Panel Smooth Transition Regression model and an application to investment under credit constraints

Andrés González
Stockholm School of Economics
Banco de la República de Colombia

Timo Teräsvirta
Stockholm School of Economics

Dick van Dijk
Econometric Institute
Erasmus University Rotterdam

Work in progress

November 2004

Abstract

We develop a non-dynamic panel smooth transition regression model with individual fixed effects. In this model the parameters are allowed to change smoothly as a function of an exogenous variable. The model can be viewed as an alternative to the threshold panel model by Hansen (1999a). We extend the modelling strategy for smooth transition models to the panel framework. Tests for parameter constancy and no remaining nonlinearity are proposed. Small-sample properties of these tests are investigated by simulation. The results indicate that the proposed tests are applicable in small samples. A panel smooth transition model is specified and fitted to the same data set as the one used in Hansen (1999a) who investigated whether financial constraints affect firms' investment decisions. The estimated model is evaluated and the findings discussed.

Keywords: Smooth transition models; Panel data; Misspecification test; Liquidity Constraints.

JEL Classification Codes: C12, C13 C23, C52.

Acknowledgements: This research has been supported by Jan Wallanders and Tom Hedelius's Foundation Grants No. J99/37 and J02-35. A part of the work for the paper was carried out when the first author was a visiting graduate student at the Department of Economics, Université de Toulouse I, whose kind hospitality is gratefully acknowledged. Material from this paper has been presented at ESF-EMM Network Annual Meeting, Rome November 2003, Common Features in Maastricht, December 2003, the 11$^{\text{th}}$ International Panel Data Conference, College Station, TX, June 2204 and EEA-ESEM, Madrid 2004.

Stockholm School of Economics, P.O. Box 6501, SE-113 83 Stockholm, Sweden

1 Introduction

A standard assumption in panel models is that regression functions are identical across all observations in a sample. In some applications this assumption may be violated. Hansen (1999a) suggested an alternative in which the individual observations are divided into classes according to an observable variable. If the regression functions are the same for all observations in a class and this holds for all classes, this assumption leads to a panel threshold regression (PTR) model. Hansen (1999a) derives econometric techniques for the PTR model. These include maximum likelihood estimation, asymptotic confidence intervals for the parameters as well as specification tests for determining the number of classes. The last problem leads to nonstandard inference because of an identification problem present in the tests. The PTR approach is motivated by an empirical example, in which there exists theory suggesting that the regression functions may not be identical across samples. In particular, the economic theory suggests that in the case of imperfect information, external finance may be limited, and already heavily indebted firms may have to use their cash flow to finance their investments. This separates them from other firms whose access to external sources of financing is not restricted. A leading article making use of this classification of firms is Fazzari, Hubbard, and Petersen (1988).

At the end of his article, Hansen (1999a) points out that one could also apply models with a smooth transition to this problem. In that case, instead of a small finite number of classes there would be a smooth transition controlled by an observable variable from one extreme regime to another. In the application to investment financing, that would mean that the degree of indebtedness would have a more subtle effect on the availability of external financing than a PTR model would allow. In this paper we introduce a nondynamic fixed effect panel model in which the regression coefficients are allowed to change smoothly as a function of an exogenous variable. In this sense, the paper offers an alternative methodology to Hansen (1999a).

This paper is organized as follows: The next section introduces the panel smooth transition model. The third section discusses estimation and model building. Section 4 contains small-sample properties of the specification procedure. Section 5 contains an illustration of the proposed methodology where it is applied to the economic problem analyzed by Hansen (1999a). Section 6 concludes.

2 Panel smooth transition regression model

The Panel Smooth Transition Regression (PSTR) model is a fixed effect model with exogenous regressors. The basic PSTR model with a single transition is defined as follows:

$$y_{it} = \mu_i + \beta_1' x_{it} + \beta_2' x_{it} g\left(q_{it}; \gamma, \mathbf{c}\right) + u_{it} \tag{1}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The dependent variable, y_{it}, is a scalar, μ_i is an unobservable time-invariant regressor, x_{it} is a $k-$dimensional vector of time-varying exogenous variables[1], q_{it} is an observable transition variable and u_{it} are the errors.

The main feature of this model is the transition function $g\left(q_{it}; \gamma, \mathbf{c}\right)$. It is a continuous and bounded function of q_{it} that allows the parameter in (1) to change smoothly as a function of

[1]Lagged values of the dependent variable are not allowed because the presence of the fixed effect would invalidate the use of the within transformation to handle the nuisance parameters; see Chamberlain (1984, p 1256).

q_{it}. In this work we follow Granger and Teräsvirta (1993), Teräsvirta (1994) and Jansen and Teräsvirta (1996) and define

$$g\left(q_{it};\gamma,\mathbf{c}\right)=\left(1+\exp\left(-\gamma\prod_{j=1}^{m}(q_{it}-c_j)\right)\right)^{-1},\gamma>0,c_1\leq,\ldots,\leq,c_m \tag{2}$$

where $\mathbf{c}=(c_1,\ldots,c_m)'$ is an m-dimensional vector of location parameters, and $\gamma\geq 0$ and $c_1\leq\cdots\leq c_m$ are identification restrictions. Parameter γ determines the slope of the transition function. When $m=1$ and $\gamma\to\infty$, (1) and (2) define the two-regime PTR model in Hansen (1999a). When $m>1$ and $\gamma\to\infty$, the number of identical regimes remains two, but the function switches between zero and one at $c_1,\ldots,c_m$. Finally, when $\gamma\to 0$, the transition function (2) becomes constant and the model is the standard linear model with fixed effects.

The transition function (2) with $m=1$ or $m=2$ is already a very flexible parametrization since it allows different types of changes in the parameters. For example, if $m=2$, $c_1=c_2=c$, (2) implies that only the Euclidean distance between q_{it} and c has an effect on y_{it}. Moreover, if $\gamma\to\infty$, transition function (2) defines a three-regime model whose outer regimes are identical and different from the mid-regime. Finally, when $m=1$, the model allows a single monotonic smooth transition whose location is controlled by c_1.

A possible generalization of the PSTR model is the general additive PSTR model

$$y_{it}=\mu_i+\beta_0'x_{it}+\sum_{j=1}^{r}\beta_j'x_{it}g_j(q_{it}^{(j)};\gamma_j,\mathbf{c}_j)+u_{it} \tag{3}$$

where the transition functions are of type (2). If $m=1$ for g_j, $j=1,\ldots,r$, $q_{it}^{(j)}\equiv q_{it}$ and $\gamma_j\to\infty$, $j=1,\ldots,r$, (3) collapses into an $(r+1)$-regime PTR model of Hansen (1999a). Consequently, the general additive PSTR model can be used as an alternative to multiple-regime PTR model. Additionally, when the larger model the investigator is willing to consider is a PSTR model (1) with $r=1$ and $m=1$ or $m=2$, model (3) can serve as an alternative in the evaluation of the estimated PSTR model (1). This possibility will be discussed in Section 3.4.2

3 Building panel smooth transition regression models

3.1 Modelling cycle

The PSTR model is a nonlinear model, and its use requires a systematic modelling strategy. Hansen (1999a) outlines a modelling cycle for the PTR model that consists of testing linearity and selecting the number of regimes using statistical tests. The latter stage also implies maximum likelihood estimation of the parameters of at least two PTR models. Hansen assumes that the threshold variable is given, but if it were not, his procedure could probably be extended to include the possibility of choosing it from a set of candidate variables as in Hansen (1999b).

In this paper, we consider a modelling cycle for PSTR models consisting of specification, estimation and evaluation stages. Specification includes testing linearity and, if it is rejected, determining the form of the transition function (2), that is, choosing between $m=1$ and $m=2$. At the evaluation stage the estimated model is subjected to misspecification tests to check whether or not it can be considered an adequate description of the data. The null hypotheses to be tested include parameter constancy, no remaining nonlinearity and no autocorrelation in the errors.

A similar cycle has been previously suggested for smooth transition autoregressive (STAR) and smooth transition regression (STR) models; see, for example, Teräsvirta (1998) or van Dijk, Teräsvirta, and Franses (2002) for description and discussion. It has inspired the techniques developed in the present work.

3.2 Testing linearity against PSTR model

The first step of the specification stage is to test linearity against PSTR. This is important for both statistical and economic reasons. Statistically, the PSTR model is not identified if the data-generating process is linear, and a linearity test is necessary to avoid the estimation of unidentified models. The PTR model has the same property. From the economics point of view, a linearity test may account for testing some economic theory suggestions. For instance, in the example on the access of firms to external financing, established theory suggests a linear model, whereas a nonlinear model is required if there are credit restrictions that depend on the degree of indebtedness of the firm.

Testing linearity in the PSTR model (1) can be carried out in two ways either by testing $\mathrm{H}_0^1 : \beta_2 = 0$ or $\mathrm{H}_0^2 : \gamma = 0$. In both cases the test will be nonstandard because under either null hypothesis, the PSTR model contains unidentified nuisance parameters. In particular, $(\gamma, \mathbf{c}')$ are not identified under H_0^1 and $(\beta_2', \mathbf{c}')$ under H_0^2. The testing problem when unidentified nuisance parameters are presented under the null was first studied by Davies (1977, 1987). Luukkonen, Saikkonen, and Teräsvirta (1988), Andrews and Ploberger (1994) and Hansen (1996) proposed alternative solutions to the problem. Recently, Hansen (1999a, 2000) applied his testing procedure in the PTR framework. We follow Luukkonen, Saikkonen, and Teräsvirta (1988) and test the linearity hypothesis as $\mathrm{H}_0 : \gamma = 0$. To circumvent the identification problem, we replace $g\left(q_{it}; \gamma, \mathbf{c}\right)$ by its first-order Taylor expansion around $\gamma = 0$ and test an equivalent hypothesis in an auxiliary regression. After replacing $g(q_{it}; \gamma, \mathbf{c})$ in (1) by its Taylor expansion and merging terms we obtain the following auxiliary regression,

$$y_{it} = \mu_i + \beta_1'^{*} x_{it} + \beta_2'^{*} x_{it} q_{it} + \cdots + \beta_m'^{*} x_{it} q_{it}^m + u_{it}^* \tag{4}$$

where the parameter vectors β_2^*,. . . ,β_m^* are multiples of γ and $r_{it} = u_{it} + O\left(\gamma^m\right) \beta_2' x_{it}$. Testing $\mathrm{H}_0 : \gamma = 0$ in (1) is equivalent to testing $\mathrm{H}_0^* : \beta_2^* = \cdots = \beta_m^* = 0$ in (4). Note that under the null hypothesis $\{u_{it}^*\} = \{u_{it}\}$, so the Taylor series approximation does not affect the asymptotic distribution theory.

We make the following assumptions about model (1) under the null hypothesis:

Assumption L1: For each t, $\{y_{it}, x_{it}, q_{it}\}$ are independently distributed (i.d.) across i.

Assumption L2: For each i, u_{it} is i.i.d over t and independent of $\{(x_{it}, q_{it})_{t=1}^T\}$, and $\mathsf{E}(u_{it}) = 0$.

Assumption L3: $\mathsf{E}|x_{its} q_{it}^m|^{1+\delta} \leq \Delta_1 < \infty$, for $i = 1, \ldots, N$, $t = 1, \ldots, T$, $s = 1, \ldots, k$, where $\delta > 1$.

