ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Random Walk Smooth Transition Autoregressive Models

Heather M. Anderson Chin Nam Low

November 2004
Revised May 2005

Working Paper 22/04

Random Walk Smooth Transition Autoregressive Models

Heather M. Anderson*
Australian National University
Canberra, ACT, 0200
AUSTRALIA

Chin Nam Low
Monash University
Clayton, Victoria 3800
AUSTRALIA

Abstract

This paper extends the family of smooth transition autoregressive (STAR) models by proposing a specification in which the autoregressive parameters follow random walks. The random walks in the parameters capture permanent structural change within a regime switching framework, but in contrast to the time varying STAR (TV-STAR) specification introduced by Lundbergh et al (2003), structural change in our random walk STAR (RW-STAR) setting follows a stochastic process rather than a deterministic function of time. We suggest tests for RW-STAR behaviour and study the performance of RW-STAR models in an empirical setting, focussing on interpretation and out of sample forecast performance.

Keywords: Forecast density evaluation, Non-constant parameters, Permanent structural change, Random walk smooth transition models, Regime switching, Time varying smooth transition models

JEL classification: C22, C51, E32

*Correspondence: E-mail: Heather.Anderson@anu.edu.au

1. Introduction

There is large empirical literature that documents nonlinear behavior in macroeconomic and financial time series. Some authors (see e.g. Kim and Nelson (1989), Stock and Watson (1996) or Clements and Hendry (1999)) have focussed on parameter instability as a means of accounting for structural change. Others (see e.g. Teräsvirta and Anderson (1992), Beaudry and Koop (1993) or Pesaran and Potter (1997)) have focussed on modelling behavioral regimes and the transition between them. An implicit distinction between these two types of nonlinearity is that the first embodies a notion of permanent change while the second embodies transitory changes between fixed states. However, this distinction is made only occasionally, because usually the researcher is interested in modelling just one type of nonlinearity, rather than discriminating between them or simultaneously accounting for both.

Recent work by Lundbergh et al (2003) has incorporated structural instability into a regime switching framework, by developing a time-varying smooth transition autoregressive (TV-STAR) model. This model allows for smooth transition between two distinct regimes, but in contrast to other regime switching models in which regime specific parameters remain constant, the regime specific parameters in the TV-STAR model evolve according to a deterministic function of time. The main advantage of this specification is that it allows the underlying dynamics of each regime to change, so that in a business cycle context the dynamics of a recession are allowed to change, as are the dynamics of an expansion. Lundbergh et al (2003) use a logistic function in time to model the variation in their regime parameters, so that all changes in these parameters are monotonic. This monotonicity seems appropriate for capturing the inevitable effects of phenomena such as global warming or the internet on the economy. However, monotonic changes are questionable if one believes that the direction of parameter variation might change at different points in time, as might be the case if exogenous stochastic shocks or government policies are primarily responsible for instigating structural change.

One might make the TV-STAR model more flexible by using a non-monotonic function of time to generate parameter changes, but such an approach would still restrict the sorts of changes that recessions (or expansions) could undergo. In particular, the TV-

STAR model would still be unable to account for parameter changes resulting from stochastic influences. Given this limitation we therefore explore an alternative approach, by introducing and studying a random walk smooth transition autoregressive (RW-STAR) model. The RW-STAR model differs from the TV-STAR model in that it allows the regime parameters to follow random walk processes, so that in a business cycle setting the characteristics of recessionary and expansionary regimes can change over time, but in a stochastic rather than a deterministic fashion. It differs from the more standard random walk autoregressive (RWAR) specification in that it allows for regime dependent behavior. Our aims in developing the RW-STAR model are to offer an alternative way of capturing permanent parameter change in a setting that is already nonlinear, and also to look at the issue of whether observed "shifts" in business cycle characteristics are deterministic or more stochastic in nature.

We organize our work as follows. In Section 2, we define our RW-STAR model and compare its main properties to standard RWAR, STAR and TV-STAR models. We discuss tests for RW-STAR behavior in Section 3, which also contains a small Monte Carlo exercise that examines the size and power of some of our tests. Section 4 illustrates our tests and modelling strategies on OECD industrial production data. Here, we also investigate and compare various aspects of the in-sample and out of sample performance of linear, RWAR, LSTAR, TV-STAR and RW-STAR models, providing forecast density evaluation as well as point forecast evaluation. Finally, in Section 5 we provide some concluding remarks.

2. The RW-STAR Model

2.1. The model

The standard STAR model of order p for a univariate time series y_t (see eg Teräsvirta (1994)) is given by

$$y_t = \pi_1' w_t + \pi_2' w_t G(s_t; \gamma, c) + \varepsilon_t \tag{2.1}$$

where $\varepsilon_t \sim nid(0, \sigma^2)$, $\pi_j = (\pi_{j0}, \pi_{j1}, ..., \pi_{jp})'$ for $j = 1, 2$, $w_t = (1, y_{t-1}, ..., y_{t-p})'$ and $G(s_t; \gamma, c)$ is a transition function which is continuous in s_t and bounded by zero and one. The argument s_t of $G(s_t;\gamma,c)$ is usually y_{t-d} with $d>0$, but s_t is sometimes an exogenous variable such as a policy or leading indicator variable. If one sets $s_t = t$ then one obtains the TV-AR model discussed by Lin and Teräsvirta (1994). Different forms of the transition function give rise to different types of regime switching behavior. The most popular choice for $G(s_t;\gamma,c)$ is the logistic function with

$$G(s_t;\gamma,c) = (1+\exp\left[-\gamma(s_t-c)\right])^{-1} \tag{2.2}$$

which gives rise to a logistic STAR (LSTAR) model, but another common choice is the exponential function given by

$$G(s_t;\gamma,c) = 1-\exp\left[-\gamma(s_t-c)^2\right], \tag{2.3}$$

which gives rise to an exponential STAR (ESTAR) model. In each case the centrality parameter c determines the location of the two regimes that correspond to extreme values of $G(s_t;\gamma,c)$ (i.e. zero or one), while the smoothing parameter $\gamma>0$ determines the speed of transition between regimes with respect to changes in the transition variable s_t. For an LSTAR model, $G(s_t;\gamma,c)=0$ and $G(s_t;\gamma,c)=1$ correspond to "lower" and "upper" regimes, and these are often interpreted as recessionary and expansionary regimes when modelling business cycles. For an ESTAR model, $G(s_t;\gamma,c)=0$ and $G(s_t;\gamma,c)=1$ correspond to "inner" and "outer" regimes. A key feature of these two STAR specifications is that although the behavior of y_t is regime dependent, there is no long-run structural change because the set of possible regimes remains constant.

