Open access • Journal Article • DOI:10.2139/SSRN.2762523

Identification and Estimation Issues in Exponential Smooth Transition Autoregressive Models — Source link

Daniel Buncic

Institutions: Stockholm University

Published on: 11 May 2016 - Social Science Research Network

Topics: Natural exponential family, Exponential function, Exponential growth, Indicator function and Function (mathematics)

Buncic, Daniel

Working Paper

Identification and estimation issues in exponential smooth transition autoregressive models

Sveriges Riksbank Working Paper Series, No. 344

Provided in Cooperation with:
Central Bank of Sweden, Stockholm

SVERIGES RIKSBANK
WORKING PAPER SERIES

344

Identification and Estimation issues in Exponential Smooth Transition Autoregressive Models

Daniel Buncic

October 2017

Identification and Estimation issues in Exponential Smooth Transition Autoregressive Models*

Daniel Buncic†

Sveriges Riksank Working Paper Series
No. 344

October 2017

Abstract

Exponential smooth transition autoregressive (ESTAR) models are widely used in the international finance literature, particularly for the modelling of real exchange rates. We show that the exponential function is ill-suited as a regime weighting function because of two undesirable properties. Firstly, it can be well approximated by a quadratic function in the threshold variable whenever the transition function parameter γ, which governs the shape of the function, is *'small'*. This leads to an identification problem with respect to the transition function parameter and the slope vector, as both enter as a product into the conditional mean of the model. Secondly, the exponential regime weighting function can behave like an indicator function (or dummy variable) for very large values of the transition function parameter γ. This has the effect of *'spuriously overfitting'* a small number of observations around the location parameter μ. We show that both of these effects lead to estimation problems in ESTAR models. We illustrate this by means of an empirical replication of a widely cited study, as well as a simulation exercise.

Keywords: Exponential STAR, non-linear time series models, identification and estimation issues, exponential weighting function, real exchange rates, simulation analysis.

JEL codes: C13, C15, C50, F30, F44.

*I am grateful to Lorenzo Camponovo, Adrian Pagan, Paolo Giordani, Timo Teräsvirta, Xin Zhang and seminar participants at the University of St. Gallen and Sveriges Riksbank for helpful discussions and comments on earlier versions of the paper. The opinions expressed in this article are the sole responsibility of the author and should not be interpreted as reflecting the views of Sveriges Riksbank.

†Research and Modelling Division, Financial Stability Department, Sveriges Riksbank, SE-103 37, Stockholm, Sweden. E-mail: daniel.buncic@riksbank.se. Web: http://www.danielbuncic.com.

'Asymptotic results often provide but cold comfort for practical econometricians, who perforce live in a finite-sample world' (Johnston and DiNardo, 2001, page 264)

1. Introduction

The Exponential Smooth Transition Autoregressive (ESTAR) model has become one of the workhorse econometric models in the international finance literature, particulary for the modelling of real exchange rates. ESTAR models were introduced by Granger and Teräsvirta (1993) and Teräsvirta (1994) into the economics literature as a generalization of the (non-linear) exponential autoregressive model of Haggan and Ozaki (1981) and threshold time series models of Tong (1983). They have been extended to multivariate and vector error correction settings, as well as to models allowing for fractional integration and time varying conditional heteroskedasticity (see, for instance, the studies by Rothman *et al.* (2001), Milas and Legrenzi (2006), Smallwood (2008), Chan and McAleer (2002), among many others).

Despite being widely used, the exponential function employed in ESTAR models is ill-suited as a regime weighting function, or when used in a general non-linear autoregressive specification as in Haggan and Ozaki (1981). The reason for this ill-suitability is due to two undesirable features of the exponential function. The first feature is that for small values of the transition function parameter γ, which governs the shape of the function, the exponential function can be well approximated by a quadratic function in the threshold variable z_t. The consequence of this is that the slope vector attached to the non-linear regime and the transition function parameter can be shown to enter as a product into the first part of the non-linear conditional mean obtained from a Taylor series approximation of the exponential function, which leads to identification issues. In the empirical and simulation examples that we show, it can be seen that there is a nearly perfect off-setting effect of these two parameters on the conditional mean. What is particularly problematic with this scenario is that it is not a small sample issue that vanishes as the sample size increases, but rather a population property of the model.

The second feature that makes the exponential function unsuitable for econometric modelling is that for extremely large values of the transition function parameter γ, the exponential weighting function will be equal to unity for nearly all values of the transition variable, except at the point where the transition variable itself is equal to the location parameter μ, that is, at $z_t = \mu$. The effect of this on the model is that only a very small number of observations around μ receive a weight different from 1. This leads to an '*outlier fitting effect*' of the exponential function, which is similar to the way that a dummy variable is used to remove the influence of aberrant observations on the conditional mean of a model. Furthermore, in *'small samples'*, searching over μ with potentially large values in the γ parameter results in an extremely ill-behaved log-likelihood function, with many local maxima and frequent abrupt changes.[1] Although this is *'only'* a small sample problem, our simulation results indicate that it can be pervasive for sample sizes as large as 500 observations, resulting in *'large'* γ estimates in over 70% of the simulations.

There exists ample evidence of these problems with ESTAR models in the empirical literature. For instance, Michael *et al.* (1997) fit ESTAR models to real exchange rate data for a number of countries on a bilateral basis. In panels (a) and (b) of Figure 1 on page 875 in their paper, one can see that for the UK-US series, the weighting function remains well below 0.3 for the entire range of the data, while for the UK-France series, only 4 data points receive a weight in excess of 0.3, with both functions being quadratic *'looking'* in shape. Taylor and Peel (2000) use monetary fundamentals to study the evolution of exchange rates and utilize the ESTAR model to capture non-linearities in the data. From the regime weighting functions plotted in Figure 2 on page 45 of their paper, it can be seen that the transition function weights remain below 0.4 over the entire range of the data and are again quadratic looking in shape. The study by Baum *et al.* (2001) provides even stronger symptoms of a weakly identified model. The estimates of the transition function parameter γ that Baum *et al.* (2001) report in Tables 4 and 5 on page 391 of their paper are — with the exception of the WPI based real exchange rate for Norway — between 0.0042 and 0.0833! The corresponding transition function plots on pages 392 and 393 show again a quadratic looking shape.

Similar issues are evident in Sarantis (1999), Taylor *et al.* (2001), Kilian and Taylor (2003), Kapetanios *et al.* (2003), Sarno *et al.* (2006), Paya and Peel (2006), Sollis (2008), Taylor and Kim (2009), Cerrato *et al.* (2010), Pavlidis *et al.* (2011), Beckmann *et al.* (2015) and many others. The study by Beckmann *et al.* (2015) is particularly noteworthy to single out here, as the estimation results reported in Table 3 of their paper provide first hand empirical evidence of both estimation problems that we outline above. Beckmann *et al.* (2015) estimate ESTAR models on gold returns, using stock returns from 23 different equity markets as regressor and threshold variables. The model is complicated by the addition of a GARCH type volatility process on the error term in the ESTAR models that are fitted. As can be seen from the results reported in Table 3 on page 22 of their paper, the estimates of the transition

[1] The danger of a *local* sharp peak in the likelihood is that it gives the false impression of having obtained very precise parameter estimates and that the model fits the data rather well.

function parameter γ hit the lower bound of 0 of the admissible parameter space for 5 out of the 23 results, with the corresponding slope parameters becoming extremely large in absolute value, reaching a magnitude of over 22 000 for Turkey, for instance.[2] For 3 of the 23 results, the γ estimates are very large, with the one for the World equity market being over 27 000! These findings show clear symptoms of the identification and estimation issues that make ESTAR models unsuitable for econometric modelling.

The objective of this study is to outline and discuss these identification and estimation issues with ESTAR models when the transition function parameter γ takes on either *'small'* or *'large'* values. We begin by showing analytically through a Taylor series approximation of the exponential function that if the higher order terms in the expansion which enter the conditional mean are zero, the model is not identified with respect to the slope vector and the transition function parameter. We show how to use existing LM type tests applied to the higher order terms in the auxiliary regression to test for identification. We then proceed to illustrate the identification and estimation problems in an empirical setting using real world data by replicating the well known and widely cited study of Taylor *et al.* (2001). Lastly, we present a simulation analysis, where we assess the severity of the above discussed issues with regards to increasing values in the transition function parameter γ and increasing sample sizes.

The remainder of the paper is organised as follows. Section 2 describes the ESTAR model and the asymptotic properties of standard estimators that are utilized to estimate STAR models in general. In Section 3, we formalise the identification issue in ESTAR models, defining also some of the tools used to measure weak identification and a simple way to test for it. In Section 4 we provide an empirical replication as well as a simulation study. Section 5 concludes the paper.

2. Smooth Transition Autoregressive Models

Let y_t be a scalar time series which follows a general Smooth Transition Autoregressive (STAR) model, taking the form:[3]

$$y_t = \boldsymbol{x}_t \boldsymbol{\alpha} + \boldsymbol{x}_t \boldsymbol{\beta} \mathcal{G}(z_t; \gamma, \mu) + \epsilon_t, \tag{1}$$

[2] The actual value that is reported for $\hat{\gamma}$ is 0.00.

[3] It is common to interpret STAR type models either within a regime switching framework, where the transition from one regime to the other is smooth, or they can be viewed as a continuum of regimes, yielding a general parametric non-linear conditional mean function. In this paper, we will stick to the regime narrative as it fits well with the economic motivation and data that STAR models have been fitted to (see also van Dijk *et al.* (2002) for a discussion of this view and Buncic (2012) for an out-of-sample forecast evaluation.

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $(k \times 1)$ vectors of regime specific slope parameters and ϵ_t is a disturbance term assumed to follow a martingale difference sequence (MDS) with respect to the information set $\mathcal{F}_{t-1}$, with mean 0 and variance σ^2. The $(1 \times k)$ vector of control (or regressor) variables at time t is denoted by $\boldsymbol{x}_t$.[4] The regime weighting function $\mathcal{G}(z_t; \gamma, \mu)$ is a continuous and smooth function, which is bounded between $[0, 1]$. When it takes the form of an exponential function defined as $\mathcal{G}(z_t; \gamma, \mu) = 1 - \exp\left\{-\gamma(z_t - \mu)^2\right\}$, the model in (1) is known as the Exponential STAR (ESTAR) model. The parameter $\gamma \in \mathbb{R}_+$ in $\mathcal{G}(z_t; \gamma, \mu)$ determines the smoothness of the transition function, while $\mu \in \mathbb{R}$ is a threshold location parameter. The transition variable z_t can be a deterministic variable, an exogenous variable known at time $t-1$, or, as is quite frequently the case in empirical studies, the lagged endogenous variable, that is, $z_t = y_{t-q}$, for some integer $q > 0$.

A plot of the exponential function is shown in Figure 1 below. In the plot, we have normalized μ to 0, and show the function over a commonly encountered interval from -0.4 to 0.4 for threshold variable z_t and an equally spaced grid of 101 points, for 5 different different γ values ($\gamma = \{0.5, 5, 50, 500, 50000\}$). The limiting properties of the exponential weighting functions are visible graphically. For instance, we have $\lim_{\gamma\to\infty} \mathcal{G}\left(z_t; \gamma, \mu\right) = 1$ except at $z_t = \mu$, which yields 0 for $\mathcal{G}\left(z_t; \gamma, \mu\right)$, and $\lim_{\gamma\to 0} \mathcal{G}\left(z_t; \gamma, \mu\right) = 0$, for all z_t. One can notice that the shape of the exponential function is the same as that of an inverted Gaussian density. Moreover, for the two '*small*' values of γ of 0.5 and 5, the exponential function $\mathcal{G}\left(z_t; \gamma, \mu\right)$ takes the shape of a quadratic looking function in z_t. For very large values of γ, $\mathcal{G}\left(z_t; \gamma, \mu\right)$ is equal to 1 for all values of z_t, except for 3 points, of which 2 are close 1, and 1 is exactly equal to 0.

ESTAR models can be estimated by non-linear least squares (NLS) by solving the following minimisation problem:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{2} \sum_{t=1}^{T} \Big(y_t - \mathcal{C}(\boldsymbol{x}_t, z_t; \boldsymbol{\theta}) \Big)^2. \tag{2}$$

where $\boldsymbol{\theta} = [\boldsymbol{\alpha}; \boldsymbol{\beta}; \gamma; \mu]$ and $\mathcal{C}(\boldsymbol{x}_t, z_t; \boldsymbol{\theta}) = \boldsymbol{x}_t\boldsymbol{\alpha} + \boldsymbol{x}_t\boldsymbol{\beta}\mathcal{G}(z_t; \gamma, \mu)$ is the conditional mean of y_t given information up to time $t-1$. The NLS estimator defined in (2) is equivalent to a

[4] Note here that, in order to simplify the description of the model, we do not use different notation for the regressors $\boldsymbol{x}_t$ in the two '*regimes*'. That is, we do not write $\boldsymbol{x}_{1,t}$ and $\boldsymbol{x}_{2,t}$ to emphasize that they could be different. It is implicitly assumed that these regressors can be a subset of the global regressor set $\boldsymbol{x}_t = [1\ \boldsymbol{y}_{t-1}\ \mathbf{w}_{t-1}\ \boldsymbol{d}_t]$, where $\boldsymbol{y}_{t-1}$ is a vector of lagged dependent variables, $\mathbf{w}_{t-1}$ is a vector of lagged exogenous variables, and $\boldsymbol{d}_t$ is a vector of predetermined or deterministic time trend polynomials, dummy variables and/or seasonal indicators known for all t. In empirical applications, the appropriate regressor set for each regime is commonly determined by variable selection procedures (see van Dijk *et al.* (2002) for additional discussion) and is allowed to differ between the two regimes.

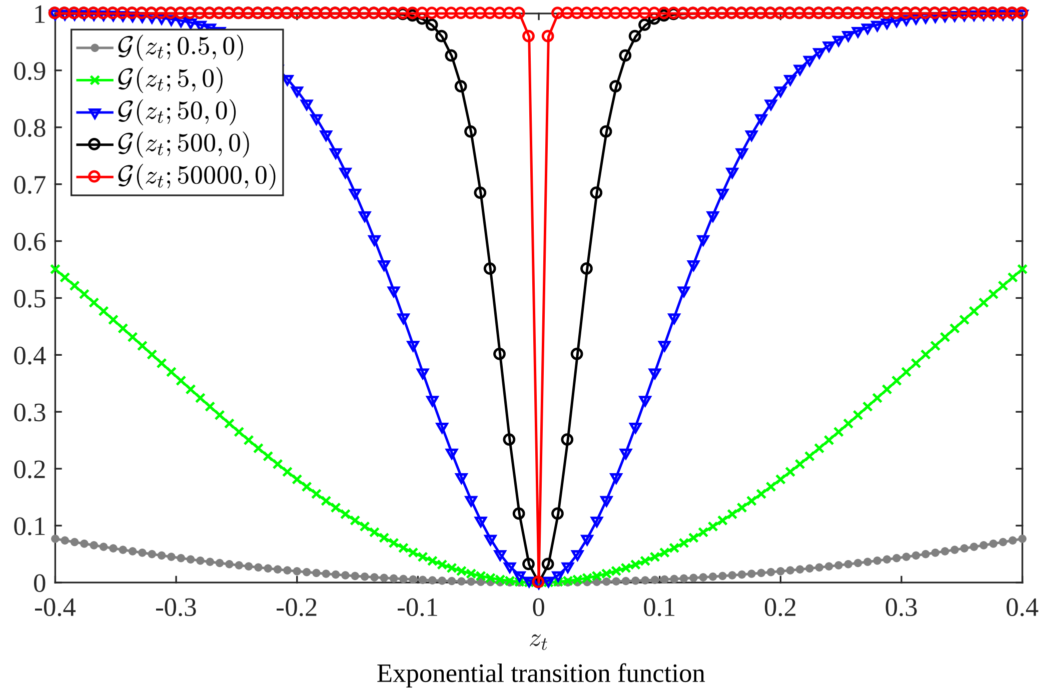


Figure 1: Transition weighting function $\mathcal{G}(z_t; \gamma, \mu)$ for a logistic function (left panel) and an exponential function (right panel) over a grid of 101 equally spaced z_t values in the interval -0.4 to 0.4, with μ fixed at 0. The transition function is evaluated over 5 different γ values. These are: $\gamma = \{0.5, 5, 50, 500, 50000\}$.

Quasi Maximum Likelihood Estimator (QMLE) if the distributional properties of ϵ_t are unknown, but are assumed to be $\mathsf{N}(0, \sigma^2)$ for estimation purposes, where $\mathsf{N}(\cdot)$ denotes a Normal (or Gaussian) random variable. Under standard regularity conditions (see Wooldridge (1994) and Pötscher and Prucha (1997) and for more general error process Chan and McAleer (2002)), the NLS/QMLE estimator $\hat{\boldsymbol{\theta}}$ in (2) will be consistent and asymptotically Normal distributed, that is, $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \rightarrow \mathsf{N}\big(0, \mathbb{V}(\hat{\boldsymbol{\theta}})\big)$, where $\boldsymbol{\theta}_0$ denotes the true parameter vector and $\mathbb{V}(\hat{\boldsymbol{\theta}})$ is the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\theta}}$.

A consistent estimate of $\mathbb{V}(\hat{\boldsymbol{\theta}})$ can be computed from a sandwich form variance-covariance matrix estimator as $\widehat{\mathbb{V}}(\hat{\boldsymbol{\theta}}) = \widehat{\mathbb{A}}^{-1}\widehat{\mathbb{B}}\widehat{\mathbb{A}}^{-1}$,where $\widehat{\mathbb{A}}$ is an estimate of the Hessian evaluated at $\hat{\boldsymbol{\theta}}$:

$$\widehat{\mathbb{A}} = T^{-1}\sum_{t=1}^{T}\left(\nabla_{\boldsymbol{\theta}}\mathcal{C}(\boldsymbol{x}_t, z_t; \hat{\boldsymbol{\theta}})\nabla_{\boldsymbol{\theta}}\mathcal{C}(\boldsymbol{x}_t, z_t; \hat{\boldsymbol{\theta}})' - \nabla^2_{\boldsymbol{\theta}}\mathcal{C}(\boldsymbol{x}_t, z_t; \hat{\boldsymbol{\theta}})\hat{\epsilon}_t\right), \tag{3}$$

and $\widehat{\mathbb{B}}$ is an estimate of the outer product of the gradient evaluated at $\hat{\boldsymbol{\theta}}$:

$$\widehat{\mathbb{B}} = T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t^2\nabla_{\boldsymbol{\theta}}\mathcal{C}(\boldsymbol{x}_t, z_t; \hat{\boldsymbol{\theta}})\nabla_{\boldsymbol{\theta}}\mathcal{C}(\boldsymbol{x}_t, z_t; \hat{\boldsymbol{\theta}})', \tag{4}$$

with $\nabla_{\boldsymbol{\theta}} = \frac{\partial\mathcal{C}}{\partial\boldsymbol{\theta}}$ and $\nabla^2_{\boldsymbol{\theta}} = \frac{\partial^2\mathcal{C}}{\partial\boldsymbol{\theta}\partial\,\boldsymbol{\theta}'}$ denoting gradient and Hessian differentiation operators, respectively, and $\hat{\epsilon}_t = y_t - \mathcal{C}(\boldsymbol{x}_t, z_t; \hat{\boldsymbol{\theta}})$ are the fitted residuals.

Given the non-linear structure of the objective function in (2), NLS/QMLE estimation of STAR models requires the use of (standard) numerical optimization procedures. Moreover, since the conditional mean function $\mathcal{C}(\boldsymbol{x}_t, z_t; \boldsymbol{\theta}) = \boldsymbol{x}_t\boldsymbol{\alpha} + \boldsymbol{x}_t\boldsymbol{\beta}\mathcal{G}(z_t; \gamma, \mu)$ is linear in the slope parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, once γ and μ are fixed at some admissible values, the estimation problem can be substantially simplified by concentrating the sum of squares function in (2) with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The numerical optimisation problem is then reduced to two dimensions only, that is, along the γ and μ parameter dimensions, which, in the simplest scenario can be obtained by a trivial two dimensional grid search (see section 5.2 in van Dijk *et al.* (2002) for a more elaborate discussion). These grid search estimates can then be used as initial values in the preferred numerical routine.[5] In all estimations, we set the upper and lower bounds on the initial grid search for γ at 1×10^{-6} and 1×10^{6}, respectively, to allow the transition parameter to take on very large and small values, and use 300 equally space points from the 10^{th} to the 90^{th} percentiles of the threshold variable z_t. Analytic first and second derivatives are used in the numerical routines that follow.

3. Identification and Estimation Issues

This section discusses identification and estimation issues that arise with the exponential function in ESTAR models when γ takes on either *'small'* or *'extremely large'* values.