Assumption L4: $\mathsf{E}|x_{its} q_{it}^m u_{it}|^{2+\delta} \leq \Delta_2 < \infty$ for $i = 1, \ldots, N$, $t = 1, \ldots, T$, $s = 1, \ldots, k$, where $\delta > 1$.

Theorem 1 *If assumptions L1 to L4 are satisfied, then the least squared estimator $\hat{\beta}$ of $\beta = (\beta_1^*, \ldots, \beta_m^*)'$ is consistent and asymptotically normal under the null hypothesis when $N \to \infty$ with T fixed.*

Proof. See Appendix A.1 ■

Even though the null hypothesis H_0^* can be tested using any of the three classical tests, we restrict ourselves to the LM test because it only requires the estimation of (4) under the null. The computation of the LM statistic involves two steps. First, eliminate the fixed effect from (4). Second, compute the LM statistic for the transformed model. The LM test and its F-version can be computed in three stages as follows:

1. Regress $\tilde{y}_{it} = y_{it} - \sum_t y_{it}/T$ on $\tilde{x}_{it} = x_{it} - \sum_t x_{it}/T$ and compute the sum of squared residuals SSR_0.

2. Regress $\tilde{y}_{it}$ on $\tilde{x}_{it}$ and $(x_{it}'q_{it} - \sum_t x_{it}'q_{it}/T, \ldots, x_{it}'q_{it}^m - \sum_t x_{it}'q_{it}^m/T)$ and compute the sum of squared residuals SSR_1.

3. Compute,

$$\mathrm{LM} = TN(\mathrm{SSR}_0 - \mathrm{SSR}_1)/\mathrm{SSR}_0 \tag{5}$$

$$\mathrm{LM_F} = \{(\mathrm{SSR}_0 - \mathrm{SSR}_1)/mk\} \,/\, \{\mathrm{SSR}_1/(TN - N - mk)\} \tag{6}$$

   Under the null hypothesis, statistic (5) is asymptotically distributed as χ^2_{mk} and the F-statistic (6) has an approximate $\mathrm{F}\,[mk, TN - N - mk]$ distribution.

Suppose that the larger model the investigator is willing to consider is the PSTR model (1) with $m = 1$ or 2 in (2). The linearity test can then be used to choose between $m = 1$ and $m = 2$. Granger and Teräsvirta (1993) and Teräsvirta (1994) proposed the use of a sequence of linearity tests for determining m. The testing sequence applied to the present situation is the following: Using the auxiliary regression (4) with $m = 3$, test the null hypothesis $\mathrm{H}_0 : \beta_3 = \beta_2 = \beta_1 = 0$. If it is rejected, test $\mathrm{H}_{04} : \beta_3 = 0$, $\mathrm{H}_{03} : \beta_2 = 0|\beta_3 = 0$ and $\mathrm{H}_{02} : \beta_1 = 0|\beta_3 = \beta_2 = 0$. Select $m = 2$ if the rejection of H_{03} is the strongest one, and otherwise select $m = 1$. For the reasoning behind this rule, see Teräsvirta (1994).

3.3 Estimation of parameters

Estimation of parameters of the PSTR model (1) is a relatively straightforward application of the fixed effect estimator and nonlinear least squares [NLS]. One has to eliminate the individual effects μ_i by removing individual-specific means and to apply NLS to the transformed model to estimate the remaining parameters. This estimating procedure can be seen as maximum likelihood where first the likelihood function is concentrated with respect to the fixed effects.

In order to discuss the asymptotic properties of the ML estimator we write (1) for individual i as,

$$Y_i = \iota\mu_i + X_{1i}\beta_1 + W_i(\beta_3)\beta_2 + U_i \tag{2.1'}$$

where ι is a $(T \times 1)$ vector of ones, $X_{1i} = (x_{i1}', \ldots, x_{iT}')'$ and $W_i(\beta_3) = (g(X_{2i}\beta_3) \odot X_{1i})$ with $X_{2i} = (\iota', q_i', \ldots, q_i'^m)'$ and $\beta_3 = \gamma(1, c_1^*, \ldots, c_m^*)'$. The dimensions of X_{1i}, X_{2i} are $(T \times k)$, $(T \times (m+1))$, respectively.

We make the following assumptions for the PSTR model (1) or (2.1'):

Assumption E1: $\{Y_i, X_{1i}, q_i\}$ is an independently identically distributed sequence of random variables and $U_i = Y_i - \mathsf{E}[Y_i|\mu_i, X_{1i}, q_i]$.

Assumption E2: $g(x'_{2it}\beta_3)\beta_2 - g(x'_{2it}\beta_3^0)\beta_2^0 \neq 0$ when $\beta_2 \neq \beta_2^0$ and/or $\beta_3 \neq \beta_3^0$.

Assumption E3: The parameter space Θ is a compact subset of $\mathbb{R}^K$ and $\beta^0 \in \Theta$.

Assumption E4: $\mathsf{E}[|u_{it}^2|^2] \leq \Delta_1 < \infty$ for $i = 1, \ldots, n$, $t = 1, \ldots, T$.

Assumption E5: $\mathsf{E}[|x_{2it,j}x_{2is,h}x_{is,r}x_{it,l}|^2] \leq \Delta < \infty$, for $j, h = 1, \ldots, m$, $r, l = 1, \ldots, k$ and $i = 1, \ldots, N$, $t = 1, \ldots, T$.

Assumption E6: $\mathsf{E}[|x_{jit,h}|^2] \leq \Delta < \infty$, for $j = 1, 2$, $i = 1, \ldots, N$, $h = 1, \ldots, k$ and $t = 1, \ldots, T$.

Assumption E7: $V \equiv \mathsf{E}\left[[X_{1i} \vdots W_i(\beta_3^0)]'Q_T[X_{1i} \vdots W_i(\beta_3^0)]\right]$ is positive definite. $Q_T = I_T - \frac{1}{T}\iota\iota'$ is the within transformation matrix.

Theorem 2 *If assumptions (E1) to (E7) are satisfied, the maximum likelihood estimator is consistent and asymptotically normal when $N \to \infty$ and T is fixed.*

Proof. See Appendix A.2 ■

The only assumption that is not standard is (E2) which is an identification assumption, assumptions (E1), (E3) to (E7) are standard in linear panel models with strictly exogenous regressor. Even though we have assumed that the observations across individuals are i.i.d it could be relaxed in order to allow for heterogeneity. Such a generalization would imply the existence of higher-order moments. [see White (1980) and White (2000) for details].

As mentioned before, the estimation of the parameters in (1) is carried out in two steps. First, we eliminate the fixed effects and then apply NLS to the transformed model. Even though the first step is standard in linear models, equation (1) calls for a more careful treatment. Specifically, note that the individual means in (1) have the form

$$\bar{y}_i = \mu_i + \beta_1\bar{x}_i + \beta_2\overline{w}_i(\gamma, \mathbf{c}) + \bar{u}_i \tag{7}$$

where $\bar{y}_i$, $\bar{x}_i$, $\bar{w}_i$ and $\bar{u}_i$ are individual means. Subtracting equation (7) from equation (1) yields

$$\tilde{y}_{it} = \beta'\tilde{x}_{it}^*(\gamma, \mathbf{c}) + \tilde{u}_{it} \tag{8}$$

where $\tilde{y}_{it} = y_{it} - \bar{y}_i$, $\tilde{u}_{it} = u_{it} - \bar{u}_i$, $\beta = (\beta_1', \beta_2')'$, $\tilde{x}_{it}^*(\gamma, \mathbf{c}) = (x'_{it} - \bar{x}'_i, x'_{it}g(q_{it}; \gamma, \mathbf{c}) - \bar{w}'_i(\gamma, \mathbf{c}))'$. Consequently, the transformed vector $\tilde{x}_{it}^*(\gamma, \mathbf{c})$ in (8) depends on $(\gamma, \mathbf{c}')'$ through both the levels and the individual means. For this reason, $\tilde{x}_{it}^*(\gamma, \mathbf{c})$ has to be recomputed at each iteration.

The iterations have the following form. First, given $\left(\gamma^{(j)}, \mathbf{c}^{(j)\prime}\right)'$ estimate $\beta^{(j)}$ by ordinary least squares, which yields

$$\hat{\beta}^{(j)} = \left(\sum_i\sum_t \tilde{x}_{it}^*\left(\gamma^{(j)}, \mathbf{c}^{(j)}\right)\tilde{x}_{it}^{*\prime}\left(\gamma^{(j)}, \mathbf{c}^{(j)}\right)\right)^{-1}\sum_i\sum_t \tilde{x}_{it}^*\left(\gamma^{(j)}, \mathbf{c}^{(j)}\right)y_{it}^* \tag{9}$$

Then, conditionally on $\beta^{(j)}$, estimate $\left(\gamma^{(j+1)}, \mathbf{c}'^{(j+1)}\right)'$ by NLS. This amounts to solving

$$\left(\hat{\gamma}^{(j+1)}, \hat{\mathbf{c}}'^{(j+1)}\right)' = \arg\min_{(\gamma,\mathbf{c})}\sum_{i=1}\sum_{t=1}\left(\bar{y}_{it} - \beta'^{(j)}\tilde{x}_{it}^*(\gamma, \mathbf{c})\right)^2 \tag{10}$$

Leybourne, Newbold, and Vougas (1998) proposed a similar procedure for STAR models; see also Teräsvirta (1998) and van Dijk, Teräsvirta, and Franses (2002) for discussion. Small-sample properties of this procedure are investigated by simulation in Section 4.1.

An issue that deserves special attention in the estimation of the PSTR model is the selection of starting-values. Good starting values may considerably facilitate the numerical optimization or, conversely, inappropriate starting-values may cause problems. A feasible method for smooth transition models is a grid search. It is seen that (8) is linear in parameters when parameters $(\gamma, \mathbf{c}')'$ are fixed. This suggest the following algorithm. First, define an array of values for $(\gamma, \mathbf{c}')'$ such that $\gamma > 0$, and $c_{j\max} < \max\{q_{it}\}$ and $c_{j\min} > \min\{q_{it}\}$, $j = 1, \ldots, m$. Calculate (9) for all these values in turn and select the vector $(\gamma^*, \mathbf{c}'^*)'$ minimizing the sum of squared residuals as starting-values of the estimation algorithm. Hansen (1999a) also applied a form of grid search in the estimation of the parameters of PTR model.

3.4 Evaluation of the estimated model

After estimating the parameters, the estimated PSTR model has to be evaluated. In this section we consider a number of misspecification tests for this purpose. One of them, the test of no remaining nonlinearity, may also be viewed as a specification test. In this test, the alternative hypothesis is a multiple PSTR model, and the test is thus a smooth transition counterpart to the test in Hansen (1999a) for determining the number of regimes in the PTR model.

The tests to be considered in this section resemble the ones that Eitrheim and Teräsvirta (1996) derived for STAR models. It turns out that they can be modified to fit the present framework. The new tests are the test of parameter constancy and that of no remaining nonlinearity. Error autocorrelation is also an indicator of misspecification. Its presence can, however, already be tested by applying the test by Baltagi and Li (1995).

3.4.1 Testing parameter constancy

Testing parameter constancy in panel data models has not received as much attention as in the time series literature. A possible explanation is that in many applications T is relatively small, which makes the assumption of parameter constancy difficult to test. However, with an increasing number of panels with relatively large T the test for parameter constancy in fixed effects models becomes feasible and important. Even though our test is developed for PSTR models, after minor modifications it can be applied to linear fixed effects models.