The RW-STAR model replaces the π_1 and π_2 parameter vectors in the standard STAR model by time-varying parameter vectors $\pi_{1,t}$ and $\pi_{2,t}$ that follow random walk processes given by

$$\pi_{j,t} = \pi_{j,t-1} + \eta_{jt}. \tag{2.4}$$

for $j=1$, 2. We assume that $\eta_{j,t} \sim nid(0,\, \Omega^2_{\eta j})$, which allows us to treat the resulting model as a conditionally Gaussian state-space model (see Harvey (1989)), with a measurement equation given by

$$y_t = \pi'_{1,t}w_t + \pi'_{2,t}w_t G(s_t;\gamma,c) + \varepsilon_t. \tag{2.5}$$

Equation (2.4) implies stochastic variation in the parameters of (2.5), so that observations taken from a RW-STAR process will differ from LSTAR observations in that they will be conditionally heteroskedastic. A process following equation (2.5) with (2.4) will differ from an autoregressive process with random walk coefficients (a RWAR process) in that it will exhibit regime dependent behavior. We will just consider using (2.2) in (2.5) in our empirical work in Section 3.2, although our above definition of RW-STAR models includes other possible specifications for G (with $0 < G < 1$). The model is completed by making an assumption about the correlation structure between ε_t and η_t. Here we make the simplifying assumptions that $\Omega_\eta^2 = I_2 \otimes \Omega_{\eta j}^2$ is a diagonal matrix and that ε_t is uncorrelated with each of the p elements in each $\eta_{j,t}$.

The primary motivation for the RW-STAR model is to allow for stochastic, but permanent changes in the autoregressive parameters in an LSTAR setting, and it is useful to consider some points relating to how (2.4) affects the time variation in $\pi_{1,t}$ and $\pi_{2,t}$. A preliminary observation is that (2.4) implies that structural change is permanent. A "return to normalcy" assumption (see eg, Lin and Teräsvirta (1999)) given by $\pi_{j,t} = \phi_j \pi_{j,t-1} + \eta_{jt}$ would be less restrictive, but would not imply permanent change.

Next, we can observe that since $\pi_{1,t}$ and $\pi_{2,t}$ follow random walks, realizations will typically wander, or follow paths that might change direction at any point in time. It is, however, possible for parameters to follow monotonic paths for an extended period of time. Movement in the π_j parameters in a business cycle context means that the dynamic characteristics of recessions and expansions can change over time. Individual elements in $\pi_{j,t}$ can remain fixed (if the relevant component in $\Omega_{\eta j}^2$ is zero), but in general the variation in each element in $\pi_{j,t}$ will become more pronounced, the larger the corresponding variance component in $\Omega_{\eta j}^2$. This contrasts with the TV-STAR model outlined below, where parameter paths are typically monotonic, and all of them move together.

A final observation is that (2.4) implies that $\pi_{1,t}$ and $\pi_{2,t}$ change in each period, in response to each η_t. This provides a convenient way of capturing structural change that *gradually evolves* over many periods, and it is appropriate for modelling changes due to the phasing in of a particular policy, or the gradual adoption of a new technology. Equation (2.4) is not well suited for capturing random but *infrequent* permanent structural change, although one could explicitly deal with this by specifying a process for η_t in which non-zero values are very rare. Equation (2.2) in Nyblom (1989) provides an example of how this might be done.

Related work by Kim and Nelson (1999) and McConnell and Perez-Quiros (2000) has studied infrequent changes in the mean and variance parameters associated with each state in a Markov switching model of US output growth. These authors wanted to determine whether the gap between recessionary and expansionary growth rates has narrowed and whether there has been a decline in the volatility of output growth. Both papers find evidence of a structural shift in 1984 (Q1), but neither allows for gradual structural change. The RW-STAR model studied here explicitly caters for gradual change in mean, and it also allows for implicit changes in variance. Thus one could couch the narrowing gap and volatility decline questions within a RW-STAR framework, allowing for gradual rather than sudden changes in business cycle characteristics.

The RW-STAR model provides a way of modelling stochastic but permanent structural change in an LSTAR setting, and it can be contrasted against the TV-STAR model introduced by Lundbergh, Teräsvirta and van Djik (2003), that provides a way of modelling deterministic permanent structural change within the same setting. This latter model combines the STAR model with the time varying autoregressive (TV-AR) model of Lin and Teräsvirta (1994) to obtain

$$y_t = \left[\pi_1' w_t + \pi_2' w_t G(s_t; \gamma_1, c_1)\right] \left[1 - G(t; \gamma_2, c_2)\right] + \left[\pi_3' w_t + \pi_4' w_t G(s_t; \gamma_1, c_1)\right] G(t; \gamma_2, c_2) + \varepsilon_t \tag{2.7}$$

where $G(t, \gamma_2, c_2)$ is a logistic transition function as in (2.2). Equation (2.7) shows that this model can be interpreted as a STAR model in which the autoregressive parameters undergo gradual and deterministic changes from π_1 to π_3, and from π_2 to π_4. Lundbergh et al (2003) test each of the 214 series analyzed in Stock and Watson's (1996) study of parameter instability, and find that their testing procedure supports a TV-STAR specification in 17% of (the differenced versions of) these series. They also develop a TV-STAR model of the help wanted index in the USA, and show that this model has good in-sample and out of sample properties.

3. Modelling procedure

A typical nonlinear model building strategy starts with a linear approximation to the DGP, and then tests for particular forms of nonlinearity that might characterize the data. LM type tests are often used when contemplating nonlinear alternatives, because they avoid the complicated exercise of having to estimate a (potentially inappropriate) nonlinear alternative, while still at the model specification stage. Teräsvirta (1994) and Lundbergh at al (2003) both base their model specification strategies on specific-to-general approaches that employ outward looking LM tests, although both also use a general-to specific approach when refining their chosen nonlinear specification.

Our model specification strategy follows the broad principles advocated by these previous authors, and involves first choosing a linear autoregressive model for the series using a model selection criterion such as AIC, and then testing outwards for various nonlinear alternatives. Our basic testing tools are adaptations of Teräsvirta's (1994) linearity test against STAR behavior, and Nyblom's (1989) test for the constancy of parameters over time. Nyblom's test allows the parameters of the AR(p) process to follow a martingale under the alternative, and therefore it is particularly appropriate for finding evidence that parameters follow random walks. Another possibly appropriate parameter constancy test might be Lin and Teräsvirta's (1999) test that allows each of the AR(p) parameters to follow AR(q) processes under the alternative. This latter test is interesting because it exploits the conditional heteroskedasticity associated with the stochastic variation in random coefficients, but we do not explore its properties here.

In the context of an AR(p) model, the LM STAR test statistic is the (F-version of the) test of $H_0 : \beta_2 = \beta_3 = 0$ in the auxiliary model

$$y_t = \beta_1' w_t + \beta_2' w_t s_t + \beta_3' s_t^3 + \varepsilon_t, \tag{3.1}$$

as discussed in Eitrheim and Teräsvirta (1996). If $s_t = y_{t-d}$ where the delay parameter d is unknown, then one typically conducts a sequence of tests with different d, and then uses the test result with the lowest p-value to guide the choice of y_{t-d}. Nyblom's (1989) test applied to an AR(p) model is based on the test statistic

$$L = T^{-2} \sum_{t=1}^{T} Z_t' (\widehat{W})^{-1} Z_t, \tag{3.2}$$

where $Z_t = \sum_{t=1}^{T} w_t e_t$ (with e_t being the residuals obtained from the AR(p)), and $\widehat{W} = T^{-1}(\sum_{t=1}^{T} w_t w_t')\widehat{\sigma}_e^2$ (with $\widehat{\sigma}_e^2 = T^{-1}(\sum_{t=1}^{T} e_t^2)$). The test statistic has a non-standard distribution under the null, but critical values are tabulated in Nyblom (1989). Luukkonen et al (1988) discuss the size and power properties of the above STAR test, while Nyblom (1989) discusses the size and power properties of his parameter constancy test. To get a feeling for how well these tests perform in different nonlinear situations, we undertake a small Monte-Carlo study in Section 3.1 below.

