Homogenous vs. Heterogenous Transition Functions in Smooth Transition Regressions – A LM-Type Test

Matei Demetrescu* Julian S. Leppin† Stefan Reitz‡

This version: August 2017

Abstract

(Panel) Smooth Transition Regressions substantially gained in popularity due to their flexibility in modeling regression coefficients as homogeneous or heterogeneous functions of transition variables. In the estimation process, however, researchers typically face a trade-off in the sense that a single (homogeneous) transition function may yield biased estimates if the true model is heterogeneous, while the latter specification is accompanied by convergence problems and longer estimation time, rendering their application less appealing. This paper proposes a Lagrange multiplier test indicating whether the homogeneous smooth transition regression model is appropriate against the competing heterogeneous alternative. The empirical size and power of the test are evaluated by Monte Carlo simulations.

Keywords: STR model; multivariate; nonlinear models, testing

JEL Classification: C52; C22; C12

*Institute for Statistics and Econometrics, University of Kiel.

†University of Kiel and Hamburg Institute of International Economics (HWWI), Heimhuder Strasse 71, 20148 Hamburg, Germany.

‡Corresponding author. We thank the participants of the 3rd International Workshop on 'Financial Markets and Nonlinear Dynamics' (FMND) for helpful comments and suggestions. We are also indebted to Annalee McWilliams for excellent research assistance. Kiel Institute for the World Economy, Germany, and Institute for Quantitative Business and Economics Research, University of Kiel, D-24118 Kiel, Heinrich-Hecht-Platz 9, email: stefan.reitz@qber.uni-kiel.de.

1 Introduction

Since its introduction, the class of smooth transition regression (STR) models has become increasingly popular in empirical research. The standard modeling framework for smooth transition models goes back to Teräsvirta (1994) originally suggesting a framework for a (homogeneous) smooth transition autoregressive model (STAR). Since then, this class of models has been extended to cover a variety of specific properties of empirical applications. Amongst others, smooth transitions functions have been introduced in Vector Autoregressive (STR-VAR) models (Camacho, 2004), GARCH (STR-GARCH) models (Lundbergh and Teräsvirta, 1998), or panel regression (PSTR) models (González et al., 2005). In fact, panel STR models have become quite popular in applied work,[1] and we focus on their specification in the following.

STR models with more than one regime-switching regressor can be built with either homogeneous or heterogeneous[2] transition functions. STR models with homogeneous transition functions use the same transition function for each regime-switching regressor. On the other hand, heterogeneous transition functions allow for different transition functions across regressors with either the same or a different set of transition variables. In our case heterogeneous transition functions across the model's regressors arise from differing parameters for a given set of transition variables. In the context of panel smooth transition models, Leppin and Reitz (2016) allow for heterogeneous transition functions. While heterogeneous STR model are not common in panel smooth transition models, the idea is not new in the context of Vector STAR models. As mentioned in van Dijk et al. (2002), p.8 "It is straightforward to generalize the model to incorporate equation-specific transition functions ... and thereby allow for equation-specific regime-switching". Applications of Vector STAR models with equation specific transition functions can be found in de Dios Tena and Tremayne (2009) or Schleer and Semmler (2015). Teräsvirta and Yang (2014) outlines linearity testing for Vector STAR models.

The standard (panel) STR model with common transition functions is nested within the heterogeneous STR model. Of course, the application of a more flexible heterogeneous STR model is the cautionary choice, because estimating the homogeneous STR model when a heterogeneous specification is appropriate generally leads to biased parameter values. However, heterogeneous transition functions come at the price of convergence problems and increased estimation time, because for each regime-switching regressor a set of parameters specifying the nonlinear transition has to be estimated. While computation

[1]On August 31st, 2017, Google Scholar reported a number of 446 citations of the original PSTR paper of González et al. (2005).

[2]The use of heterogeneous nonlinearity in the context of STR model is not totally unambiguous. Another characterization of heterogeneous nonlinearity is used by Anderson and Vahid (1998). They derive a test against common nonlinear transition in multivariate regressions.

time is not a concern in small samples (small number of units and time observations), it becomes an important factor with increasing sample size and increasing number of regime-switching parameters.

Therefore, we suggest to augment the specification step of the heterogeneous STR estimation procedure by a test for homogeneity against the heterogeneous STR alternative. To this end, we propose a Lagrange Multiplier (LM) test and show how to implement it in both pure time-series and panel situations. Our test is intended to inform the research whether the parsimonious homogeneous (panel) STR is sufficient without sacrificing unbiased parameter estimates.

The model specification step of STR models generally consists of a set of linearity tests against the nonlinear alternative. The usual test is the Taylor expansion-based linearity test from Luukkonen et al. (1988). Monte Carlo-based tests for linearity are available as well, see for example Hansen (1996) and Hansen (1999). A comparison of the tests can be found in González and Teräsvirta (2006). The logic of these tests also allows for the post-estimation model evaluation. For instance, Eitrheim and Teräsvirta (1996) proposed a Taylor expansion-based test for no-remaining nonlinearity.[3]

Our test may be applied in the final model evaluation stage as a misspecification test after the estimation of a homogeneous (panel) STR model. In this case, the test determines the adequateness of the estimated model specification from ex post. Since the test for multivariate transition functions requires the estimation of the model under the null of homogeneity it can be seen as a misspecification test after the estimation of a standard STR model with a common transition function.

The reminder of the paper is structured as follows. Section 2 introduces the proposed test. Section 3 presents a Monte Carlo evaluation of the properties of our test. Here, we examine the power and size of the test based on estimates from nonlinear least squares (NLS). Section 4 concludes, and the appendix gathers some technical results.

2 Testing for homogeneity

2.1 The Langrange Multiplier approach

We now motivate our test statistic on the case of a single unit.

The homogeneous STR model for time $t = 1, ..., T$ with $K > 1$ regime-switching

[3]Eitrheim and Teräsvirta also provide size and power simulations. Another evaluation of the test's properties can be found in Lundbergh and Teräsvirta (1998).

regressor variables $x_{t,k}$ and transition function g is given as

$$y_t = \sum_{k=1}^{K} \beta_{0,k} x_{t,k} + \sum_{k=1}^{K} \beta_{1,k} x_{t,k} g(q_t; \gamma, c) + u_t, \tag{1}$$

with a unique logistic transition function

$$g(q_t; \gamma, c) = \frac{1}{1 + \exp\left(-\gamma(q_t - c)\right)}, \tag{2}$$

which is used for all K regime-switching variables. The regimes are taken to be well-separated, $\beta_{1,k} \neq 0$. The parameter c is a location parameter, γ is the speed of transition between regimes and q_t is the single transition variable. The heterogeneous model differs from the previous model by allowing for regressor-specific transition functions

$$g(q_t; \gamma_k, c_k) = \frac{1}{1 + \exp\left(-\gamma_k(q_t - c_k)\right)}, \tag{3}$$

where the parameters γ_k and c_k are regressor-specific. Note that the transition variable q_t is restricted to be the same across all transition functions,[4] and the transition functions all belong to the same family g indexed by the respective location and transition parameters. The heterogeneous STR model boils down to the homogeneous null model, if we restrict the parameters γ_k and c_k to be equal across all transition functions. In other words, the homogeneous STR model is nested in the heterogeneous STR model, which allows the construction of a formal test of the null hypothesis

$$H_0 : \gamma = \gamma_1 = ... = \gamma_k, \;\; c = c_1 = ... = c_k.$$

To derive the LM-type test we assume that the data generating process of the variable y_t is a smooth transition-type function with $K > 1$ exogenous regressors $x_{t,k}$, $k = 1, \ldots, K$, and iid Gaussian errors u_t, possibly with K different transition functions $g_{t,k}$ as specified in eqs. (1) and (3). This is only for deriving a test; for establishing its limiting null distribution, we specify less restrictive conditions in the following subsection.