Our alternative to parameter constancy is that the parameters in (1) change smoothly from one regime to another. The model under the alternative may be called the Time Varying Panel Smooth Transition regression [TV-PSTR] model. It can be written as follows:

$$\begin{aligned} y_{it} &= \mu_i + \left[\beta_{11}' x_{it} + \beta_{12}' x_{it} g\left(q_{it}; \gamma_1, \mathbf{c}_1\right)\right] \\ &+ f\left(t/T; \gamma_2, \mathbf{c}_2\right)\left[\beta_{21}' x_{it} + \beta_{22}' x_{it} g\left(q_{it}; \gamma_1, \mathbf{c}_1\right)\right] + u_{it} \end{aligned} \tag{11}$$

where $g\left(q_{it}; \gamma_1, \mathbf{c}_1\right)$ and $f\left(t/T; \gamma_2, \mathbf{c}_2\right)$ are transition functions as defined in (2) and $(\gamma_1, \mathbf{c}_1')'$ and $(\gamma_2, \mathbf{c}_2')'$ are the parameter vectors. Equation (11) has a structure similar to the time-varying STAR model discussed in Lundbergh, Teräsvirta, and van Dijk (2003). One may also write

(11) as follows:

$$y_{it} = \mu_i + [\beta_{11} + \beta_{12} f\,(t/T;\gamma_2,\mathbf{c}_2)]'\,x_{it} \\ + [\beta_{21} + \beta_{22} f\,(t/T;\gamma_2,\mathbf{c}_2)]'\,x_{it} g\,(q_{it};\gamma_1,\mathbf{c}_1) + u_{it}. \tag{12}$$

Equation (12) shows how the parameters of the model vary between β_{11} and $\beta_{11}+\beta_{12}$ and β_{21} and $\beta_{21}+\beta_{22}$, respectively, smoothly and deterministically over time.

The alternative model (11) allows multiple alternatives to parameter constancy depending on the specification of $f\,(t/T;\gamma_2,\mathbf{c}_2)$. The general specification of $f\,(t/T;\gamma_2,\mathbf{c}_2)$ is

$$f\,(t/T;\gamma_2,\mathbf{c}_2) = \left(1+\exp\left(-\gamma_2\prod_{j=1}^{h}(t/T-c_{2j})\right)\right)^{-1}, t=1,\ldots,T \tag{13}$$

where $\mathbf{c}_2 = (c_{21},\ldots,c_{2h})'$ is an h-dimensional vector of location parameters. As before, the value of h determines the alternative hypothesis. In particular, it controls the form of switches in parameters. When $h=1$, the TV-PSTR model allows monotonic change in parameters. Equivalently, when $h=2$, the parameters change symmetrically around $(c_{21}+c_{22})/2$. Finally, γ_2 measures the smoothness of the change: when $\gamma_2\to\infty$ in (13), $f\,(t/T;\gamma_2,\mathbf{c}_2)$ becomes a step function, so structural breaks are included in the alternative as special cases. On the other hand, when $\gamma_2=0$ in (13), model (11) has constant parameters.

Note that it is assumed that the parameters in the transition function $g(q_{it};\gamma,c_1)$ are constant over time. This assumption is a practical one: such "second-order" nonconstancy is considerably harder to detect than nonconstancy in the regression coefficients, in particular as T may not be large in applications. We also assume a common transition function $f\,(t/T;\gamma_2,\mathbf{c}_2)$ for all individuals.

The null hypothesis of constant parameters in model (11) can be stated as $\mathrm{H}_0:\gamma_2=0$. However, under this hypothesis $(\beta_{12}',\beta_{22}',\mathbf{c}_2')'$ are not identified. To circumvent this problem we follow Eitrheim and Teräsvirta (1996) and replace (13) in (11) by its first-order Taylor expansion around $\gamma_2=0$. After merging terms we get the following auxiliary regression

$$\begin{aligned} y_{it} &= \mu_i + x_{it}'\beta_{11}^* + x_{it}'(t/T)\beta_1^* + \cdots + x_{it}'(t/T)^h\beta_h^* \\ &\quad + \left\{x_{it}'\beta_{21}^* + x_{it}'(t/T)\beta_{h+1}^* + \cdots + x_{it}'(t/T)^h\beta_{2h}^*\right\} g\,(q_{it};\gamma_1,\mathbf{c}_1) + u_{it}^* \end{aligned} \tag{14}$$

where $u_{it}^* = u_{it} + R\,(t/T,\gamma_2,\mathbf{c}_2)$ and $R\,(t/T,\gamma_2,\mathbf{c}_2)$ is the approximation error in the Taylor expansion. In (14), $\beta_j^* = \gamma_2\beta_j$ for $j=1,2,\ldots,h,h+1,h+2,\ldots,2h$. Then the original null hypothesis, $\mathrm{H}_0:\gamma_2=0$ can be tested in the auxiliary regression (14) as $\mathrm{H}_0':\beta_j^*=0$ for $j=1,2,\ldots,h,h+1,h+2,\ldots,2h$. Finally, note that under H_0' $u_{it}^*=u_{it}$, $i=1,\ldots,N$, so the Taylor series approximation does not affect the distribution assumptions. Therefore, under the null hypothesis and standard regularity conditions, the NLS estimator of $\mathbf{b} = (\beta_{11}^{*\prime},\beta_1^{*\prime},\ldots,\beta_{21}^{*\prime},\beta_{h+1}^{*\prime},\ldots,\beta_{2h},\gamma,\mathbf{c}')'$ is consistent and asymptotically normal for fixed T and $N\to\infty$.

In this context it is convenient to use the LM test because it only requires the estimation of (14) under the null hypothesis. In order to compute the LM statistic and its F-version we need to define the following vectors:

$$
\begin{aligned}
\hat{v}_{it} =& \left(\tilde{w}_{it}', \tilde{z}_{it}', (\partial\tilde{z}_{it}/\partial\gamma_1)'\,\hat{\beta}_{12}, (\partial\tilde{z}_{it}/\partial c_{11})'\,\hat{\beta}_{12}, \ldots, (\partial\tilde{z}_{it}/\partial c_{1m})'\,\hat{\beta}_{12}\right)' \\
\hat{\xi}_{it} =& (\tilde{x}_{it}', \tilde{\psi}_{it}', (\partial\tilde{\psi}_{it}/\partial\gamma_1)', (\partial\tilde{\psi}_{it}/\partial c_{11})', \ldots, (\partial\tilde{\psi}_{it}/\partial c_{1m})')' \\
\tilde{\psi}_{it} =& x_{it} g\left(q_{it}, \gamma_1, \mathbf{c}_1\right) - 1/T \sum_{t=1}^{T} x_{it} g(q_{it}, \gamma_1, \mathbf{c}_1) \\
\tilde{w}_{it} =& x_{it}(t/T)^j - 1/T \sum_{t=1}^{T} x_{it}(t/T)^j, j = 1, \ldots, h
\end{aligned}
$$

and

$$
\tilde{z}_{it}' = x_{it} g\left(q_{it}; \gamma_1, \mathbf{c}_1\right)(t/T)^j - 1/T \sum_{t=1}^{T} x_{it} g\left(q_{it}; \gamma_1, \mathbf{c}_1\right)(t/T)^j, j = 1, \ldots, h
$$

The χ^2 and F version of the test can be computed in three stages as follows[2]:

1. Estimate the PSTR model and compute the residual sum of squares SSR_0.
2. Regress $\tilde{y}_{it}$ on $\hat{v}_{it}$ and $\hat{\xi}_{it}$ and compute the residual sum of squares SSR_1.
3. Compute the χ^2 and F-versions of the tests as follows;

$$
\begin{aligned}
\mathrm{LM} &= TN(\mathrm{SSR}_0 - \mathrm{SSR}_1)/SSR_0 \\
\mathrm{LM}_{\mathrm{F}} &= \{(\mathrm{SSR}_0 - \mathrm{SSR}_1)/2hk\} \,/\, \{\mathrm{SSR}_1/(TN - N - 2hk)\}
\end{aligned} \tag{15}
$$

   Under the null hypothesis LM is asymptotically distributed as $\chi^2_{(2hk)}$ and LM_{F} is approximately distributed as $F(2hk, TN - N - 2hk)$.

Small-sample properties of LM_{F} will be investigated by simulation in Section 4.2.2.

3.4.2 Testing the hypothesis of no remaining nonlinearity

The purpose of this test is twofold. First, if the basic PSTR model (1) with (2) is the largest model that one wants to consider, the test is a misspecification test. A rejection indicates that the specification is not satisfactory. If the investigator is willing to consider a multiple PSTR model, the test serves as a test for the number of transition functions in the model. When the test is carried out, it is not necessary to assume that the new transition function in the alternative has the same transition variable as the one in the estimated model. As already mentioned, the test bears resemblance to the test of a similar hypothesis in Eitrheim and Teräsvirta (1996).

The model we consider in this subsection is the general additive PSTR model

$$
y_{it} = \mu_i + x_{it}'\beta_0 + \sum_{j=1}^{r} g_j(q_{it}^{(j)}; \gamma_j, c_j) x_{it}'\beta_j + u_{it} \tag{16}
$$

[2] See Appendix C for the mathematical derivation of the test.

where r is the number of regimes. This model can be written in a way that resembles the Multiple Regime Threshold model presented by Hansen (1999a). In fact, after adding and subtracting the appropriate elements, (16) becomes an $(r+1)$-regime STR panel model,

$$y_{it} = \mu_i + \left(1 - \sum_{j=1}^{r} g\left(q_{it}^{(j)};\gamma_j, c_j\right)\right) x_{it}'\beta_0 + \sum_{j=1}^{r} x_{it}'\left(\beta_0+\beta_j\right) g\left(q_{it}^{(j)};\gamma_j,\mathbf{c}_j\right) + u_{it}$$

where $g\left(q_{it}^{(j)};\gamma_j,\mathbf{c}_j\right)$ are transition functions as defined in (2) with $m=1$.

Consider the case in which $r=2$ and $q_{it}^{(j)} = q_{it}$ for $j=1,2$. Equation (16) then becomes three regime STR panel model of the form:

$$y_{it} = \mu_i + \left(1-g_1-g_2\right) x_{it}'\beta_0 + g_1 x_{it}'\left(\beta_0+\beta_1\right) + g_2 x_{it}'\left(\beta_0+\beta_2\right) + u_{it} \tag{17}$$

Two regimes in (17) are associated with $g_1 = g_2 = 0$ and $g_1 = g_2 = 1$, respectively, and there is an intermediate regime associated with $g_1 = 0$ and $g_2 = 1$ or $g_1 = 1$ and $g_2 = 0$. When $\gamma_j \to \infty$, $j = 1,2$, model (17) collapses into a three-regime PTR model of Hansen (1999a).

The test of no remaining nonlinearity can also be used as a test for determining the number of transition functions in the PSTR model. In applications a linear panel model is often applied by estimating several models into which the observations are allocated by a classifier q_{it}. In general, the number of models or regimes and the values of q_{it} that define the different models are selected on an *ad hoc* basis. Here we show how one can select the number of regimes sequentially.