Each of the above tests look for just one type of nonlinearity, but if both types are found, then one might view this as evidence of RW-STAR behavior. A "procedure" that rejected linearity in favour of a RW-STAR specification when each of the STAR and Nyblom tests rejected linearity would have distorted size if one were to view it as a "test", but in principle one can adjust critical values to account for this. Alternatively, one might consider a single test with an RW-STAR alternative, but the development of a direct test of linearity against RW-STAR behavior is beyond the scope of this paper.

As a compromise we note that if a STAR model has been estimated, then one can test STAR nonlinearity against RW-STAR nonlinearity, by applying the Nyblom parameter constancy test to the linearised version of the STAR model. To operationalize this, one fixes the parameters γ and c in the STAR model at $\widehat{\gamma}$ and $\widehat{c}$ to evaluate the variables in $w_t G(s_t; \gamma, c)$, and then treats (2.1) as the linear regression which holds under the null of parameter constancy. The calculation of the test statistic L in (3.2) then involves setting $Z_t = \sum_{t=1}^{T} x_t \epsilon_t$ (with x_t being $(w_t,\ w_t G(s_t; \widehat{\gamma}, \widehat{c}))$, and ϵ_t being the residuals obtained from the linearised STAR(p) model), and $\widehat{W} = T^{-1}(\sum_{t=1}^{T} x_t x_t')\widehat{\sigma}_\epsilon^2$ (with $\widehat{\sigma}_\epsilon^2 = T^{-1}(\sum_{t=1}^{T} \epsilon_t^2)$). This procedure is somewhat informal (especially since we use $w_t G(s_t; \widehat{\gamma}, \widehat{c})$ to approximate $w_t G(s_t; \gamma, c)$), but we call it our RW-STAR test because high values of L will be consistent with a RW-STAR specification.

When considering the estimation of a RW-STAR specification, our suggested modelling strategy is to start by performing STAR tests on an AR(p) model chosen by AIC (or some other model selection procedure). Typically, one will perform a sequence of STAR tests, corresponding to a sequence of possible transition variables ($s_t = y_{t-d}$, for $d = 1, 2, \ldots, p$), and then focus on the test result (and corresponding d^*) that gives rise to the lowest p-value. If that result is statistically significant (one may want to adjust critical values to account for the fact that one has conducted a sequence of tests rather than just one test), then it is worthwhile looking for evidence of parameter instability. This can be done in one of several ways. To look for random walk variation one could simply perform Nyblom's test on the AR(p) model, and/or one might estimate the STAR(p) model (with transition variable $s_t = y_{t-d^*}$) implied by the STAR tests, and then test for RW-STAR behavior conditional on that estimated STAR(p) model. Alternatively one might simply use the TV-STAR test developed by Lundbergh et al (2003) to look for evidence of parameter instability within a STAR framework. Rejection of the null for any of these parameter stability tests suggests that it will be worthwhile attempting the estimation of a RW-STAR(p) model (with transition variable $s_t = y_{t-d^*}$). We discuss estimation below. Obviously if the sequence of STAR tests does not lead to a rejection of the null hypothesis, then attempts to estimate a RW-STAR model are unlikely to be successful.

3.1. Performance of the nonlinearity tests

The specification procedure described above relies on a sequence of tests rather than a single direct test, to provide evidence of RW-STAR behavior. Given this, it is potentially useful to determine how the various nonlinearity tests embodied in (3.1) and (3.2) will behave in a RW-STAR setting, and what these tests are likely to find when related forms of nonlinearity are present in the data. We therefore undertake a small Monte Carlo study to explore these issues.

Our study involves five DGPs, which include two models without regimes (an AR(4) and a RWAR(4)), and three STAR models (STAR(4), TV-STAR(4) and RW-STAR(4)) in which the transition variable is y_{t-1}. We include the TV-STAR DGP in our experiments because researchers might be considering both RW- and TV-STAR models as possible alternatives, and it is potentially useful to know how our tests perform in each case. We use four lags in each of our models to ensure that our DGPs have non-trivial dynamics, and we (roughly) calibrate our DGP coefficients to estimated models based on an aggregated index of seven OECD7 data described below. These DGPs are reported in Appendix A, and they are taken to represent the sorts of series that might be encountered in practice.

We consider seven sets of nonlinearity tests in our study. The first is a Nyblom (1989) test for constancy of parameters, and this is followed by tests for STAR and TV-STAR behavior in situations where the delay parameter d is known. The TV-STAR test is outlined in Lundbergh et al (2003), and we include it in our experiments because we expect it to have power against RW-STAR specifications. The fourth test rejects linearity when both the Nyblom and STAR tests reject this null, and it provides information on when both sorts of nonlinearity (i.e. regime switching and non constancy of parameters) appear to be present in the data. The final three tests repeat the previous three tests, but in a more realistic setting where the researcher does not know the delay parameter. In these last three cases, the test conclusion is based on the lowest p-value found for $d = 1, \ldots, 4$. Of the full set of seven tests, only the first three are likely to have nominal size, but since the remaining "tests" may provide the researcher with information that might aid subsequent model specification, it is useful to know the extent of any size distortion. We do not include our RW-STAR test in our Monte Carlo study, due to the practical difficulties involved in estimating an LSTAR model for each replication[1], but we illustrate its use in our empirical application below.

We report rejection frequencies for the thirty five experiments (7 tests on each of 5 DGPs) in Table 1. Each experiment is based on 10,000 replications of samples of 400, and all tests are conducted on the last 300 observations at the 5% level of significance. Key findings are as follows:

(i) The Nyblom test has reasonable power (57.54%) against the RW-STAR DGP, and little power against the other STAR DGPs.

(ii) The STAR test has moderate power against the RW-STAR DGP (52.77% in a known transition delay setting), although it also has similar power against other nonlinear DGPs. Thus, it appears that in contrast to the Nyblom test, the STAR test is sensitive to different sorts of nonlinearity, rather than just the STAR form

[1]Estimation of LSTAR models often involves the maximization of badly behaved likelihoods when the true DGP is not LSTAR. Under such circumstances, an automated estimation program that does not involve human intervention is very unlikely to find the global maximum.

of nonlinearity.

(iii) The joint use of both nonlinearity tests is unreliable, in the sense that given a RW-STAR DGP, the probability that both tests will reject linearity is only about 30%.

(iv) Not surprisingly, the rejection frequencies for the STAR tests that minimize p-values over $d = 1, \dots, 4$ are considerably higher than those using the correct delay. This indicates that one needs to use results from these sets of tests quite conservatively.

(v) The TV-STAR tests had good power (76.16% and 86.52%) against the RW-STAR processes.