Estimation under the null is conducted by ignoring the potentially heterogeneous nature of the transition functions, i.e. a standard STR model with a single transition function g is estimated by imposing the restrictions that $\gamma = \gamma_1 = ... = \gamma_K$ and $c = c_1 =$

[4]Otherwise no test for parameter equality is needed, because the transition functions would differ by definition.

$... = c_K$. This reduces the model to

$$y_t = \sum_{k=1}^{K} \beta^*_{0,k} x_{t,k} + \sum_{k=1}^{K} \beta^*_{1,k} x_{t,k} g(q_t; \gamma, c) + u_t. \tag{4}$$

In general, estimating Equation (4) when there is heterogeneity results in biased parameter estimates since $\beta^* \neq \beta$, therefore testing homogeneity is an important model specification step.

Let θ be the parameter vector, $\theta = (\beta_{0,1}, \ldots, \beta_{0,K}, \beta_{1,1}, \ldots, \beta_{1,K}, \gamma_1, \ldots, \gamma_K, c_1, \ldots, c_K)'$, and by $\hat{\theta}_{\text{H0}}$ the ML estimator under the null hypothesis, i.e. in the model from Equation (4). When the error term in Equation (4) is normally distributed with mean 0 and variance σ^2, the (conditional) log-likelihood is given by

$$\begin{aligned} l(\theta) = & -\frac{T}{2}\log 2\pi - \frac{T}{2}\log\sigma^2 \\ & -\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(y_t - \sum_{k=1}^{K}\beta_{0,k}x_{t,k} - \sum_{k=1}^{K}\beta_{1,k}x_{t,k}g(q_t;\gamma_k, c_k)\right)^2, \end{aligned} \tag{5}$$

where, without loss of generality, we may treat σ^2 to be known. The gradient of the log-likelihood, evaluated at $\theta = \hat{\theta}_{\text{H0}}$ under the null, should be insignificantly different from zero and the LM test is based on this property.

Following Engle (1984) we reformulate the test of parameter equality to a test for omitted variables, where the omitted variables are simply the derivatives of the log likelihood with respect to the restricted parameters. This can be derived directly with the use of Taylor series approximations. Starting from the unrestricted, heterogenous model

$$y_t = \sum_{k=1}^{K} \beta_{0,k} x_{t,k} + \sum_{k=1}^{K} \beta_{1,k} x_{t,k} g(q_t; \gamma_k, c_k) + u_t, \tag{6}$$

expand Equation (6) with the terms $\beta_{1,k} x_{t,k} g(q_t; \gamma, c)$ to derive

$$y_t = \sum_{k=1}^{K} \beta_{0,k} x_{t,k} + \sum_{k=1}^{K} \beta_{1,k} x_{t,k} g(q_t; \gamma, c) + \sum_{k=1}^{K} \beta_{1,k} x_{t,k} \left[g(q_t; \gamma_k, c_k) - g(q_t; \gamma, c)\right] + u_t. \tag{7}$$

If the transition functions are equal across the regressors (i.e. $\gamma_k = \gamma$, $c_k = c$ for $k = 1, \ldots, K$), the terms $\beta_{1,k} x_{t,k} \left[g(q_t; \gamma_k, c_k) - g(q_t; \gamma, c)\right]$ cancel out, which leaves us with the homogeneous STR model. In case of different transition functions the two terms $\beta_{1,k} x_{t,k} g(q_t; \gamma, c)$ in Equation (7) cancel out and we are back at the heterogenous model from Equation (6). Therefore, if the terms $\beta_{1,k} x_{t,k} \left[g(q_t; \gamma_k, c_k) - g(q_t; \gamma, c)\right]$ are significant, at least one of the parameters γ_k or c_k is not equal to the others across transition functions.

With given estimates of parameters γ and c under the null we can reformulate Equation (7) to test for the joint hypothesis of $\gamma = \gamma_1 = ... = \gamma_k$ and $c = c_1 = ... = c_k$ without estimating the model under the alternative. We approximate the terms $g(q_t;\gamma_k,c_k) - g(q_t;\gamma,c)$ linearly by Taylor series expansions with respect to c and γ. The approximation is located around the values of c and γ estimated under the H_0, $\hat{c}_{\text{H0}}$ and $\hat{\gamma}_{\text{H0}}$. This linearization serves two purposes. First, the parameters γ_k and c_k are not identified under the null hypothesis, and the Taylor approach side-steps this; see Luukkonen et al. (1988). Second, the approach also leads to (approximate) linear regression models, which eases the computation of the proposed test. We therefore obtain

$$y_t \approx \sum_{k=1}^{K} \beta_{0,k} x_{t,k} + \sum_{k=1}^{K} \beta_{1,k} x_{t,k} g(q_t;\gamma,c) + \sum_{k=1}^{K} \beta_{1,k} x_{t,k} \Bigg[g(q_t;\hat{\gamma}_{\text{H0}},\hat{c}_{\text{H0}}) + \frac{\partial g}{\partial \gamma_k}\bigg|_{\substack{\hat{\gamma}_{\text{H0}}\\ \hat{c}_{\text{H0}}}} (\gamma_k - \hat{\gamma}_{\text{H0}})$$
$$+ \frac{\partial g}{\partial c_k}\bigg|_{\substack{\hat{\gamma}_{\text{H0}}\\ \hat{c}_{\text{H0}}}} (c_k - \hat{c}_{\text{H0}}) - g(q_t;\hat{\gamma}_{\text{H0}},\hat{c}_{\text{H0}}) - \frac{\partial g}{\partial \gamma}\bigg|_{\substack{\hat{\gamma}_{\text{H0}}\\ \hat{c}_{\text{H0}}}} (\gamma - \hat{\gamma}_{\text{H0}}) - \frac{\partial g}{\partial c}\bigg|_{\substack{\hat{\gamma}_{\text{H0}}\\ \hat{c}_{\text{H0}}}} (c - \hat{c}_{\text{H0}}) \Bigg] + u_t. \quad (8)$$

Since we approximate both $g(q_t;\gamma_k,c_k)$ and $g(q_t;\gamma,c)$ around the same values $(\hat{\gamma}_{\text{H0}},\hat{c}_{\text{H0}})$, the terms $g(q_t;\hat{\gamma}_{\text{H0}};\hat{c}_{\text{H0}})$ above cancel out. We fill in the required derivatives of the transition function, which can be found in Appendix A, and set $w_t = \exp\left(-\hat{\gamma}_{\text{H0}}(q_t - \hat{c}_{\text{H0}})\right)$ to obtain[5]

$$\begin{aligned} y_t \approx\ & \sum_{k=1}^{K} \beta_{0,k} x_{t,k} + \sum_{k=1}^{K} \beta_{1,k} x_{t,k} g(q_t;\gamma,c) + \sum_{k=1}^{K} \beta_{1,k} x_{t,k} \Big[-\frac{w_t}{(1+w)_t^2}(-q_t + \hat{c}_{\text{H0}})(\gamma_k - \hat{\gamma}_{\text{H0}}) \\ & - \frac{w_t}{(1+w)_t^2} \hat{\gamma}_{\text{H0}} (c_k - \hat{c}_{\text{H0}}) + \frac{w_t}{(1+w)_t^2}(-q_t + \hat{c}_{\text{H0}})(\gamma - \hat{\gamma}_{\text{H0}}) \\ & + \frac{w_t}{(1+w)_t^2} \gamma_{\text{H0}} (c - \hat{c}_{\text{H0}}) \Big] + u_t, \end{aligned}$$

which can be further aggregated to

$$\begin{aligned} y_t \approx\ & \sum_{k=1}^{K} \beta_{0,k} x_{t,k} + \sum_{k=1}^{K} \beta_{1,k} x_{t,k} g_t(q_t;\gamma,c) + \sum_{k=1}^{K} \beta_{1,k} x_{t,k} \Big\{ \big[(\gamma - \hat{\gamma}_{\text{H0}}) + (\hat{\gamma}_{\text{H0}} - \gamma_k) \big] \\ & \frac{w_t}{(1+w_t)^2}(-q_t + \hat{c}_{\text{H0}}) + \big[(c - \hat{c}_{\text{H0}}) + (\hat{c}_{\text{H0}} - c_k) \big] \frac{w_t}{(1+w_t)^2} \hat{\gamma}_{\text{H0}} \Big\} + u_t. \end{aligned}$$