In order to demonstrate the procedure, we assume that a PSTR model (16) with $r=1$ has been estimated and adding another transition function is considered. The extended model can be written as follows:

$$y_{it} = \mu_i + x_{it}'\beta_0 + g_1\left(q_{it};\gamma_1,c_1\right) x_{it}'\beta_1 + g_2\left(q_{it};\gamma_2,c_2\right) x_{it}'\beta_2 + u_{it} \tag{18}$$

and the null hypothesis of no additional transition in (18) is $H_0 : \gamma_2 = 0$. Under H_0, the parameters in (18) cannot be estimated consistently because the model is not identified. As before, the identification problem can be circumvented by replacing $g_2\left(q_{it};\gamma_2,c_2\right)$ in (18) by a Taylor expansion around $\gamma_2 = 0$. Choosing a first-order Taylor approximation leads to testing the hypothesis $\mathrm{H}_0' : \beta_{22}^* = \cdots = \beta_{2m}^* = 0$ in the following auxiliary regression:

$$\begin{aligned} y_{it} &= \mu_i + x_{it}'\beta_0^* + g_1\left(q_{it};\hat{\gamma}_1,\hat{c}_1\right) x_{it}'\beta_1 \\ &\quad + \left(x_{it}q_{it}\right)'\beta_{22}^* + \cdots + \left(x_{it}q_{it}^m\right)'\beta_{2m}^* + e_{it}^* \end{aligned} \tag{19}$$

where $\left(\gamma_1, c_1'\right)'$ is the parameter vector estimated under the null hypothesis.

In order to compute the χ^2 and F-versions of the test we set $\hat{\tilde{z}}_{it} = \left(\tilde{x}_{it}', \underbrace{\tilde{\omega}_{it}'}\left(\gamma_1,\mathbf{c}_1\right)\right)'$ and $\tilde{v}_{it} = \left(\widetilde{x_{it}'q_{it}},\ldots,\widetilde{x_{it}'q_{it}^m}\right)'$ where $\tilde{w}_{it}(.) = x_{it}g(q_{it},\gamma,\mathbf{c}) - \sum_t x_{it}g(q_{it},\gamma,\mathbf{c})/T$ and $\widetilde{x_{it}q_{it}^j} = x_{it}q_{it}^j - \sum_{t=1}^{T} x_{it}q_{it}^j/T$, $j=1,\ldots,m$. The test can be computed in three stages as follows:

1. Estimate the PSTR model (1) and compute the residual sum of squares SSR_0.

2. Regress $\tilde{y}_{it}$ on $\hat{\tilde{z}}_{it}$ and $\tilde{v}_{it}$, and compute the residual sum of squares SSR_1.

3. Compute the χ^2 and F-versions of the test as follows:

$$\mathrm{LM} = TN(\mathrm{SSR}_0 - \mathrm{SSR}_1)/\mathrm{SSR}_0$$
$$\mathrm{LM}_\mathrm{F} = \{(\mathrm{SSR}_0 - \mathrm{SSR}_1)/mk\} \, / \, \{\mathrm{SSR}_1/(TN - N - mk)\}$$

   Statistic LM has an asymptotic χ^2_{mk} distribution under H_0, whereas LM_F is approximately $F[mk, TN - N - 2mk]$, when H_0' holds.

This testing procedure can be used to determine the number of regimes in the general additive PSTR model. The selection can be done by using the following sequence of hypothesis: Given an estimated PSTR model with $r = r^*$, test the null hypothesis $\mathrm{H}_0 : r = r^*$ against $\mathrm{H}_1 : r = r^* + 1$. If H_0 is not rejected, the testing procedure ends. Otherwise, the null hypothesis $\mathrm{H}_0 : r = r^* + 1$ is tested against the model with $r = r^* + 2$. The testing procedure continues until the first acceptance of H_0.

The sequence of tests for specifying a general additive PSTR model can be carried out as follows:

1. Estimate the linear model and test linearity at significance level α.

2. If linearity is rejected, estimate a single transition PSTR model.

3. Test the hypothesis of no remaining nonlinearity for this model. If the hypothesis of no remaining nonlinearity is rejected at significance level $\tau\alpha$, $0 < \tau < 1$, estimate a double-transition model. The significance level is reduced by a factor τ in order to favour parsimony. Estimation of this model can be carried out in two stages. First, use a grid search of (γ_2, c_2) to find the initial values and then estimate the model by NLS. The grid search is conditional on the estimated values $(\hat{\gamma}_1, \hat{\mathbf{c}}_1')'$ from the previous stage of the process.

4. Continue until the first acceptance of the hypothesis of no remaining nonlinearity.

We investigate by simulation the small sample properties of the algorithm. The results indicate that the algorithm works well in small samples when the noise to signal ratio is small and/or the sample size is relatively large, $N \geq 40$ and $T \geq 10$. In other cases, the results indicate that the algorithm tends to identify a lower number of regimes than the true one and only rarely it selects a larger model[3].

Eitrheim and Teräsvirta (1996) pointed out potential numerical problems in the computation of the misspecification tests. In particular, when the estimates of γ_j are relatively large, such that the transition between regimes is rapid, the partial derivatives of the transition functions $g_j\left(q_{it}; \hat{\gamma}_j, \hat{\mathbf{c}}_j\right)$, $j = 1, \ldots, r$, with respect to $\left(\hat{\gamma}_j, \hat{\mathbf{c}}_j'\right)'$ obtain about the same value with few exceptions. As a result, the moment matrix of the auxiliary regression becomes near-singular. However, the contribution of the terms involving these partial derivatives to the test statistic is negligible at large values of γ_j. They can simply be omitted from the auxiliary regression without influencing the empirical size (and power) of the test statistic.

[3]The simulation results not presented here are available upon request

4 Simulation study

In this section we investigate the small-sample properties of the specification strategy for the PSTR model by simulation. The section is divided in three subsections. The first subsection deals with small sample properties of the NLS estimators. In the second subsection we investigate the size and power properties of the tests of linearity, parameter constancy and no remaining nonlinearity.

The design of the Monte Carlo experiment is the following. The number of replications equals 1000. Every experiment is carried out for the following sample sizes: $N = 10, 40, 160$ and $T = 5, 10$. The vector of exogenous variables $(x'_{it}, q_{it})'$ is generated independently for each individual following a VAR(1) model. The first 100 observations for each generated sample are discarded to avoid initialization effects. The endogenous variable is generated from

$$y_{it} = \mu_i + x'_{it}\beta_0 + \sum_{j=1}^{r} g\left(q_{it}; \gamma_j, \mathbf{c}_j\right) x'_{it}\beta_i + e_{it} \tag{20}$$

where $g\left(q_{it}; \gamma_j, \mathbf{c}_j\right)$ is defined as in (2), $\mu_i = \sigma_\mu u_i$ with $\sigma_\mu = 10$, u_i and e_{it} are a standard normal variables. The values of r, m and $(\gamma_j, \mathbf{c}'_j)'$ vary from one design to another and $(\beta'_0, \ldots, \beta'_r)'$, is obtained as follows: $\beta_0 = (1.0, 3.0)'$, and $\beta_j = 0.7\beta_0$ for $j = 1, \ldots, r$ in Section 4.1 and $\beta_j = 0.1\beta_0$ for $j = 1, \ldots, r$ in Section 4.2.

4.1 Estimation of the PSTR model

In order to investigate small-sample properties of our estimation algorithm we carry out two experiments. In the first one, we simulate model (20) with $r = 1$ and $m = 1$. We set the value of location vector equal $c = \mathsf{E}(q_i) = 3.5$ where $\mathsf{E}(q_i)$ is the unconditional mean of the transition variable. In the second experiment we generate the data from model (20) with $r = 1$ and $m = 2$, with location vector $\mathbf{c} = (3, 4)'$. Finally, to find out the effect of γ on the estimation we consider values $\gamma = 4$, 50. In all cases the estimation is carried out as described in Section 3.3.

The results for the experiments where $m = 1$ are in Table 1 and the ones where $m = 2$ in Table 2. The reported statistics are Monte Carlo mean and standard deviation of $(\hat{\gamma}, \hat{c})'$ and the bias in estimating $(\beta'_0, \beta'_1)'$. The tables are divided in two panels the upper one contains the results for $\gamma = 4$, whereas the results of the second experiment with $\gamma = 50$ can be found in the lower panel. The results indicate that the parameter estimates $(\beta'_0, \beta'_1)'$ and c of the PSTR model can be estimated with reasonable accuracy. However, by comparing the estimated standard deviations of $\mathbf{c}$ between the panels in each table, it is seen that the accuracy is higher when γ is large than when it is small. In particular, for $\gamma = 4$ and $m = 1$ the average $\hat{\sigma}_c$ equals 0.04, whereas for $\gamma = 50$ the average $\hat{\sigma}_c$ is 0.02. This may be due to the fact that for $\gamma = \infty$ the elements of $\mathbf{c}$ are estimated superconsistently. Consequently, when γ is finite but sufficiently large, $\mathbf{c}$ is estimated more accurately than is the case when γ is small. This result contrasts with the fact that small values of γ can be estimated with greater accuracy than large values. For example, for $N = 160$ and $T = 5$, $\hat{\sigma}_\gamma$ equals 0.26 while for $\gamma = 50$ is 7.2. But then, when γ is large the magnitude of the $\hat{\sigma}_\gamma$ does not play a role because as long as the estimate is also large the estimation error does not matter very much. At the other end, if γ is close to zero estimation may be difficult because then the model is close to being unidentified and the sequence of estimates may not converge. A similar conclusion holds for models that are close

Table 1: Small sample properties of the estimating procedure for the Panel LSTR model

$\gamma = 4, c_1 = 3.5$						
N	T	$\hat{\gamma}$	$\hat{\sigma}_\gamma$	$\hat{c_1}$	$\hat{\sigma}_{c_1}$	bias$(\beta) \times 10^{-2}$
10	5	4.517	3.054	3.501	0.084	0.400
	10	4.088	0.757	3.500	0.049	0.054
40	5	4.027	0.541	3.501	0.036	0.793
	10	4.011	0.337	3.500	0.023	0.243
160	5	4.014	0.257	3.499	0.018	0.179
	10	4.005	0.166	3.500	0.011	0.425
$\gamma = 50, c_1 = 3.5$						
N	T	$\hat{\gamma}$	$\hat{\sigma}_\gamma$	$\hat{c_1}$	$\hat{\sigma}_{c_1}$	bias$(\beta) \times 10^{-2}$
10	5	140.44	290.8	3.500	0.027	0.342
	10	94.92	172.0	3.500	0.014	0.029
40	5	58.17	54.7	3.500	0.007	0.887
	10	52.03	10.3	3.500	0.004	0.327
160	5	51.43	7.2	3.500	0.003	0.272
	10	50.37	4.5	3.500	0.002	0.260

to be linear, that is models in which β_1 in (20) is relatively small.

Finally, by comparing $\hat{\sigma}_\gamma$ across values of N for given T, one can notice the gains from the panel in estimating γ. For instance, when $m = 1$, $T = 10$ and $\gamma = 4$; $\hat{\sigma}_\gamma$ for $N = 10, 40, 160$ equal 0.75, 0.33 and 0.16, respectively. This is not an unexpected outcome because the accuracy of the estimate should increase with the number of observed transitions.

4.2 Small-sample properties of the misspecification tests

In this section we investigate the small-sample properties of our misspecification tests. In particular, we investigate the size and power of these tests for different values of N and T. The section is divided in three subsections and each subsection contains the results for a particular test.