Overall the simulations suggest that RW-STAR behavior can be detected relatively easily, although without the additional information supplied via the Nyblom tests, it seems quite possible that the researcher might confuse RW-STAR and TV-STAR DGPs.

4. Modelling industrial production of selected OECD countries

In this section we provide a detailed analysis of the nonlinearity properties of GDP indicators for various countries. Our aim is to compare the abilities of different nonlinear specifications with respect to capturing both long-run and short-run structural change, and to compare the forecasting abilities of each type of model.

4.1. The data

Our data consists of seasonally adjusted values of the logarithmic monthly indices of industrial production for US, Japan, France, Germany, UK and an aggregated index of seven OECD countries, abstracted from the OECD Main Economic Indicators. Observations from January 1962 to December 1999 are used for testing and in-sample estimation, and observations from January 2000 to December 2003 are used for forecasting. The French data has been adjusted for the effects of strikes in 1968, prior to undertaking any analysis, and two outlying observations are removed from each of the

German and United Kingdom series. All analysis is based on (100×) annual growth rates (first differenced data) which are approximately stationary. The six data series are illustrated in Figure 1, which shows the cyclical behavior of each series. Preliminary analysis finds that 12 lag AR(p) specifications are appropriate for all countries except for the OECD (where an AR(10) specification seems more appropriate), and all subsequent testing and modelling is based on the 12 lag specification.

4.2. Linearity tests

We first undertake a set of tests to assess the extent of nonlinearity in each series. We focus on looking for evidence of regime switching, structural change and combinations of these characteristics, and report the results of our tests in Table 2.

In Table 2, the p-values for the LSTAR tests (column 5) relate to the minimum p-value obtained when using Teräsvirta's (1994) third order LSTAR tests for $d = 1, \ldots, 12$. We choose to present this test in addition to the STAR test in (3.1) because of its superior power against LSTAR processes, and our belief that as business cycle indicators, indices of industrial production were more likely to follow LSTAR processes than ESTAR processes. Our reported p-values, ranging from 0.0000 (for the United Kingdom and OECD) to 0.0403 (for Japan), show clear evidence of regime switching in each of the six series, supporting the estimation of an LSTAR specification for each country. In each case, the reported value of d_L (in column 4) corresponds to the delay giving rise to the minimized p-value, but we note that for each country there are always several values of d_L that gave rise to a statistically significant test statistic. The STAR tests in (3.1) (see column 8 for p-values and column 7 for the delay that achieves the minimum p-value) reflect very similar findings. The reported results suggest that each country index follows a STAR type process, and further testing (not reported) shows that there are several choices of d_S that lead to statistically significant test results.

We next test for non-constancy in parameters. The p-values obtained for the Nyblom tests provide strong evidence of non constant parameters in the Japanese series, and weaker evidence of this for the United Kingdom. These tests do not assume regime switching and are not really tests of RW-STAR behavior, but the observation that both

STAR and Nyblom tests are statistically significant for Japan and the UK suggests that parameter non-constancy and regime switching are present in each of these two series. Since the LSTAR tests have found strong evidence of a LSTAR type nonlinearity in each of the six series, we next estimate LSTAR models (with the delay reported in Table 4), and base our RW-STAR tests on these models. P-values for these RW-STAR tests are reported in column 6. These tests do not find evidence of RW-STAR characteristics, except for possibly in the Japanese case. We note, however, that the Lundbergh et al (2003) tests for TV-STAR behavior (in column 9 in Table 2) all find very strong evidence of parameter change in an LSTAR framework, and we use these latter tests to justify the estimation of both TV-STAR and RW-STAR models.

4.3. Development of baseline models

We start by estimating the linear and RWAR models that we will use as reference points for assessing the relative merits of RW-STAR models. All linear models are initially based on AR(12) specifications, and for each country we remove the least statistically significant coefficients, one at a time, until such time as AIC stops improving. Summary results for these linear models are reported in Table 3.

We base our RWAR models on our (reduced) linear models and equation (2.4), and estimate them by using the Kalman filter to compute the prediction error decomposition of the likelihood. The AR coefficient state variables π_t are obtained by filtering, conditional on the last observation y_{t-1} for $t = 1, 2, \ldots, T$, and their estimation is improved by smoothing through a backwards recursion algorithm so that they are re-estimated using all past observations up to y_{t-1}. We use the estimated $\widehat{\pi}$ from our (reduced) linear models as seeds for the random walk processes by making the assumption that $\pi_{t=0}$ is distributed N($\widehat{\pi}$, Q_1) for large Q_1. We initially set the variance σ_{ε}^2 equal to 1, and as the state variances, which reflect the degree of variation of the time varying parameters in the state vector are likely to be very small, we transform them into log variances and initilise them at -5 (and then later we use the delta method to obtain asymptotic standard errors when we convert log variances back to into variances). Details relating to the use of the Kalman filter can be found in Harvey (1989), Kim and Nelson (1999) and Durbin and Koopman (2000).

The estimated RWAR models for Germany, UK and the OECD are very similar to their linear counterparts, in that their random coefficients stay very close to the corresponding linear (and constant) parameter estimates. In contrast, the random coefficients for the intercept and first lag terms in the RWAR models for the United States, Japan and France all show considerable variation, leading to a noticeable reduction in both the sum of squared errors (ESS) and AIC. Figure 2 illustrates/compares the relevant OLS and RWAR coefficients for these last three countries, while summary statistics relating to all countries are presented in Table 3. The reported ESS for the RWAR models are based on smoothed estimates, and they are scaled up to account for the observations that are lost when starting up the estimation of the random coefficients. The reported AIC measures for the RWAR models are also adjusted to allow direct comparability with AIC for the linear (and other) models.

The LSTAR models are estimated using nonlinear least squares. Given that the joint estimation of γ and c is often difficult (see Teräsvirta (1994)), we use a preliminary grid search to find good starting values for these parameters and we standardize the transition function during estimation. For each country except the OECD aggregate, we use the transition variable that minimized the p-value for the LSTAR tests, first estimating equation (2.1) with (2.2) and $p = 12$, and then imposing coefficient restrictions on the model, one at a time, until AIC stopped improving. In the first instance, we were unable to obtain satisfactory estimates for the OECD case by following this procedure, but we had no problems once we used y_{t-2} as the transition variable, rather than y_{t-1}. An LSTAR test supported this alternative transition variable (the p-value is 0.0003). Most of the restrictions that were imposed during the reduction of the general LSTAR(12) model were simple exclusion restrictions (imposed when a coefficient estimate was clearly statistically insignificant), but on several occasions restrictions of the form $\pi_{1k} = -\pi_{2k}$ were imposed (after testing that this was appropriate). This latter sort of restriction implies that some of the dynamic features that characterize recessionary phases of the business cycle, disappear during expansionary phases of the cycle.