This leads to the auxiliary model with possibly omitted variables

$$y_t \approx \sum_{k=1}^{K} \beta_{0,k} x_{t,k} + \sum_{k=1}^{K} \beta_{1,k} x_{t,k} g(q_t;\gamma,c) + \sum_{k=1}^{K} z_{t,k}^{a} \delta_k^{a} + \sum_{k=1}^{K} z_{t,k}^{b} \delta_k^{b} + u_t$$

[5]Note that w is the same irrespectively of whether the transition function with common parameters (γ, c) or k individual parameters (γ_k, c_k) is considered. Both are evaluated at the estimated values under the H_0.

where

$$
\begin{aligned}
z^a_{t,k} &= \beta_{1,k}\frac{w_t}{(1+w)^2_t}\hat{\gamma}_{\text{H0}} x_{t,k} \\
z^b_{t,k} &= \beta_{1,k}\frac{w_t}{(1+w_t)^2}(\hat{c}_{\text{H0}} - q_t)x_{t,k} \\
\delta^a_k &= (c - c_k) \\
\delta^b_k &= (\gamma - \gamma_k).
\end{aligned}
$$

The null of homogeneity translates into $\delta^a_k = \delta^b_k = 0$, $k = 1, \ldots, K$. In fact, we may only include up to $2(K-1)$ omitted regressors $z^a_{t,k}$ and $z^b_{t,k}$ since one restriction is redundant. To see this, note that we could easily reformulate H_0 to $\gamma_1 = \gamma_K, \gamma_2 = \gamma_K, ..., \gamma_{K-1} = \gamma_K$ and $c_1 = c_K, c_2 = c_K, ..., c_{K-1} = c_K$.

Therefore, we shall test the significance of the artificial regressors $z^a_{t,k}$ and $z^b_{t,k}$ in

$$
y_t = \sum_{k=1}^{K}\beta_{0,k}x_{t,k} + \sum_{k=1}^{K}\beta_{1,k}x_{t,k}g(q_t;\gamma,c) + \sum_{k=1}^{K-1}\delta^a_k z^a_{t,k} + \sum_{k=1}^{K-1}\delta^b_k z^b_{t,k} + u_t \tag{9}
$$

Testing for equality of parameters in the transition function can be carried out by testing the joint significance of the $2(K-1)$ additional parameters δ^a_k and δ^b_k in Equation (9). The null hypothesis is thus reformulated as an omitted variable test, which is asymptotically equivalent to the LM test based on the Gaussian likelihood.

We shall test the equivalent null hypothesis $\delta^a_k = \delta^b_k = 0$ using the corresponding LM test which is computed using OLS regressions only as follows. Denoting by $\hat{u}$ the $T \times 1$ vector of residuals obtained from the model estimation under the null hypothesis, by $\hat{V}$ the matrix of gradients of the regression function at the restricted estimators, and by $\hat{Z}$ the $T \times 2(K-1)$ matrix stacking $z^a_{t,k}$ and $z^b_{t,k}$, the test can be conducted with the help of the LM auxiliary regression of $\hat{u}$ on $\hat{V}$ and $\hat{Z}$,

$$
\hat{u} = \hat{V}\tilde{\pi} + \hat{Z}\tilde{\delta} + \tilde{v},
$$

where $\tilde{\cdot}$ denotes OLS estimation. More precisely, $\hat{V} = \left[X, \hat{D}, \hat{A}_1, \hat{A}_2\right]$ with $X = [x_1; \ldots; x_K]$ where x_k are $(T \times 1)$ vectors, $\hat{D} = [x_1 g(q, \hat{\gamma}_{\text{H0}}, \hat{c}_{\text{H0}}); \ldots; x_K g(q, \hat{\gamma}_{\text{H0}}, \hat{c}_{\text{H0}})]$ and, furthermore, $\hat{A}_1 = [\sum_{k=1}^{K}\beta_{1,k}x_k(\partial\hat{g}/\partial\gamma)]$ and $\hat{A}_2 = [\sum_{k=1}^{K}\beta_{1,k}x_k(\partial\hat{g}/\partial c)]$, where $\partial\hat{g}/\partial\gamma$ stands for the vector of derivatives of g evaluated at q_t and $\hat{\gamma}_{\text{H0}}, \hat{c}_{\text{H0}}$. The matrix

$$
Z = \left[\beta_{1,1}x_1\frac{\partial\hat{g}}{\partial\gamma}; \ldots; \beta_{1,K-1}x_{K-1}\frac{\partial\hat{g}}{\partial\gamma}; \beta_{1,1}x_1\frac{\partial\hat{g}}{\partial c}; \ldots; \beta_{1,K-1}x_{K-1}\frac{\partial\hat{g}}{\partial c}\right]
$$

contains the relevant partial derivatives of the unrestricted model (i.e. the omitted variables as suggested by Equation (9)).

The hull hypothesis of homogeneity translates into the null of $\delta = 0$ in the LM auxiliary regression. E.g. the usual LM statistic is of the form $\hat{\sigma}^{-2}\hat{u}'\hat{Z}(\hat{Z}'\hat{Z}-\hat{Z}'\hat{V}(\hat{V}'\hat{V})^{-1}\hat{V}'\hat{Z})^{-1}\hat{Z}'\hat{u}$, where $\hat{\sigma}^2 = \hat{u}'\hat{u}/T$.[6] Regularity conditions assumed, the limiting null distribution of the above statistic is χ^2_{2K-2}; this is a particular instance of the panel case we address in the following subsection and do not discuss it separately. Note that this version excludes heteroskedasticity, which is more likely to appear in panels. In the panel case we shall consider more robust tests, also to cross-unit correlation.

2.2 The Panel Case

We now consider the problem of testing the homogeneity of the transition functions within each unit, while keeping homogeneity for each regressor k across the panel. We find this to be the relevant question for several reasons. Firstly, it is more likely that transition functions attached to different variables are different than transition functions attached to the same variable but in different units. Secondly, while a completely heterogenous panel is of course conceivable, it requires different estimation techniques (e.g. random coefficient models) unless N is fairly small. Therefore, it is a more reasonable approach to first decide for or against within-homogeneity, and then to deal with across-homogeneity, if necessary. We thus view our panel test as one step in specifying a panel STR model, and we leave the rest to further research.[7]

Consider therefore the following panel STR model,

$$y_{i,t} = \sum_{k=1}^{K} \beta_{0,k} x_{i,t,k} + \sum_{k=1}^{K} \beta_{1,k} x_{i,t,k} g(q_{i,t}; \gamma_k, c_k) + u_{i,t}, \tag{10}$$

for

$$t = 1, \ldots, T \text{ and } i = 1, \ldots, N.$$

The null hypothesis to be tested is the same,

$$H_0 : \gamma = \gamma_1 = ... = \gamma_k,\; c = c_1 = ... = c_k.$$

To do so, we estimate the model in (10) under the null hypothesis of homogenous transition functions and compute the $NT \times 1$ residual vector $\hat{u}$ based on the restricted

[6] The test can equivalently be performed using the following steps: 1. Estimate the model under the H_0, compute the residuals $\widehat{u}$ and the sum of squared residuals SSR_0; 2. Regress $\hat{u}$ on $X, \hat{D}, \hat{A}_1, \hat{A}_2$ and $\hat{Z}$ as defined above and compute the sum of squared residuals SSR_1; 3. Compute the LM-test, $LM = T(SSR_0 - SSR_1)/SSR_0$.