4.2.1 Linearity test

Previous studies have documented that the F-version of the test has better size properties in small samples than the asymptotic χ^2-based statistic. For this reason, we only report the results for the F-version. [See Eitrheim and Teräsvirta (1996) and van Dijk, Teräsvirta, and Franses (2002) for details]. Additionally, since the auxiliary regression (4) with $m = 3$ has power against the alternative model (1) with $m = 1$ and 2 in (2), we compute the test statistic using $m = 1, 2, 3$ in (4)[4]. We denote by m^* the maximum power of q_{it} in the auxiliary regression

[4]The test based on the auxiliary regression (4) assuming $m^* = 3$ has power against the alternative models with $m = 1$ and $m = 2$ because it can be interpreted as having been derived by replacing the transition (2) in (1) by its second order Taylor expansion. When $m = 1$ in (2) the second order Taylor expansion yields an auxiliary regression with $m^* = 2$, and when $m = 2$ the second order Taylor expansion of transition function yields an auxiliary regression with $m^* = 4$. Consequently, the test based on $m^* = 3$ has power against both alternative models, $m = 1$ and $m = 2$.

Table 2: Small sample properties of the estimating procedure for the Panel LSTR model with $m = 2$

		$\gamma = 4, c_1 = 3, c_1 = 4$						
N	T	$\hat{\gamma}$	$\hat{\sigma}_\gamma$	$\hat{c_1}$	$\hat{\sigma}_{c_1}$	$\hat{c_2}$	$\hat{\sigma}_{c_2}$	bias(β) $\times 10^{-2}$
10	5	10.932	56.670	3.095	0.270	3.906	0.273	0.674
	10	4.547	1.808	3.083	0.232	3.917	0.231	2.558
40	5	4.238	1.147	3.066	0.194	3.935	0.196	2.554
	10	4.099	0.763	3.039	0.143	3.960	0.143	2.741
160	5	4.032	0.592	3.026	0.106	3.974	0.106	2.936
	10	4.012	0.391	3.008	0.051	3.992	0.051	2.157
		$\gamma = 50, c_1 = 3, c_1 = 4$						
N	T	$\hat{\gamma}$	$\hat{\sigma}_\gamma$	$\hat{c_1}$	$\hat{\sigma}_{c_1}$	$\hat{c_2}$	$\hat{\sigma}_{c_2}$	bias(β) $\times 10^{-2}$
10	5	114.25	221.85	2.999	0.037	4.002	0.041	0.181
	10	72.65	95.98	3.000	0.014	4.000	0.013	0.124
40	5	54.80	21.59	3.000	0.008	4.000	0.008	1.266
	10	51.30	8.03	3.000	0.005	4.000	0.005	0.085
160	5	50.57	5.40	3.000	0.003	4.000	0.003	0.327
	10	50.39	3.50	3.000	0.002	4.000	0.002	0.229

Table 3: Empirical size of the linearity test at 0.05 nominal level

N	T	$m^* = 1$	$= 2$	$= 3$
10	5	0.053	0.054	0.056
	10	0.059	0.052	0.052
40	5	0.039	0.050	0.054
	10	0.042	0.054	0.059
160	5	0.048	0.050	0.052
	10	0.048	0.046	0.042

The m^* in the name of the columns indicates the maximum value of m in the auxiliary regression (4)

(4). The order m^* of the auxiliary regression (4) may affects both the size and power of the test. For instance, approximations to the alternative model based on large values of m^* imply a loss of degrees of freedom which may cause problems, but they may add power to the test because the approximation to the nonlinear component improves with increasing m^*.

In order to investigate the empirical size of our linearity test we generate 1000 samples of a linear panel with fixed effects. The results are based on the nominal significance level 0.05 and can be found in Table 3. Each column contains results for the test statistic based on one auxiliary regression with m^* =1, 2, 3. It is seen that the empirical size of the test is close to the nominal size at all sample sizes. The loss of degrees of freedom associated with large values of m^* compared to small ones does not seem to affect the size of the test.

To investigate the power of the test we generate samples from the PSTR model (20) with $r = 1$ and for $m = 1$, 2. The model under the alternative is thus a standard PSTR model with either a monotonically increasing ($m = 1$) or symmetric ($m = 2$) transition function. To

Table 4: Empirical power of the linearity test at the 5% significance level

		$\gamma = 2, c = 3.5$			$\gamma = 4, c = 3.5$		
N	T	$m^* = 1$	$= 2$	$= 3$	$= 1$	$= 2$	$= 3$
10	5	0.16	0.12	0.10	0.23	0.16	0.14
	10	0.35	0.26	0.24	0.53	0.43	0.40
40	5	0.54	0.43	0.37	0.75	0.65	0.63
	10	0.92	0.85	0.81	1.00	0.98	0.98
160	5	0.99	0.99	0.98	1.00	1.00	1.00
	10	1.00	1.00	1.00	1.00	1.00	1.00
		$\gamma = 2, c = (3, 4)$			$\gamma = 4, c = (3, 4)$		
N	T	$m^* = 1$	$= 2$	$= 3$	$= 1$	$= 2$	$= 3$
10	5	0.06	0.08	0.08	0.06	0.10	0.09
	10	0.07	0.15	0.14	0.07	0.20	0.19
40	5	0.06	0.24	0.19	0.06	0.32	0.28
	10	0.06	0.49	0.41	0.06	0.64	0.60
160	5	0.05	0.75	0.69	0.05	0.88	0.85
	10	0.05	0.99	0.97	0.05	1.00	1.00

The m^* in the name of the columns indicates the maximum value of m in the auxiliary regression (4)

find out the extent to which the value of γ affects the power we set $\gamma = 2$ and $\gamma = 4$. Table 4 contains the estimated powers of the test. The upper panel shows the power for the PSTR model with $m = 1$. The results for $m = 2$ appear in the lower panel. In all cases, we report the power of the test based on three auxiliary regressions (4) with $m^* = 1, 2, 3$. It is seen that the test based on the auxiliary regression with $m^* = m$ has best power for both alternative models $m = 1$ and $m = 2$. Moreover, from the right-hand panel it appears that when $m = 2$ the test based on $m^* = 1$ does not have power, while the test based on $m^* = 3$ has only little less power than the test based on $m^* = 2$. The results from this section indicate, not unexpectedly, that the test based on the auxiliary regression with $m^* = 2$ is preferable to the others, especially in situations where the transition is symmetric around $(c_1 + c_2)/2$ $(m = 2)$. Finally, the power of the test seems to depend positively on the value of γ.

4.2.2 Parameter constancy test

Next we investigate the size and power properties of a parameter constancy test. To estimate the size of test we generate the replications with the PSTR model (20) with constant parameters, setting $r = 1$, $m = 1$, $\gamma_1 = 3$ and $c_1 = 3.5$, $(\beta'_{11}, \beta'_{12})' = (1.0, 3.0, 2.0, 1.0)'$. For easy of presentation we denote by h^* the maximum power of (t/T) in the auxiliary regression (14). As before, the auxiliary regression (14) with $h^* = 3$ has power against alternative models with $h = 1$ and $h = 2$ because it nests the ones based on $h^* = 1$ and $h^* = 2$.

The power of the test is computed by simulating the model under the alternative hypothesis (11). We consider two alternative models. In the first one, we allow one transition in the parameters in the middle of the sample: $h = 1$ and $c_2 = 0.5$. In the second one, the transition is symmetric around $(c_{21} + c_{22})/2 = 0.5$ with $c_{21} = 0.3$ and $c_{22} = 0.7$. In both cases $\gamma_2 = 4$ in (13) and $(\beta'_{21}, \beta'_{22})' = 0.1(\beta'_{11}, \beta'_{12})'$. We assume that the transition occurs at the same time

Table 5: Empirical size and power of the parameter constancy test at the 0.05 level of significance

		Size			$h = 1$			$h = 2$		
N	T	$h^* = 1$	$= 2$	$= 3$	$= 1$	$= 2$	$= 3$	$= 1$	$= 2$	$= 3$
10	5	0.061	0.057	0.039	0.098	0.084	0.073	0.062	0.052	0.051
	10	0.056	0.053	0.049	0.169	0.148	0.122	0.081	0.101	0.087
40	5	0.044	0.049	0.046	0.342	0.245	0.188	0.050	0.093	0.089
	10	0.044	0.035	0.049	0.697	0.580	0.485	0.125	0.229	0.201
160	5	0.054	0.050	0.045	0.948	0.902	0.835	0.096	0.316	0.246
	10	0.050	0.053	0.058	1.000	1.000	0.997	0.484	0.850	0.792

for all N individuals and that the change in the parameters is the same for all of them. This setup is consistent with the alternative hypothesis but it excludes other interesting options also nested within the alternative model as is the case in which only a proportion of units present non constancy in parameters. If this is the case, the power of the test may be affected and in particular it may have no power if this proportion is small. But then, in this situation the model for all units can be characterized by a PSTR model with constant parameters.

Table 5, with three panels, contains the estimated size and power of the test. The leftmost panel contains the empirical size of the test, the other two have the power results. All three panels are divided in three columns, one for each value of h^*. The results indicate that the F-version of the test has the correct size for all sample sizes and values of h^*. Similarly, with the results in linearity test the test based on $h^* = h$ has better power that the others. However, in contrast to the results in the linearity test, the test based on the auxiliary regression with $h^* = 1$ when $h = 2$ has not trivial power. In this situation, it seems more appropriate to base the decision about h as suggested in Granger and Teräsvirta (1993) and Teräsvirta (1994).

4.2.3 Test of no remaining nonlinearity and properties of the procedure for determining the number of regimes in a multiple regime PSTR panel

The size simulations of the test of no remaining nonlinearity are carried out using (20) and setting $r = 1$, $m = 1$, $\gamma_1 = 4$ and $c_1 = 3.5$. In order to estimate the power of the test we generate the samples from the same model with $r = 2$. The parameters in the first transition function are the same as in the size simulations. The second transition function is of type (2) with $m = 1$ and $\gamma = 4$, $c = 2.3$. As before, we consider only the F-version of the test and use three different auxiliary regressions (19) with $m^* = 1, 2, 3$. The estimated size and power at the 0.05 nominal level are shown in Table 6. The results indicate that the test based on the auxiliary regression with $m^* = 2$ has the best properties. First, its size is close to the nominal size and second, it has better power than the other alternatives.

5 An application: Investment and financial constraints

In this section we present an application of the modelling cycle for PSTR models. We choose the same economic problem and data set as Hansen (1999a)[5] and compare our results to the ones

[5]The data set is available in Bruce Hansen web page http://www.ssc.wisc.edu/~bhansen/

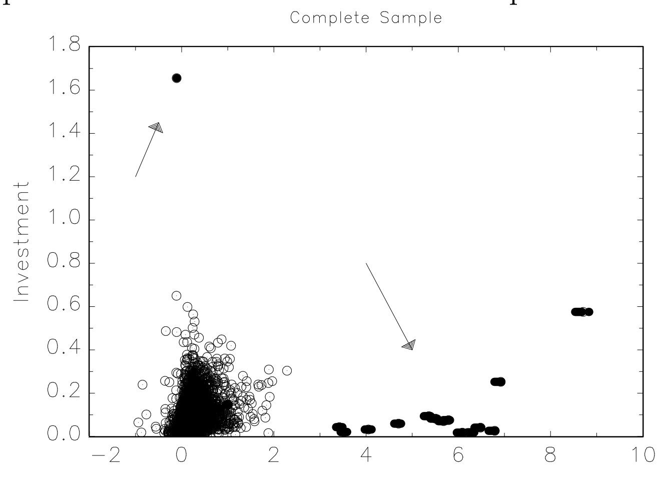
Table 6: Empirical size and power of the test for no remaining nonlinearity at the 0.05 level of significance

		Size			$h=1$			$h=2$		
N	T	$h^*=1$	$=2$	$=3$	$=1$	$=2$	$=3$	$=1$	$=2$	$=3$
10	5	0.051	0.059	0.060	0.069	0.068	0.066	0.062	0.073	0.080
	10	0.052	0.054	0.049	0.123	0.114	0.119	0.061	0.085	0.071
40	5	0.043	0.054	0.052	0.226	0.217	0.192	0.067	0.132	0.104
	10	0.049	0.051	0.052	0.508	0.510	0.470	0.084	0.244	0.210
160	5	0.050	0.047	0.042	0.716	0.756	0.742	0.134	0.459	0.375
	10	0.056	0.053	0.050	0.973	0.992	0.990	0.248	0.806	0.750

in that article. The economic question under investigation is whether firms that are financially constrained behave differently from financially unconstrained firms when it comes to financing investment. Fazzari, Hubbard, and Petersen (1988) argued that when the capital market is imperfect, the firm's financial structure is not irrelevant to its investment decisions because internal and external funds are not perfect substitutes. In their view, when a firm is facing credit constraints, investment may depend upon the availability of internal funds, such as the firm's cash flow. For instance, in presence of asymmetric information between the firm and the provider of external finance, the cost of external funding is higher than the cost of internal funding, which favours the use of internal resources to finance investment. Furthermore, when the debt level is high the firm is likely to be financially constrained and consequently, its cash flow will be positively correlated with the investment rate.