Key features of our estimated LSTAR models are reported in Table 4, and the implied transition functions are illustrated in Figure 3. The transition functions are steep for all countries excepting Japan and Germany, so that for most countries, small changes in the transition variable can cause quite sharp changes in dynamics. The upper regimes for Japan and Germany contain relatively few observations, but estimates are precise, presumably because transition between regimes is slow in this case. Diagnostic testing finds no serial correlation in the residuals of these LSTAR models and simulations show that all of them are stationary. As should be the case, the sum of squared residuals for the LSTAR models are lower than those for the corresponding linear models, but AIC is also lower for each country, suggesting that LSTAR models provide a better representation of the data than the linear models. Comparing the fit of the LSTAR and RWAR models, we find that although the RWAR model fits better in the US and Japanese cases, the converse is true in the remaining cases.

The procedures that we use to estimate our TV-STAR models are very similar to those used for the LSTAR models, although we use simulated annealing rather than a grid search to find good starting values for c_1, c_2, γ_1 and γ_2. See Brooks and Morgan (1999) for discussion on simulated annealing. For each country, we use the transition variable (in column 7 of Table 2) that minimized the p-value for the TV-STAR tests, and as above, we first estimated a very general version of equation (2.7) (with $p = 12$), and then removed insignificant autoregressive variables one by one until such time as AIC stopped improving.

The important features of the final TV-STAR specifications are reported in Table 5, and transition functions in y_{t-d} and t are respectively illustrated in Figures 4 and 5. With the exceptions of the United Kingdom and the OECD, the transition functions in y_{t-d} were steeper than those for the LSTAR models, and the centers of all transition functions in y_{t-d} also moved. The TV-STAR lower regime for the UK contains relatively few observations, but estimates are precise, presumably because transition between regimes is slow in this case. The transition functions in time are steep, and with the exception of Germany, all indicate structural change in about the mid seventies (about a third of the way through the sample). All models pass serial correlation tests and appear to be stochastically stationary. Not surprisingly, the sums of squared errors associated with these models are lower than those for the corresponding LSTAR models, but AIC is also lower. The TV-STAR model also has better fit than the RWAR model.

4.4. Estimation of RW-STAR models

Our estimation of the RW-STAR models is based on the Kalman filter and is analogous in many respects to our estimation of the RWAR models. In this case our time varying parameters are $(\pi_{1,t},\ \pi_{2,t})'$, and we start estimation by assuming that the starting values for $(\pi_{1,t},\ \pi_{2,t})'$ are distributed N($(\pi_1,\ \pi_2)'\ P_1$) where $(\pi_1,\ \pi_2) = (\widehat{\pi}_1,\ \widehat{\pi}_2)$ are the (reduced) estimated LSTAR(12) coefficients and P_1 is diagonal with large elements. As before, we initially set the variance σ_ε^2 equal to 1, and after transforming the variances of the state vector into logs, we initilise them at -5. For each country, the RW-STAR models uses the same transition variable (i.e. y_{t-d}) as in the LSTAR model. The key differences between the estimation of these two models is that the RW-STAR model has additional parameters (γ and c) associated with transition function, and that this model contains two sets of time varying parameters rather than just one.

We specified γ and c in our RW-STAR models as constant rather than time varying parameters, because of the well known problems in jointly estimating these parameters in LSTAR contexts and our desire to keep this "first pass" at estimating RW-STAR models relatively simple. This specification embodies the palatable assumption that demarcation between recessions and expansions stays constant through-out the sample, but it also assumes (perhaps unrealistically) that the speed of transition between these phases is always the same. It turned out that our treatment of γ and c as constant parameters did not sufficiently simplify the estimation problem, and although we would have liked to let γ be a free parameter, we resorted to conditioning on γ after experiencing difficulties with convergence. For each country, we conditioned on the value of γ obtained from the corresponding LSTAR model. In general, we were able to estimate c as a free parameter, although for the US model we had to restrict c to ensure that each regime contained sufficient observations to allow the estimation of $(\pi_{1,t},\ \pi_{2,t})'$.

As noted above, the estimation of time varying parameters in a RW-STAR model is quite different from that in a RWAR model because there are now two sets of coefficients to estimate rather than one. This is not an innocuous difference, because all sample observations influence the estimation of each π_t in the RWAR case, while the value of $G(y_{t-d};\gamma,c)$ in the RW-STAR case determines whether an observation will influence the estimated $\widehat{\pi}_{1,t}$, the estimated $\widehat{\pi}_{2,t}$, or both $\widehat{\pi}_{1,t}$ and $\widehat{\pi}_{2,t}$. Insufficient variation in $G(y_{t-d}; \gamma, c)$ in RW-STAR models can lead to estimation problems that resemble those that are sometimes experienced when estimating the π parameters in LSTAR models. In particular, identification problems can occur if $G(y_{t-d}; \gamma, c) \simeq 0$ for most of the sample, (such a sample will contain very little information about the upper regime parameters), and a similar lack of information can affect the identification and estimation of the lower regime parameters if $G(y_{t-d}; \gamma, c) \simeq 1$ for most of the sample. The location of c (and to a lesser extent the value of γ) influences the values that $G(y_{t-d}; \gamma, c)$ can take, with estimation problems being more severe, when c implies just a few observations in one of the regimes. The problem is alleviated if γ is quite small (because a small γ often allows $G(y_{t-d}; \gamma, c)$ to take values that are well away from 0 and 1), and it would also be alleviated if we impose cross regime restrictions on the π parameters, because then observations that fall in one of the regimes could influence the estimation of parameters associated with both regimes. Given that one is estimating how parameters *move* in a RW-STAR model (as opposed to estimating fixed parameters in an LSTAR model), one might expect this sort of identification problem to be much more severe in a RW-STAR setting. Thus, it is important to pay close attention to the estimated values of c (and also γ) for RW-STAR models, and to check that the implied transition function is likely to allow the identification of $(\pi_{1,t}, \pi_{2,t})'$.

We report summary results for our RW-STAR models in Tables 6a, and in general these models fit better than RWAR models (US is an exception), but not as well as TV-STAR models (Japan is an exception). Table 6a also contains details about the estimated transition functions and they are illustrated in Figure 6. In general the RW-STAR transition functions have centers that are closer to the median of the data than their LSTAR counterparts, although this is not true in the Japanese case. Here, although the centre (2.91%) is near upper end of the distribution of the Japanese data, we are able to identify movement in the upper regime parameters, because the small γ ensures that $G(y_{t-d}; \gamma, c)$ is sufficiently often away from 0. See Figure 6. The centre of the German RW-STAR transition function is also quite large (2.21%), but again a graph of the German transition functions show that $G(y_{t-d}; \gamma, c)$ is often away from 0, so we are not concerned about a lack of identification. The interesting point with respect to the Japanese and German cases is that the lower regimes in these cases are better interpreted as periods when growth is not strong, rather than recessionary regimes. Only when growth gets very strong do these economies venture into their upper regimes. As mentioned above, in the US case unrestricted estimation of c caused identification problems (indeed the estimation algorithm placed c outside the range of the data), but once we imposed the restriction that $-2 < c < 2$, the estimation algorithm converged to a solution that was likely to allow precise estimation.