[7] See e.g. Breitung et al. (2016) and the references therein for coefficient homogeneity tests in linear panel models.

pooled panel estimator $\hat{\theta}_{\text{H0}}$. Then, for each unit $i = 1, \ldots, N$, we construct the matrices $\hat{V}_i$ and $\hat{Z}_i$ like for the single-unit case. Finally, stack $\hat{V} = \left[\hat{V}_1'; \ldots; \hat{V}_N'\right]'$, $\hat{Z} = \left[\hat{Z}_1'; \ldots; \hat{Z}_N'\right]'$, and test the above null of homogeneity in the auxiliary regression

$$\hat{u} = \hat{V}\tilde{\pi} + \hat{Z}\tilde{\delta} + \tilde{v} \tag{11}$$

by checking whether $\tilde{\delta}$ is significantly different from 0. The difference to the derivation in the previous subsection is that the model in (10) is now a pooled panel regression.

Let us now spell out the conditions under which we work. They will be considerably more general than those under which the test is derived, so we actually have a quasi-LM test.

Consider the error terms first. In panel macroeconometrics, cross-sectional dependence is a critical feature; see e.g. the overview article of Breitung (2015). We therefore make the following

Assumption 1. *Let $u_{i,t} = \lambda_i' f_t + \varepsilon_{i,t}$ where the r factors satisfy $f_t \sim iid\,(0, I_r)$ and are independent of the idiosyncratic components $\varepsilon_{i,t}$ at all times and units. The loadings λ_i are constant and uniformly bounded. Furthermore, $\varepsilon_{i,t} \sim iid\,(0, \sigma_i^2)$ are independent sequences across the panel. Finally, $\varepsilon_{i,t}$ and f_t are uniformly L_4 bounded.*

Formally, the independence conditions imply that the regression errors follow a strict factor model in the terminology of Chamberlain and Rothschild (1983). The model thus exhibits cross-sectionally dependent errors. Note however that we do not impose further restrictions on the loadings, such that, in addition to the typical strong dependence generated by classical factor models, we allow for moderate cross-dependence in the sense of Bailey et al. (2016) or even for weak cross-dependence (where most loadings are zero or close to zero). The model is in fact very general in what concerns error cross-dependence and one important result is to show that we may account for it by means of panel-robust standard errors (Beck and Katz, 1995) without having to specify the type of cross-dependence (weak, moderate, or strong). The independence assumption may be relaxed at the cost of more involved moment requirements and high-level convergence conditions; see e.g. Bai (2003). We maintain them however to somewhat simplify the proofs.

Turning our attention to the regressors, let V and Z be the counterparts of $\hat{V}$ and $\hat{Z}$ evaluated at θ_{ESTN} rather than $\hat{\theta}_{ESTN}$. We assume strict exogeneity as specified in the following

Assumption 2. *Let $x_{i,t,k}$ and $q_{i,t}$ be deterministic such that $[V; Z]$ has bounded elements and $\frac{1}{NT}[V; Z]'[V; Z] \to Q$, a positive definite matrix.*

Assuming deterministic regressors is a fast way of ensuring exogeneity. Another would have been to require the r.h.s. variables to be independent of the errors $u_{i,t}$ for all times and units, but this would have required additional moment and ergodicity conditions so we prefer the former. It should be emphasized that such strict exogeneity is not essential for the results, since the asymptotic null distribution of the proposed test is derived by means of a CLT for martingale difference arrays. Therefore, allowing for weakly exogenous regressors, say in the form of lagged dependent variables, would only have required additional moment and stability conditions (which further complicate the model). On the other hand, a dynamic panel structure would ease dealing with serial error correlation, which we excluded by means of Assumption 1; otherwise, serial correlation would have to be captured by more complex standard errors; see Driscoll and Kraay (1998).

The panel LM-type test with robustness is then computed as a significance test for the parameter vector δ,

$$\mathcal{T} = \tilde{\delta}' \left(\widehat{\mathrm{Cov}(\tilde{\delta})} \right)^{-1} \tilde{\delta}, \tag{12}$$

where the panel-robust estimator $\widehat{\mathrm{Cov}(\tilde{\delta})}$ is a Huber/Eicker/White estimator in the pooled panel regression (11). Its expression is given in the proof of the following

Proposition 1. *Under Assumptions 1 and 2, it holds as $N, T \to \infty$ that*

$$\mathcal{T} \xrightarrow{d} \chi^2_{2K-2}.$$

Proof: *See Appendix C.*

Examining the proof, it may be seen that the result holds for fixed N as well, such that the single-unit result follows as a particular case $N = 1$.

Working with the pooled panel regression from Equation 11 may be seen as suboptimal, should there be across-heterogeneity, as unaccounted heterogeneity may lead to losses of power. Simulation results reported in the following section suggest that this is not necessarily the case. Should one fear that across-heterogeneity is an issue, one may conduct the LM homogeneity test unit-wise, and then use multiple testing techniques to identify the units where heterogeneity is significant; see e.g. Simes (1986) and Benjamini and Hochberg (1995) for suitable multiple testing methods.

3 Monte Carlo Simulation

In this section, we assess the finite-sample behavior of the LM-type test against heterogenous transition functions. To this end, we report size and power simulations of the proposed test. All simulations are run in Stata.

We compare size and power for the following specification of our test: a standard test without robustification (χ^2_a), a heteroscedasticity-robust version (χ^2_b), an individual-cluster version robust against individual-level correlation (χ^2_c) (see Beck and Katz, 1995), as well as a time-cluster version robust against cross-sectional correlation (χ^2_d) (see Driscoll and Kraay, 1998). The heteroskedasticity-robust version of our test (χ^2_b) uses the Huber/Eicker/White estimator with the assumption of independent observations.[8] The time-cluster version of the test is over-specified given our assumed model from the previous section and the simulation design below; we include them to see if there are costs in terms of size or power of using more complex covariance matrix estimators than required. An outline and formulas for the robust variance estimator in the Stata context can be found in the User's Guide in StataCorp (2017b), pp. 322-327 or the Programming Reference Manual in StataCorp (2017a) pp. 550-573.

For the homogeneous null scenario, we use the following setting:

$$y_{i,t} = \beta_{0,1}x_{1,i,t} + \beta_{0,2}x_{2,i,t} + \beta_{1,1}x_{1,i,t}g_t(q_t;\gamma_1,c_1) + \beta_{1,2}x_{2,i,t}g_t(q_t;\gamma_2,c_2) + u_{i,t} \tag{13}$$

where

$$\begin{aligned}
x_{k,i,t} &\sim N(0,1)\\
q_t &\sim U(10,20)\\
c_1 &\sim U(12,18)\\
c_2 &= c_1\\
\gamma_1 &\sim U(1,3)\\
\gamma_2 &= \gamma_1\\
\beta_{0,k} &\sim U(0.1,0.5)\\
\beta_{1,k} &= -\beta_{0,k}\\
u_{i,t} &= e_{i,t}\\
e_{i,t} &\sim iidN(0,0.1)
\end{aligned}$$

With this design of c and q,[9] we ensure that the location parameter c will never hit the outer ranges of the distribution of the transition variable q which would imply just one regime. We set the transition parameter γ such that the transition occurs in a somewhat smooth manner. For the heterogeneous scenario, we consider the case of differences in γ and differences in c. The heterogeneous scenario with $c_1 \neq c_2$ deviates as follows from the

[8]We are aware of the fact that the usual sandwich estimator is invalid for fixed T, large N for fixed effects estimator as pointed by Stock and Watson (2008).