This situation calls for a PSTR model such that the transition variable measures the financial position of each firm and period. A natural candidate, and the one Hansen (1999a) used, is the debt-to-asset ratio. It measures the existing debt level and can be treated as a proxy for availability of external funds to the firm. This is because external providers of financing may be reluctant to lend capital to strongly indebted firms. The data set is extracted from the one used

Figure 1: Scatter plot of the cash flow and investment pairs for the complete sample



by Hall and Hall (1993) and consists of 565 firms observed from 1973 to 1987. The variables include the ratio of investment to capital I_{it}, the ratio of total market value to assets Q_{it}, the ratio of long-term debt to assets D_{it} and the ratio of cash flow to assets CF_{it}. We delete two firms from the original sample because they have atypical investment-cash flow relationship. Figure 1 shows the scatter plot of investment *versus* cash flow. The observations for the deleted firms are represented by solid circles. As is seen, these firms have either extremely large levels of cash flow with low levels of investment or extremely large levels of investment with very low levels of cash flow.

An application of the PSTR modelling strategy Our starting point is that investment can be characterized by a Multiple Regime PSTR model (16). Our maintained model for investment thus has the form

$$\begin{aligned} I_{it} = \mu_i + \theta_1 Q_{it-1} + \theta_2 Q_{it-1}^2 + \theta_3 Q_{it-1}^3 + \theta_4 D_{it-1} \\ +\theta_5 Q_{it-1} D_{it-1} + \beta_0 CF_{it-1} + \sum_{j=1}^{r} \beta_j CF_{it-1} g\left(D_{it-1}; \theta_j\right) + e_{it} \end{aligned} \tag{21}$$

where the transition functions $g\left(D_{it-1}; \theta_j\right)$ $j = 1, \ldots, r$ are of type (2) with $m = 1$. Following Hansen (1999a) we include the terms Q_{it-1}^j for $j = 2, 3$, D_{it-1} and $Q_{it-1} D_{it-1}$ to account for possible omitted variables.

The multiple-regime PSTR model is specified as suggested in Section 3.4.2. The results of the linearity test are presented in Table 7. Since linearity is rejected at 5% significance level we estimate a Multiple PSTR model with $r = 1$ and $m = 1$ and test whether another transition is required. The next null hypothesis of $r = 1$ is not rejected, so our final model has a single transition.

Table 7: Determining the number of regimes using the test for no remaining nonlinearity

		$\mathrm{H}_0 : r = 0$ *vs* $\mathrm{H}_0 : r = 1$
		Actual Significance level: $\alpha = 0.05$
$F2$	8.58	2×10^{-10}
$F3$	6.00	4×10^{-11}
		$\mathrm{H}_0 : r = 1$ *vs* $\mathrm{H}_0 : r = 2$
		Actual Significance level: $\alpha = 0.025$
$F2$	2.31	0.10
$F3$	2.27	0.08

We use 0.05 as significance level and assume that $\tau = 0.5$.

Table 8 contains the parameter estimates of the single transition model. It is seen from the table and Figure 2 that the transition is quite sharp and the model thus is close to a two-regime PTR model. The first regime contains the firms with low debt levels, $D_{it-1} < 0.0154$ and the second regime the more strongly indebted firms, $D_{it-1} \geq 0.0154$.

The combined "parameter" of the cash flow as a function of the debt level equals

$$\beta^{CF}\left(D_{it-1}\right) = \beta_1 + \beta_2 g\left(D_{it-1}; \gamma_1, c_1\right)$$

where β_1 and β_2, are the coefficients of the cash flow variable CF_{it-1} in Table 8. Figure 2 shows the estimated parameter of the cash flow as a function of the debt level. Our results suggest a positive relationship between cash flow and investment which conforms to the prediction of Fazzari, Hubbard, and Petersen (1988).

Table 8: Parameter estimates for the final PSTR model

Regressor	Coefficient estimate	Standard error
Q_{it-1}	0.0118	0.0010
$Q_{it-1}^2/10^3$	-0.2602	0.0365
$Q_{it-1}^3/10^6$	1.4500	0.2700
D_{it-1}	-0.0218	0.0026
$Q_{it-1}D_{it-1}$	0.0017	0.0016
CF_{it-1}	0.0539	0.0054
$CF_{it-1}g\left(D_{it-1};\gamma_1,c_1\right)$	0.0355	0.0047
	Transition Functions	
Parameter	Coefficient estimate	Standard error
γ_1	0.508×10^5	139.28
c_1	0.01554	0.0055

In order to compare the results of our methodology with those obtained following Hansen's approach we estimate (21) as a multiple PTR model. To determine the number of thresholds we apply Hansen's sequential procedure which is similar to ours. The difference is that the candidate models are PTR models. The procedure works as follows. First, estimate a linear model and test it against a model with one threshold (two regimes). If the null hypothesis is rejected, estimate a single-threshold model and test it against a double-threshold one. The procedure is continued until the hypothesis no additional threshold is not rejected.

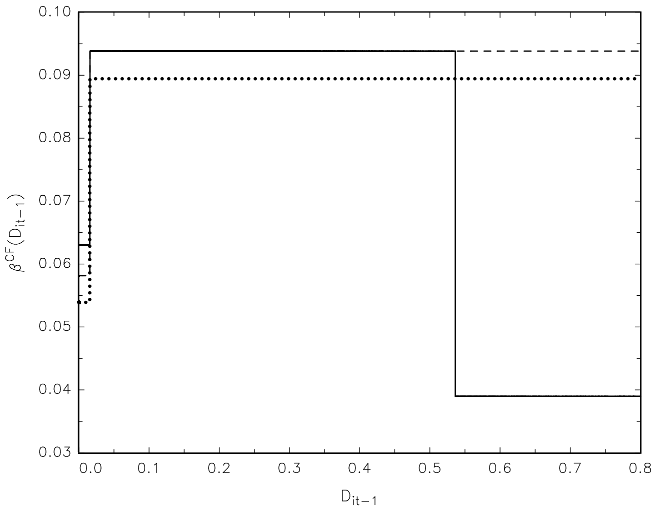
Table 9 contains the results of the linearity test. They indicate that a two-regime model is enough to characterize the nonlinearity in the data. The final estimated PTR model is reported in Table 10. The threshold value $c_1 = 0.0157$, which is very close to the estimated one in the PSTR model. This is expected since the estimate of γ in the PSTR model is large. Figure 2 contains the coefficient of the cash flow as a function of the debt level in the estimated PTR model. The resulting graph is similar to the one obtained with the PSTR model.

Table 9: Test for threshold effects using Hansen (1999a)

Test a for single threshold	
F_1	44.3
P-value	0.0000
(10%,5%,1% critical values)	(13.9,18.4,25.8)
Test a for double threshold	
F_2	9.1
P-value	0.26
(10%,5%,1% critical values)	(12.4 ,15.5 ,19.7)

Figure 2 also contains the coefficient of the cash flow as a function of the debt level in the original PTR model of Hansen (1999a). Hansen used the complete data set and concludes that a three-regime threshold model characterized the data. His results indicated that the relationship between investment and cash flow is nonmonotonic in the sense that the connection between these variables is weaker for very highly indebted firms. He argued that there is considerable uncertainty in the estimate of the cash flow coefficient in the last regime. From our results it

Figure 2: Estimates of the cash flow parameter as a function of the lagged debt level, D_{it-1}



Dotted line: PSTR.
Dashed line: Panel threshold estimates with restricted sample.
Solid line: Panel threshold estimates with complete sample.

becomes clear that this uncertainty is due to the presence of the two outliers in the original sample. When they are removed, our results agree with the ones obtained using the PTR approach.

Table 10: Parameter estimates for the two-regime model of Hansen (1999a)

Regressor	Coefficient estimate	Standard errors
Q_{it-1}	0.0117	0.0009
$Q_{it}^2/10^3$	-0.2540	0.0285
$Q_{it}^3/10^6$	1.4028	0.2151
D_{it-1}	-0.0268	0.0046
$Q_{it-1}D_{it-1}$	0.0022	0.0014
$CF_{it-1}I\left(D_{it-1} \leq c_1\right)$	0.0582	0.008
$CF_{it-1}I\left(D_{it-1} \geq c_1\right)$	0.0938	0.007
Threshold Values		
Parameter	Coefficient estimate	
c_1	0.0157	

6 Conclusions

We introduce a panel model in which the parameters can change as a function of an exogenous variable. The model is thus an alternative to the Panel Threshold model by Hansen (1999a). In the present context however, standard asymptotic theory can be used as the likelihood function in the PSTR model is a continuous function of the parameters. We present tests for linearity, parameter constancy and remaining non linearity. These tests, serve as both specification and misspecification tests. The small sample properties of the proposed statistics was investigated by simulation and the results indicate that they behave well even in panels with small N and T.

References

Amemiya, T. (1985): *Advanced Econometrics*. Hardvard University Press.

Andrews, D. W. K., and W. Ploberger (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 62, 1383–1414.

Baltagi, B. H., and Q. Li (1995): "Testing AR(1) against MA(1) disturbances in an error component model," *Journal of Econometrics*, 68, 133–151.

Chamberlain, G. (1984): "Panel data," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. 2, pp. 1248–1318, Amsterdam:. North-Holland.

Davies, R. B. (1977): "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 64, 247–254.

——— (1987): "Hypothesis testing when a nuisance parameter is present only under the alternative.," *Biometrika*, 74, 33–43.

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

Fazzari, S., G. Hubbard, and B. Petersen (1988): "Financial constrains and corporate investments," *Brokings Papers on Economic Activity*, pp. 141–195.

Granger, C. W. J., and T. Teräsvirta (1993): *Modelling nonlinear economic relationships*. Oxford: Oxford University Press.

Hall, B. H., and R. E. Hall (1993): "The value and performance of U.S. corporations.," *Brookings Papers on Economic Activity*, pp. 1–34.

Hansen, B. E. (1996): "Inference when a nuisance parameter is not identified under the null hypothesis.," *Econometrica*, 64, 413–430.

——— (1999a): "Threshold effects in non-dynamic panels: Estimation, testing and inference," *Journal of Econometrics*, 93, 345–368.

——— (1999b): "Testing for linearity," *Journal of Economic Surveys*, 13, 551–576.