We report the start values, end values and variation in each of our smoothed RW-STAR coefficients in Tables 6b to 6g, and provide graphs of the some of these coefficients in Figure 7. The tables show that many of the coefficients experience very little fluctuation, but we see substantial fluctuation in at least three or four of the coefficients for each country. In all countries, we see variation in at least one of the coefficients for each regime, and usually it is one of the intercepts (π_{10t} or π_{20t}) or low order AR coefficients that varies the most. Figure 7 plots the three most variable coefficients for each country. Here, we see that movement in the coefficients is not always monotonic, as would be predicted by a TV-STAR model. For the US, it is interesting to note that coefficient movements are very similar to those in Figure 2.

4.5. Forecast performance

This section evaluates the RWAR, LSTAR, TV-STAR and RW-STAR specifications by comparing their one-step-ahead forecast performance with AR models. The forecasting sample covers 48 months from January 2000 to December 2003, and all forecasts are bona fide in the sense that model specifications are not updated over the forecasting period. Given recent claims (see e.g. Clements and Smith (2000)) that the forecasting advantages associated with nonlinear models may not become apparent if evaluation is based solely on forecast root mean squared error (RMSE) criteria, we provide a broad forecast analysis that includes Diebold Mariano (1995) tests for equality of forecasting accuracy (DM tests), Pearson goodness of fit tests (GOF tests) as modified by Anderson (1994), and the forecast density evaluation methods (FDE methods) suggested by Diebold, Gunther and Tay (1998). Our analysis follows recent work by Boero and

Marrocu (2004), who provide an accessible overview of GOF tests and FDE methods.

Table 7 provides an evaluation of point forecasts for each of our models. Here, we see that when assessing RMSE, there are no clear winners or losers. The AR model is best (i.e. lowest) for two countries but also worst (highest) for two countries, and the same is true for the TV-STAR model. The RW-STAR model is neither best nor worse for any country. DM tests applied to the mean squared forecasting errors indicate some statistically significant differences, with the forecasts from RW-STAR models outperforming RWAR forecasts for the United States, and outperforming the LSTAR and TV-STAR models for Japan.

The remainder of our evaluation techniques are based on probability integral transforms (z_t) of the actual realization (y_t) with respect to a model's forecast density $p_t(y_t)$,

$$\text{i.e.} \qquad z_t = \int_{-\infty}^{y_t} p_t(u)du = P_t(y_t)$$

The set of probability integral transforms will be i.i.d. $U(0,1)$ random variables if the forecast density follows the true density, and the forecast density evaluation methods that we use are all based on this fact. We plot our (transformed) empirical forecast densities to gain a visual impression of whether or not they are uniform, and use Lilliefor's (1967) critical values of the Kolmogorov Smirnov tests to draw 95% confidence bands for the observed z_t under the null that that they are i.i.d. $U(0,1)$. The independence aspect of the $i.i.d.U(0,1)$ hypothesis are tested by applying Ljung-Box tests to the autocorrelations of $(z_t - \overline{z})$, $(z_t - \overline{z})^2$ and $(z_t - \overline{z})^3$, while the GOF tests look for departures from the uniform distribution by looking at features such as the location, scale and skewness of the (transformed) forecast distribution. We use six autocorrelations for our Ljung-Box tests, and eight partitions of our forecast density for our GOF tests.

We obtain our one-step-ahead forecast densities by using a modified bootstrap method which draws from a $U(0,1)$ distribution that has been indexed to the cumulative distribution of the in-sample errors. When obtaining the bootstrap sample, the cumulative distributions are made continuous by scaling the intermediate points between any two consecutive indexed errors that have been drawn. One step ahead forecast densities for RWAR and RW-STAR models are simulated by sampling both smoothed state errors and smoothed observation errors.

Plots of the empirical distribution function (versus the theoretical distribution) of the z_t series are provided in Figure 8. The interesting feature of these plots is that all six TV-STAR models perform well, while most other forecasts (excepting those for Germany) deviate outside the bands that should contain an i.i.d. $U(0,1)$ distribution with 95% confidence. In most cases this deviation is not severe, with only the US LSTAR and UK RWAR straying a long way from the confidence bands.

The Ljung-Box tests for the autocorrelation of the z_t series and its powers (see the first three columns of Table 8) find little evidence of dependence in the transformed forecast errors. This is consistent with the i.i.d. property that will characterize the forecast density if it follows the true density, and it also indicates that the i.i.d. assumption that is needed to use the 95% confidence intervals in Figure 8, is roughly satisfied.

The last three columns of Table 8 show why most of the empirical forecast densities for LSTAR RWAR and RW-STAR models do not follow the true forecast densities. The main problem appears to be with the empirical median, which differs from the true median in nearly all cases where the (transformed) forecasting densities stray from the $U(0,1)$ distribution. This is consistent with Clements and Smith's (2000) observation that point forecasts from nonlinear models can be misleading. The statistics in the "location" column indicate that the RW-STAR forecasts track the median a little better than AR and LSTAR forecasts, and about the same as the RWAR forecasts. The last two columns of Table 8 shows that there are only a few problems with scale and skewness.

5. Conclusion

The RW-STAR model proposed in this paper provides a potentially useful framework for studying parameter change in regime switching environment. Using industrial production data for several countries, we find evidence of non-constant parameters in a setting where there is also evidence of regime switching, and we also find that RW-STAR models seem to be able to capture this behavior. Some of our estimated random walk coefficients move very little, consistent with behavior implied by standard STAR models, while some of our estimated random walk coefficients gradually increase (or decrease) over time, consistent with behavior implied by TV-STAR models. However, some of our RW-STAR models have coefficients that change in ways that neither LSTAR nor TV-STAR models can capture, and we therefore conclude that RW-STAR models can provide a potentially useful way of capturing time variation in regime specific parameters.

Forecast density evaluations show that the TV-STAR models outperform the RW-STAR models, and this suggests that TV-STAR models may have captured permanent structural change better. The interesting thing here is that the TV-STAR models all embody a sudden structural change in the early to mid seventies (and not in the early eighties as documented by Kim and Nelson (1999a)), so that the TV-STAR predictions are essentially LSTAR predictions based on data from the mid seventies onwards. It is noteworthy that some of our RW-STAR coefficients for the US, France and the UK seem to track this structural change in the "seventies", and that some also change direction at the time of the well known "volatility decline" in 1984. See Figure 7. The "bump" in the US π_{11} coefficient in 1990 might also be associated with the first Gulf war. It is apparent that RW-STAR coefficients can sometimes track historical episodes, and thus it seems useful to undertake further research on the applications of RW-STAR models.

The RW-STAR models presented here are quite primitive, but in sample diagnostics based on standardizes residuals find serial correlation only in the OECD model. This might be removed by estimating, rather than fixing the transition parameter. Further, after estimating the RW-STAR model and determining which parameters "move" and which ones stay constant, it might also be useful to undertake a second estimation stage that fixes and then reestimates those parameters that remained constant during the initial estimation, despite their random walk specification. We leave these refinements for future research.

6. References

Anderson, G. (1994), "Simple tests of distributional form", *Journal of Econometrics*, 62, 265-276.

Beaudry, P. and G. Koop (1993), "Do recessions permanently change output?", J*ournal of Monetary Economics*, 31, 149-163.

Boero, G. and E. Marrocu (2004), "The performance of SETAR models: a regime conditional evaluation of point, interval and density forecasts", *International Journal of Forecasting*", 20, 305-320.