[9]Note that we generated the data with $q_{i,t} = q_t$, which is, of course, covered by our assumptions.

homogeneous scenario:

$$c_2 = \begin{cases} c_1 + 0.025c_1 & \text{if } c_1 < 15 \\ c_1 - 0.025c_1 & \text{if } c_1 \geq 15 \end{cases}$$

which ensures that the second location parameter will be lower/higher than the upper/lower bound. The scenario with $\gamma_1 \neq \gamma_2$ uses

$$\gamma_2 = 0.5\gamma_1$$

Furthermore, we consider size and power scenarios with cross-sectional correlation of the error term. The scenarios differ from the above outlined settings in the following way:

$$\begin{aligned} u_{i,t} &= \kappa_i \eta_t + e_{i,t} \\ e_{i,t} &\sim iidN(0, 0.1) \\ \eta_t &\sim iidN(0, 0.1) \\ \kappa_i &\sim U(0.3, 0.7) \end{aligned}$$

For the simulation, we estimate the parameters by nonlinear least squares (NLS) using Stata. The NLS estimator includes fixed effects because we suppose that most practitioners will include them in their panel estimation. We use a grid search procedure to find starting values for the NLS estimation. We simulate 1000 replications for each scenario, which differ by the previously outlined data generating schemes and by different combinations of N and T.

We start with the empirical size for the scenario without cross-sectional correlations. Table 1 lists the empirical sizes for our test version with different combinations of N and T. The results show that the test without any robustification (χ^2_a) performs fairly well in the case of non-correlated error terms. Only in small samples, we see a small distortion. For a larger combination of N and T, this disappears and the empirical size of the test appears to converge to the nominal one. The test versions with robustification against heteroscedasticity χ^2_b or individual-level correlation (χ^2_c) show a similar behaviour compared to the standard test. The serial correlation-robust version of the χ^2-test (χ^2_d) are undersized, at least for small N.

The results for the simulations when error cross-sectional correlation is present are listed in Table 2. The changes compared to the scenario without cross-sectional correlation are only moderate. At least in these settings, it seems that error cross-sectional correlation does not affect the empirical size of the different test versions in a significant way (this is because the regressors $x_{i,t,k}$ are not correlated across the units).

Table 1: Empirical size: H_0 without cross-sectional correlation

N	T	χ^2_a-Test	χ^2_b-Test	χ^2_c-Test	χ^2_d-Test
20	20	0.050	0.054	0.046	0.017
	50	0.052	0.047	0.041	0.038
	100	0.067	0.064	0.046	0.060
	200	0.061	0.061	0.056	0.051
50	20	0.054	0.054	0.040	0.017
	50	0.051	0.058	0.049	0.028
	100	0.057	0.056	0.048	0.051
	200	0.052	0.050	0.042	0.047
100	20	0.060	0.061	0.044	0.026
	50	0.059	0.064	0.050	0.045
	100	0.050	0.054	0.051	0.043
	200	0.056	0.054	0.052	0.055
200	20	0.062	0.061	0.051	0.022
	50	0.065	0.068	0.059	0.041
	100	0.054	0.055	0.051	0.047
	200	0.049	0.049	0.051	0.039

Table 2: Empirical size: H_0 with cross-sectional correlation

N	T	χ^2_a-Test	χ^2_b-Test	χ^2_c-Test	χ^2_d-Test
20	20	0.070	0.066	0.045	0.023
	50	0.050	0.045	0.034	0.029
	100	0.052	0.041	0.037	0.049
	200	0.054	0.054	0.043	0.055
50	20	0.063	0.060	0.051	0.026
	50	0.057	0.055	0.047	0.044
	100	0.043	0.045	0.037	0.040
	200	0.041	0.043	0.038	0.041
100	20	0.085	0.077	0.067	0.029
	50	0.052	0.050	0.047	0.048
	100	0.050	0.045	0.050	0.043
	200	0.043	0.043	0.040	0.047
200	20	0.055	0.053	0.042	0.019
	50	0.051	0.055	0.053	0.047
	100	0.053	0.053	0.051	0.048
	200	0.049	0.051	0.051	0.044

Table 3: Empirical power: differences in c, no cross-sectional correlation

N	T	χ^2_a-Test	χ^2_b-Test	χ^2_c-Test	χ^2_d-Test
20	20	0.436	0.407	0.325	0.169
	50	0.731	0.721	0.643	0.578
	100	0.844	0.852	0.807	0.812
	200	0.933	0.934	0.912	0.924
50	20	0.660	0.650	0.612	0.247
	50	0.871	0.872	0.846	0.743
	100	0.952	0.954	0.951	0.952
	200	0.991	0.992	0.991	0.989
100	20	0.811	0.805	0.794	0.360
	50	0.944	0.943	0.936	0.858
	100	0.992	0.991	0.988	0.987
	200	0.998	0.998	0.999	0.998
200	20	0.899	0.895	0.887	0.418
	50	0.984	0.986	0.985	0.927
	100	0.998	0.998	0.998	0.996
	200	1.000	1.000	1.000	1.000

Table 4: Empirical power: differences in c, with cross-sectional correlation

N	T	χ^2_a-Test	χ^2_b-Test	χ^2_c-Test	χ^2_d-Test
20	20	0.387	0.359	0.281	0.123
	50	0.651	0.638	0.570	0.524
	100	0.805	0.801	0.752	0.761
	200	0.917	0.912	0.886	0.909
50	20	0.574	0.568	0.568	0.230
	50	0.856	0.852	0.837	0.742
	100	0.944	0.944	0.942	0.933
	200	0.979	0.978	0.974	0.977
100	20	0.791	0.780	0.768	0.334
	50	0.940	0.942	0.941	0.849
	100	0.980	0.981	0.980	0.979
	200	0.997	0.998	0.997	0.998
200	20	0.896	0.902	0.883	0.405
	50	0.976	0.975	0.974	0.913
	100	0.997	0.998	0.997	0.995
	200	1.000	1.000	1.000	0.999

Table 5: Empirical power: differences in γ, no cross-sectional correlation

N	T	χ^2_a-Test	χ^2_b-Test	χ^2_c-Test	χ^2_d-Test
20	20	0.499	0.481	0.379	0.282
	50	0.765	0.757	0.695	0.704
	100	0.899	0.898	0.854	0.886
	200	0.968	0.968	0.954	0.968
50	20	0.787	0.780	0.753	0.503
	50	0.915	0.919	0.897	0.894
	100	0.985	0.985	0.985	0.983
	200	0.999	0.998	0.997	0.996
100	20	0.875	0.877	0.867	0.657
	50	0.987	0.987	0.983	0.975
	100	0.980	0.980	0.980	0.980
	200	1.000	1.000	1.000	1.000
200	20	0.955	0.954	0.947	0.767
	50	0.997	0.997	0.995	0.987
	100	1.000	1.000	1.000	1.000
	200	1.000	1.000	1.000	1.000