3.1. Outline of the identification problem

We begin by illustrating that identification problems arises when $\mathcal{G}(z_t; \gamma, \mu)$ is well approximately by a quadratic function in z_t over the range of the observable threshold variable z_t. Consider the general ESTAR specification for y_t in (1), where for simplicity and without loss of generality, we can restrict $\boldsymbol{\alpha}$ to $\mathbf{0}$, to yield:[6]

$$y_t = \boldsymbol{x}_t\boldsymbol{\beta}\mathcal{G}(z_t; \gamma, \mu) + \epsilon_t, \tag{5a}$$

$$\mathcal{G}(z_t; \gamma, \mu) = 1 - \exp\left\{-\gamma(z_t - \mu)^2\right\}. \tag{5b}$$

[5] Note here also that it is quite common to standardize the γ parameter in $\mathcal{G}(q_{t-1}; \gamma, \mu)$ by deflating it by the standard deviation (or variance) of the threshold variable. The motivation for doing this is to have a better overview of the initial values to choose when starting the numerical optimisation procedure. We do not do this in our implementation as, firstly, we specify a fairly wide grid of γ and μ values to get initial estimates, and secondly, we want to maintain transparency with the values of γ that are used, as they are directly comparable to those parameter values estimated in the international finance literature.

[6] Alternatively, one can restrict $\boldsymbol{\alpha}$ at $\bar{\boldsymbol{\alpha}}$ and then define $\tilde{y}_t = y_t - x_t\bar{\boldsymbol{\alpha}}$, and then proceed as in (5) above, but now with $\tilde{y}_t$ in place of y_t.

Expanding $\mathcal{G}(z_t;\gamma,\mu)$ around $\gamma = 0$ by a second order Taylor series (denoted by $\widetilde{\mathcal{G}}_2(z_t;\gamma,\mu)$), yields:

$$\widetilde{\mathcal{G}}_2(z_t;\gamma,\mu) = \gamma(z_t-\mu)^2 - \frac{1}{2}\gamma^2(z_t-\mu)^4 + \mathcal{R}_2(z_t;\gamma,\mu), \tag{6}$$

where $\mathcal{R}_2(z_t;\gamma,\mu)$ is the remainder term, taking the form:

$$\mathcal{R}_2(z_t;\gamma,\mu) = \sum_{i=3}^{\infty}(-1)^{j-1}\frac{1}{j!}\gamma^j(z_t-\mu)^{2j}. \tag{7}$$

Replacing $\mathcal{G}(z_t;\gamma,\mu)$ in (5) with $\tilde{\mathcal{G}}_2(z_t;\gamma,\mu)$ from the expansion in (6) leads to the relation:

$$y_t = \boldsymbol{x}_t\boldsymbol{\beta}\,\widetilde{\mathcal{G}}_2(z_t;\gamma,\mu) + \epsilon_t \tag{8}$$

$$y_t = \boldsymbol{x}_t\boldsymbol{\beta}\left(\gamma(z_t-\mu)^2 - \frac{1}{2}\gamma^2(z_t-\mu)^4\right) + \underbrace{\boldsymbol{x}_t\boldsymbol{\beta}\mathcal{R}_2(z_t;\gamma,\mu) + \epsilon_t}_{\nu_t} \tag{9}$$

$$y_t = (z_t-\mu)^2\boldsymbol{x}_t\boldsymbol{\beta}\gamma - \frac{1}{2}(z_t-\mu)^4\boldsymbol{x}_t\boldsymbol{\beta}\gamma^2 + \nu_t. \tag{10}$$

The relation in (10) becomes a regression model with two sets of regression vectors, ie., $(z_t-\mu)^2\boldsymbol{x}_t$ and $(z_t-\mu)^4\boldsymbol{x}_t$, and two sets of slope parameters, $\boldsymbol{a} = \boldsymbol{\beta}\gamma$ and $\boldsymbol{b} = -\frac{1}{2}\boldsymbol{\beta}\gamma^2$, which can be written as:

$$y_t = (z_t-\mu)^2\boldsymbol{x}_t\boldsymbol{a} + (z_t-\mu)^4\boldsymbol{x}_t\boldsymbol{b} + \nu_t, \tag{11}$$

or in compact matrix form:

$$\mathcal{Y} = \mathcal{X}\mathcal{B} + \mathcal{V}, \tag{12}$$

where $\mathcal{Y}$ is the (time dimension) stacked vector form of y_t, the parameter matrix $\mathcal{B} = [\boldsymbol{a};\ \boldsymbol{b}]$ is of dimension $(2k \times 1)$, $\mathcal{X}$ is the $(T \times 2k)$ stacked matrix form of $[(z_t-\mu)^2\boldsymbol{x}_t\ (z_t-\mu)^4\boldsymbol{x}_t]$, and $\mathcal{V}$ is the $(T \times 1)$ stacked vector form of ν_t. Notice here that $\boldsymbol{\beta}$ and γ enter as a product into $\boldsymbol{a}$.

Suppose now that $\boldsymbol{a}$ and $\boldsymbol{b}$ are known. To be able to identify the separate effects from γ and $\boldsymbol{\beta}$ in (10) on y_t, we need to be able to recover γ and $\boldsymbol{\beta}$ from $\boldsymbol{a}$ and $\boldsymbol{b}$. This is the model source of identification. We can see that

$$\boldsymbol{a} = \boldsymbol{\beta}\gamma \tag{13}$$

$$-\frac{1}{2}\boldsymbol{a}\gamma = \underbrace{-\frac{1}{2}\boldsymbol{\beta}\gamma^2}_{=\boldsymbol{b}}$$

$$\begin{aligned} -\frac{1}{2}\boldsymbol{a}\gamma &= \boldsymbol{b} \\ \boldsymbol{a}\gamma &= -2\boldsymbol{b}, \end{aligned} \tag{14}$$

where the last relation in (14) is a system of k equations of the form

$$\begin{aligned} a_1\gamma &= -2b_1 \\ a_2\gamma &= -2b_2 \\ &\vdots \\ a_k\gamma &= -2b_k, \end{aligned} \tag{15}$$

so that γ can be recovered from either one of the following relations:

$$\gamma = -2\frac{b_i}{a_i}, \forall i = 1, \ldots, k. \tag{16}$$

Once γ is known, $\boldsymbol{\beta}$ can be obtained by inverting the $\boldsymbol{a} = \boldsymbol{\beta}\gamma$ relation in (13), that is, from

$$\boldsymbol{\beta} = \gamma^{-1}\boldsymbol{a}. \tag{17}$$

Hence, provided that at least one of the b_i is non-zero (and $|a_i| < \infty$) $\forall i = 1, \ldots, k$, we will be able to recover γ from (16), and then plug that value into (17) to find $\boldsymbol{\beta}$. Nevertheless, if $b_i = 0$ $\forall i = 1, \ldots, k$, then $\gamma = 0$, and $\boldsymbol{\beta} \to \infty$. When this is the case, we will not able to identify the separate effects of $\boldsymbol{\beta}$ and γ on the conditional mean $\mathcal{C}(\boldsymbol{x}_t, z_t; \boldsymbol{\theta})$ of y_t, and the parameters of the ESTAR model in (5) will not be estimable.

In practice, $\boldsymbol{a}$ and $\boldsymbol{b}$ are not known and need to be estimated from (11). For $\boldsymbol{a}$ and $\boldsymbol{b}$ to be estimable, we need the design matrix $\mathcal{X}'\mathcal{X}$ required to solve for $\mathcal{B}$ in (12) to be non-singular, so that the inverse $(\mathcal{X}'\mathcal{X})^{-1}$ exist. From the empirical examples that we will discuss below, it is clear that the inverse $(\mathcal{X}'\mathcal{X})^{-1}$ can be found without any numerical difficulties. For instance, the smallest eigenvalue is around 0.01 for the real exchange rate data used in Taylor *et al.* (2001), with the correlation between the two column entries in $\mathcal{X}$ being around 0.95, which is somewhat high, but does not pose any numerical difficulties. There generally thus seems to exist enough structure in the data for $\boldsymbol{a}$ and $\boldsymbol{b}$ to be estimable.

Nevertheless, even if $\boldsymbol{a}$ and $\boldsymbol{b}$ are estimable, the key question for identification of the γ and $\boldsymbol{\beta}$ parameters when working with empirical data is to determine if $\boldsymbol{b}$ is *statistically* different from zero. Such a hypothesis can be easily tested in the given set-up, as it can be implemented in the same manner that standard tests of linearity against ESTAR (or Logistic STAR in the broader context) non-linearity are implemented (see Saikkonen and Luukkonen (1988) for the general testing framework and the suggestion in Escribano and Jordá (1999) for ESTAR models). That is, the standard approach to test the null hypothesis of linearity in ESTAR models, is to take the formulation in (1), ie.,

$$y_t = \boldsymbol{x}_t\boldsymbol{\alpha} + \boldsymbol{x}_t\boldsymbol{\beta}\mathcal{G}(z_t;\gamma,\mu) + \epsilon_t, \tag{18}$$

where we again replace $\mathcal{G}(z_t;\gamma,\mu)$ by its second-order Taylor series expansion around $\gamma = 0$, to yield as before

$$y_t = \boldsymbol{x}_t\boldsymbol{\alpha} + \overbrace{(z_t-\mu)^2\boldsymbol{x}_t\boldsymbol{a} + \underbrace{(z_t-\mu)^4\boldsymbol{x}_t\boldsymbol{b}}_{\text{Part II: higher order terms}}}^{\text{Part I: standard test for non-linear}} + \nu_t. \tag{19}$$

The expression in (19) and the relations for $\boldsymbol{a}$, $\boldsymbol{b}$, and ν_t are the same as in (11) above, however with the $\boldsymbol{\alpha}$ parameter on the linear part left unrestricted. A test of linearity that determines if $\gamma = 0$ is then formulated as a Lagrange Multiplier (LM) type test in the auxiliary regression model in (19), where the null hypothesis is $\mathcal{H}_0 : \boldsymbol{a} = \boldsymbol{b} = \boldsymbol{0}$ with the alternative that at least one is non-zero (Saikkonen and Luukkonen, 1988). To warrant a non-linear model specification, at least one of the terms under Part I in (19) has to be non-zero. For the ESTAR model parameters to be identifiable, however, we need the higher order terms in Part II of (19) to be statistically different from zero, that is, $\boldsymbol{b}$ has to be non-zero in population. Given the current set-up, it is again possible to use the LM testing framework to test for identification in ESTAR models with the null hypothesis of interest now being $\mathcal{H}_0 : \boldsymbol{b} = \boldsymbol{0}$.[7]

To formalise the concept of identification, let $\boldsymbol{\theta}$ be defined as the parameter vector of interest (ie., $\boldsymbol{\theta} = [\boldsymbol{\alpha}; \boldsymbol{\beta}; \gamma; \mu]$ with dimension $[(2k+2) \times 1]$ as before), and let $\mathbf{Y}$ denote the full vector of observable random variables needed to formulate a probabilistic model for $\{y_t\}_{t=1}^T$, that is, $\mathbf{Y} = \{y_t, \boldsymbol{x}_t, z_t\}_{t=1}^T$. The likelihood function that describes the probabilistic model is then denoted by $p(\mathbf{Y};\boldsymbol{\theta})$. Further, let $\boldsymbol{\Theta}$ represent the admissible parameter space for $\boldsymbol{\theta}$, so that $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Following from Definition 2 in Rothenberg (1971, page 578), a parameter $\boldsymbol{\theta}_1 \in \boldsymbol{\Theta}$ is said to be globally identifiable if there exists no other $\boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$ which is observationally equivalent. More specifically, $\boldsymbol{\theta}_1 \in \boldsymbol{\Theta}$ is said to be globally identifiable if for any other $\boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$, we have that $p(\mathbf{Y};\boldsymbol{\theta}_1) \neq p(\mathbf{Y};\boldsymbol{\theta}_2)$ for some observable data $\mathbf{Y}$. To obtain

[7] Note here that standard inference is valid under the null hypothesis, because the remainder term $\mathcal{R}_2(z_t;\gamma,\mu)$ that enters ν_t is zero under the null hypothesis of $\gamma = 0$. Also, the only way that $\boldsymbol{b} = -\frac{1}{2}\boldsymbol{\beta}\gamma^2$ can be zero is if $\gamma = 0$, given that $\boldsymbol{\beta}$ is non-zero and a non-linear model is justified.

local identification, we only require $p(\mathbf{Y};\boldsymbol{\theta}_1)$ to be unique around some neighbourhood of $\boldsymbol{\theta}_1$.

From Theorem 1 in Rothenberg (1971, page 579), it follows that a necessary and sufficient condition for local identifiability of $\boldsymbol{\theta}$ is that the Fisher information matrix $\mathcal{I}(\boldsymbol{\theta})$ is non-singular when evaluated at the true $\boldsymbol{\theta}_0$ parameter value, where

$$\mathcal{I}(\boldsymbol{\theta}) = -\mathbb{E}\left[\left(\frac{\partial^2 \mathscr{L}(\boldsymbol{\theta}|\mathbf{Y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)\right], \tag{20}$$

$\mathbb{E}[\cdot]$ is again the expectation operator, and $\mathscr{L}(\boldsymbol{\theta}|\mathbf{Y}) = \log\left[p(\mathbf{Y};\boldsymbol{\theta})\right]$ is the log-likelihood function. Local identification of the ESTAR model can then be determined by checking the rank of $\mathcal{I}(\boldsymbol{\theta})$ for all admissible points in the parameter space $\boldsymbol{\Theta}$.

Before we discuss how to numerically assess identification issues when working with empirical data, it will be useful here to highlight that, in general, there will be two types of unidentifiable parameter scenarios to consider. The first is the case where we change one parameter (or a subset of parameters) $\boldsymbol{\theta}^{(i)}$ and there is no change in the log-likelihood function, that is,

$$\frac{\partial \mathscr{L}(\boldsymbol{\theta}|\mathbf{Y})}{\partial\boldsymbol{\theta}^{(i)}} = \mathbf{0} \tag{21}$$

for all observable data $\mathbf{Y}$, where $\boldsymbol{\theta}^{(i)}$ denotes the subset vector of parameters of interest. This is a common scenario when nuisance parameters are present in the testing model.[8] The second scenario occurs when the change in the log-likelihood function due to a change in one parameter (or set of parameters) $\boldsymbol{\theta}^{(i)}$ can be offset entirely by a combination (function) of changes in the remaining parameters $\boldsymbol{\theta}^{(-i)}$, that is:

$$\frac{\partial \mathscr{L}(\boldsymbol{\theta}|\mathbf{Y})}{\partial\boldsymbol{\theta}^{(i)}} = f\left(\frac{\partial \mathscr{L}(\boldsymbol{\theta}|\mathbf{Y})}{\partial\boldsymbol{\theta}^{(-i)}}\right), \tag{22}$$

for all observable $\mathbf{Y}$, where $f(\cdot)$ denotes this off-setting function. In our context, the off-setting case in (22) is the problem, because if $\boldsymbol{b}$ in (11) is equal to zero (and $\boldsymbol{a}$ is non-zero), then $\boldsymbol{a} = \boldsymbol{\beta}\gamma$, and we will not be able to identify the separate effects from $\boldsymbol{\beta}$ and γ on the conditional mean $\boldsymbol{x}_t\boldsymbol{\alpha} + (z_t - \mu)^2\boldsymbol{x}_t\boldsymbol{a}$ of the model, and thus its impact on the likelihood function.

[8] For instance, when wanting to formulate a test for linearity in the ESTAR model, there are two ways the ESTAR model collapses to a linear one. In the formulation that we have in (1), that would be either if $\boldsymbol{\beta} = \mathbf{0}$ or if $\gamma = 0$. Under either of these two scenarios, the other parameter that is not restricted to 0, ie., $\boldsymbol{\beta}$ or γ, together with μ, becomes a nuisance parameter, which under the null can take on any value in the admissible parameter space.

3.1.1. Identification in models fitted to empirical data

When working with models fitted to empirical data, the relations in (21) and (22) will not hold exactly, but rather only approximately, as they depend also on the random variation in $\mathbf{Y}$. In terms of the information matrix, this means that $\mathcal{I}(\boldsymbol{\theta})$ will be nearly singular or rank deficient. For a given observable sample $\mathbf{Y}$, the likelihood function $\mathscr{L}(\boldsymbol{\theta}|\mathbf{Y})$ will be flat and un-informative over certain regions of the permissable parameter space $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Consequences of near singularity will be that small changes in the sample data $\mathbf{Y}$ or the parameter vector $\boldsymbol{\theta}$ will lead to substantial changes in the estimates of $\boldsymbol{\theta}$. Moreover, as is well known with weakly identified models, parameter estimates will be highly correlated, resulting in large standard errors, and finite sample distributions will be very different from their asymptotic approximations (see Pagan and Robertson (1998)).

Weak identification is generally much harder to characterise, as it is parameter as well as data specific. To assess issues related to weak identification, it is necessary to formulate a specific parameter scenario $\boldsymbol{\theta}_0$ and then assess *numerically* how close to singularity $\mathcal{I}(\boldsymbol{\theta})$ is, conditional on the observed data $\mathbf{Y}$. How close $\mathcal{I}(\boldsymbol{\theta})$ is to singularity, and thus how weakly identified our parametric model is, can then be measured by the size of the condition number of the information matrix $\mathcal{I}(\boldsymbol{\theta})$. For a non-singular matrix $\mathcal{A}$, the condition number is defined as:

$$\text{cond}(\mathcal{A}) = \|\mathcal{A}\|\|\mathcal{A}^{-1}\|, \tag{23}$$

where $\|\cdot\|$ denotes the norm. If the Euclidean or 2-norm is used, then (23) becomes

$$\text{cond}_2(\mathcal{A}) = \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}}, \tag{24}$$

where λ denotes the singular value of $\mathcal{A}$. When λ_{min} is small and λ_{max} is large, then $\text{cond}_2(\mathcal{A})$ can become extremely large, indicating that the inverse of matrix $\mathcal{A}$ is badly conditioned. In general, when $\text{cond}_2(\mathcal{A})$ is large, then $\mathcal{A}$ will be nearly singular, which in our case will mean weak identification. The closer $\text{cond}_2(\mathcal{A})$ is to its minimum value of 1, the further away $\mathcal{A}$ is from singularity, and the stronger identified the model is.[9]

How large a condition number is needed for $\mathcal{A}$ to be considered singular? There exist not direct guidelines in the numerical computing literature to narrow down the magnitude of values, as it depends on the numerical precision of the computing environment.

[9]For a positive definite matrix $\mathcal{A}$, all the singular values will be positive and so the smallest value that $\text{cond}_2(\mathcal{A})$ in (24) can take is 1.

For instance, in Matlab (64bit), the machine precision is double, that is, 2.2204×10^{-16}.[10] Whenever the inverse or reciprocal of the 1-norm is less then machine precision, matrix $\mathcal{A}$ is deemed to be poorly conditioned and a warning of the form: '`Warning: Matrix is close to singular or badly scaled. Results may be inaccurate`' will be displayed.[11] In the econometrics literature, Greene (2011, page 999) suggests that if the 2-norm $\text{cond}_2(\mathcal{A})$ in (24) is greater than $20^2 = 400$, then the condition number is considered to be extremely large and the matrix $\mathcal{A}$ is singular numerically, which is indicative of a weakly identified model.[12] We will use the condition number of the information matrix to determine how well identified the ESTAR model is for different values of the γ parameter, and observed data $\mathbf{Y}$.

Another quantity that is frequently computed to examine identification issues in the literature on DSGE models (see for instance Iskrev (2010), among others) is the correlation of the information matrix. This matrix is defined as:

$$\widetilde{\mathcal{I}}(\boldsymbol{\theta}) = \mathcal{D}(\boldsymbol{\theta})^{-1/2}\mathcal{I}(\boldsymbol{\theta})\mathcal{D}(\boldsymbol{\theta})^{-1/2}, \tag{25}$$

where $\mathcal{D}(\boldsymbol{\theta}) = \text{diag}[\mathcal{I}(\boldsymbol{\theta})]$. The correlation matrix is a useful quantity to examine in the current setting, as it provides extra insights into which components of the parameter vector $\boldsymbol{\theta}$ are the ones causing identification problems. Having computed the condition number condenses the information down to determine whether identification in a model could be a problem, but does not reveal which parameters are affected.

3.2. Some other issues

Two other issues that frequently arise when estimating ESTAR models in finite samples are the spurious overfitting of outlier observations resulting from extremely large γ values and the severe finite sample bias of the estimates of the transition function parameter γ, when the true parameter is close to its boundary of 0. The first issue is of particular concern when working with empirical financial and/or macroeconomic series. It nevertheless also arises when estimating ESTAR models on well behaved simulated data. The problem is that the

[10] The function `eps` in Matlab provides this numerical value.

