.28	81.80
Kurtosis	15.60	6.52	0.68	21.64	11.16	1.84

Panel C: <i>Three groups</i>						
Size	$N = 75$			$N = 300$		
	10%	5%	1%	10%	5%	1%
Joint	72.00	58.04	18.84	99.12	97.04	90.84
White	81.96	73.20	45.92	99.16	98.20	93.84
Skewness	17.48	10.80	3.00	43.72	33.76	16.60
Kurtosis	12.20	4.40	0.32	30.40	14.32	1.04

Panel D: <i>Linear regression with SGED errors</i>						
Size	$N = 300$			$N = 300$		
	10%	5%	1%	10%	5%	1%
Joint	54.52	46.16	27.20	78.72	69.84	45.92
White	26.00	16.92	5.80	23.56	14.92	5.36
Skewness	46.92	37.88	21.16	59.84	49.12	30.88
Kurtosis	58.00	51.40	31.92	85.08	77.60	54.84

Notes: Monte Carlo empirical rejection rates based on 2,500 replications. Test statistics computed using the feasible version that makes use of the theoretical expressions replacing the true parameter values ϕ_0 by their MLEs $\hat{\phi}_N$. Rejection rates based on a parametric bootstrap procedure in which we simulate $B = 99$ samples from the switching regression model estimated under the null.

Table 3: Empirical Application: Convergence of real per capita income

Panel A: Period 1870–2003							
	Parameter Estimates			Test statistic		Bootstrap p-values	
	$k = 1$	$k = 2$		OPS	IM	OPS	IM
	Intercept	7.026		-0.299	Joint	15.822	8.515
Slope	-0.688	0.252	White	4.517	6.040	0.663	0.141
Variance	0.008	0.152	Skewness	4.641	1.987	0.717	0.565
Probability	0.232	0.768	Kurtosis	1.777	0.228	0.805	0.867

Panel B: Period 1950–2003							
	Parameter Estimates			Test statistic		Bootstrap p-values	
	$k = 1$	$k = 2$		OPS	IM	OPS	IM
	Intercept	16.859		4.418	Joint	26.754	12.984
Slope	-1.652	-0.316	White	7.002	5.021	0.463	0.191
Variance	0.031	0.981	Skewness	8.417	6.328	0.434	0.087
Probability	0.252	0.748	Kurtosis	0.977	0.296	0.900	0.829

Notes: As in Di Vaio and Enflo (2011), we use real per capita income data for 64 countries over the period 1870–2003, but obtained from the 2010 Maddison database. OPS refers to the version of the statistic proposed by Chesher (1983) and Lancaster (1984), while IM to the feasible version that makes use of the theoretical expressions replacing the true parameter values ϕ_0 by their MLEs $\hat{\phi}_N$. p-values based on a parametric bootstrap procedure in which we simulate $B = 9999$ samples from the switching regression model estimated under the null.

Figure 1:

Figure 1a: Period 1870–2003

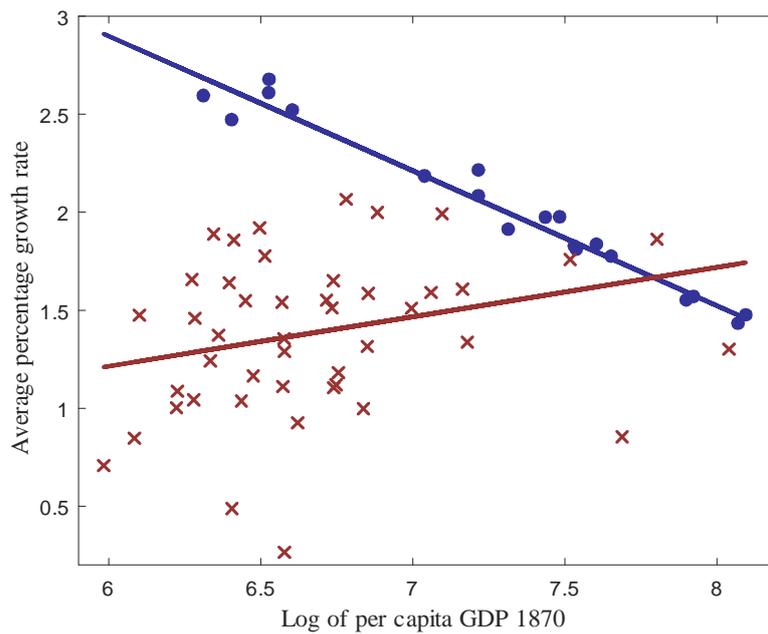
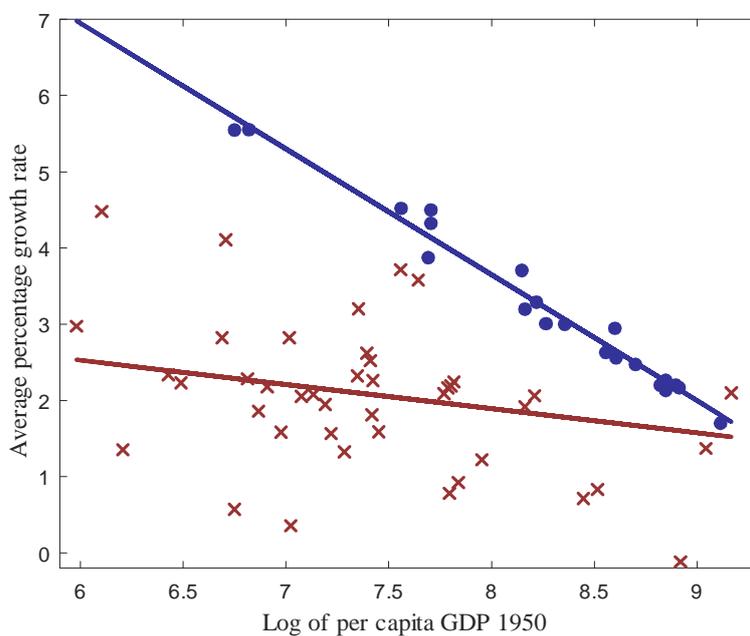


Figure 1b: Period 1950–2003



Notes: As in Di Vaio and Enflo (2011), we use real per capita income data for 64 countries over the period 1870-2003, but obtained from the 2010 Maddison database.