Table 6: Empirical power: differences in γ, with cross-sectional correlation

N	T	χ^2_a-Test	χ^2_b-Test	χ^2_c-Test	χ^2_d-Test
20	20	0.428	0.414	0.307	0.252
	50	0.742	0.730	0.658	0.677
	100	0.878	0.878	0.830	0.857
	200	0.944	0.943	0.923	0.936
50	20	0.703	0.698	0.670	0.485
	50	0.904	0.900	0.887	0.875
	100	0.971	0.970	0.967	0.962
	200	0.993	0.994	0.989	0.993
100	20	0.841	0.843	0.829	0.599
	50	0.974	0.974	0.971	0.956
	100	0.995	0.996	0.994	0.993
	200	1.000	1.000	1.000	1.000
200	20	0.939	0.937	0.933	0.752
	50	0.995	0.993	0.993	0.988
	100	1.000	1.000	1.000	1.000
	200	1.000	1.000	1.000	1.000

Empirical power is listed in Table 3. We start with differences in the location parameter c which are generally easier to detect than differences in the speed of transition. For low combinations of N and T, the test has considerable problems detecting heterogeneity in the transition functions but with increasing N or increasing T, the test reliably rejects the null hypothesis. Like for the empirical size, the three version χ^2_a, χ^2_b and χ^2_c show a somewhat similar power. The test with robustification against cross-sectional correlation (χ^2_d) needs more observations than the others to reliable reject the null but performs well for greater samples.

The results for empirical power when cross-sectional correlation is present can be found in Table 4 and are mainly unchanged compared to the scenario without cross-sectional correlation.

The power results for detecting differences in γ are listed in Tables 5 and 6. As mentioned before, differences in γ are harder to detect and have to be more pronounced to result in a rejection of the null hypothesis. A difference of 50% between γ_1 and γ_2 is not reliably detected for a low combination of N and T like $N = 20$ and $T = 20$. With increasing numbers, all versions of the test always reject the null hypothesis which holds for the Monte-Carlo scenario with and without cross-sectional correlation. Again, the χ^2_d version needs more observations to always reject the null when the alternative is true.

4 Conclusion

This paper proposes a Lagrange multiplier test for checking the homogeneity of smooth transition regressions against regressors-specific transition functions.

We consider panel versions of this procedure that are robust against cross-sectional correlation. The test statistic is regression-based and follows a χ^2 null distribution asymptotically for $N, T \to \infty$ jointly. The test is intended to serve as a model diagnostic tool: after estimation of a model with homogenous transition functions, the test allows one to double-check the homogeneity assumptions.

In Monte Carlo experiments, the four different version of the test (standard, robustification against heteroskedasticity, robustification against panel correlation, robustification against cross-sectional correlation) show good size control and satisfactory power. Due to its simple implementation, we expect our test to provide a useful tool in model evaluation against heterogeneous smooth transition regression functions.

Appendix

A Derivatives of the logistic function

Let $w = \exp(-\gamma(q-c))$. The derivatives of the (logistic) transition function are

$$\begin{aligned}\frac{\partial g}{\partial \gamma} =& -1[1+\exp\{-\gamma(q-c)\}]^{-2}\exp\{-\gamma(q-c)\}(-q+c)\\ =&(1+w)^{-2}w(q-c)\end{aligned}$$

$$\begin{aligned}\frac{\partial g}{\partial c} =& -1[1+\exp\{-\gamma(q-c)\}]^{-2}\exp\{-\gamma(q-c)\}\gamma\\ =& -(1+w)^{-2}w\gamma\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 g}{\partial \gamma^2} =&2(1+w)^{-3}w^2(q-c)^2-(1+w)^{-2}w(q-c)^2\\ =&(1+w)^{-2}w(q-c)^2\Big(2(1+w)^{-1}w-1\Big)\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 g}{\partial c^2} =&2(1+w)^{-3}w^2\gamma^2-(1+w)^{-2}w\gamma^2\\ =&(1+w)^{-2}w\gamma^2\Big(2(1+w)^{-1}w-1\Big)\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 g}{\partial c\partial \gamma} = \frac{\partial^2 g}{\partial \gamma\partial c} =& -2(1+w)^{-3}w^2(q-c)\gamma+(1+w)^{-2}w(q-c)\gamma-(1+w)^{-2}w\\ =&(1+w)^{-2}w\Big(-2(1+w)^{-1}w(q-c)\gamma+(q-c)\gamma-1\Big)\end{aligned}$$

$$\begin{aligned}\frac{\partial^3 g}{\partial \gamma^3} =&6(1+w)^{-4}w^3(q-c)^3-6(1+w)^{-3}w^2(q-c)^3+(1+w)^{-2}w(q-c)^3\\ =&(1+w)^{-2}w(q-c)^3\Big(6(1+w)^{-2}w^2-6(1+w)^{-1}w+1\Big)\end{aligned}$$

$$\frac{\partial^3 g}{\partial c^3} = -6(1+w)^{-4}w^3\gamma^3 + 6(1+w)^{-3}w^2\gamma^3 - (1+w)^{-2}w\gamma^3$$
$$= (1+w)^{-2}w\gamma^3\Big(-6(1+w)^{-2}w^2 + 6(1+w)^{-1}w - 1\Big)$$

$$\frac{\partial^3 g}{\partial\gamma^2\partial c} = \frac{\partial^3 g}{\partial\gamma\partial c\partial\gamma} = \frac{\partial^3 g}{\partial c\partial\gamma^2} = -6(1+w)^{-4}w^3(q-c)^2\gamma + 6(1+w)^{-3}w^2(q-c)^2\gamma$$
$$- 4(1+w)^{-3}w^2(q-c) - (1+w)^{-2}w(q-c)^2\gamma + 2(1+w)^{-2}w(q-c)$$

$$= (1+w)^{-2}w(q-c)\Big(-6(1+w)^{-2}w^2(q-c)\gamma + 6(1+w)^{-1}w(q-c)\gamma$$
$$- 4(1+w)^{-1}w - (q-c)\gamma + 2\Big)$$

$$\frac{\partial^3 g}{\partial\gamma\partial c^2} = \frac{\partial^3 g}{\partial c\gamma\partial c} = \frac{\partial^3 g}{\partial c^2\partial\gamma} = 6(1+w)^{-4}w^3(q-c)\gamma^2 - 6(1+w)^{-3}w^2(q-c)\gamma^2$$
$$+ (1+w)^{-2}w(q-c)\gamma^2 - 2(1+w)^{-2}w\gamma + 4(1+w)^{-3}w^2\gamma$$

$$= (1+w)^{-2}w\gamma\Big(6(1+w)^{-2}w^2(q-c)\gamma - 6(1+w)^{-1}w(q-c)\gamma$$
$$+ 4(1+w)^{-1}w + (q-c)\gamma - 2\Big)$$

B Auxiliary results

In the following, $\|\cdot\|$ denotes the Euclidean vector norm and the corresponding induced matrix norm.