[11] In Matlab, the LAPACK reciprocal condition number is called by the function `rcond`.

[12] In Greene (2011), the condition number is formulated as the square root of the ratio of the maximum eigenvalue to the minimum eigenvalue. If this condition number is greater than 20, then the matrix is deemed nearly singular and thus difficult to invert numerically. Since we do not take the square root of the ratios, we take 20^2 as the threshold value. Note here also that the $\text{cond}_2(\mathcal{A})$ definition in (24) in terms of the singular values is more general. For positive semi-definite matrices such as the information matrix $\mathcal{I}(\boldsymbol{\theta})$, eigenvalues and singular values are identical.

exponential regime weighting function $\mathcal{G}(z_t;\gamma,\mu)$ will be equal to 1 in the limit when $\gamma \to \infty$ for all values of z_t, except at the single point where $z_t = \mu$. In finite samples, it will nearly always be possible to find a data point (or set of points) that needs to be downweighed to improve on the fit of a linear model. In such an instance, the exponential regime weighting function acts as on outlier or dummy variable fitting function, attempting to accommodate as good as possible the features of only a small number of observations around μ. We illustrate how this is omnipresent in the empirical real exchange rate series analysed in Taylor *et al.* (2001), but also when fitting ESTAR models to simulated data, which are *'well behaved'* and by construction free from outliers.

The severe finite sample bias in the estimate of γ arises due to a decisive positive skew in its sampling distribution. This bias remains even for well identified models when $\boldsymbol{\beta}$ in (18) is known or fixed. Intuitively, the positive skew is due to the true γ value that determines the data generating process being close to its lower bound of 0 of the admissible parameter space. It is well known in the econometrics literature that parameters that are close to their boundaries are likely to have finite sample distributions that are highly skewed and non-normal in general (see, for instance, the papers by Berry *et al.* (1995) and Abrevaya and Shen (2014) in the literature on random coefficient models).

4. Empirical and simulation evidence

We begin with a replication of the well known and widely cited study by Taylor *et al.* (2001) on the speed of mean reversion in real exchange rates. We then proceed with a simulation study and examine how varying the magnitude of γ and the sample size impacts on the strength of identification in ESTAR models. To keep this section as concise as possible, we have delegated the description of the real exchange rate data of Taylor *et al.* (2001) and some additional estimation results for all 4 countries to the Appendix of this paper. In Section A.2. of the Appendix, we repeat our analysis and illustrate these same problems in another study by Teräsvirta and Anderson (1992) using ESTAR models for industrial production data.

4.1. Replication of Taylor *et al.* (2001)

Taylor *et al.* (2001) fit the following ESTAR model to the real exchange rate q_t:

$$(q_t-\mu) = \alpha(q_{t-1}-\mu) + \beta(q_{t-1}-\mu)\mathcal{G}(q_{t-1};\gamma,\mu) + \epsilon_t \tag{26a}$$

$$\mathcal{G}(q_{t-1};\gamma,\mu) = 1-\exp\left\{-\gamma(q_{t-1}-\mu)^2\right\}, \tag{26b}$$

where ϵ_t is a zero mean disturbance term with variance σ^2. The relation in (26) is a simple first order version of the general form that was given in (1), with $y_t = (q_t - \mu), x_t = (q_{t-1} - \mu)$ and threshold variable $z_t = q_{t-1}$. Through standard statistical tests, Taylor *et al.* (2001, page 1030) conclude that: *"... in no case could we reject at the five percent significance level the restrictions that $\alpha = -\beta = 1$."*[13] The final restricted model that is estimated is:

$$\Delta q_t = -(q_{t-1} - \mu)\mathcal{G}(q_{t-1}; \gamma, \mu) + \epsilon_t. \tag{27}$$

Our replicated results for all 4 currency pairs are reported below in Table 1. These estimates are extremely close to the ones listed in Table 3 on page 1029 in Taylor *et al.* (2001).

Table 1: Replicated parameter estimates of the restricted model reported in Taylor *et al.* (2001).

Parameter Estimates	UK	Germany	France	Japan
$\hat{\gamma}$	0.50500170	0.29408812	0.35362630	0.18286403
$\hat{\mu}$	−0.11243280	0.00895991	−0.00591922	−0.51083432
$\hat{\sigma}$	0.03320309	0.03448975	0.03286561	0.03331077
Log-like. ESTAR	570.03185026	559.12033546	572.96391675	569.10258492
Log-like. Cubic model	570.06613111	559.14958988	572.99204741	569.13408243
Log-like. AR(1) model	568.99987094	557.92278405	571.80528152	567.46500061

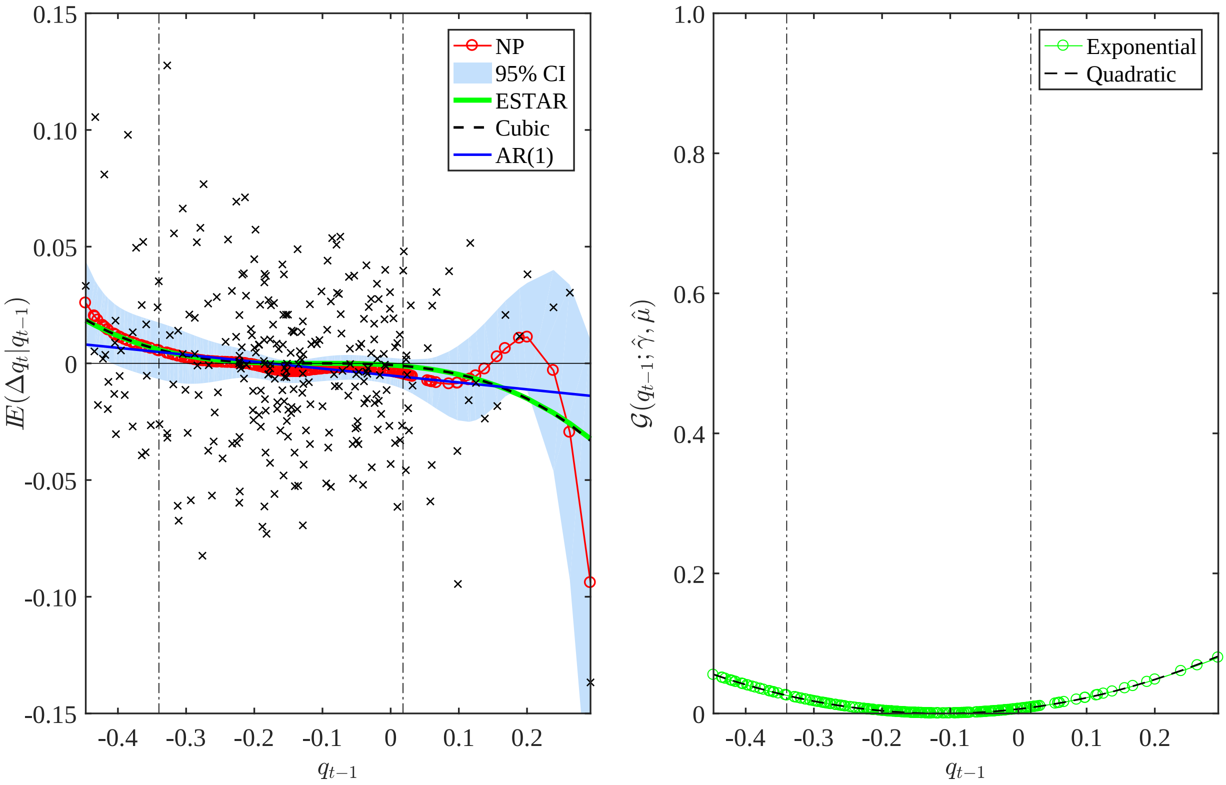
Notes: This table reports our replicated parameter estimates of the restricted ESTAR model of Taylor *et al.* (2001) (see their estimates reported in Table 3 on page 1029 for comparison). The restricted ESTAR model takes the form: $\Delta q_t = -(q_{t-1} - \mu)\mathcal{G}(q_{t-1}; \gamma, \mu) + \epsilon_t$. Log-likelihood functions of the AR(1) and Cubic models are based on the two regression equations: $\Delta q_t = \delta_1(q_{t-1} - \mu) + \epsilon_t$ and $\Delta q_t = \delta_3(q_{t-1} - \mu)^3 + \epsilon_t$.

As can be seen from Table 1, the estimates of γ are rather small. To understand how the conditional mean and the regime weighting function look like for these parameter estimates over the range of the data that we observe, we show plots of the conational mean and the weighting function for the UK series in Figure 2.[14] The green line (with circles) shows the implied conditional mean and weighting function at the ESTAR parameter estimates, while the black dashed lines show corresponding cubic and quadratic fits in $(q_{t-1} - \mu)$.[15] Examining the plot of the weighting function in Panel (b) of Figure 2, we can see how flat and weakly curved the weighting function is over the range of the threshold variable q_{t-1}. Such a shape is well approximated by a quadratic function in $(q_{t-1} - \mu)$. The fitted model is thus likely to be unidentified.

[13] In their notation, the text on page 1030 is $\beta_1 = -\beta_1^* = 1$.

[14] We use the UK real exchange rate series results here as a representative to illustrate our point, and report results for all 4 series in the Appendix. In the conditional mean plot in the left hand Panel (a), we also superimpose a linear AR(1) fit, a non-parametric fit using a local linear kernel regression estimate with 95% confidence bands, as well as a scatter of the data.

[15] The log-likelihood of a cubic regression of the form $\Delta q_t = \delta_3(q_{t-1} - \mu)^3 + \epsilon_t$ is reported in the second last row of Figure 2 for all 4 series. The quadratic fit of the exponential weighting function in Panel (b) is obtained as $-\hat{\delta}_3(q_{t-1} - \mu)^2$, while that of the cubic plotted in Panel (a) is from the fit $\hat{\delta}_3(q_{t-1} - \mu)^3$.



(a) Conditional mean

(b) Exponential weighting function

Figure 2: Plots of the conditional mean and the weighting function for the UK series.

4.1.1. Identification in Taylor *et al.* (2001)

To formally examine how *strongly* identified the model is with respect to the γ and β parameters at their estimates, we compute the condition number of the information matrix as outlined in Section 3. To illustrate this computation, let $\mathcal{I}_3(\hat{\boldsymbol{\theta}})$ be the (3×3) dimensional information matrix defined by $\widehat{\mathbb{A}}_3$, where $\hat{\boldsymbol{\theta}} = [\bar{\beta};\, \hat{\gamma};\, \hat{\mu}]$, with $\bar{\beta} = -1$ and $\hat{\gamma}$, $\hat{\mu}$ the ML estimates reported in Table 1. The Hessian matrix $\widehat{\mathbb{A}}$ is given in (3).[16] Computing the condition number of $\mathcal{I}_3(\hat{\boldsymbol{\theta}})$ as the ratio of the largest to the smallest singular values as defined in (24) for the four countries of interest yields values of 833, 1436, 1507, and 1875 for the UK, Germany, France and Japan, respectively.[17] These condition numbers are substantially larger than the threshold value of $20^2 = 400$ suggested in Greene (2011). The correlation matrix of the information matrix as defined in (25) at $\hat{\boldsymbol{\theta}} = [\bar{\beta};\, \hat{\gamma};\, \hat{\mu}]$, is shown in Table 2. The correla-

[16] The 3 subscript on $\widehat{\mathbb{A}}_3$ signifies that the first row and first column entries of $\widehat{\mathbb{A}}$ pertaining to the α parameter that is considered fixed or known at $\alpha = 0$ have been removed. Also, to be exact, considering a Normal likelihood function, the information matrix $\mathcal{I}(\hat{\boldsymbol{\theta}})$ is equal to $-H(\hat{\boldsymbol{\theta}})$, where $H(\hat{\boldsymbol{\theta}})$ is the Hessian defined as $\frac{1}{\sigma^2}\sum_{t=1}^{T}\left(\nabla^2_{\boldsymbol{\theta}}\mathcal{C}(\boldsymbol{x}_t, z_t; \hat{\boldsymbol{\theta}})\hat{\epsilon}_t - \nabla_{\boldsymbol{\theta}}\mathcal{C}(\boldsymbol{x}_t, z_t; \hat{\boldsymbol{\theta}})\nabla_{\boldsymbol{\theta}}\mathcal{C}(\boldsymbol{x}_t, z_t; \hat{\boldsymbol{\theta}})'\right)$ evaluated at the ML estimate $\hat{\boldsymbol{\theta}}$. Given the way that $\widehat{\mathbb{A}}$ in (3) is defined, that is, from the NLS minimisation as opposed to log-likelihood maximisation, we have that $\mathcal{I}(\hat{\boldsymbol{\theta}}) = \widehat{\mathbb{A}}$, ignoring the scaling by σ^2. We use the analytic Hessian in all calculations of the information matrix.

[17] These condition numbers were computed after re-scaling the columns of $\mathcal{I}_3(\hat{\boldsymbol{\theta}})$ by each column's norm to have unit length. This is done to avoid any issues related to the scaling of the information matrix (see also the discussion on page 999 in Greene (2011)).

tion between the first two elements of $\hat{\theta}$ corresponding to the β and γ parameters exceeds 0.99 in absolute value for all 4 real exchange rate series. Movements in these two parameters off-set one another (nearly) one for one.

Table 2: Correlation matrix $\widetilde{\mathcal{I}}_3(\hat{\theta})$.

UK	Germany	France	Japan
$\begin{bmatrix} 1 & - & - \\ -0.9948 & 1 & - \\ -0.2346 & -0.2371 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & - & - \\ -0.9949 & 1 & - \\ -0.4365 & 0.4412 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & - & - \\ -0.9952 & 1 & - \\ -0.4132 & -0.4171 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & - & - \\ -0.9959 & 1 & - \\ -0.4010 & -0.4053 & 1 \end{bmatrix}$

Notes: This table reports the correlation matrix $\widetilde{\mathcal{I}}_3(\hat{\theta})$ of the parameter estimates with $\bar{\beta} = -1$ and γ and μ evaluated at their ML estimates. The parameter ordering of the $\widetilde{\mathcal{I}}_3(\hat{\theta})$ matrix is $[\bar{\beta};\, \hat{\gamma};\, \hat{\mu}]$.

We can now more formally test for identification by examining the importance of the higher order terms on the conditional mean using the LM framework discussed in Section 3. Substituting $\mathcal{G}(q_{t-1};\gamma,\mu)$ in (26) with $\widetilde{\mathcal{G}}_2(q_{t-1};\gamma,\mu)$ defined in (6) yields the auxiliary regression model:

$$\begin{aligned} \Delta q_t &= \beta(q_{t-1}-\mu)\left[\gamma(q_{t-1}-\mu)^2 - \frac{1}{2}\gamma^2(q_{t-1}-\mu)^4\right] + \nu_t \\ &= a(q_{t-1}-\mu)^3 + \underbrace{b(q_{t-1}-\mu)^5}_{\text{Part II: higher order terms}} + \nu_t, \end{aligned} \tag{28}$$

where $a = \beta\gamma$ and $b = -\frac{1}{2}\beta\gamma^2$. For the ESTAR model to be identified, the higher order terms above Part II in (28) need to be significantly contributing to the likelihood function, that is, b has to be non-zero in population. Using the LM testing framework, we compute the χ^2 version of the LM test as LM $= T(SSR_0 - SSR_1)/SSR_0$, where SSR_0 and SSR_1 are the sum of squared residuals (SSR) from the restricted and unrestricted models of (28). These test results are reported in Table 3 below. Comparing the magnitudes of the SSR_0 and SSR_1 values, it is clear that the higher order term adds very little to the conditional mean and hence the likelihood of the model, suggesting that the null hypothesis of a lack of identification cannot be rejected.[18]

4.1.2. Can we estimate the model without the $\beta = -1$ restriction?

We predict that $\gamma \to 0$, and β becomes very large in absolute value to accommodate the cubic structure of the unidentified conditional mean of the model. Estimation results for the

[18] Note here, that since we are only testing one restriction (ie., $b = 0$), we have one degree of freedom in the χ^2 critical values. This test coincides with a standard $t-$test, whose values can be recovered from the square root of the LM test values provided in the second last row of Table 3.

Table 3: LM test results of significance of higher order terms.

Test statistic	UK	Germany	France	Japan
SSR_0	0.346801	0.369092	0.335987	0.353700
SSR_1	0.345935	0.368579	0.335496	0.353126
LM	0.717104	0.399049	0.418899	0.465982
$p-$value	0.397096	0.527581	0.517487	0.494842

Notes: This table reports the LM test significance results of the higher order terms in (28). The terms SSR_0 and SSR_1 are the sum of squared residuals of the restricted and unrestricted models, respectively, and LM is the Lagrange-Multiplier test statistic computed as $T(SSR_0 - SSR_1)/SSR_0$. The last row reports asymptotic $p-$values of the LM test, with 1 degree of freedom.

model:

$$\Delta q_t = \beta(q_{t-1} - \mu)\mathcal{G}(q_{t-1}; \gamma, \mu) + \epsilon_t. \tag{29}$$

are reported in Table 4. As can be seen from the β and γ estimates, there is an offsetting effect. The estimate of the γ parameter converges towards 0, while $\hat{\beta}$ goes towards a large negative value. Note that the lower bound in the initial grid search for γ was set at 1×10^{-6}, so the values that are estimated are only marginally lower than that. What is important to point out here is that the log-likelihood function of the ESTAR and the cubic models are numerically identical up to 6 decimal points. Relating this to the identification discussion of Section 3, the situation that we have here is $\hat{\gamma} \to 0$, $\hat{\beta} \to -\infty$, with $\hat{\beta}\hat{\gamma} \to \hat{\delta}_3 \neq 0$, where $\hat{\delta}_3$ is the coefficient obtained from a cubic regression, ie., of Δq_t on $(q_{t-1} - \mu)^3$. The higher order terms in the second order approximation listed in (28) related to $\hat{\beta}\hat{\gamma}^2$ go to 0 due to $\hat{\gamma} \to 0$. Given the information in the data, the conditional means of the cubic regression and the ESTAR model provide the same fit. This is also evident visually from Panel (a) of Figure 2.

Table 4: Replicated parameter estimates of the Taylor *et al.* (2001) model with the unit-root restriction only.

Parameter Estimates	UK	Germany	France	Japan
$\hat{\beta}$	−504662.45687917	−305284.18535728	−368394.40126473	−184561.06430017
$\hat{\gamma}$	0.00000098	0.00000096	0.00000095	0.00000098
$\hat{\mu}$	−0.11230079	0.00924026	−0.00571998	−0.51121351
$\hat{\sigma}$	0.03319913	0.03448624	0.03286239	0.03330712
Log-like. ESTAR	570.06613104	559.14958979	572.99204734	569.13408227
Log-like. Cubic model	570.06613111	559.14958988	572.99204741	569.13408243
Log-like. AR(1) model	568.99987094	557.92278405	571.80528152	567.46500061

Notes: This table reports the replicated parameter estimates of the ESTAR model of Taylor *et al.* (2001) which only imposes the unit-root restriction on the inner regime. This ESTAR model takes the form: $\Delta q_t = \beta(q_{t-1} - \mu)\mathcal{G}(q_{t-1}; \gamma, \mu) + \epsilon_t$. Log-likelihood functions of the AR(1) and Cubic models are based on the two regression equations: $\Delta q_t = \delta_1(q_{t-1} - \mu) + \epsilon_t$ and $\Delta q_t = \delta_3(q_{t-1} - \mu)^3 + \epsilon_t$.

4.1.3. Estimating the model without any restrictions

To illustrate that the exponential weighting function has a propensity to overfit outliers by acting like a dummy variable function, we estimate the unrestricted ESTAR model of Taylor *et al.* (2001) defined in (26a), and re-parameterised here as:

$$\Delta q_t = (\alpha - 1)(q_{t-1} - \mu) + \beta(q_{t-1} - \mu)\mathcal{G}(q_{t-1}; \gamma, \mu) + \epsilon_t, \tag{30}$$

where $(q_{t-1} - \mu)$ is subtracted from both sides of the original relation in (26) to visualise the conditional mean $\mathbb{E}(\Delta q_t | q_{t-1})$ in the same way as before. These estimation results are reported in Table 5. As can be seen from the estimates reported in Table 5, $\hat{\gamma}$ attains extremely large values. The slope coefficients $(\hat{\alpha} - 1)$ and $\hat{\beta}$ are mirror images, in the sense that their absolute magnitudes are very similar, differing only in their sign. All log-likelihoods are larger than the estimates from the restricted models of Taylor *et al.* (2001).

Table 5: Replicated parameter estimates of the entirely unrestricted Taylor *et al.* (2001) model.