——— (2000): "Sample splitting and threshold estimation," *Econometrica*, 68, 575–603.

Jansen, E. S., and T. Teräsvirta (1996): "Testing parameter constancy and super exogeneity in econometric equations," *Oxford Bulletin of Economics and Statistics*, 58, 735–763.

Leybourne, S., P. Newbold, and D. Vougas (1998): "Unit roots and smooth transitions," *Journal of Time Series Analysis*, 19, 83–97.

Lundbergh, S., T. Teräsvirta, and D. van Dijk (2003): "Time-varying smooth transition autoregressive model," *Journal of Business and Economic Statistics*, 21, 104–121.

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

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

TERÄSVIRTA, T. (1998): "Modelling economic relationships with smooth transition regressions," in *Handbook of applied economic statistics*, ed. by A. Ullah, and D. E. A. Giles, pp. 507–552. New York: Marcel Dekker.

VAN DIJK, D., T. TERÄSVIRTA, AND P. H. FRANSES (2002): "Smooth transition autoregressive models - a survey of recent developments," *Econometric Reviews*, 21, 1–47.

WHITE, H. (1980): "Nonlinear regression on cross-section data," *Econometrica*, 48, 721–746.

——— (2000): *Asymptotic theory for econometricians.* 2nd edition - Academic Press.

A Asymptotic properties

This Appendix contains the proof of Theorem 1. The following Lemma is a standard result but it is stated to clarify the notation in this section and Appendix B.

Lemma 1 *Let $W_i = (w'_{i1}, \ldots, w'_{iT})'$ be a $(T \times k)$ matrix of random variables and let $Q_T = (c_1, \ldots, c_T) = I_T - \frac{1}{T}\iota\iota'$ be the within transformation matrix where ι is a $(T \times 1)$ vector of ones. Then $W'_i Q'_T Q_T W_i = \sum_{s=1}^{T}\sum_{t=1}^{T} c_{st} w_{it} w'_{is}$ where $c_{st} = c'_s c_t$ is finite non random function of T.*

A.1 Linearity test: Consistency and asymptotic normality of the auxiliary regression

As explained in Section 3.2, the linearity test is based on the auxiliary regression (4) which for observation i can be written as

$$Y_i = \mu_i \iota + X_i^* \beta + R_i \tag{1}$$

where $Y_i = (y_{i1}, \ldots, y_{iT})'$, X_i^* is a $(T \times km)$ matrix with $X_i^* = (X_i, (X_i \odot q_{it}), \ldots, (X_i \odot q_i^m))$ and R_i is the error vector. After eliminating the fixed effects, the least squares estimator of β has the form

$$\begin{aligned}\hat{\beta} &= (\sum_{i=1}^{N} X_i^{*\prime} Q_T X_i^*)^{-1} \sum_{i=1}^{N} X_i^{*\prime} Q_T Y_i \\ &= \beta^0 + (\sum_{i=1}^{N} X_i^{*\prime} Q_T X_i^*)^{-1} \sum_{i=1}^{N} X_i^{*\prime} Q_T R_i \end{aligned} \tag{2}$$

Under the null hypothesis, $R_i = U_i$ for $i = 1, \ldots, N$. Moreover, some elements of β^0 equal zero under H_0, but the idea behind the linearity test is that these elements can be estimated consistently. Thus, under H_0,

$$\begin{aligned}\hat{\beta} &= \beta^0 + (\sum_{i=1}^{N} X_i^{*\prime} Q_T X_i^*)^{-1} \sum_{i=1}^{N} X_i^{*\prime} Q_T U_i \\ &= \beta^0 + \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} c_{ts} x_{it}^* x_{is}^{*\prime} \right)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} c_{ts} x_{it}^* u_{is} \end{aligned} \tag{3}$$

The second equality follows from Lemma 1. Formula (3) provides a convenient form for analyzing the asymptotic properties of $\hat{\beta}$ when $N \to \infty$ and T is fixed. In fact, consistency and asymptotic normality can be established in the usual way by applying the Markov law of large numbers and the Lyapunov central limit theorem; See Theorems 3.15 and 5.13 in White (2000).

B Consistency and asymptotic normality of the maximum likelihood estimator in the Panel Smooth Transition model

In this appendix we sketch the proof of Theorem 2 in the paper.

B.1 Logarithmic likelihood, score and Hessian

As described in Section 3.3 the estimation of PSTR models is carried out in two steps. First, the fixed effects are eliminated by multiplying (2.1') with the within transformation matrix Q_T. In the second step, the log-likelihood, after eliminating the fixed effects, is maximized with respect to $\beta = (\beta_1', \beta_2', \beta_3')'$. The concentrated log-likelihood can be written as follows:

$$L_N(\beta) = c - \frac{1}{2N} \sum_{i}^{N} (Y_i - X_i\beta_1 - W_i(\beta_3)\beta_2)' Q_T (Y_i - X_i\beta_1 - W_i(\beta_3)\beta_2). \tag{4}$$

The elements in score vector $S_N(\beta)$ for (4), evaluated at the true value $\beta = \beta^0$, are,

$$\begin{aligned}
\frac{\partial L(\beta)}{\partial \beta_1}|_{\beta=\beta^0} &= \frac{1}{N}\sum_{i=1}^{N} X_i' Q_T U_i \\
\frac{\partial L(\beta)}{\partial \beta_2}|_{\beta=\beta^0} &= \frac{1}{N}\sum_{i=1}^{N} W(\beta_3^0)' Q_T U_i \\
\frac{\partial L(\beta)}{\partial \beta_{3j}}|_{\beta=\beta^0} &= \frac{1}{N}\sum_{i=1}^{N} \beta_2^{0\prime} \frac{\partial W_i(\beta_3^0)'}{\partial \beta_{3j}} Q_T U_i \quad j = 1, \ldots, m
\end{aligned}$$

The elements of the Hessian matrix have the form

$$\frac{\partial^2 L(\beta)}{\partial\beta_1\beta_1'} = -\frac{1}{N}\sum_{i=1}^{N} X_i' Q_T X_i \tag{5}$$

$$\frac{\partial^2 L(\beta)}{\partial\beta_1\beta_2'} = -\frac{1}{N}\sum_{i=1}^{N} X_i' Q_T W_i(\beta_3) \tag{6}$$

$$\frac{\partial^2 L(\beta)}{\partial\beta_1\beta_{3j}} = -\frac{1}{N}\sum_{i=1}^{N} X_i' Q_T \frac{\partial W_i(\beta_3)}{\partial\beta_{3j}}; \quad j = 1,\ldots,m \tag{7}$$

$$\frac{\partial^2 L(\beta)}{\partial\beta_2\beta_2'} = -\frac{1}{N}\sum_{i=1}^{N} W_i(\beta_3)' Q_T W_i(\beta_3) \tag{8}$$

$$\frac{\partial^2 L(\beta)}{\partial\beta_2\beta_{3j}} = -\frac{1}{N}\sum_{i=1}^{N}\left[\beta_2'\frac{\partial W_i(\beta_3)'}{\partial\beta_{3j}} Q_T W_i(\beta_3) - \beta_2'\frac{\partial W_i(\beta_3)'}{\partial\beta_{3j}} Q_T U_i\right], j = 1,\ldots,m \tag{9}$$

$$\frac{\partial^2 L(\beta)}{\partial\beta_{3j}\beta_{3h}} = -\frac{1}{N}\sum_{i=1}^{N}\left[\beta_2'\frac{\partial W_i(\beta_3)'}{\partial\beta_{3j}} Q_T \frac{\partial W_i(\beta_3)}{\partial\beta_{3j}}\beta_2 - \beta_2'\frac{\partial^2 W_i(\beta_3)'}{\beta_{3j}\beta_{3h}} Q_T U_i\right], \tag{10}$$

$$j, h = 1,\ldots,m.$$

B.2 Consistency

In order to proof consistency of the maximum likelihood estimator $\hat{\beta}$ we apply Theorem 4.1.1 in Amemiya (1985). The maximum likelihood estimator is consistent if (i) the parameter space Ω is compact, (ii) the objective function is continuous in $\beta \in \Omega$ and (iii) the objective function converges to a nonstochastic function uniformly in probability in $\beta \in \Omega$ as $N \to \infty$. Conditions (i) and (ii) are satisfied given assumptions (E1) to (E7). In what follows we verify that (iii) is also satisfied.

Using Assumption (E1) and the equality

$$W_i(\beta_3^0)\beta_2^0 - W_i(\beta_3)\beta_2 = -W_i(\beta_3)(\beta_2 - \beta_2^0) - (W_i(\beta_3) - W_i(\beta_3^0)\beta_2^0$$

one can write the likelihood function (4) in deviations from the mean as,

$$
\begin{aligned}
L_N =&\frac{1}{N}\sum_{i=1}^{N}(\beta_1-\beta_1^0)'X_i'Q_TX_i(\beta_1-\beta_1^0)\\
&-\frac{2}{N}\sum_{i=1}^{N}(\beta_1-\beta_1^0)'X_i'Q_TW_i(\beta_3)(\beta_2-\beta_2^0)\\
&-\frac{2}{N}\sum_{i=1}^{N}(\beta_1-\beta_1^0)'X_i'Q_T(W_i\left(\beta_3)-W_i(\beta_3^0)\right)\beta_2\\
&-\frac{2}{N}\sum_{i=1}^{N}(\beta_1-\beta_1^0)'X_i'Q_TU_i\\
&+\frac{1}{N}\sum_{i=1}^{N}(\beta_2-\beta_2^0)'W_i(\beta_3)'Q_TW_i(\beta_3)(\beta_2-\beta_2^0)\\
&-\frac{2}{N}\sum_{i=1}^{N}(\beta_2-\beta_2^0)'W_i(\beta_3)'Q_T(W_i\left(\beta_3)-W_i(\beta_3^0)\right)\beta_2\\
&-\frac{2}{N}\sum_{i=1}^{N}(\beta_2-\beta_2^0)'W_i(\beta_3)'Q_TU_i\\
&+\frac{1}{N}\sum_{i=1}^{N}\beta_2'\left(W_i(\beta_3)-W_i(\beta_3^0)\right)'Q_T\left(W_i(\beta_3)-W_i(\beta_3^0)\right)\\
&-\frac{2}{N}\sum_{i=1}^{N}\beta_2'\left(W_i(\beta_3)-W_i(\beta_3^0)\right)'Q_TU_i\\
&+\frac{1}{N}\sum_{i=1}^{N}U_i'Q_TU_i
\end{aligned}
$$

In a compact parameter space uniform convergence of the likelihood function follows from uniform convergence of the following moment matrices:

$$\tfrac{1}{N}\textstyle\sum_{i=1}^{N} X_i'Q_TX_i \rightarrow \mathsf{E}\left[X_i'Q_TX_i\right] \tag{11}$$

$$\tfrac{1}{N}\textstyle\sum_{i=1}^{N} X_i'Q_TW_i(\beta_3) \rightarrow \mathsf{E}\left[X_i'Q_TW_i(\beta_3)\right] \tag{12}$$

$$\tfrac{1}{N}\textstyle\sum_{i=1}^{N} W_i(\beta_3)'Q_TW_i(\beta_3) \rightarrow \mathsf{E}\left[W_i(\beta_3)'Q_TW_i(\beta_3)\right]. \tag{13}$$

Convergence in probability as $N\rightarrow\infty$ of (11) follows from Assumption (E6) and the law of large numbers for i.i.d random variables. In order to proof convergence in probability of the moment matrices (12) and (13) we use the uniform law of large number for i.i.d processes (Theorem 4.5.2 in Amemiya). We have to show that there exists a dominant function $h(X_1, X_2)$ such that the absolute value of the elements of the moment matrices (12) and (13) are less than $h(X_1, X_2)$ for all $\beta\in\Omega$ and that $\mathsf{E}\left[h(X_1, X_2)\right]<\infty$.