Brooks, S.P. and B.J.T. Morgan, (1999), "Optimisation using simulated annealing", *The Statistician*, 44.

Clements, M.P. and D.F. Hendry, (2001) "Forecasting: non-stationary economic time series" **MIT Press.**

Clements, M.P. and J.P. Smith (2000), "Evaluating the forecast densities from linear and nonlinear models: applications to output growth and unemployment", *Journal of Forecasting*, 19, 255 - 276.

Diebold, F.X., T.A. Gunther, and A.S. Tay (1998), "Evaluating density forecasts with applications to financial risk management", *International Economic Review*, 39(4), 863-883.

Durbin, J. and S.J. Koopman (2000), "Time series analysis by state space methods", **Oxford University Press**.

Eitreheim, Ø. and T. Teräsvirta (1996), "Testing the adequacy of smooth transition autoregressive models", *Journal of Econometrics*, 74, 59 - 76.

Harvey, A.C. (1989),"Forecasting, structural time series models and the Kalman filter", **Cambridge University Press**.

Kim, C.J. and C.R. Nelson, (1989) "The time-varying-parameter model as an alternative to ARCH for modelling changing conditional variance: the case of Lucas hypothesis", *Journal of Business and Economic Statistics*, 7, 433 - 440

Kim, C.J. and C.R. Nelson (1999a), "Has the US economy become more stable? A Bayesian approach based on a Markov switching model of the US business cycle", *Review of Economics and Statistics*, 81, 608 - 616.

Kim, C.J. and C.R. Nelson (1999b), "State space models with regime switching", **MIT Press**.

Lilliefors, H.W. (1967), "On the Kolmogorov-Smirnov test for normality with the mean and variance unknown", *Journal of the American Statistical Association*, 62, 399-402.

Lin, C.F. and T. Teräsvirta (1994),"Testing constancy of regression parameters against continuous structural change", *Journal of Econometrics*, 62, 211-228.

Lin, C.F. and T. Teräsvirta (1999),"Testing parameter constancy in linear models against stochastic stationary parameters", *Journal of Econometrics*, 90, 193-213.

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

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

McConnell, M. M. and G. Perez Quiros, (2000), "Output fluctuations in the United States" What has changed since the early 1980s?, *American Economic Review*, 90, 1464-1476.

Nyblom, J. (1989), "Testing for the constancy of parameters over time", *Journal of the American Statistical Association*, 84, #405, 223 - 230.

Pesaran, M.H. and S.M. Potter (1997), "A floor and ceiling model of US output", *Journal of Economic Dynamics and Control*, 21, 661-695.

Stock, J.H. and M.W. Watson (1996), "Evidence on structural instability in macroeconomic time series relations", *Journal of Business and Economic Statistics*, 14 (1996), 11-30.

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

Teräsvirta, T. and H.M. Anderson (1992), " Characterising nonlinearities in business cycles using smooth transition autoregressive models", *Journal of Applied Econometrics*, 7, 208-218.

APPENDIX A: DGPS FOR THE POWER SIMULATIONS

The coefficients in these DGP are taken from estimated versions of these models applied to the twelfth differenced (logs of) the OECD data. The first four simulated values for all DGPs are obtained from the generating process N(1.3,1.7), and then subsequent values are based on following. We generate 10,000 replications of samples of 400 observations and conduct our tests on the last three hundred simulated observations.

AR(4) MODEL:

$y_t = 0.08 + 0.8y_{t-1} + 0.2y_{t-2} - 0.1y_{t-3} - 0.3y_{t-4} + \varepsilon_t$, where $\varepsilon_t \sim N(0.0.35)$.

AR(4) MODEL WITH RANDOM COEFFICIENTS:

$\phi_0 = [0.03, -0.6, 0.4, 0.4, -0.2]$

$\phi_t = \phi_{t-1} + \eta_t$,where $\eta_{1,t} \sim N(0, 0.01)$

$y_t = x_t^{'}\phi_t + \varepsilon_t$,

where $\varepsilon_t \sim N(0.0.35)$, and $x_t = [1, y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}]$

LSTAR(4) MODEL:

$y_t = 0.03 + 0.8y_{t-1} - 0.2y_{t-2} - 0.3y_{t-3} - 0.1y_{t-4}+$

$(0.07 - 0.6y_{t-1} + 0.4y_{t-2} + 0.4y_{t-3} - 0.2y_{t-4})\,(1 + \exp(-2.72(y_{t-1} + 0.2)))^{-1} + \varepsilon_{t,}$

where $\varepsilon_t \sim N(0.0.35)$.

TV-STAR(4) MODEL

$\phi_1 = [0.34, 0.8, -0.1, -0.6, -0.5]$

$\phi_2 = [0.27, 0.96, 0.04, -0.1, 0.02]$

$\phi_3 = [0.03, 0.63, 0.06, 0.3, 0.06]$

$\phi_4 = [0.06, 0.8, 0.4, 0.3, -0.6]$

$y_t = x_t^{'}\phi_1(1 - G(t))(1 - G(y_{t-1})) + x_t^{'}\phi_2(1 - G(t))G(y_{t-1})+$

$x_t^{'}\phi_3 G(t)(1 - G(y_{t-1})) + x_t^{'}\phi_4 G(t)G(y_{t-1}) + \varepsilon_t$

where $\varepsilon_t \sim N(0.0.35)$, $x_t = [1, y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}]$, $G(y_{t-1}) = (1 + \exp(-16(y_{t-1} - 0.35)))^{-1}$

and $G(t) = (1 + \exp(-2.72(t/300 - 0.3)))^{-1}$

RW-STAR(4) MODEL

Initial values of parameters

$\phi_{1,0} = [0.03, 0.8, -0.2, -0.3, -0.1]$

$\phi_{2,0} = [0.07, -0.6, 0.4, 0.4, -0.2]$

$\phi_{1,t} = \phi_{1,t-1} + \eta_{1,t}$,where $\eta_{1,t} \sim N(0, 0.01)$ and

$\phi_{2,t} = \phi_{2,t-1} + \eta_{2,t}$,where $\eta_{2,t} \sim N(0, 0.01)$.

$$y_t = \begin{bmatrix} x_t^{'} & x_t^{'}G(y_{t-1}) \end{bmatrix} \begin{bmatrix} \phi_{1,t} \\ \phi_{2,t} \end{bmatrix} + \varepsilon_t,$$

where $\varepsilon_t \sim N(0.0.35)$, $x_t = [1, y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}]$

and $G(y_{t-1}) = (1 + \exp(-2.72(y_{t-1} + 0.74)))^{-1}$

Table 1: Power of Nonlinearity Tests (% rejections)
(10,000 replications of samples of 300, nominal test size is 5%)

DGP	TEST						
	NYBLOM	STAR (kd)	TVSTAR (kd)	N&STAR (kd)	STAR (ud)	TVSTAR (ud)	N&STAR (ud)
AR(4)	2.13	5.00	4.77	0.15	15.46	12.82	0.51
RWAR(4)	68.24	52.52	84.75	31.77	69.18	88.70	46.63
STAR(4)	5.37	51.54	34.24	3.48	84.32	67.54	4.82
TVSTAR(4)	13.73	48.66	62.30	7.68	96.09	93.62	13.50
RWSTAR(4)	57.54	52.77	76.16	29.95	73.21	86.52	43.18

The DGPs are given in Appendix A. The symbol (kd) indicates that the test assumes knowledge of the delay (i.e. $d = 1$), while the symbol (ud) indicates that the test minimises the p-value over $d = 1, \ldots, 4$.