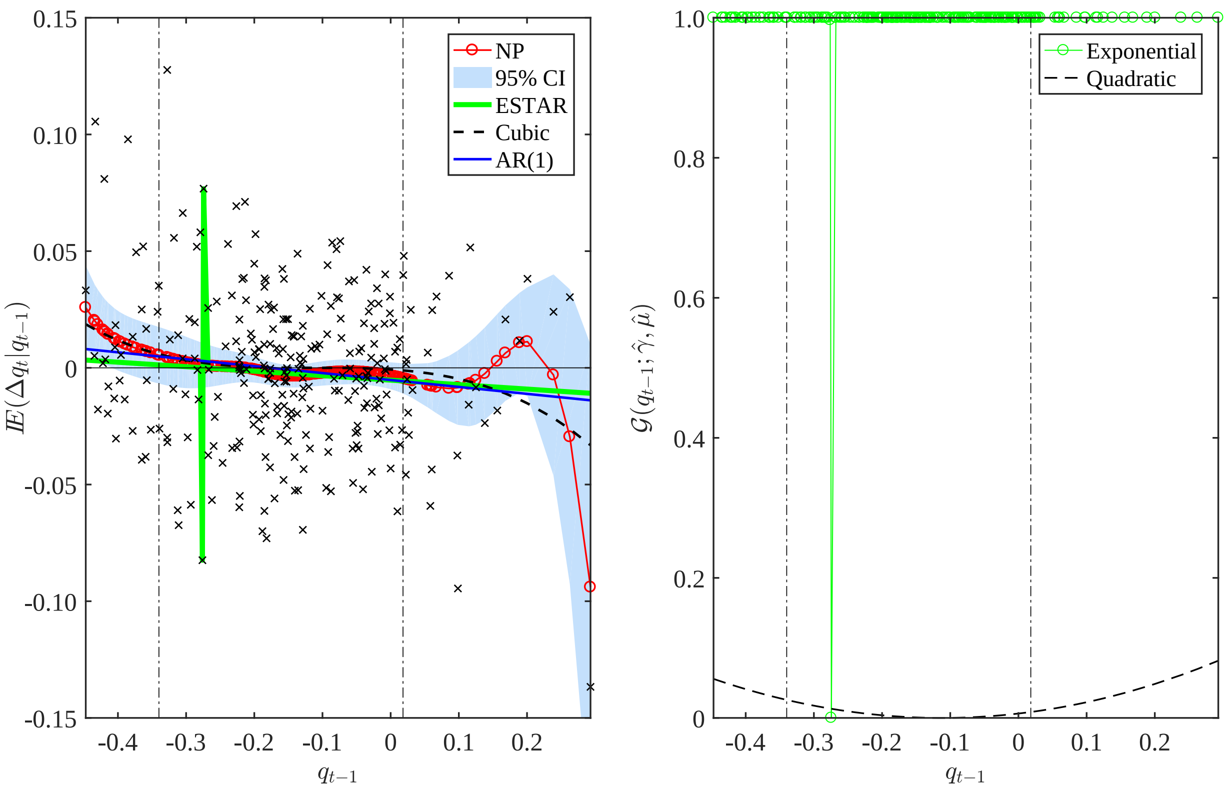
Parameter Estimates	UK	Germany	France	Japan
$(\hat{\alpha} - 1)$	14183.23931549	−80.43652137	4895.09914479	340594.58805192
$\hat{\beta}$	−14183.25851544	80.41534978	−4895.11461596	−340594.59946710
$\hat{\gamma}$	1937108.34779812	220354.95060881	1129966.10535952	1587721.65109557
$\hat{\mu}$	−0.27437231	−0.15258840	−0.02174839	−0.70815477
$\hat{\sigma}$	0.03272230	0.03387141	0.03223191	0.03256410
Log-like. ESTAR	574.21811145	564.31246531	578.55178003	575.60897466
Log-like. Cubic model	570.06613111	559.14958988	572.99204741	569.13408243
Log-like. AR(1) model	568.99987094	557.92278405	571.80528152	567.46500061

Notes: This table reports the replicated parameter estimates of the unreported (entirely) unrestricted ESTAR model of Taylor *et al.* (2001). The unrestricted ESTAR model takes the form: $\Delta q_t = (\alpha - 1)(q_{t-1} - \mu) + \beta(q_{t-1} - \mu)\mathcal{G}(q_{t-1}; \gamma, \mu) + \epsilon_t$. Log-likelihood functions of the AR(1) and Cubic models are based on the two regression equations: $\Delta q_t = \delta_1(q_{t-1} - \mu) + \epsilon_t$ and $\Delta q_t = \delta_3(q_{t-1} - \mu)^3 + \epsilon_t$.

What do these models fit? We plot the implied conditional means $\mathbb{E}(\Delta q_t | q_{t-1})$ as well as weighting functions $\mathcal{G}(q_{t-1}; \gamma, \mu)$ at the estimates of Table 5 to obtain a visual feel for the models. In Figure 3 we show these for the UK real exchange rate series, where the figure has the same format as Figure 2.[19] The fitted exponential weighting function for the UK series in Panel (b) of Table 5 shows that $\mathcal{G}(q_{t-1}; \gamma, \mu)$ is numerically equal to 1 (up to 8 decimal places) for all but two observations. One is 0.99668679, which is also quite close to 1, with the second being 0.00005427. The function thus assigns a zero weight to one observation, effectively acting as a dummy indicator for that observation. The effect of this on the implied conditional mean is visible from Panel (a) of Table 5. The conditional

[19] To conserve space and avoid repetition, plots for all 4 series are shown in the Appendix.

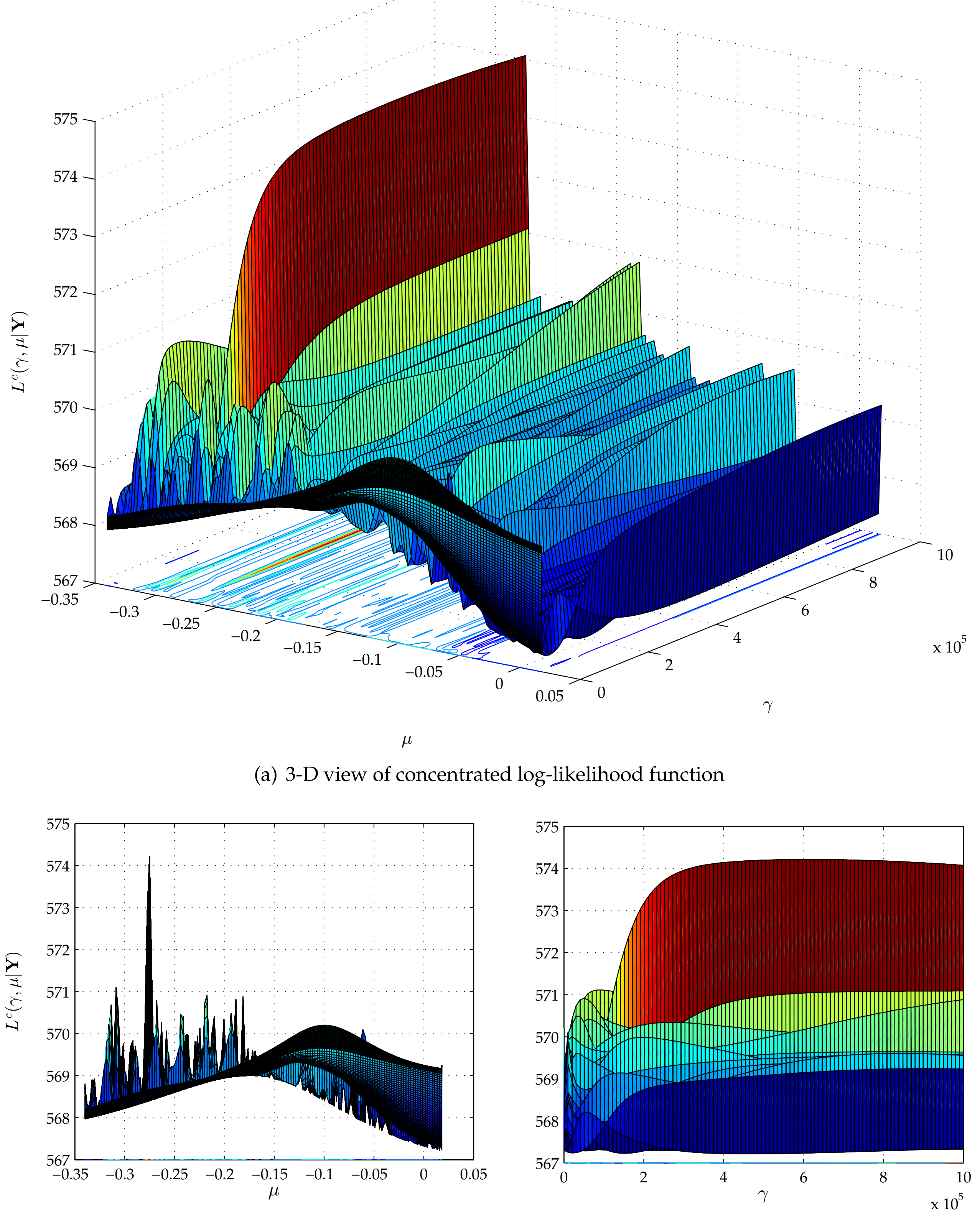
mean is a linear function of q_{t-1} over the entire range of the threshold variable, with the only exception being two data points around $\hat{\mu}$, where it spikes up and down. The AR(1) conditional mean (shown by the solid blue line in Panel (a) of Table 5) largely overlaps with the ESTAR model's fit.



(a) Conditional mean (b) Exponential weighting function

Figure 3: Conditional mean and transition weighting function plots for the estimated unrestricted ESTAR model defined in (30) for the real exchange rate series of the UK and Germany, shown respectively in panels (a) and (b).

In Figure 4 we show surface plots from various angles of the concentrated log-likelihood function over the γ and μ grids that we consider in the initialisation of the numerical routines for the UK real exchange rate series. What is evident from these plots is that the concentrated log-likelihood function is extremely ill-behaved, with a large number of local maxima. This ill-behaviour is not only visible for extremely large values of γ, but also for more moderately sized values. Searching over the threshold location parameter μ allows for large γ values and therefore admits the fitting of a very few extreme observations, which results in the highly irregular shape of the concentrated log-likelihood function and the conditional means that we observe.



(a) 3-D view of concentrated log-likelihood function

(b) μ-axis view

(c) γ-axis view

Figure 4: Plot of the concentrated log-likelihood function of the unrestricted ESTAR models defined in (30) for the UK real exchange rate series over the μ and γ search grids. The top Panel (a) shows the 3-D axis view, the two bottom Panels (b) and (c) show the $\mu-$axis and $\gamma-$axis views.

4.2. Simulation evidence

We proceed to study the above outlined problems within a controlled simulation experiment to abstract from issues that could arise due to particular features in the empirical real exchange rate data. To keep the considered calibrations limited, we take the UK parameter estimates as our baseline *'true'* model, and then consider increasing values of γ as additional parameter scenarios to examine its effect on identification. The model that we simulate from is given in (26). The true set of parameters used in the simulation are: $\mu_0 = -0.1124$, $\alpha_0 = 1, \beta_0 = -1$, and $\sigma_0 = 0.0332$. The error term ϵ_t is drawn from a standard Normal distribution. The 6 different γ values that we consider are $\gamma_0 = \{0.505, 5, 50, 250, 500, 1000\}$. The 0.505 value is the empirical estimate. The other γ values approximate various shapes in the transition function and also the implied conditional mean of the ESTAR model. In total, we simulate $S = 10\,000$ sequences, with samples of size $T = \{288, 500, 1500, 5000\}$.

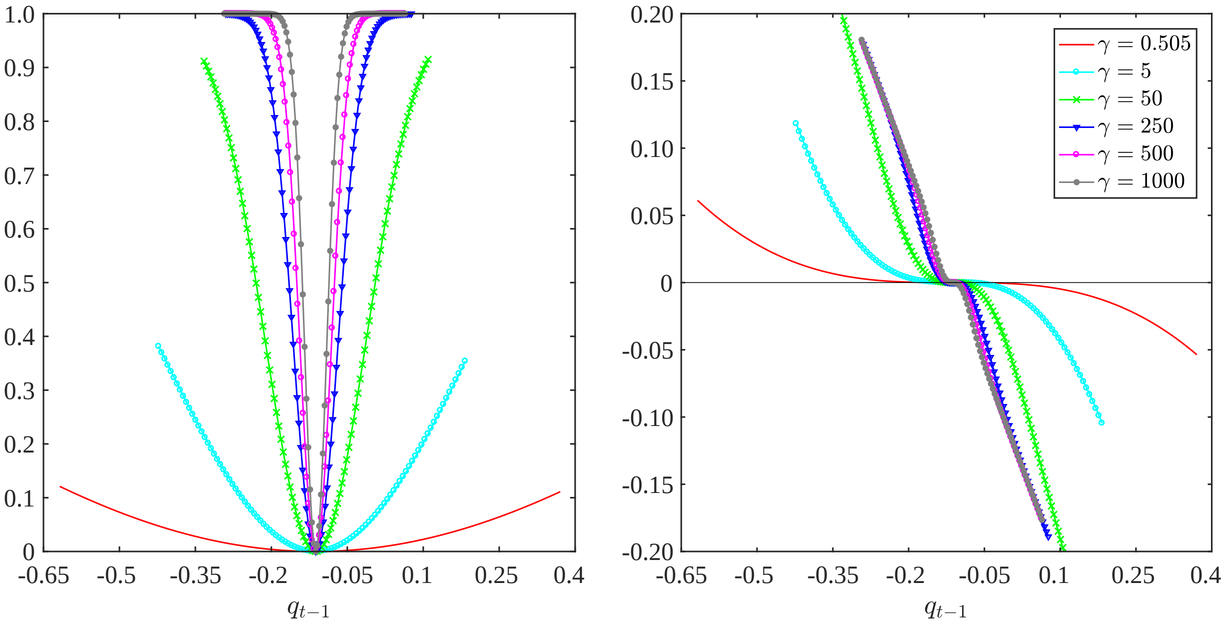
As the shape of the transition weighting function is important for understanding identification issues in ESTAR models, we find it informative here to provide a visual overview of the various shapes that $\mathcal{G}(q_{t-1}; \gamma, \mu)$ and the conditional mean can take under the above considered true model calibrations.[20] These plots are shown in the left and right hand panels of Figure 5, respectively. Each of the weighting function and conditional mean plots under the different γ values are drawn over the maximal range of simulated q_{t-1} values obtained under that particular γ_0 simulation. The overall $x-$axis range in Figure 5 is set to the widest range of (simulated) q_{t-1} values obtained from the $\gamma_0 = 0.505$ calibration. This range is from about -0.65 to 0.4, which is somewhat wider than the range observed in the empirical UK real exchange rate series from -0.45 to 0.3.

From the plots in Figure 5 we can see that the lowest γ value generates the widest range for q_{t-1}, while for larger γ values, the q_{t-1} range becomes narrower. This is particularly evident for γ values of $250, 500$ and 1000, where the transition weighting function covers all feasible values in the $[0, 1]$ interval of $\mathcal{G}(q_{t-1}; \gamma, \mu)$, with the q_{t-1} range being rather narrow between -0.35 and 0.1. The key point to take away from these plots is that the calibrations that we choose for the simulations cover a wide range of possible shapes of the transition weighting function $\mathcal{G}(q_{t-1}; \gamma, \mu)$. As the shape of $\mathcal{G}(q_{t-1}; \gamma, \mu)$ determines how weakly identified the ESTAR model is, we can see that there is a fairly wide coverage of different shapes and hence identification scenarios.

4.2.1. Identification results

In Table 6 we report arithmetic averages of the correlations between the β and γ parameters and the condition numbers of the (3×3) information matrix $\mathcal{I}_3(\hat{\theta})$, where we follow again

[20]To plot the different possible shapes, we generate an equally spaced grid of 100 data points over the range of the simulated q_{t-1} series, ie., from $\min(q_{t-1})$ to $\max(q_{t-1})$, for the 6 considered calibrations for γ. We then evaluate $\mathcal{G}(q_{t-1}; \gamma, \mu)$ as well as the conditional mean $\mathbb{E}(\Delta q_t | q_{t-1}) = -(q_{t-1} - \mu)\mathcal{G}(q_{t-1}; \gamma, \mu)$ over this equally spaced grid.



(a) Transition weigthing function

(b) Implied conditional mean

Figure 5: Exponential transition function weights $\mathcal{G}(z_t; \gamma, \mu)$ and implied conditional means of the ESTAR models that we simulate from. These are evaluated and plotted over an equally spaced grid of 101 points from $\min(q_t)$ to $\max(q_t)$ for each of the considered γ grid values. Each line in the plots is drawn over the maximum range of possible q_t values from the simulated data for a given γ value.

the same computational approach as in Section 4.1 with the empirical data. The estimated ESTAR model is the restricted model in (27). The (3×3) information matrix is denoted by $\mathcal{I}_3(\hat{\boldsymbol{\theta}})$ with $\hat{\boldsymbol{\theta}} = [\bar{\beta}; \hat{\gamma}; \hat{\mu}]$, where $\bar{\beta} = -1$ is the Taylor *et al.* (2001) restriction, and $\hat{\gamma}$ and $\hat{\mu}$ are the estimates obtained from the simulated data. All averages are computed over $S = 10\,000$ simulated sequences. Row entries in Table 6 correspond to the six different γ calibrations that are used, ie., $\{0.505, 5, 50, 250, 500, 1000\}$, while column entries list the considered sample sizes $T = \{288, 500, 1500, 5000\}$.

Table 6: Correlation and condition numbers.

T / γ_0	288		500		1500		5000	
	corr(β,γ)	cond$_2(\mathcal{I}_3)$	corr(β,γ)	cond$_2(\mathcal{I}_3)$	corr(β,γ)	cond$_2(\mathcal{I}_3)$	corr(β,γ)	cond$_2(\mathcal{I}_3)$
0.505	−0.9970	7241.19	−0.9978	4780.68	−0.9987	10686.76	−0.9993	14630.15
5	−0.9939	1607.43	−0.9953	2201.88	−0.9971	7583.23	−0.9983	5614.71
50	−0.9780	606.54	−0.9818	844.88	−0.9867	520.60	−0.9889	349.06
250	−0.9011	57.34	−0.9065	40.80	−0.9086	23.21	−0.9089	21.50
500	−0.8064	15.10	−0.8108	13.00	−0.8125	10.12	−0.8127	9.80
1000	−0.6634	6.60	−0.6679	5.66	−0.6711	5.23	−0.6717	5.14

Notes: This table reports averages of the correlation between the β and γ parameter estimates and the condition number of the (3×3) information matrix $\mathcal{I}_3(\hat{\boldsymbol{\theta}})$ for various sample sizes and γ_0 values. We consider γ_0 values over the grid $\in \{0.505, 5, 50, 250, 500, 1000\}$ and sample sizes of $T \in \{288, 500, 1500, 5000\}$. These results are based on arithmetic averages computed over $10\,000$ simulations. The correlation between β and γ, denoted by corr(β,γ), is computed as the {2,1} element of the correlation matrix $\widetilde{\mathcal{I}}_3(\hat{\boldsymbol{\theta}})$ defined in (25), with $\hat{\boldsymbol{\theta}} = [\bar{\beta}; \hat{\gamma}; \hat{\mu}]$ and $\bar{\beta} = -1$. The condition number, denoted by cond$_2(\mathcal{I}_3)$ in the table, is defined in (24) and is also evaluated at $\hat{\boldsymbol{\theta}}$.

From the correlation and condition number entries in Table 6 we can see that for the smaller γ_0 values, the identification problem is not a finite sample issue, but persists for all sample sizes we consider. For instance, when γ takes the value 0.505 obtained in Taylor *et al.* (2001), the average correlation between β and γ increases (in absolute value) steadily toward 1 as the sample size increases from 288 observations to 5000. At the same time, the condition number more than doubles, suggesting that the model becomes *'less identified'* as the sample size increases. With an increasing γ_0 size, the correlations as well as the condition numbers decrease steadily. This result is most clearly seen from the change in the size of the correlations and condition numbers when moving from a γ_0 value of 50 to 250. The correlations decrease in absolute value from around 0.98 to about 0.90, with the condition numbers dropping more than ten fold to values of about 57 or less, which are well below the threshold value of $20^2 = 400$ suggested by Greene (2011). For γ_0 values of 5 and 50, the correlations and condition numbers are still rather large, while for higher γ_0 values of 500 and 1000, these numbers decrease, suggesting that β and γ can now be identified from the data. Overall, these results confirm our hypothesis of identification problems in ESTAR models being linked to the magnitude of the γ parameter, which determines the shape of the transition weighting function $\mathcal{G}(q_{t-1}; \gamma, \mu)$.