In order to find the dominant function it is convenient to write explicitly the typical (j,h)-element of (12) and (13). Using Lemma 1, it is seen that the (j,h)-element, $j,h=1,\ldots,k$ in

(12) can be written as,

$$\frac{1}{N}\sum_{i=1}^{N} z_i^{(j,h)}(\beta_3) \tag{14}$$

where

$$z_i^{(j,h)}(\beta_3) = \sum_{t=1}^{T}\sum_{s=1}^{T} c_{st}\, g(x'_{2t}\beta_3)x_{is,h}x_{it,j}. \tag{15}$$

One dominant function $h(X_1, X_2)$ for $z_i^{(j,h)}(\beta_3)$ is $\sup_{\beta\in\Omega}|z_i^{(j,h)}(\beta_3)|$. In fact, given that $g(x'_{2t}\beta_3) < \eta$, $\eta \in [0,1]$. We have that, for any $T < T_0 < \infty$,

$$\begin{aligned}
\mathsf{E}[\sup_{\beta\in\Omega}|z_i(\beta_3)|] \leq{}& \eta\sum_{s=1}^{T}|c_{st}||x_{is,h}x_{it,j}| \\
\leq{}& \eta\sum_{s=1}^{T}\sum_{t=1}^{T}|c_{st}|\mathsf{E}[|x_{is,h}x_{it,j}|] \\
\leq{}& \eta\sum_{s=1}^{T}\sum_{t=1}^{T}|c_{st}|(\mathsf{E}[|x_{is,h}|^2])^{1/2}(\mathsf{E}[|x_{it,j}|^2]])^{1/2} \\
\leq{}& \eta\Delta\sum_{s=1}^{T}\sum_{t=1}^{T}|c_{st}| < \infty.
\end{aligned}$$

where $\Delta < \infty$. A similar argument can be used to show that the sample moment (13) converges uniformly to its population moment.

Consequently, conditions (i),(ii) and (iii) of the Theorem 4.1.1 in Amemiya (1985) are satisfied and the maximum likelihood estimator is consistent for $N \to \infty$ and fixed T.

B.3 Asymptotic normality

To prove asymptotic normality we apply Theorem 4.1.3 in Amemiya (1985). It states that if (*i*) $\hat{\beta}$ is consistent, (*ii*) the score vector evaluated at the true value of the parameters is asymptotically normal (*iii*) the Hessian matrix is continuous and (*iv*) the average Hessian converges in probability to a nonsingular matrix for any estimator $\beta^* \to \beta^0$ then $\sqrt{N}(\hat{\beta}-\beta^0) \xrightarrow{d} \mathsf{N}[0, \sigma^2 V^{-1}]$.

Normality of the score evaluated at β^0 follows from the central limit theorem for i.i.d random variables and assumptions (E1) - (E7). The continuity condition (*iii*) is satisfied given the specification for the panel smooth transition model. To prove the convergence in probability of the average Hessian we apply Theorem 4.2.1 in Amemiya (1985). That is, convergence of the Hessian follows from uniform convergence of its elements and the fact that $\hat{\beta}$ is consistent.

The proof of uniform convergence of the elements (5) -(8) of Hessian is the same as the proof of uniform convergence of the likelihood function. In order to prove uniform convergence of (9) and (10), use Lemma 1 and write (9) extensively for each element as,

$$\frac{1}{N}\sum_{i=1}^{N}\left[\beta_2^{*\prime}\frac{\partial W_i(\beta_3^*)'}{\partial\beta_{3j}}Q_T W_i(\beta_3^*) - \beta_2^{*\prime}\frac{\partial W_i(\beta_3^*)'}{\partial\beta_{3j}}Q_T U_i^*\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left[z_{1i}^{(j,h)}(\beta^*) - z_{2i}^{(j,h)}(\beta^*)\right] \tag{16}$$

where

$$z_{1i}^{(j,h)}(\beta^*) = \sum_{t=1}^{T}\sum_{s=1}^{T} c_{st}\lambda_{is}(\beta_3^*)g(x_{2t}'\beta_3^*)\beta_{2k}^* x_{2is,j}x_{is,k}x_{it,h} \tag{17}$$

$$z_{2i}^{(j,h)}(\beta^*) = \sum_{t=1}^{T}\sum_{s=1}^{T} c_{st}\lambda_{is}(\beta_3^*)\beta_{2k}^* x_{2is,j}x_{is,k}u_{it}^* \tag{18}$$

for $j = 1,\ldots,m; h, k = 1,\ldots,K$. The term $\lambda_{it}(\beta_3)$ denotes the first derivative of $g(\upsilon)$ with respect to $\upsilon \equiv x_{2t}'\beta_3$, where $x_{2t} = (1, q_{it}, q_{it}^2, \ldots, q_{it}^m)'$. $\lambda_{it}(\beta_3) = (1 - g(x_{2t}'\beta_3))g(x_{2t}'\beta_3)$.

The uniform convergence of (16) requires the existence of a dominant function for (17) and (18). For (17) one such dominant functions is $\sup_{\beta\in\Omega}|z_{1i}(\beta^*)|$ which is well defined because the parameter space is compact and the function $\lambda_{it}(\beta_3)$ takes values in the interval [0,1]. Moreover, $\mathsf{E}[\sup_{\beta\in\Omega}|z_{1i}^{(j,h)}(\beta^*)|]$ is finite. In fact,

$$\begin{aligned}
\mathsf{E}[\sup_{\beta\in\Omega}|z_{1i}^{(j,h)}(\beta^*)|] \leq& \sum_{t=1}^{T}\sum_{s=1}^{T}|c_{st}|\mathsf{E}[|\lambda_{is}(\beta_3^*)g(x_{2t}'\beta_3^*)||\beta_{2k}^*||x_{2is,j}x_{is,k}x_{it,h}|]\\
\leq& \Delta_1\sum_{t=1}^{T}\sum_{s=1}^{T}|c_{st}|(\mathsf{E}[|x_{2is,j}x_{is,k}x_{it,h}|^2])^{1/2}(\mathsf{E}[|\lambda_{is}(\beta_3^*)g(x_{2t}'\beta_3^*)|^2)^{1/2}\\
\leq& \Delta_1\Delta_2\Delta_3\sum_{t=1}^{T}\sum_{s=1}^{T}|c_{st}| < \infty
\end{aligned}$$

where $|\beta_{2k}| \leq \Delta_1$ and $\mathsf{E}[|\lambda_{is}(\beta_3^*)g(x_{2t}'\beta_3^*)|^2] \leq \Delta_2$ and $\mathsf{E}[|x_{2is,j}x_{is,k}x_{it,h}|^2] \leq \Delta_3$. $\Delta_i, i = 1, 2, 3$ are finite constants. Uniform convergence of $z_{2i}(\beta^*)$ in (16) follows from similar arguments.

The proof of uniform convergence of (10) is the same but one needs additional moments and notation. In fact, writing the second partial derivative of $W_i(\beta_3)$ as

$$\frac{\partial^2 W_i(\beta_3)}{\partial\beta_{3j}\beta_{3h}} = [1 - 2g(x_{2t}'\beta_3)]\lambda_{it}(\beta_3)x_{2it,j}x_{2it,h} \tag{19}$$

one can see that the dominant function for (10) exist because the summands of (10) are bounded by the $\sup_{\beta\in\Omega}$. Moreover, the supreme is well defined because, the parameter space is compact and the function $|[1 - 2g(x_{2t}'\beta_3)]\lambda_{it}(\beta_3)|$ only takes values in the interval [0,0.1]. For uniform convergence then one needs that $\mathsf{E}[|x_{2it,j}x_{2ist,h}x_{is,r}x_{it,l}|^2] \leq \Delta < \infty$ for $j, h = 1,\ldots,m$, $r, l = 1,\ldots,k$ and $t, s = 1,\ldots,T$.

Given that conditions (i) to (iv) of the Theorem 4.1.3 in Amemiya (1985) are satisfied then we have that

$$\sqrt{N}(\hat{\beta} - \beta^0) \xrightarrow{d} \mathsf{N}[0, \sigma^2 V^{-1}] \tag{20}$$

C Derivation of the LM test for parameter constancy in PSTR models

We derived the LM statistic by using the local approximation of the concentrated pseudo-likelihood function. That is, we first eliminated the fixed effect from (14) and then test the constancy of the remaining parameters. The concentrated pseudo-log likelihood function for observation i is,

$$L = c - \frac{1}{\sigma^2 NT} \sum_{i=1}^{N} e_i^{*\prime} Q_T e_i^* \tag{21}$$

where $e_i^* = (Y_i - \iota\mu - X_i\beta_{11} - X_i(\gamma_1, c_i)\beta_{2i} - Z_i(\gamma_1, c_1)\beta^*)$, $X_i(\gamma_1, c_1) = X_i \odot g(q_i, \gamma_1, c_1)$, $Z_i(\gamma_1, c_1) = [W_i \vdots W_i \odot g(q_i, \gamma_1, c)]$, $W_i = X_i \odot S_i$, $S_i = (1/T, \ldots, T/T)'$ and $X_i = (x_{i1}', \ldots, x_{iT}')'$. As before, Q_T denotes the within transformation matrix. The null hypothesis of parameter constancy is $\mathrm{H}_0 : \beta^* = 0$.

The average score evaluated at the null can be written as,

$$\begin{aligned} S_N =& \frac{1}{NT\hat{\sigma}^2} \sum_{i=1}^{N} [\hat{V}_i \vdots Z_i(\hat{\gamma}_1, \hat{c}_1)]' Q_T \hat{e}_i^* \\ =& \frac{1}{NT\hat{\sigma}^2} \sum_{i=1}^{N} [0 \vdots Z_i(\hat{\gamma}_1, \hat{c}_1)' Q_T \hat{e}_i^*]' \end{aligned} \tag{22}$$

where $\hat{V}_i = [X_i \vdots X_i(\hat{\gamma}_1, \hat{c}_i) \vdots \frac{\partial X_i(\hat{\gamma}_1, \hat{c}_1)}{\partial \gamma_1} \hat{\beta}_{21} \vdots \frac{\partial X_i(\hat{\gamma}_1, \hat{c}_1)}{\partial c_1} \hat{\beta}_{21}]$.

Using the OP estimator for the covariance matrix the LM test is,

$$LM = \frac{1}{\hat{\sigma}^2} \sum_{i=1}^{N} \hat{e}_i' Q_T Z_i(\hat{\gamma}_1, \hat{c}_1) \hat{\Sigma}^{-1} Z_i(\hat{\gamma}_1, \hat{c}_1) Q_T \hat{e}_i \tag{23}$$

where

$$\hat{\Sigma} = \frac{1}{NT} \sum_{i=1}^{N} [Z_i(\hat{\gamma}_1, \hat{c}_1)' Q_T Z_i(\hat{\gamma}_1, \hat{c}_1) - Z_i(\hat{\gamma}_1, \hat{c}_1)' Q_T \hat{V}_i (V_i' Q_T V_i)^{-1} V_i' Q_T Z_i(\hat{\gamma}_1, \hat{c}_1)] \tag{24}$$