Lemma 1. *Under the conditions of Proposition 1, it holds for Ω_T and $\xi_{t,T}$ defined in the proof of Proposition 1 that*

1. $\max_{t=1,\dots,T}\left\|\Omega_T^{-1/2}\xi_{t,T}\right\|$;
2. $\Omega_T^{-1}\left(\sum_{t=1}^T \xi_{t,T}\xi_{t,T}' - \Omega_T\right) \xrightarrow{p} 0.$

Proof: *See Appendix C.*

C Proofs

Proof of Lemma 1

To show item 1, write

$$\begin{aligned} P\left(\max_{t=1,\ldots,T}\left\|\Omega_T^{-1/2}\xi_{t,T}\right\|>C\right) &\leq P\left(\max_{t=1,\ldots,T}\|\xi_{t,T}\|>C\sqrt{\|\Omega_T\|}\right)\\ &= P\left(\cup_{t=1,\ldots,T}\left(\|\xi_{t,T}\|>C\sqrt{\|\Omega_T\|}\right)\right)\\ &\leq \sum_{t=1}^{T}P\left(\|\xi_{t,T}\|>C\sqrt{\|\Omega_T\|}\right)\\ &\leq \sum_{t=1}^{T}\frac{\mathrm{E}\left(\|\xi_{t,T}\|^4\right)}{C^4\left\|\Omega_T\right\|^2}\end{aligned}$$

where the Bonferroni and the generalized Markov inequalities have been used. Since

$$\xi_{t,T}=\sum_{i=1}^{N}\begin{pmatrix} v_{i,t}\\ z_{i,t}\end{pmatrix}\lambda_i f_t+\sum_{i=1}^{N}\begin{pmatrix} v_{i,t}\\ z_{i,t}\end{pmatrix}\varepsilon_{i,t},$$

we note that the factor structure of $u_{i,t}$ implies under Assumption 1 that

$$\mathrm{Cov}\left(\xi_{t,T}\right)=\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i'\lambda_j\begin{pmatrix} v_{i,t}\\ z_{i,t}\end{pmatrix}\left(v_{j,t}';z_{j,t}'\right)+\sum_{i=1}^{N}\begin{pmatrix} v_{i,t}\\ z_{i,t}\end{pmatrix}\left(v_{i,t}';z_{i,t}'\right)\sigma_i^2.$$

Therefore, one may find under Assumption 2 positive constants C_1 and C_2 such that $\|\mathrm{Cov}\left(\xi_{t,T}\right)\|\sim C_1\left(\sum_{i=1}^N\lambda_i\right)'\left(\sum_{i=1}^N\lambda_i\right)+C_2N$ and, thanks to the serial independence of the errors $u_{i,t}$, $\Omega_T=\sum_{t=1}^T\mathrm{Cov}\left(\xi_{t,T}\right)\sim C_1T\left(\sum_{i=1}^N\lambda_i\right)'\left(\sum_{i=1}^N\lambda_i\right)+C_2NT$. Moreover, due to the same factor structure, $\mathrm{E}\left(\|\xi_{t,T}\|^4\right)\leq\|\mathrm{Cov}\left(\xi_{t,T}\right)\|^2\leq C_1^*\left(\left(\sum_{i=1}^N\lambda_i\right)'\left(\sum_{i=1}^N\lambda_i\right)\right)^2+C_2^*N^2$. such that

$$P\left(\max_{t=1,\ldots,T}\left\|\Omega_T^{-1/2}\xi_{t,T}\right\|>C\right)=O\left(\frac{T\left(\left(\left(\sum_{i=1}^N\lambda_i\right)'\left(\sum_{i=1}^N\lambda_i\right)\right)^2+N^2\right)}{C^4T^2\left(\left(\left(\sum_{i=1}^N\lambda_i\right)'\left(\sum_{i=1}^N\lambda_i\right)\right)^2+N^2\right)}\right)=o\left(1\right)$$

as required, since this holds for any choice of C.

To show item 2, it suffices to establish that, elementwise,

$$\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\begin{pmatrix} v_{i,t} \\ z_{i,t}\end{pmatrix}\left(v'_{j,t}; z'_{j,t}\right)\left(u_{i,t}u_{j,t} - \mathrm{E}\left(u_{i,t}u_{j,t}\right)\right) = o_p\left(\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\begin{pmatrix} v_{i,t} \\ z_{i,t}\end{pmatrix}\left(v'_{j,t}; z'_{j,t}\right)\mathrm{E}\left(u_{i,t}u_{j,t}\right)\right);$$

this is a straightforward exercise given that $(u_{1,t}, \ldots, u_{N,t})'$ has a factor structure whose components are iid sequences with finite 4th order moments, and that $\left(v'_{j,t}; z'_{j,t}\right)'$ are deterministic and bounded under Assumption 2, and we omit the details.

Proof of Proposition 1

We provide the proof for the case with common intercept (which is allowed to vary across two regimes like the other regressors) to save space; the extension to fixed-effects is straightforward given that we work with deterministic regressors (up to a negligible term) and we omit the details.

Begin by noting that, under the null,

$$\hat{\theta}_{\text{H0}} - \theta_0 \approx \left(V'V\right)^{-1} V'u$$

where u stacks the NT panel errors and R_{1T} is negligible. This is a standard application of optimization estimator theory and we omit the details to save space. Moreover, we linearize the regression function to obtain under the null

$$\begin{aligned} \hat{u} - u &\approx V\left(\hat{\theta}_{\text{H0}} - \theta\right) \\ &\approx V\left(V'V\right)^{-1} V'u. \end{aligned}$$

As a consequence, we have, up to a negligible term,

$$\begin{aligned} Z'\hat{u} &= Z'u + Z'V\left(V'V\right)^{-1} V'u \\ &= \left(-Z'V\left(V'V\right)^{-1}; I\right)\begin{pmatrix} V'u \\ Z'u \end{pmatrix} \end{aligned}$$

Then, the statistic of interest is based on the OLS estimators of the auxiliary regression, for which we obtain using the Frisch-Waugh-Lovell theorem

$$\tilde{\delta} = \left(Z'Z - Z'V\left(V'V\right)^{-1} V'Z\right)^{-1} Z'\hat{u}$$

(given that $V'\hat{u} = 0$ are the f.o.c. for the restricted estimation step). The sandwich covariance matrix estimator is given by

$$\widehat{\mathrm{Cov}(\tilde{\delta})} = \left(Z'Z - Z'V\left(V'V\right)^{-1}V'Z\right)^{-1} M_T \left(Z'Z - Z'V\left(V'V\right)^{-1}V'Z\right)^{-1}$$

where

$$M_T = \left(I; -Z'V\left(V'V\right)^{-1}\right)\hat{\Omega}\left(I; -Z'V\left(V'V\right)^{-1}\right)'$$

and $\hat{\Omega}$ is an estimate of the covariance matrix of $\begin{pmatrix} V'u \\ Z'u \end{pmatrix}$. We note that this is (up to normalization) the expression for panel-robust covariance matrix estimators. The estimator $\hat{\Omega}$ is constructed such that cross-sectional correlation is accounted for (Beck and Katz, 1995),

$$\hat{\Omega} = \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\begin{pmatrix} v_{i,t} \\ z_{i,t} \end{pmatrix}\left(v'_{j,t}; z'_{j,t}\right)\hat{u}_{i,t}\hat{u}_{j,t}$$

where $z'_{j,t}$ and $v'_{j,t}$ are the lines of Z_i and V_i.

Therefore, the behavior of the test can be reduced using a standard linear transformation involving the relevant restrictions to the behavior of the score under the null, i.e. of

$$\sum_{t=1}^{T}\sum_{i=1}^{N}\begin{pmatrix} v_{i,t} \\ z_{i,t} \end{pmatrix} u_{i,t} := \sum_{t=1}^{T}\xi_{t,T}$$

with $\xi_{t,T} = \sum_{i=1}^{N}\begin{pmatrix} v_{i,t} \\ z_{i,t} \end{pmatrix} u_{i,t}$. The result is then obtained via two steps. First, we show that the score $\sum_{t=1}^{T}\xi_{t,T}$, properly standardized, follows a $4K$-dimensional multivariate normal distribution with identity covariance matrix. Second, we show that panel-robust covariance matrix estimation in the LM auxiliary regression deliver the proper standardization in the limit.