4.2.2. Finite sample distribution and bias

In Figure 6 and Table 7 we plot the finite sample distribution, and report the finite sample bias $\hat{\gamma}$ under the 6 different γ_0 calibrations and 4 sample sizes that we consider. Figure 6 is arranged in 4 rows and 6 columns corresponding to the considered sample sizes and γ_0 values, respectively. In each plot in Figure 6, we show histogram as well as Kernel density estimates of $\hat{\gamma}$ over the 10 000 $\{\hat{\gamma}_s\}_{s=1}^{S}$ sequences of simulated data. We also superimpose a Normal distribution centered at the sample average and scaled by the sample variance of $\{\hat{\gamma}_s\}_{s=1}^{S}$. Vertical red lines mark the location of the sample means. As can be seen from the plots in Figure 6, for small γ_0 values of 0.505 and 5, the sampling distributions of $\hat{\gamma}$ are severely skewed to the right. The skew can persist for sample sizes as large as 1500 observations. For larger γ_0 values and sample sizes, the sampling distributions become more symmetric and Normal looking. What is interesting to see here, nevertheless, is that the skew is still quite noticeable for even the two largest γ_0 values of 500 and 1000, and samples of size 288 and 500 observations.

Table 7 is arranged in 6 rows and 4 columns, with each column reporting the bias in absolute terms, as well as percentage of γ_0. Bias is computed as: $\text{Bias}_{\gamma_0}[\hat{\gamma}] = \mathbb{E}_{\gamma}[\hat{\gamma}] - \gamma_0$,

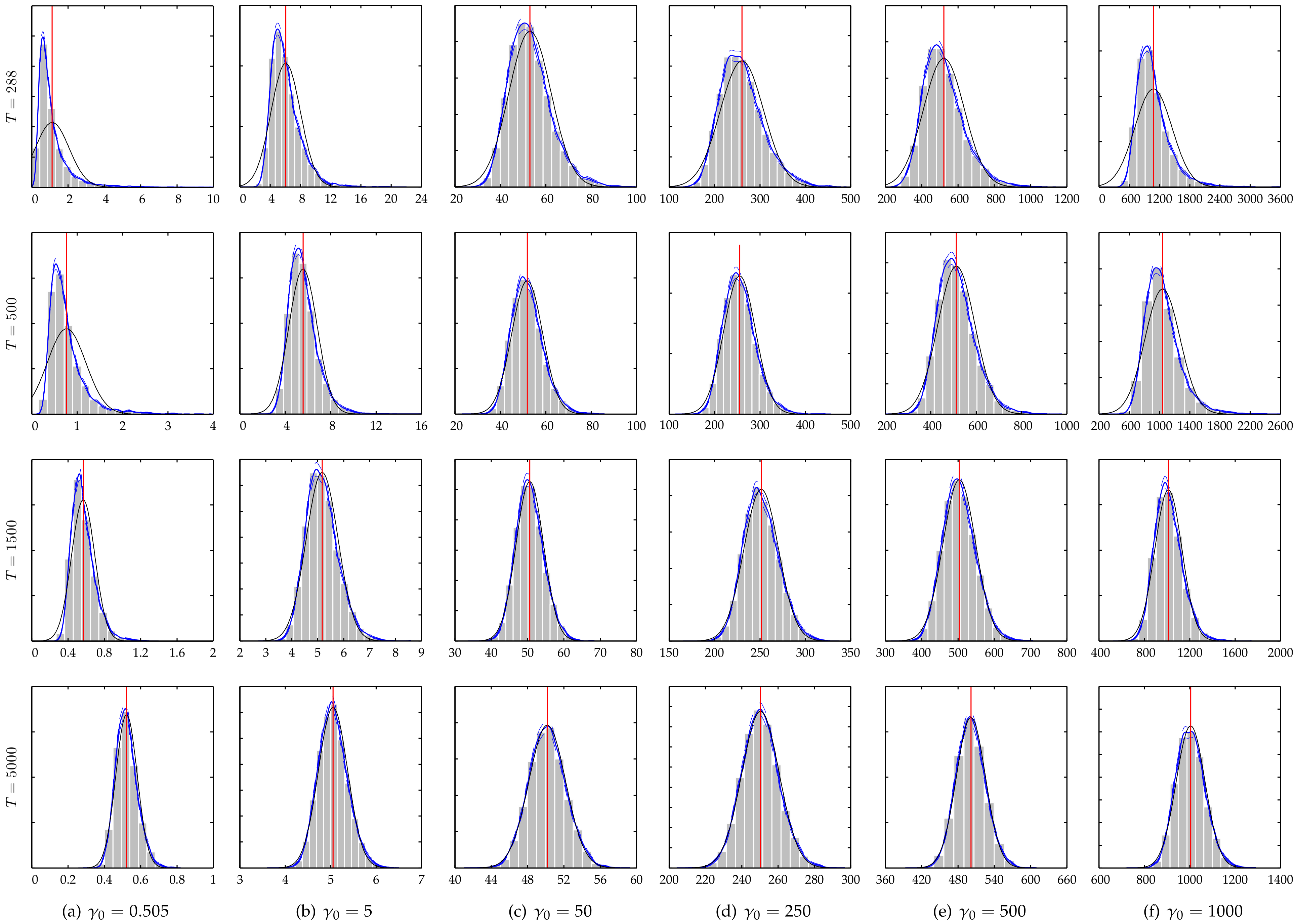


Figure 6: Finite sample distribution of $\hat{\gamma}$ for various sample sizes and γ_0 values. Columns show the results for $\gamma_0 \in \{0.505, 5, 50, 250, 500, 1000\}$. Rows show the corresponding sample sizes that are considered, with $T \in \{288, 500, 1500, 5000\}$. Vertical red lines mark the sample mean. These are based on 10 000 simulations.

where $\mathbb{E}_{\gamma}[\hat{\gamma}]$ is approximated by $\frac{1}{S}\sum_{s=1}^{S}\hat{\gamma}_s$, with $\hat{\gamma}_s$ the estimate of the transition function parameter from the s^{th} simulated ESTAR model. Bias in percent is computed as $100 \times (\text{Bias}_{\gamma_0}[\hat{\gamma}]/\gamma_0)$. The results in Table 7 show that bias can be extremely large. For the smallest γ_0 calibration (the estimate found in Taylor *et al.* (2001)), and a sample size of 288 observations, it can be as high as 0.62 in absolute terms or 122 percent, dropping to 0.26 (or 51 percent) when the sample size increases to 500 observations. Even at 1500 observations, bias remains at well over 10%. As the sample size and γ_0 increase, bias reduces gradually towards 0.

Table 7: Small sample bias of γ estimates.

γ_0 \ T	288		500		1500		5000	
	$\text{Bias}_{\gamma_0}[\hat{\gamma}]$	(in %)	$\text{Bias}_{\gamma_0}[\hat{\gamma}]$	(in %)	$\text{Bias}_{\gamma_0}[\hat{\gamma}]$	(in %)	$\text{Bias}_{\gamma_0}[\hat{\gamma}]$	(in %)
0.505	0.6180	122.37	0.2613	51.74	0.0614	12.15	0.0170	3.36
5	1.0561	21.12	0.5633	11.27	0.1699	3.40	0.0517	1.03
50	3.0524	6.10	1.8237	3.65	0.6013	1.20	0.1628	0.33
250	10.4064	4.16	5.3328	2.13	1.4249	0.57	0.3383	0.14
500	22.3257	4.47	12.7391	2.55	3.6151	0.72	1.3355	0.27
1000	80.3915	8.04	40.7419	4.07	11.8851	1.19	3.9210	0.39

Notes: This table reports the small sample bias of the estimates of the regime weighting function parameter γ for various sample sizes and γ_0 values. We consider values of $\gamma_0 \in \{0.505, 5, 50, 250, 500, 1000\}$ and sample sizes of $T \in \{288, 500, 1500, 5000\}$. These results are based on arithmetic averages computed over 10 000 simulations. The bias is computed as: $\text{Bias}_{\gamma_0}[\hat{\gamma}] = \mathbb{E}_{\gamma}[\hat{\gamma}] - \gamma_0$, where $\mathbb{E}_{\gamma}[\hat{\gamma}]$ is approximated by $\frac{1}{S}\sum_{s=1}^{S}\hat{\gamma}_s$, with $S = 10\,000$ being the number of simulations that are averaged over. The column with the heading (in %) shows the bias as a percentage of the size of γ. This is computed as: $100 \times (\text{Bias}_{\gamma_0}[\hat{\gamma}]/\gamma_0)$.

4.2.3. Estimating the unrestricted models

We now illustrate what happens when attempting to estimate the two unrestricted models of Taylor *et al.* (2001), that is,

$$\Delta q_t = \beta(q_{t-1} - \mu)\mathcal{G}(q_{t-1}; \gamma, \mu) + \epsilon_t, \tag{31}$$

$$\Delta q_t = (\alpha - 1)(q_{t-1} - \mu) + \beta(q_{t-1} - \mu)\mathcal{G}(q_{t-1}; \gamma, \mu) + \epsilon_t, \tag{32}$$

on the simulated data. We again use the same simulation set-up as above. Since the transition function parameter is key in determining the shape of $\mathcal{G}(q_{t-1}; \gamma, \mu)$, and therefore the stability as well as identification in the ESTAR model, we report estimation results for γ only.

In Figure 7 and Figure 8 we show relative frequency plots or '*histograms*' of the $\log_{10}$ (log to base 10) transformed γ estimates from fitting the two unrestricted models in (31) and (32) above, respectively.[21] The figures are again arranged in 4 rows and 6 columns corresponding

[21]We plot $\log_{10}$ transformed histograms of $\hat{\gamma}$ to be able to better illustrate graphically the large dispersion as

to the 4 different sample sizes and 6 γ_0 values that we simulate from. We superimpose a thin green vertical line to mark the true γ_0 value that the models were simulated from. The histogram plots in Figure 7 show that for about 70% of the simulated ESTAR processes, the estimates of the γ parameter hit the lower bound of the grid of 1×10^{-6} when estimating (31) and $\gamma_0 = 0.505$, for all sample sizes that are considered. When $\gamma_0 = 5$, this drops to about 57%. For $\gamma_0 = 50$, between 20% and 13% of the simulations hit the lower parameter bound for sample sizes of 288 and 500 observations, while for the two larger sample sizes, the estimates are concentrated closer to the true parameter value (in $\log_{10}$ scale). For larger γ_0 values (and all considered sample sizes) this also holds true, without a single one of the estimated models hitting the lower or upper search bounds.

Figure 8 shows analogous histogram plots of $\hat{\gamma}$ obtained from the various simulation scenarios that we consider, but now with the unrestricted ESTAR model in (32) being the model that is estimated.[22] From these histograms, we can see that under the $\gamma_0 = 0.505$ scenario, the majority of estimates are in the 1×10^4 to 1×10^7 range, even for sample sizes as large as $T = 1500$. At $T = 5000$, over 55% of the γ estimates hit the lower bound of 1×10^{-6}. When the true parameter is equal to 5, there remain a large number of extreme values that are obtained for $\hat{\gamma}$, that is, either extremely low are high estimates, with $\hat{\gamma}$ converging to 0 about 50% of the time for samples of size 1500 and 5000. As the sample size and true γ_0 values increase, the estimates tend to stabilize in the sense that they are away from the bounds and center at γ_0. These results highlight our earlier findings based on the empirical real exchange rate data that it is extremely difficult to estimate unrestricted ESTAR models due to the problematic shape the exponential function can take for small and large γ values.

One final point that we would like to make here is that the concentrated log-likelihood function of the unrestricted ESTAR model in (32) remains extremely ill-behaved with many local maxima and abrupt changes, even when the model is estimated on simulated data. In Figure 9 we show graphically the evolution of the concentrated log-likelihood surface to provide some visual evidence of this finding. The smoothness of the log-likelihood function tends to increases not only with the sample size, but also with the magnitude of *'true'* γ_0 used to generate the data.

---

well as the clustering at the boundaries of the estimates, and how these vanish with increasing γ_0 values and sample sizes. We have therefore intentionally left the axis scaling the same across the histogram plots.

[22] This is the starting model in Taylor *et al.* (2001), which was used to statistically test the unit-root inner regime restriction as well as the $\beta = -1$ constraint on the slope of the outer regime.

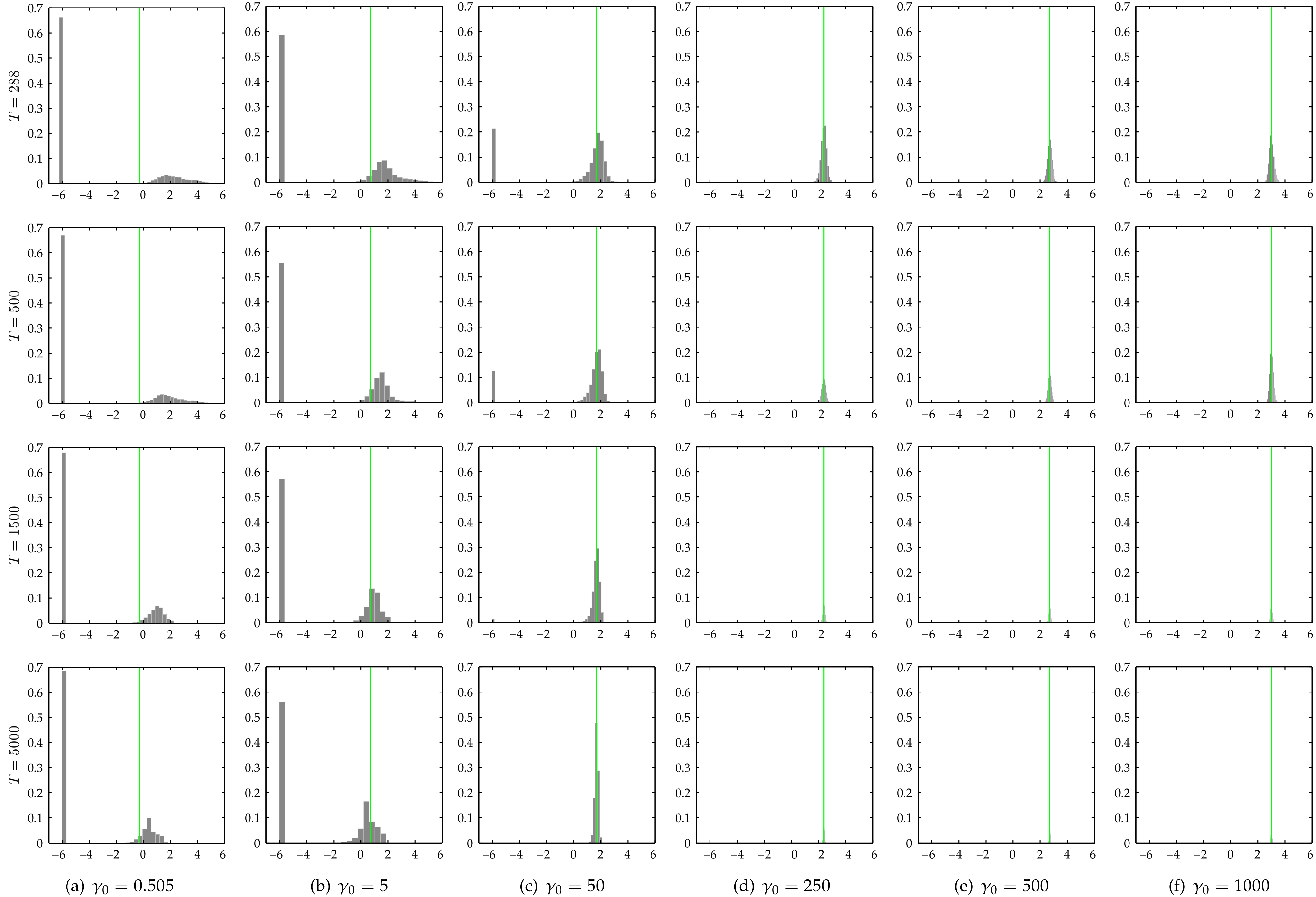


Figure 7: Relative frequency plots of $\log_{10}$ transformed $\hat{\gamma}$ estimated from the unrestricted model in (31) for various sample sizes and γ_0 values. Columns and rows show the results for $\gamma_0 \in \{0.505, 5, 50, 250, 500, 1000\}$ and $T \in \{288, 500, 1500, 5000\}$, respectively. Vertical green lines mark the true γ_0 value. All results are based on 10 000 simulations. The x−axis is log to base 10 transformed, with a reading of -6 on the x−axis meaning 1×10^{-6}.

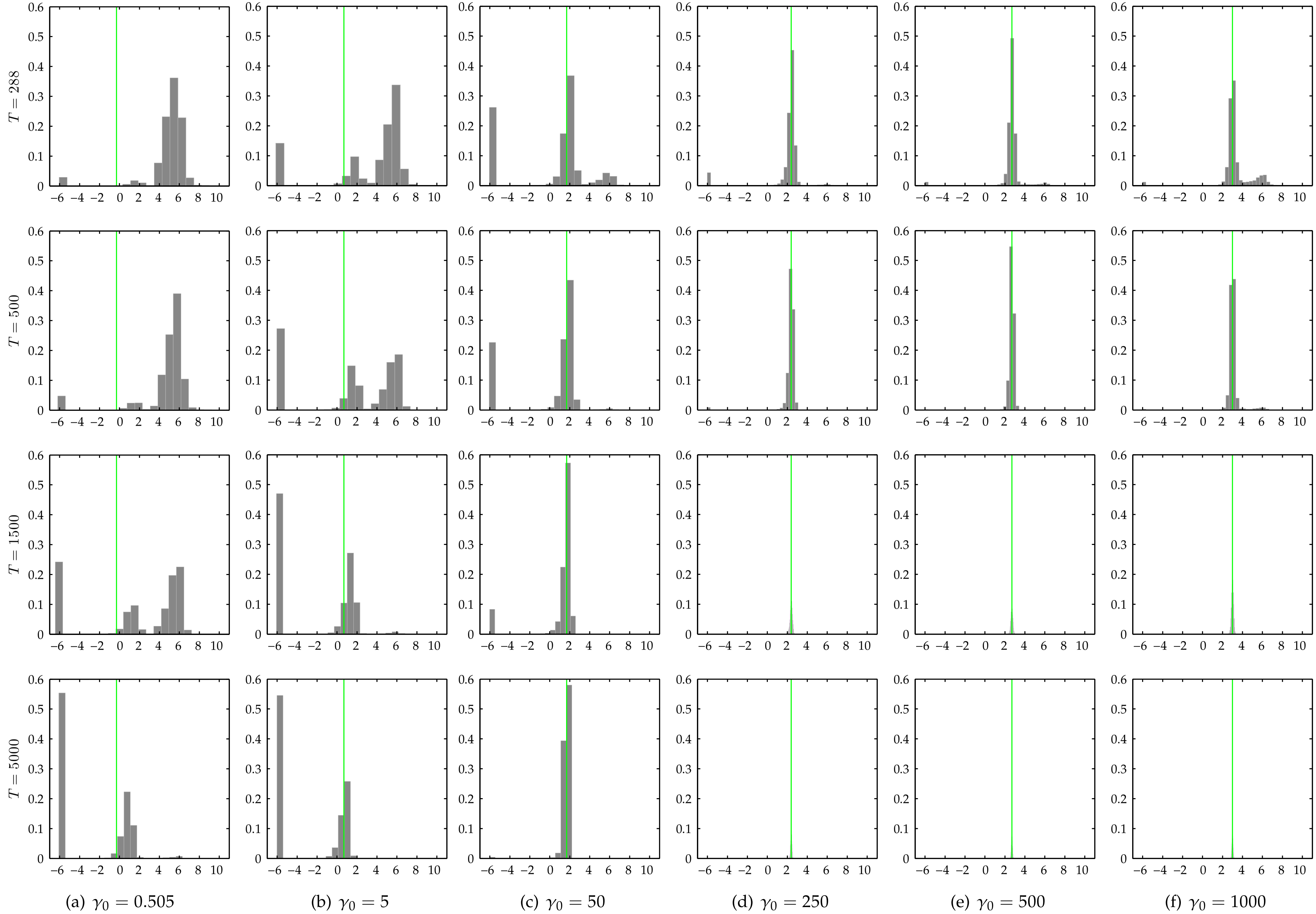
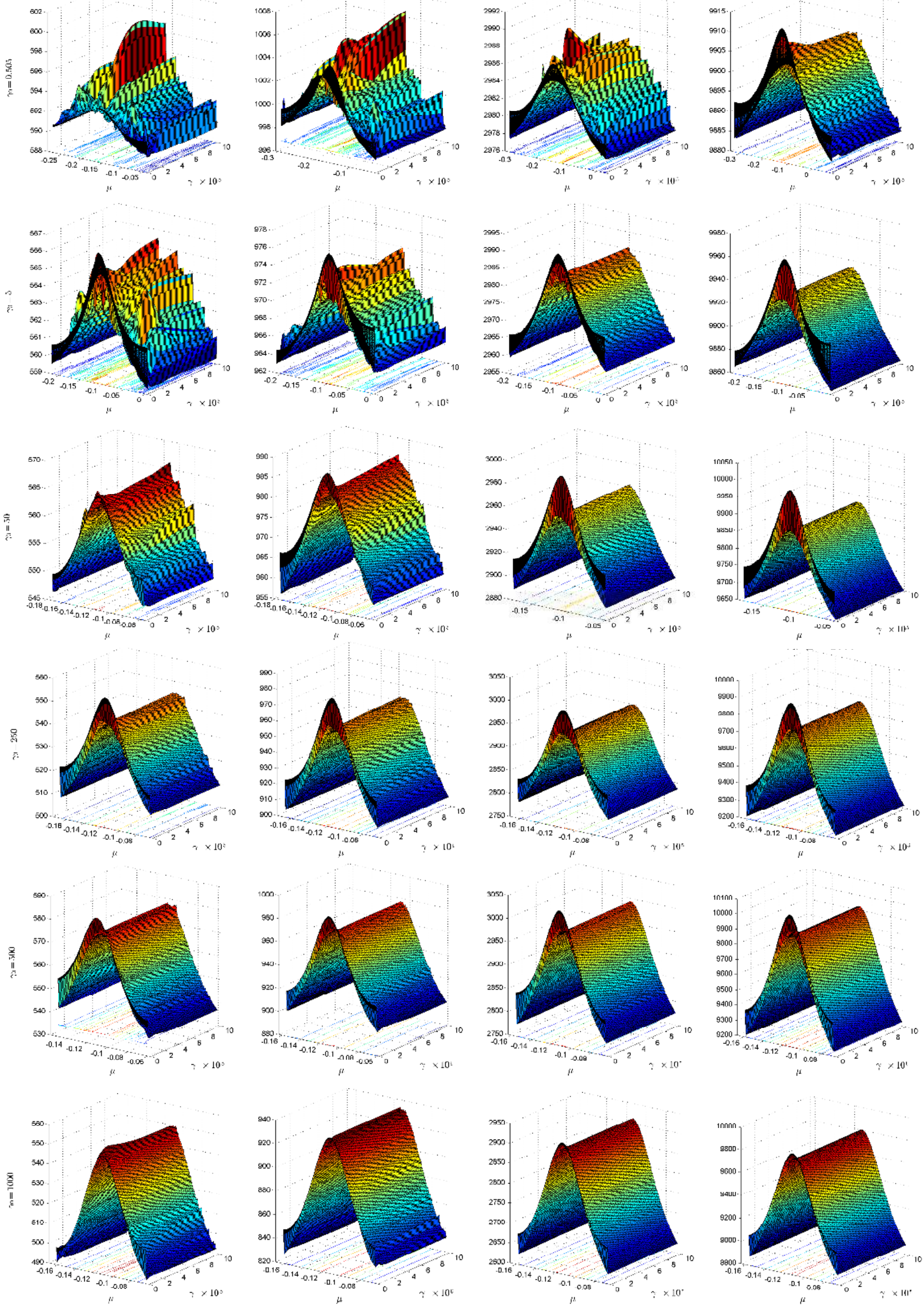


Figure 8: Relative frequency plots of $\log_{10}$ transformed $\hat{\gamma}$ estimated from the unrestricted model in (32) for various sample sizes and γ_0 values. Columns and rows show the results for $\gamma_0 \in \{0.505, 5, 50, 250, 500, 1000\}$ and $T \in \{288, 500, 1500, 5000\}$, respectively. Vertical green lines mark the true γ_0 value. All results are based on 10 000 simulations. The $x-$axis is log to base 10 transformed, with a reading of -6 on the $x-$axis meaning 1×10^{-6}.



(a) $T = 288$ (b) $T = 500$ (c) $T = 1500$ (d) $T = 5000$

Figure 9: Log-likelihood surface plots from the simulated date for various sample sizes and γ_0 values.

5. Conclusion

Exponential Smooth transition autoregressive models have been widely used in the international finance literature, particulary for the modelling of real exchange rates.

In this paper we show that the exponential function is ill-suited as a regime weighting function because of two undesirable properties. First, the exponential function can be well approximated by a quadratic function in the threshold variable z_t whenever the transition function parameter γ takes on *'small'* values. The consequence of this is that the slope vector attached to the non-linear regime and the transition function parameter will enter as a product into the conditional mean of the model, which leads to identification issues. Using an empirical example and an extensive simulation analysis, we show that there is a nearly perfect off-setting effect of these two parameters on the conditional mean when the quadratic approximation of the exponential function is *'good'*. What is particularly problematic with this scenario is that it is not a small sample issue that vanishes as the sample size increases, but rather a population property of the model.

Second, the exponential function can behave like an *'outlier fitting function'*. That is, for extremely large values of the transition function parameter γ, the exponential function will be equal to one for all values of the transition variable, except at $z_t = \mu$. The effect of this on the conditional mean of the model is that only a very small number of observations around the location parameter receive a weight that is different from one. The exponential function can thus act in the same way as a dummy variable which is designed to remove the influence of aberrant observations. From our empirical replication exercise we see that this is precisely the case for the real exchange rate data that is analysed in Taylor *et al.* (2001). The unrestricted ESTAR model always fits an extremely large γ estimate, rendering the conditional mean to be a linear function of the threshold variable for nearly the entire z_t range, with the only exception being the part that is close to μ. Using simulated data, we show that this occurs well over 70% of the time for the two smaller (true) γ_0 values that we consider, and for sample sizes as large as 500 observations. Contrary to the identification issue, the simulation results indicate that this is a *'small sample'* problem and vanish as the sample size increases.

References

Abrevaya, Jason and Shu Shen (2014): "Estimation of Censored Panel Data Models with Slope Heterogeneity," *Journal of Applied Econometrics*, **29**(4), 523–548.

Baum, Christopher F., John T. Barkoulas and Mustafa Caglayan (2001): "Nonlinear Adjustment to Purchasing Power Parity in the Post-Bretton Woods Era," *Journal of International Money and Finance*, **20**(3), 379–399.

Beckmann, Joscha, Theo Berger and Robert Czudaj (2015): "Does gold act as a hedge or a safe haven for stocks? A smooth transition approach," *Economic Modelling*, **48**, 16–24.

Berry, Steven, James Levinsohn and Ariel Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, **63**(4), 841–90.

Buncic, Daniel (2012): "Understanding forecast failure of ESTAR models of real exchange rates," *Empirical Economics*, **34**(1), 399–426.

Cerrato, Mario, Hyunsok Kim and Ronald Macdonald (2010): "Three-Regime Asymmetric STAR Modeling and Exchange Rate Reversion," *Journal of Money, Credit and Banking*, **42**(7), 1447–1467.

Chan, Felix and Michael McAleer (2002): "Maximum Likelihood Estimation of STAR and STAR-GARCH Models: Theory and Monte Carlo Evidence," *Journal of Applied Econometrics*, **17**(5), 509 – 534.

Escribano, Alvaro and Oscar Jordá (1999): "Improved Testing and Specification of Smooth Transition Regression Models," in *Nonlinear Time Series Analysis of Economic and Financial Data*, edited by Philip Rothman, New York: Springer Verlag, 289–319.

Granger, Clive W. J. and Timo Teräsvirta (1993): *Modelling Nonlinear Economic Relationships*, Oxford: Oxford University Press.

Greene, William H. (2011): *Econometric Analysis*, 7^{th} *Edition*, Prentice Hall.

Haggan, Valerie and Tohru Ozaki (1981): "Modelling Nonlinear Random Vibrations Using an Amplitude-Dependent Autoregressive Time Series Model," *Biometrika*, **68**(1), 189–196.

Iskrev, Nikolay (2010): "Local identification in {DSGE} models," *Journal of Monetary Economics*, **57**(2), 189–202.

Johnston, Jack and John DiNardo (2001): *Econometric Methods*, 4^{th} *Edition*, McGraw-Hill.

Kapetanios, George, Yongcheol Shin and Andy Snell (2003): "Testing for a Unit Root in the Nonlinear STAR Framework," *Journal of Econometrics*, **112**(2), 359–379.

Kilian, Lutz and Mark P. Taylor (2003): "Why Is It So Difficult to Beat the Random Walk Forecast of Exchange Rates?" *Journal of International Economics*, **60**(1), 85–107.

Michael, Panos, A. Robert Nobay and David A. Peel (1997): "Transaction Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation," *Journal of Political Economy*, **105**(4), 862–879.

Milas, Costas and Gabriella Legrenzi (2006): "Non-Linear Real Exchange Rate Effects in the UK Labour Market," *Studies in Nonlinear Dynamics and Econometrics*, **10**(1), Article 4.

Pagan, Adrian R. and John C. Robertson (1998): "Structural Models of the Liquidity Effect," *The Review of Economics and Statistics*, **80**(2), 202–217.

Pavlidis, Efthymios G., Ivan Paya and David A. Peel (2011): "Real exchange rates and time-varying trade costs," *Journal of International Money and Finance*, **30**(6), 1157–1179.

Paya, Ivan and David A. Peel (2006): "A New Analysis Of The Determinants Of The Real Dollar-Sterling Exchange Rate: 1871-1994," *Journal of Money, Credit and Banking*, **38**(8), 1971 – 1990.

Pötscher, Benedikt M. and Ingmar R. Prucha (1997): *Dynamic Nonlinear Econometric Models: Asymptotic Theory*, Springer-Verlag Berlin Heidelberg.

Rothenberg, Thomas J. (1971): "Identification in Parametric Models," *Econometrica*, **39**(3), 577–591.

Rothman, Philip, Dick van Dijk and Philip Hans Franses (2001): "Multivariate STAR Analysis of Money and Output Relationship," *Macroeconomic Dynamics*, **5**(4), 506–532.

Saikkonen, Pentti and Ritva Luukkonen (1988): "Lagrange Multiplier Tests for Testing Non-Linearities in Time Series Models," *Scandinavian Journal of Statistics*, **15**(1), 55–68.

Sarantis, Nicholas (1999): "Modeling Non-Linearities in Real Effective Exchange Rates," *Journal of International Money and Finance*, **18**(1), 27–45.

Sarno, Lucio, Hyginus Leon and Giorgio Valente (2006): "Nonlinearity in Deviations from Uncovered Interest Parity: An Explanation of the Forward Bias Puzzle," *Review of Finance*, **10**(3), 321–482.

Smallwood, Aaron D. (2008): "Measuring the persistence of deviations from purchasing power parity with a fractionally integrated STAR model," *Journal of International Money and Finance*, **27**(7), 1161 – 1176.

Sollis, Robert (2008): "U.S. dollar real exchange rates: Nonlinearity revisited," *Journal of International Money and Finance*, **27**(4), 516–528.

Taylor, Mark P. and Hyeyoen Kim (2009): "Real Variables, Nonlinearity, and European Real Exchange Rates," in *NBER International Seminar on Macroeconomics 2008*, edited by Jeffrey Frankel and Christopher Pissarides, University of Chicago Press, 157–193.

Taylor, Mark P. and David A. Peel (2000): "Nonlinear Adjustment, Long-Run Equilibrium and Exchange Rate Fundamentals," *Journal of International Money and Finance*, **19**(1), 33–53.

Taylor, Mark P., David A. Peel and Lucio Sarno (2001): "Nonlinear Mean-Reversion in Real Exchange Rates: Towards a Solution to the Purchasing Power Parity Puzzles," *International Economic Review*, **42**(4), 1015–1042.

Teräsvirta, Timo (1994): "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, **89**(425), 208–218.

Teräsvirta, Timo and Heather M. Anderson (1992): "Characterizing Nonlinearities in Business Cycles using Smooth Transition Autoregressive Models," *Journal of Applied Econometrics*, **7**(Supplement), 119–136.

Tong, Howell (1983): *Threshold Models in Non-Linear Time Series Analysis*, New York: Springer Verlag.

van Dijk, Dick, Timo Teräsvirta and Philip Hans Franses (2002): "Smooth Transition Autoregressive Models - A Survey of Recent Developments," *Econometric Reviews*, **21**(1), 1–47.

Wooldridge, Jeffrey M. (1994): "Estimation and inference for dependent processes," in *Handbook of Econometrics*, edited by Robert F. Engle and Daniel McFadden, Elsevier, Volume 4 of *Handbook of Econometrics*, chap. 45, 2639–2738.

Appendix to:

'Identification and Estimation issues in Exponential Smooth Transition Autoregressive Models'

Daniel Buncic

October 19, 2017

This appendix provides additional details on the replication of the study by Taylor *et al.* (2001). We also add a second empirical example, that is, the one by Teräsvirta and Anderson (1992) that utilizes industrial production data to study the dynamics of business cycles using smooth transition autoregressive models.

A.1. Replication of Taylor *et al.* (2001)

Taylor *et al.* (2001) estimate non-linear ESTAR models for the real exchange rates of the UK, Germany, France and Japan, relative to the US, where the real exchange q_t is defined as:

$$q_t = s_t - p_t + p_t^*, \tag{A.1}$$

with p_t and p_t^* being respectively the logarithms of the US and foreign CPIs, and s_t is the US dollar price of one unit of the foreign currency of interest. The full sample period is from January 1973 to December 1996. As in Taylor *et al.* (2001), we normalize the log real exchange rate series to be equal to 0 in January 1973. Following Taylor *et al.* (2001), we obtain all CPI and exchange rate data from the IMF's international financial statistics database. For descriptive purposes, we show a time series plot of the normalised real exchange rate data in Figure A.1.

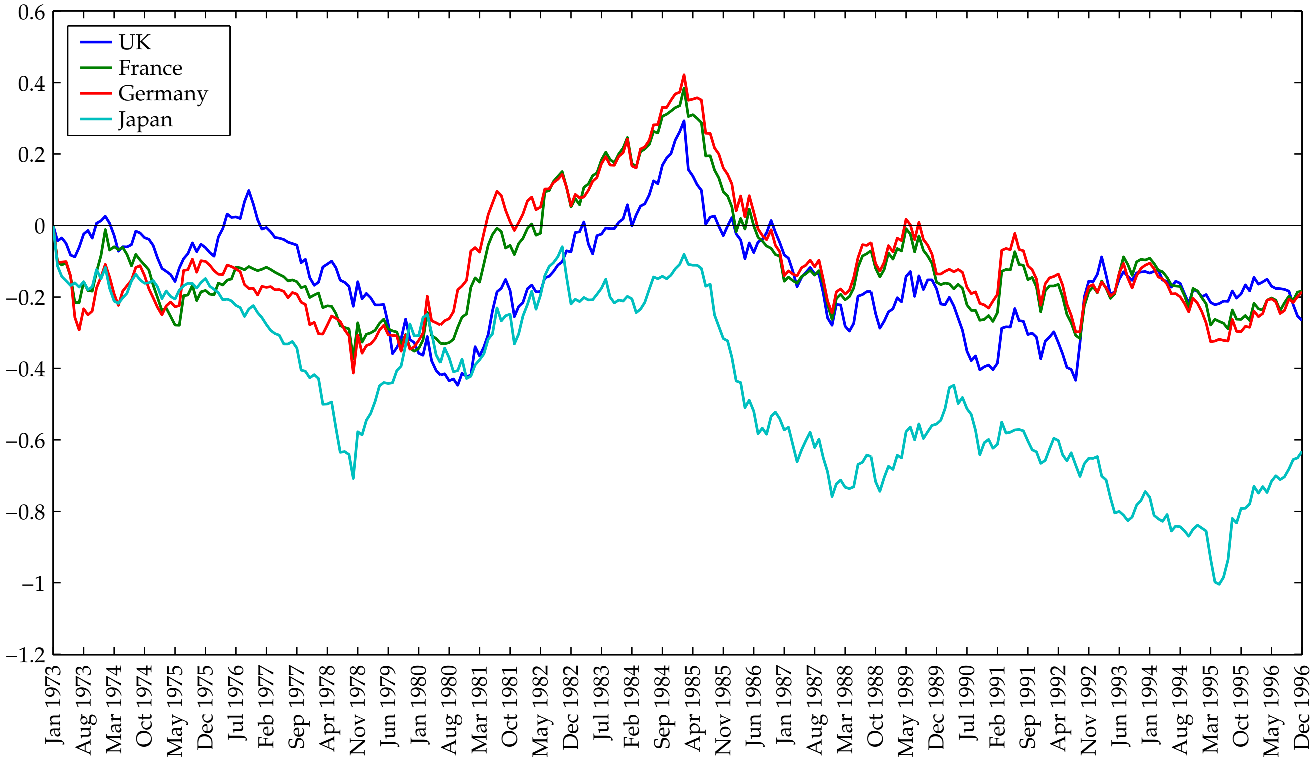


Figure A.1: Time series plots of the (normalised) real exchange rate date from Taylor *et al.* (2001).

Plots of the conditional means $\mathbb{E}(\Delta q_t|q_{t-1}) = -(q_{t-1} - \mu)\mathcal{G}(q_{t-1}; \gamma, \mu)$ and the weighting functions $\mathcal{G}(q_{t-1}; \gamma, \mu)$ of the restricted model in (27) for all 4 real exchange rate series, at the parameter estimates reported in Table 1, are shown in Figure A.2 and Figure A.3, respectively.

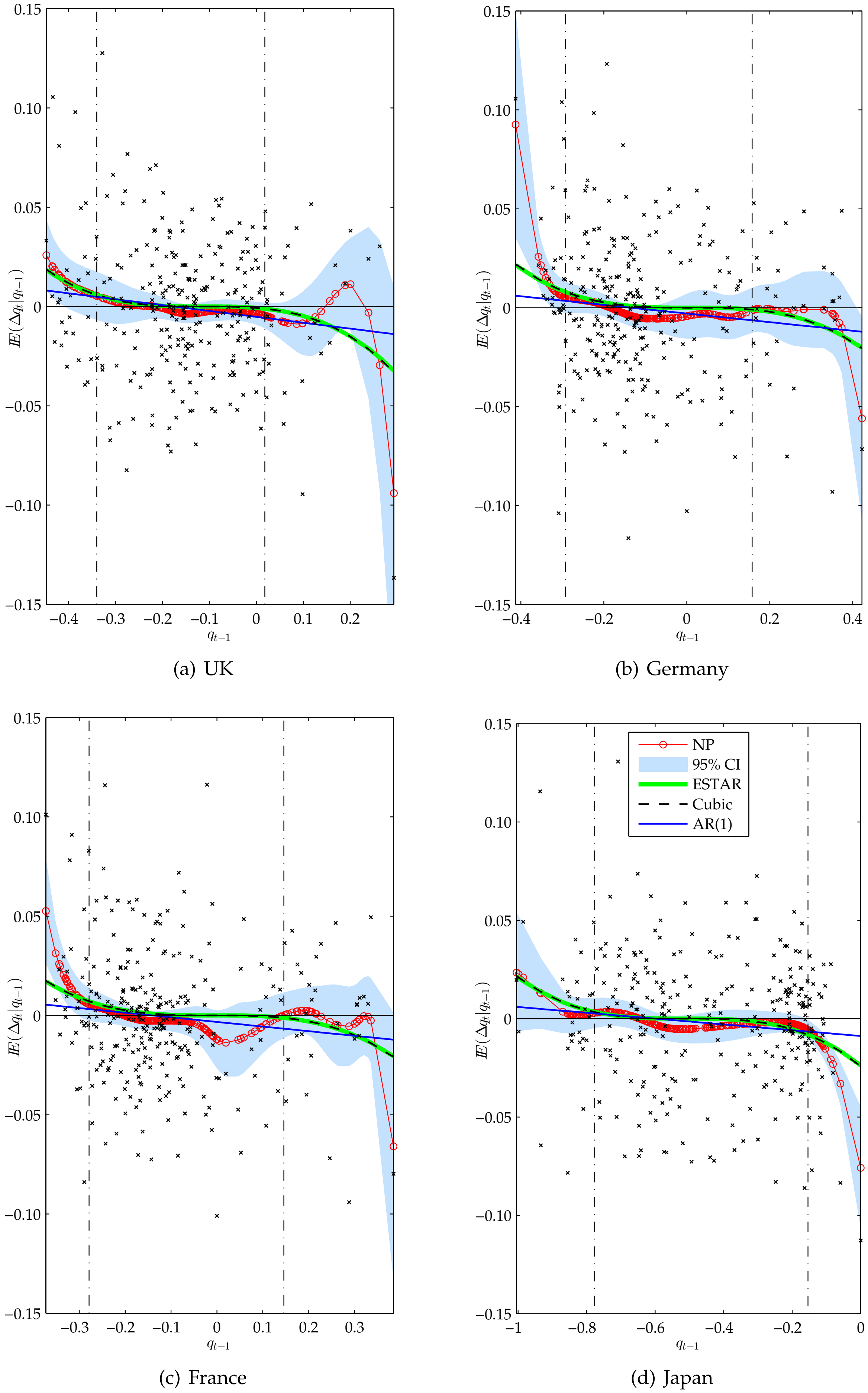


Figure A.2: Plots of the conditional means $\mathbb{E}(\Delta q_t|q_{t-1})$ of the ESTAR models in (27).