To this end, let first

$$\Omega_T = \sum_{t=1}^{T}\left(\sum_{i=1}^{N}\begin{pmatrix} v_{i,t} \\ z_{i,t} \end{pmatrix}\left(v'_{i,t}; z'_{i,t}\right)\sigma_i^2 + \sum_{i=1}^{N}\sum_{j=1}^{N}\begin{pmatrix} v_{i,t} \\ z_{i,t} \end{pmatrix}\left(v'_{i,t}; z'_{i,t}\right)\lambda'_i\lambda_j\right),$$

which is the covariance matrix of $\sum_{t=1}^{T}\xi_{t,T}$. Note that $\|\Omega_T\| \sim C_1 T\left(\sum_{i=1}^{N}\lambda_i\right)'\left(\sum_{i=1}^{N}\lambda_i\right) + C_2 NT$ for suitable constants C_1 and C_2; see the proof of Lemma 1. This implies in turn

$$\hat{\theta}_{\text{H0}} - \theta_0 = O_p\left(\frac{1}{N\sqrt{T}}\max\left\{\sqrt{\left(\sum_{i=1}^{N}\lambda_i\right)'\left(\sum_{i=1}^{N}\lambda_i\right)}; \sqrt{N}\right\}\right).$$

We now establish that

$$\Omega_T^{-1/2}\sum_{t=1}^{T}\xi_{t,T}\xrightarrow{d}\mathcal{N}\left(0,I_{4K}\right) \tag{14}$$

by applying a central limit theorem for martingale difference arrays (see Theorem 24.3 in Davidson, 1994, and the Wold-Cramér device). To this end, note that $\Omega_T^{-1/2}\xi_{t,T}$ build the lines of a martingale difference array thanks to the serial independence of the errors $u_{i,t}$ given $x_{i,t,k}$ and $q_{i,t}$, and we only need to check

1. that $\max_{t=1,\ldots,T}\left\|\Omega_T^{-1/2}\xi_{t,T}\right\|$ vanishes in probability, and
2. that $\sum_{t=1}^{T}\Omega_T^{-1/2}\xi_{t,T}\left(\Omega_T^{-1/2}\xi_{t,T}\right)'\xrightarrow{p}I_{4k}$.

The first condition is proved in Lemma 1, which also establishes that $\Omega_T^{-1}\left(\sum_{t=1}^{T}\xi_{t,T}\xi_{t,T}'-\Omega_T\right)\xrightarrow{p}0$, which implies the second condition.

Turning our attention to the panel-robust covariance matrix estimator, it is then tedious, yet straightforward, to show that

$$\Omega_T^{-1}\hat{\Omega}_T=\Omega_T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\begin{pmatrix}v_{i,t}\\ z_{i,t}\end{pmatrix}\left(v_{i,t}';z_{i,t}'\right)u_{i,t}u_{j,t}+o_p\left(1\right):=\Omega_T^{-1}\tilde{\Omega}_T+o_p\left(1\right).$$

Having shown in Lemma 1 that $\Omega_T^{-1}\left(\tilde{\Omega}_T-\Omega_T\right)=\Omega_T^{-1}\left(\sum_{t=1}^{T}\boldsymbol{\xi}_{t,T}\boldsymbol{\xi}_{t,T}'-\Omega_T\right)\xrightarrow{p}0$, it follows immediately that

$$\Omega_T^{-1}\tilde{\Omega}_T\xrightarrow{p}I_{4K},$$

which, together with the convergence in 14 implies that

$$\hat{\Omega}_T^{-1/2}\sum_{t=1}^{T}\boldsymbol{\xi}_{t,T}\xrightarrow{d}\mathcal{N}\left(\mathbf{0},I_{4K}\right)$$

thanks to Slutsky's theorem. Given the condition that $\frac{1}{NT}\left[V;Z\right]'\left[V;Z\right]\to Q$, analogous results hold for linear combinations of $\sum_{t=1}^{T}\boldsymbol{\xi}_{t,T}$ and the corresponding covariance matrix estimator, and the result follows.

References

Anderson, H. M. and Vahid, F. (1998). Testing multiple equation systems for common nonlinear components. *Journal of Econometrics*, 84(1):1 – 36.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bailey, N., Kapetanios, G., and Pesaran, M. H. (2016). Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics*, 31(6):929–960.

Beck, N. and Katz, J. N. (1995). What to do (and not to do) with time series cross-section data. *American Political Science Review*, 89(3):634–647.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, 57(1):289–300.

Breitung, J. (2015). The analysis of macroeconomic panel data. In Baltagi, B. H., editor, *The Oxford Handbook of Panel Data*, chapter 15, pages 453–492.

Breitung, J., Roling, C., and Salish, N. (2016). Lagrange multiplier type tests for slope homogeneity in panel data models. *The Econometrics Journal*, 19(2):166–202.

Camacho, M. (2004). Vector smooth transition regression models for US GDP and the composite index of leading indicators. *Journal of Forecasting*, 23(3):173–196.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304.

Davidson, J. (1994). *Stochastic Limit Theory.* Oxford University Press.

de Dios Tena, J. and Tremayne, A. (2009). Modelling monetary transmission in UK manufacturing industry. *Economic Modelling*, 26(5):1053 – 1066.

Driscoll, J. C. and Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics*, 80(4):549–560.

Eitrheim, O. and Teräsvirta, T. (1996). Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics*, 74(1):59–75.

Engle, R. F. (1984). Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. In Griliches, Z. and Intriligator, M. D., editors, *Handbook of Econometrics*, volume 2 of *Handbook of Econometrics*, chapter 13, pages 775–826. Elsevier.

González, A. and Teräsvirta, T. (2006). Simulation-based finite sample linearity test against smooth transition models. *Oxford Bulletin of Economics and Statistics*, 68:797–812.

González, A., Teräsvirta, T., and van Dijk, D. (2005). Panel smooth transition regression models. Working Paper Series in Economics and Finance 604, Stockholm School of Economics.

Hansen, B. (1999). Testing for linearity. *Journal of Economic Surveys*, 13(5):551–576.

Hansen, B. E. (1996). Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis. *Econometrica*, 64(2):413–30.

Leppin, J. S. and Reitz, S. (2016). The Role of a Changing Market Environment for Credit Default Swap Pricing. *International Journal of Finance and Economics*, 21(3):209–23.

Lundbergh, S. and Teräsvirta, T. (1998). Modelling economic high-frequency time series with STAR-STGARCH models. SSE/EFI Working Paper Series in Economics and Finance 291, Stockholm School of Economics.

Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, 75(3):491–499.

Schleer, F. and Semmler, W. (2015). Financial sector and output dynamics in the euro area: Non-linearities reconsidered. *Journal of Macroeconomics*, 46:235 – 263.

Simes, R. J. (1986). An improved bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754.

StataCorp (2017a). *Stata 15 Programming Reference Manual*. StataCorp, College Station, TX: Stata Press.

StataCorp (2017b). *Stata 15 User's Guide*. StataCorp, College Station, TX: Stata Press.

Stock, J. H. and Watson, M. W. (2008). Heteroskedasticity-robust standard errors for fixed effects panel data regression. *Econometrica*, 76(1):155–174.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89(425):208–218.

Teräsvirta, T. and Yang, Y. (2014). Linearity and Misspecification Tests for Vector Smooth Transition Regression Models. CREATES Research Papers 2014-04, School of Economics and Management, University of Aarhus.

van Dijk, D., Teräsvirta, T., and Franses, P. H. (2002). Smooth Transition Autoregressive Models – A Survey Of Recent Developments. *Econometric Reviews*, 21(1):1–47.