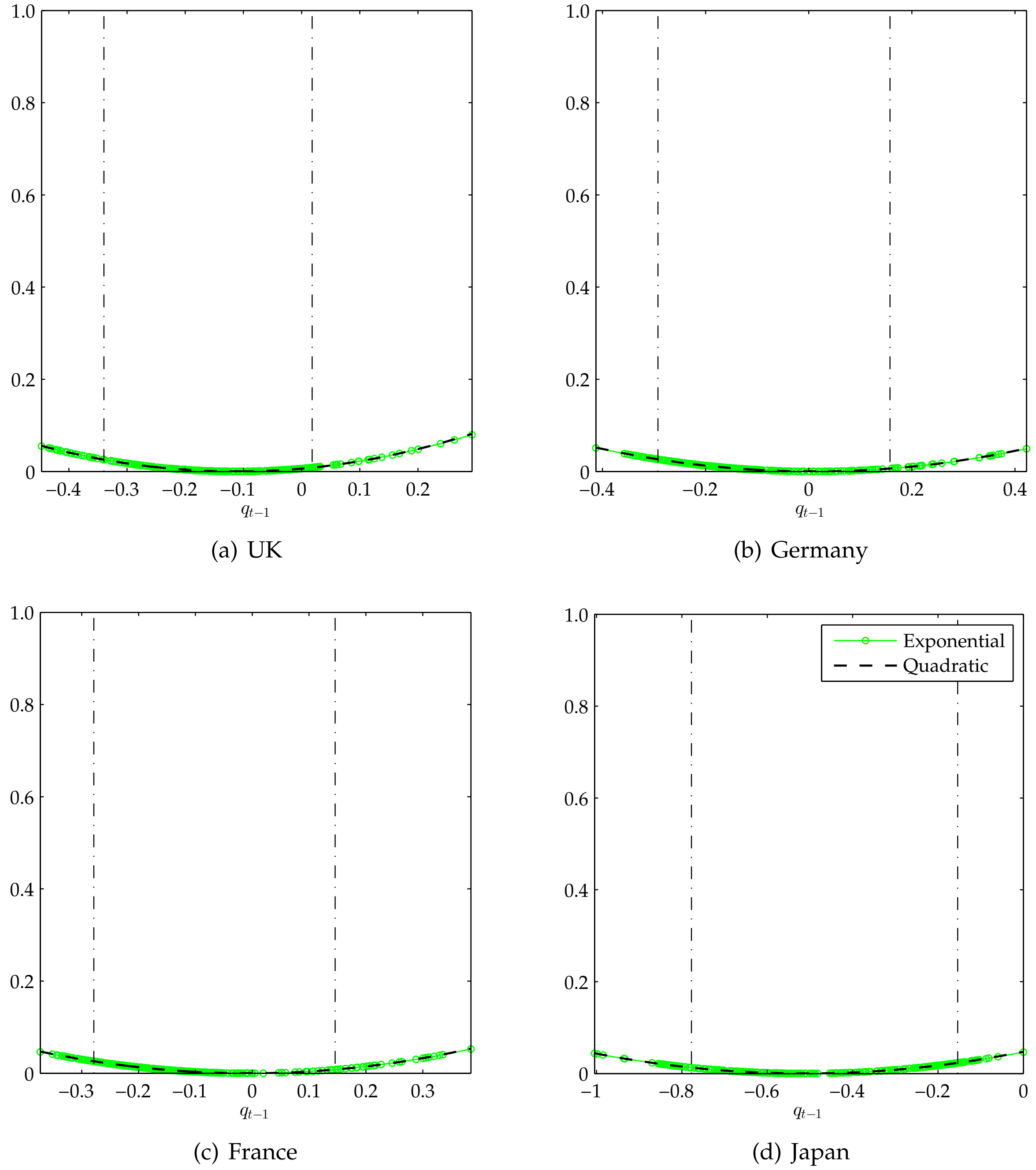
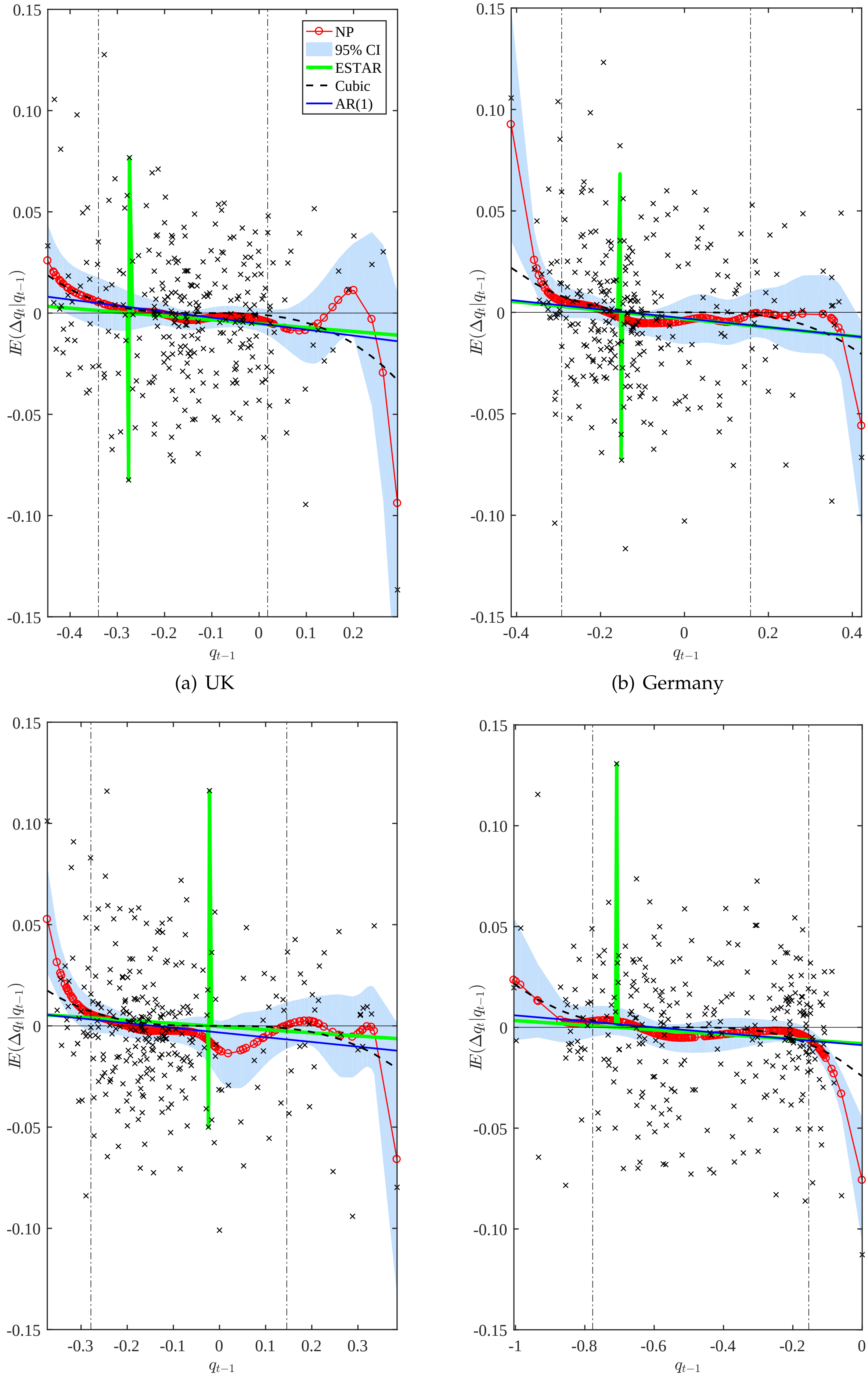


Figure A.3: Plots of the transition function weights of the estimated ESTAR models in (27).

Corresponding conditional mean and weighting function plots of the unrestricted model in (30) for all 4 real exchange rate series, at the parameter estimates reported in Table 5, are shown in Figure A.4 and Figure A.5, perspectively.



(a) UK (b) Germany

(c) France (d) Japan

Figure A.4: Plots of the estimated conditional means $I\!E(\Delta q_t|q_{t-1})$ of the unrestricted ESTAR model in (30).

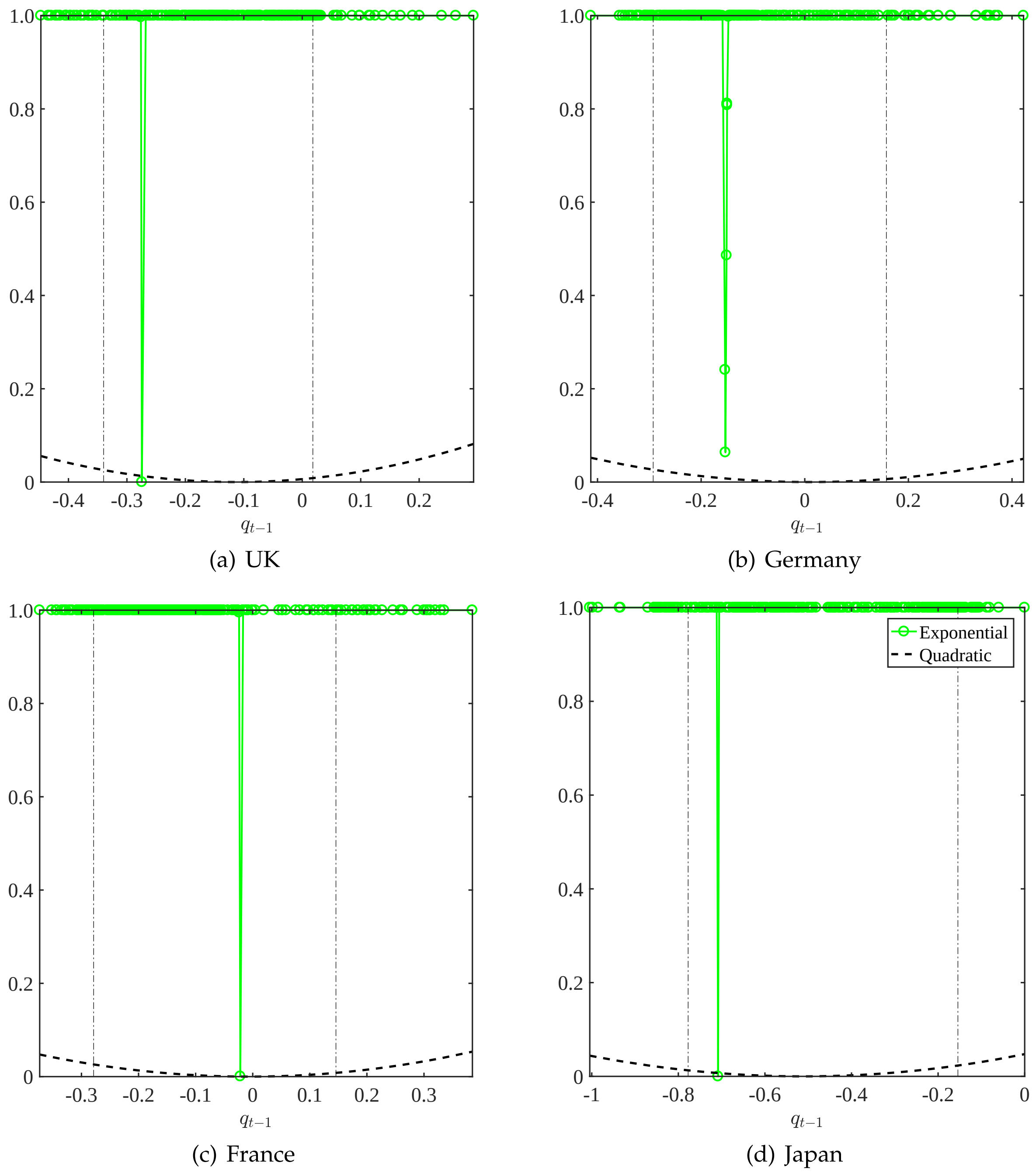
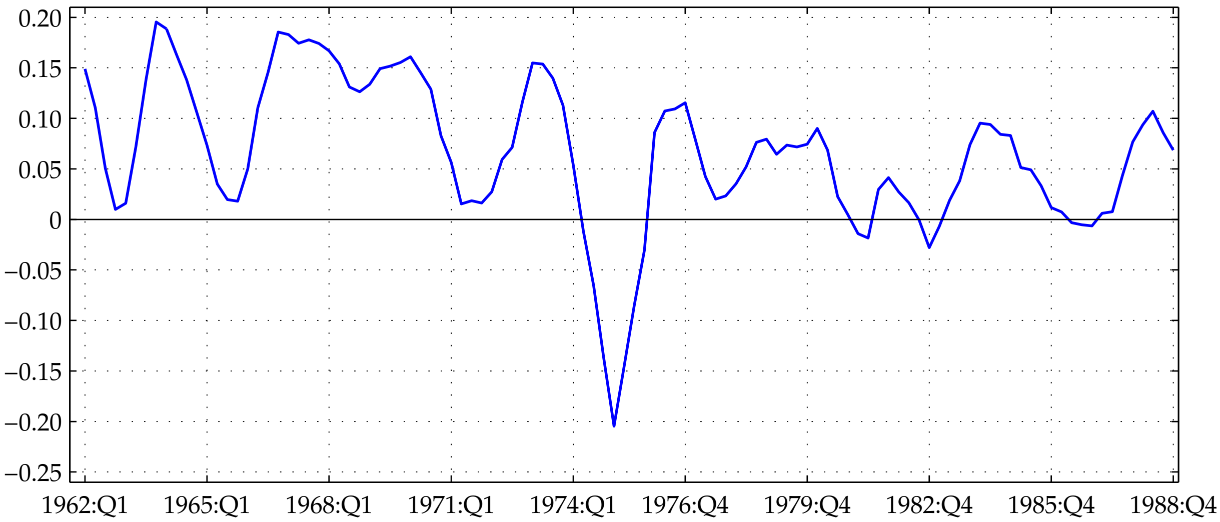


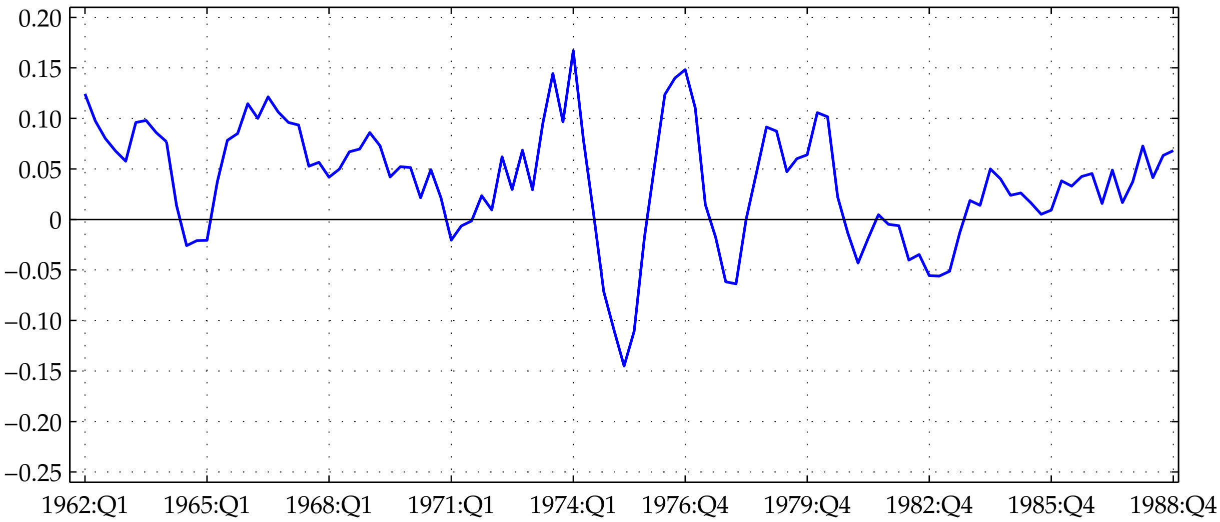
Figure A.5: Plots of the estimated transition function weights of the unrestricted ESTAR model in (30).

A.2. Teräsvirta and Anderson (1992) ESTAR model for industrial production

Teräsvirta and Anderson (1992) apply non-linearities time series models to international business cycle data. The study is interesting as it uses a mix of logistic as well as exponential regime weighting functions to model the dynamics of industrial production. We focus on replicating the ESTAR models fitted to Japanese and Italian industrial production (IP). We obtain industrial production data from the St. Louis FRED Database, using the mnemonics `ITAPROINDQISMEI` and `JPNPROINDQISMEI` for Japan and Italy, respectively. Following Teräsvirta and Anderson (1992), we transform the series using fourth differences of logged industrial production. The sample period is from 1962:Q1 to 1988:Q4, yielding around 100 observations for econometric analysis. Time series plots of the constructed series are shown in Figure A.6.[1] Comparing these plots visually to Figures 6 and 8 of Teräsvirta and Anderson (1992), it is evident that they are very similar.



(a) Japan



(b) Italy

Figure A.6: Four-quarter log-differences of industrial production from 1962:Q2 to 1988:Q4.

To avoid any additional uncertainty related to specification search (ie., finding the appropriate threshold variable and/or the dynamics in each regime) when replicating the models, we estimate the same formulations as in equations (12) and (16) in Teräsvirta and Anderson (1992). Our estimates (without stan-

[1] For Italy, due to widespread industrial action, 1970:Q1, is classified as an outlier observation by Teräsvirta and Anderson (1992) and *'adjusted'*. Since it is not clear how the adjustment was performed, we simply used linear interpolation to replace the series as the arithmetic average of the series from one quarter earlier and one quarter later. Also, note that macroeconomic data are frequently revised over time, resulting in different vintages of data being available. It is thus unlikely that the data that we obtained from FRED will be exactly the same as the one used in Teräsvirta and Anderson (1992).

dard errors) are:

$$y_t = 0.0047 + 4.08y_{t-1} - 1.87y_{t-2} - 0.53\Delta y_{t-4} \quad \text{(A.2a)}$$
$$+ (-2.67y_{t-1} + 1.39y_{t-2} - 0.32\Delta y_{t-8})\,\widehat{\mathcal{G}} + \hat{u}_t \quad \text{(A.2b)}$$
$$\widehat{\mathcal{G}} = [1 - \exp\{-919(y_{t-1} + 0.0984)^2\}], \;\; s = 0.0156. \quad \text{(A.2c)}$$

for Japan, and

$$y_t = 0.55y_{t-1} + 0.53y_{t-2} + (0.0092 + 0.79y_{t-1} \quad \text{(A.3a)}$$
$$- 0.92y_{t-2} - 1.06y_{t-4} + 0.88y_{t-5} - 0.29y_{t-8} + 0.14y_{t-9})\,\widehat{\mathcal{G}} + \hat{u}_t \quad \text{(A.3b)}$$
$$\widehat{\mathcal{G}} = [1 - \exp\{-294.23(y_{t-3} - 0.0290)^2\}], \;\; s = 0.0251. \quad \text{(A.3c)}$$

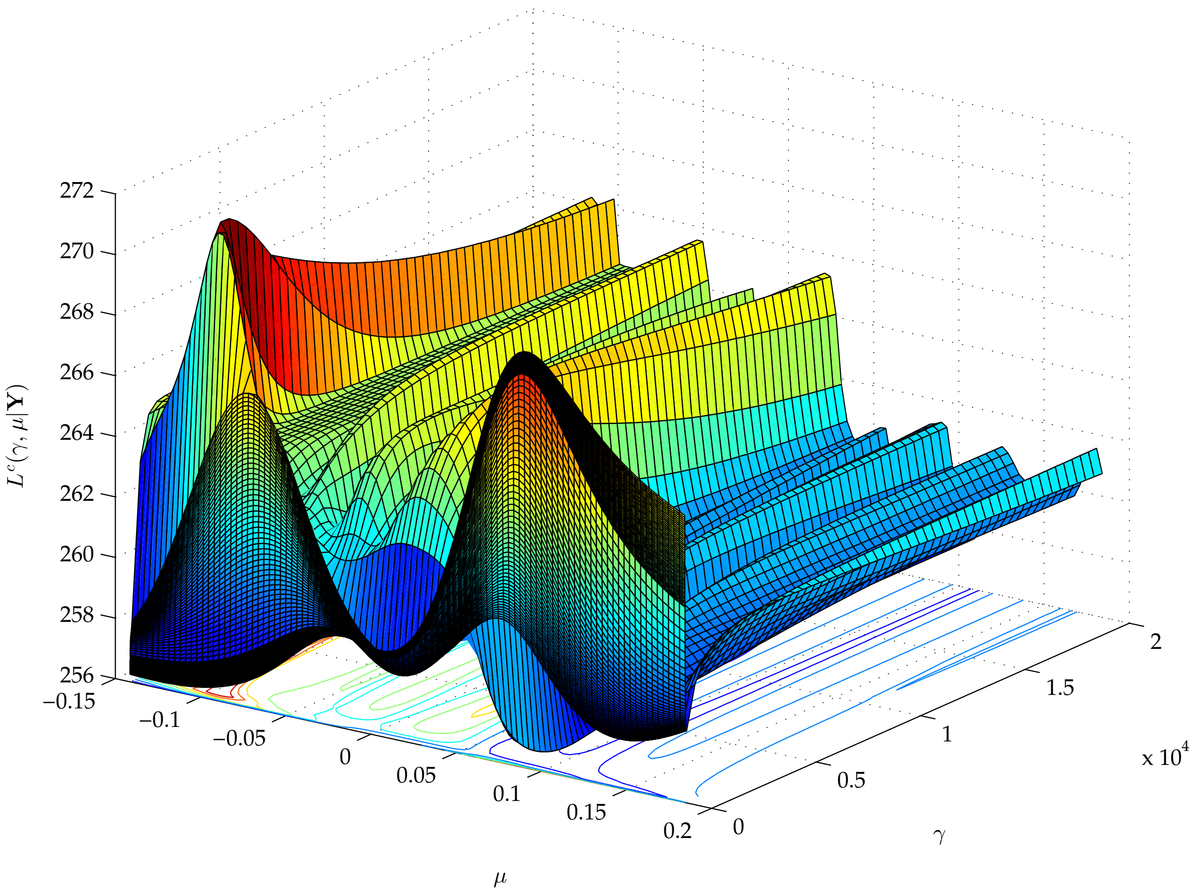
for Italy. Comparing our estimates to those in Teräsvirta and Anderson (1992), we can see that they are largely in line with their estimates reported in equations (12) and (16).[2]

Since our motivation for the replication of the study by Teräsvirta and Anderson (1992) is to highlight that the same type of ill-behaved likelihood function as with the real exchange rate data is obtained, we now show plots of the concentrated log-likelihood surfaces corresponding to the fitted ESTAR models in (A.2) and (A.3). These are show in Figure A.7. As can be seen from these plots, the likelihood functions contain again many local maxima. Although the obtained *'global'* maxima seem to be away from the extremes of the γ bounds, we should stress here that this is not guaranteed and appears to be a matter of *'chance'*. For instance, with the Italian IP dataset that is available to us, using y_{t-1} instead of y_{t-3} as the threshold variable in the transition function $\mathcal{G}(\cdot)$ in (A.3) leads to a γ estimate of over 84658, while producing a better fit (that is, a larger value of the log-likelihood function). We can see, therefore, that the same type of problems as with the real exchange rate data are encountered when estimating ESTAR models as specified in Teräsvirta and Anderson (1992) for industrial production data.

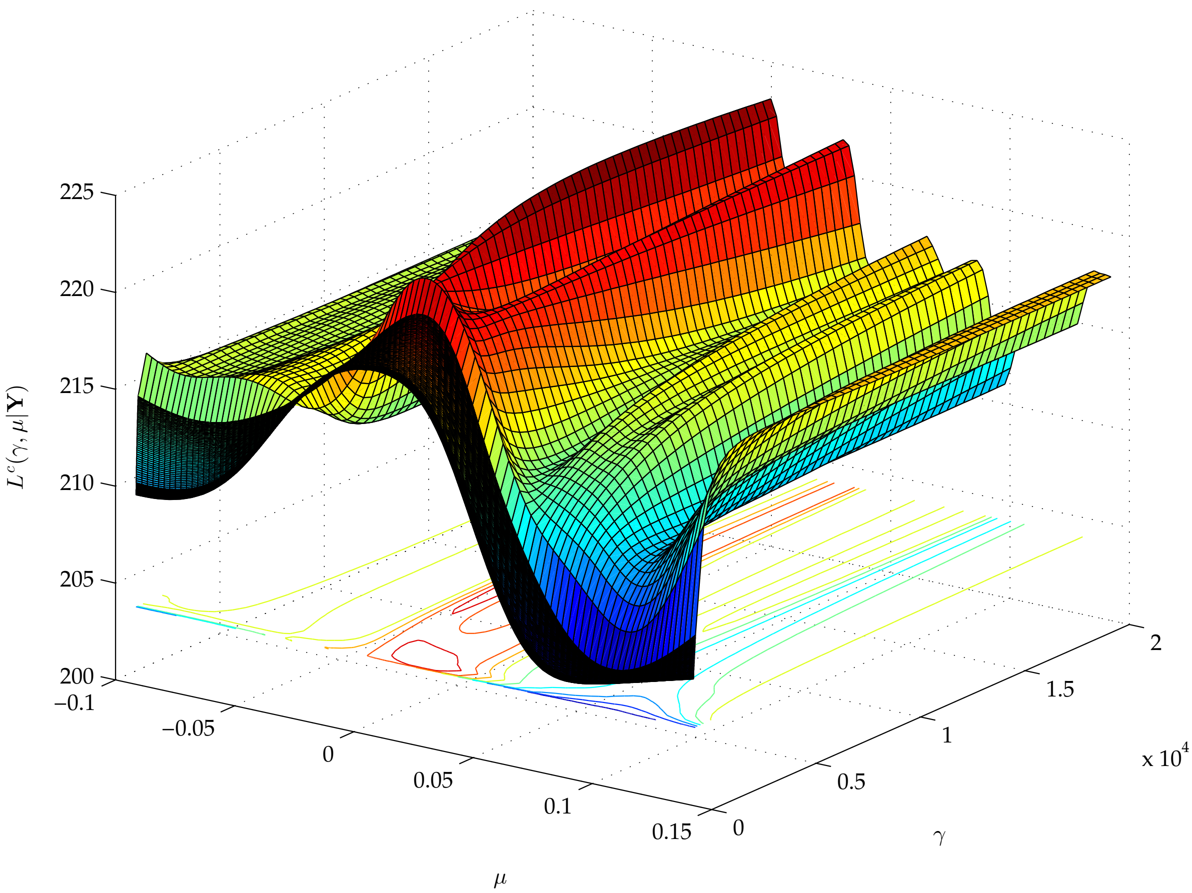
To provide some additional evidence that the ill-behaviour is a property of the exponential function, we implement again a simulation exercise following the same structure as for the real exchange rate data. That is, we take the estimated parameter values form the original estimates in (12) and (16) of Teräsvirta and Anderson (1992) as the data generating process and then estimate ESTAR models with exactly the same lag and transition variable specification to the simulated data. We consider sample sizes of $T = \{100, 200, 500, 1000\}$, with a total of 10 000 simulated sequences. In Figure A.8 we show, as was done earlier for the real exchange rate simulations, relative frequency plots of the (log to base 10 transformed) estimates of γ from the Japanese ESTAR parameterisations for the four different sample sizes that we consider. For the sake of brevity, we do not report results from the Italian simulations, which are qualitatively the same. Also, typical log-likelihood surface plots based on the simulated data for the 4 different sample sizes that we consider are shown in Figure A.9. These are shown for illustrative purposes and are not discussed.

As can be seen from theses plots, there is once again a sizable portion of γ estimates that become very large for the two smaller sample sizes that we consider in the simulations. For instance, for $T = 100$, 20% of the simulations return a γ estimate in excess of 5000. Also, 0.5% of the estimates hit the lower bound on the γ threshold of 1×10^{-6}. At $T = 200$, these proportions drop to 10% and 0.25%, respectively. With a sample size of $T = 500$ observations, only 1% of the simulations return a γ estimate of over 5000, while only 1 out of the 10 000 estimates hits the 1×10^{-6} lower bound. At $T = 1000$, such extreme estimates for γ are not attained anymore. Although these distortions are smaller than the ones encountered earlier with the ESTAR parameterisations for real exchange rates, it should be kept in mind that we have assumed the true lag structure capturing the dynamics in each regime as well as the transition variable to be known. Results are considerably worse when these are determined by the data.

[2]The largest differences are for the γ estimates for the Japanese series (reported value in Teräsvirta and Anderson (1992) is $1.54 \times 196 = 301.84$, as well as the estimates on y_{t-1} in both regimes, which are 3.03 and -1.68, instead of our estimated values of 4.08 and -2.67.



(a) Japan



(b) Italy

Figure A.7: Concentrated log-likelihood function plots of the fitted ESTAR models in (A.2) and (A.3).

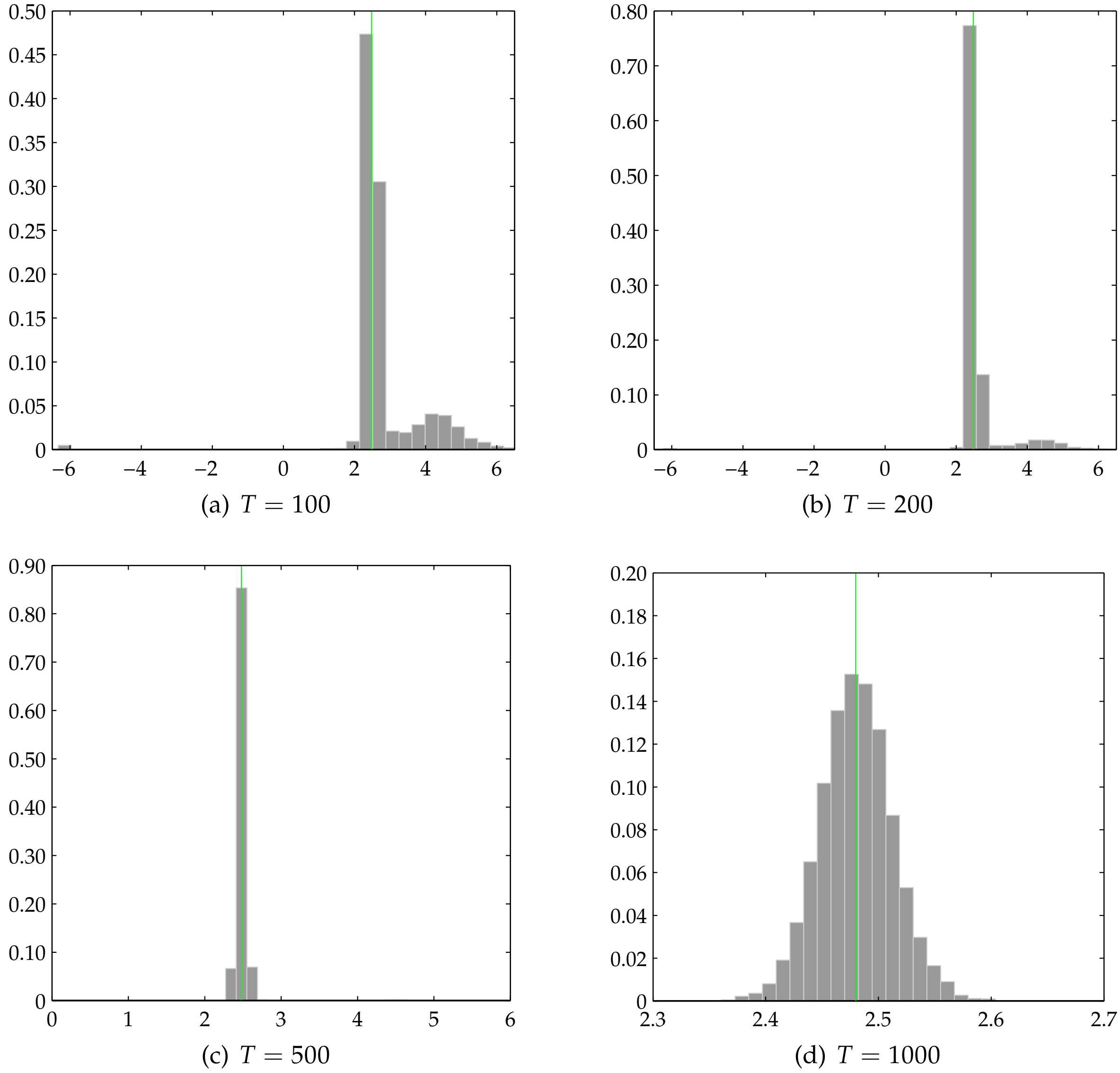


Figure A.8: Relative frequency plots of $\log_{10}$ transformed $\hat{\gamma}$ estimated on data simulated from the ESTAR model for Japanese IP data reported in equation (12) on page S130 in Teräsvirta and Anderson (1992) for samples of size $T \in \{100, 200, 500, 1000\}$. Vertical green lines mark the true value. All results are based on 10 000 simulations. The $x-$axis is log to base 10 transformed, with a reading of 4 on the $x-$axis meaning 1×10^4.

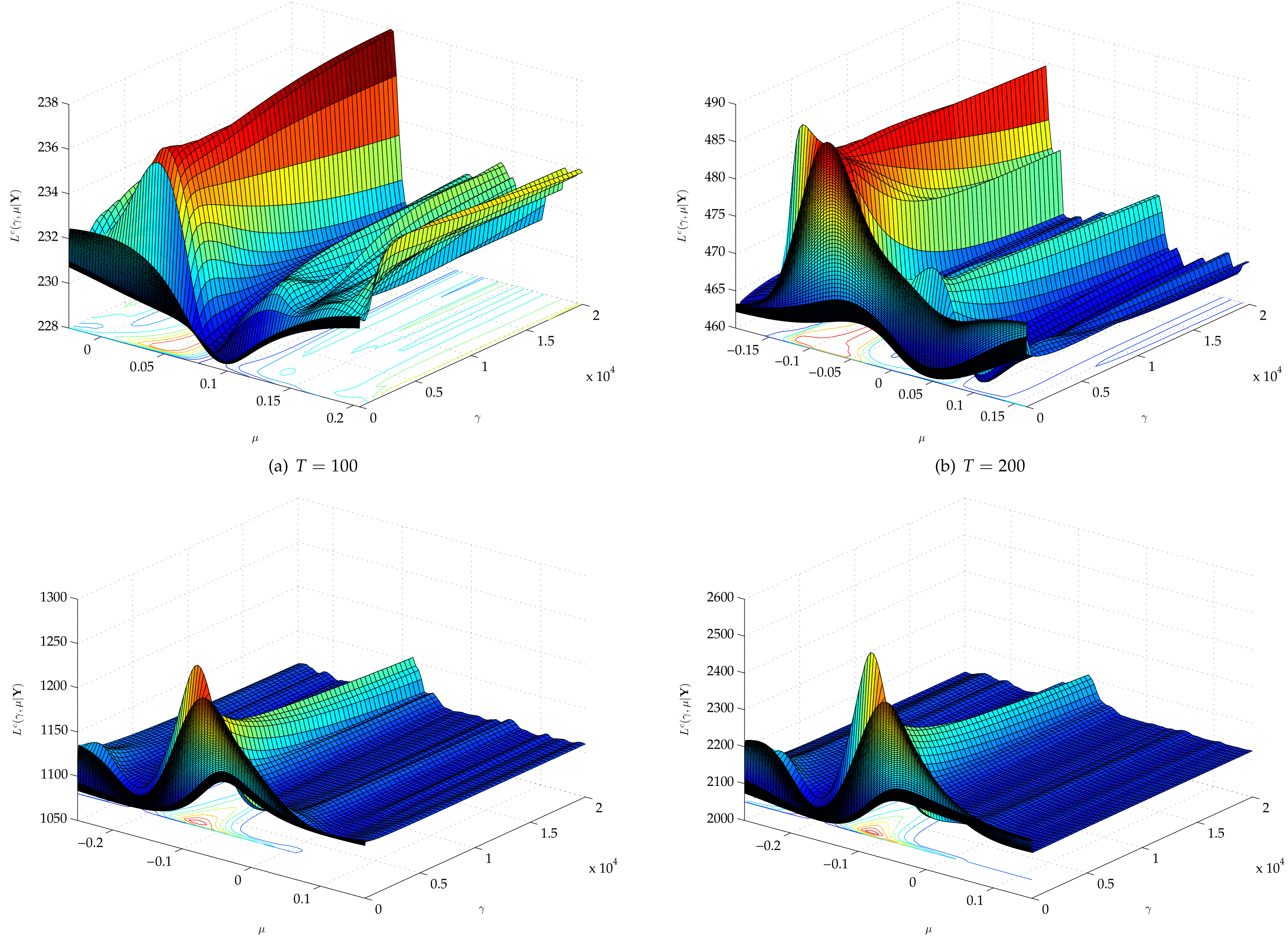


Figure A.9: Log-likelihood function surface plots from the simulated date for various sample sizes. Data are simulated from the ESTAR model estimates of the Japanese IP series reported in equation (12) on page S130 in Teräsvirta and Anderson (1992). Sample sizes $T \in \{100, 200, 500, 1000\}$.

Earlier Working Papers:

For a complete list of Working Papers published by Sveriges Riksbank, see www.riksbank.se

Estimation of an Adaptive Stock Market Model with Heterogeneous Agents *by Henrik Amilon*	2005:177
Some Further Evidence on Interest-Rate Smoothing: The Role of Measurement Errors in the Output Gap *by Mikael Apel and Per Jansson*	2005:178
Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through by Malin Adolfson, Stefan Laséen, Jesper Lindé and Mattias Villani	2005:179
Are Constant Interest Rate Forecasts Modest Interventions? Evidence from an Estimated Open Economy DSGE Model of the Euro Area *by Malin Adolfson, Stefan Laséen, Jesper Lindé and Mattias Villani*	2005:180
Inference in Vector Autoregressive Models with an Informative Prior on the Steady State *by Mattias Villani*	2005:181
Bank Mergers, Competition and Liquidity *by Elena Carletti, Philipp Hartmann and Giancarlo Spagnolo*	2005:182
Testing Near-Rationality using Detailed Survey Data *by Michael F. Bryan and Stefan Palmqvist*	2005:183
Exploring Interactions between Real Activity and the Financial Stance *by Tor Jacobson, Jesper Lindé and Kasper Roszbach*	2005:184
Two-Sided Network Effects, Bank Interchange Fees, and the Allocation of Fixed Costs *by Mats A. Bergman*	2005:185
Trade Deficits in the Baltic States: How Long Will the Party Last? *by Rudolfs Bems and Kristian Jönsson*	2005:186
Real Exchange Rate and Consumption Fluctuations follwing Trade Liberalization *by Kristian Jönsson*	2005:187
Modern Forecasting Models in Action: Improving Macroeconomic Analyses at Central Banks *by Malin Adolfson, Michael K. Andersson, Jesper Lindé, Mattias Villani and Anders Vredin*	2005:188
Bayesian Inference of General Linear Restrictions on the Cointegration Space *by Mattias Villani*	2005:189
Forecasting Performance of an Open Economy Dynamic Stochastic General Equilibrium Model *by Malin Adolfson, Stefan Laséen, Jesper Lindé and Mattias Villani*	2005:190
Forecast Combination and Model Averaging using Predictive Measures *by Jana Eklund and Sune Karlsson*	2005:191
Swedish Intervention and the Krona Float, 1993-2002 *by Owen F. Humpage and Javiera Ragnartz*	2006:192
A Simultaneous Model of the Swedish Krona, the US Dollar and the Euro *by Hans Lindblad and Peter Sellin*	2006:193
Testing Theories of Job Creation: Does Supply Create Its Own Demand? *by Mikael Carlsson, Stefan Eriksson and Nils Gottfries*	2006:194
Down or Out: Assessing The Welfare Costs of Household Investment Mistakes *by Laurent E. Calvet, John Y. Campbell and Paolo Sodini*	2006:195
Efficient Bayesian Inference for Multiple Change-Point and Mixture Innovation Models *by Paolo Giordani and Robert Kohn*	2006:196
Derivation and Estimation of a New Keynesian Phillips Curve in a Small Open Economy *by Karolina Holmberg*	2006:197
Technology Shocks and the Labour-Input Response: Evidence from Firm-Level Data *by Mikael Carlsson and Jon Smedsaas*	2006:198
Monetary Policy and Staggered Wage Bargaining when Prices are Sticky *by Mikael Carlsson and Andreas Westermark*	2006:199
The Swedish External Position and the Krona *by Philip R. Lane*	2006:200

Sveriges Riksbank
Visiting address: Brunkebergs torg 11
Mail address: se-103 37 Stockholm

Website: www.riksbank.se
Telephone: +46 8 787 00 00, Fax: +46 8 21 05 31
E-mail: registratorn@riksbank.se