Table 2: P-Values for Nonlinearity Tests

Country	AR BASED TESTS		STAR BASED TESTS					
	Nyblom	TV-AR	d_L	LSTAR	RW-STAR	d_S	STAR	TVSTAR
United States	0.2090	0.3376	1	0.0075	0.2553	11	0.0328	0.0868
Japan	0.0003	0.0000	1	0.0403	0.1068	3	0.0055	0.0000
France	0.1780	0.2499	9	0.0107	0.7100	9	0.0252	0.0217
Germany	0.2380	0.0090	1	0.0131	0.9497	5	0.0061	0.0074
United Kingdom	0.0891	0.7120	1	0.0000	0.8826	1	0.0009	0.0017
OECD7	0.2380	0.0668	1	0.0000	0.1553	1	0.0000	0.0000

d_L is the transition variable lag that minimises the p-value for the LSTAR test and d_S is the transition variable lag that minimises the p-value for the STAR and TVSTAR tests

Table 3: Summary Details of Linear and RWAR models
(effective sample of 444 observations from 1963:1 to 1999:12)

Country	LINEAR MODELS			RWAR MODELS		
	Parameters	ESS	AIC	Parameters	ESS	AIC
United States	8	196.86	2.06	16	174.15	1.97
Japan	8	634.17	3.23	16	553.89	3.13
France	6	691.05	3.31	12	662.13	3.29
Germany	9	944.45	3.63	18	919.93	3.64
United Kingdom	5	742.26	3.37	10	726.87	3.37
OECD7	6	155.18	1.81	12	152.05	1.82

Linear models are AR(12) models after statistically insignificant coefficients have been removed. RWAR models are the (reduced) AR models with random walk coefficients.

Table 4: Summary Details of LSTAR models
(effective sample of 444 observations from 1963:1 to 1999:12)

Country	Parameters	d	$\widehat{c}$	$\widehat{\gamma}$	ESS	AIC
United States	12	1	-0.43	78.56	184.76	2.01
Japan	19	1	2.72	2.06	573.26	3.18
France	17	9	-1.51	∞	633.71	3.27
Germany	20	1	2.81	1.77	855.67	3.58
United Kingdom	15	1	0.91	∞	659.61	3.30
OECD7	10	2	-0.28	∞	136.98	1.71

Reported data relates to LSTAR(12) models, after statistically insignificant coefficients have been removed. A ∞ reported in the $\widehat{\gamma}$ column means that the LSTAR model is effectively a threshold model.

Table 5: Summary Details of TV-STAR models
(effective sample of 444 observations from 1963:1 to 1999:12)

Country	Parameters	d	$\widehat{c}_1$	$\widehat{\gamma}_1$	$\widehat{c}_2$	$\widehat{\gamma}_2$	ESS	AIC
United States	28	11	-0.22	∞	1975:03	∞	161.01	1.95
Japan	24	3	-0.41	∞	1976:10	∞	512.90	3.09
France	27	9	-1.01	106	1976:04	∞	577.24	3.22
Germany	30	5	-0.89	∞	1968:01	∞	793.24	3.55
United Kingdom	33	1	-2.77	0.63	1972:11	∞	578.84	3.26
OECD7	28	1	-0.24	2.67	1974:12	∞	119.55	1.65

Reported data relates to TVSTAR(12) models, after statistically insignificant coefficients have been removed. A ∞ reported in the $\widehat{\gamma}_1$ column means that the LSTAR features of the model are effectively threshold features. A ∞ reported in the $\widehat{\gamma}_2$ structural change in the model is effectively a structural shift.

Table 6a: Summary Details of RW-STAR models
(effective sample of 444 observations from 1963:1 to 1999:12)

Country	Parameters	d	$\widehat{c}$	$\widehat{\gamma}$	ESS	AIC
United States	22	1	0.97	78.56	172.82	1.99
Japan	36	1	2.91	2.06	479.35	3.08
France	32	9	-1.02	∞	597.28	3.28
Germany	38	1	2.21	1.77	807.19	3.61
United Kingdom	28	1	0.78	∞	653.86	3.35
OECD7	18	2	-0.28	∞	130.76	1.70

RW-STAR models are random walk versions of the reduced LSTAR models. A ∞ reported in the $\widehat{\gamma}$ column means that the LSTAR features of the model are effectively threshold features.

Table 6b: State Variables for RW-STAR model for the United States

Parameter	Start Value	End Value.	$\sigma_{\eta\pi}$
π_{10t}	0.2872	0.2923	0.0156
π_{11t}	-0.0543	-0.1337	0.0337
π_{12t}	0.0649	0.0658	0.0007
π_{13t}	0.1172	0.1172	0.0000
π_{14t}	0.2168	0.0248	0.0195
π_{19t} $(= -\pi_{19t})$	0.0755	0.0755	0.0000
π_{110t} $(= -\pi_{210t})$	0.0285	0.0285	0.0000
π_{112t} $(= -\pi_{212t})$	-0.0267	0.0899	0.0153
π_{20t}	0.3339	0.3339	0.0069
π_{21t}	0.2394	0.2394	0.0046

$\widehat{\sigma}_{\varepsilon} = 0.6601$

Coefficient π_{jk} refers to the coefficient for the kth lag in the jth regime.

Table 6c: State Variables for RW-STAR model for Japan

Parameter	Start Value	End Value	$\sigma_{\eta\pi}$
π_{11t}	-0.1532	-0.3579	0.0181
π_{12t}	0.1069	0.0770	0.0106
π_{13t}	0.5523	0.1200	0.0259
π_{14t}	0.1101	0.3029	0.0134
π_{15t}	0.0991	0.1016	0.0011
π_{18t}	0.0937	0.1077	0.0071
π_{19t}	0.0979	-0.0628	0.0147
π_{110t}	-0.1084	-0.1046	0.0022
π_{112t}	-0.0956	-0.0943	0.0034
π_{20t}	12.6554	12.6555	0.0046
π_{21t}	-3.7354	-3.7343	0.0041
π_{26t}	1.1913	-0.0160	0.0583
π_{27t}	-0.8388	-0.8388	0.0001
π_{28t}	-1.1028	-0.4503	0.0462
π_{29t}	-0.7151	-0.7219	0.0061
π_{210t}	0.8301	0.8420	0.0056
π_{211t}	-0.9470	-0.9516	0.0035

$\widehat{\sigma}_{\varepsilon} = 1.1266$

Coefficient π_{jk} refers to the coefficient for the kth lag in the jth regime

Table 6d: State Variables for RW-STAR model for France

Parameter	Start Value	End Value	$\sigma_{\eta\pi}$
π_{11t}	-1.8134	-0.4309	0.0000
π_{13t} $(= -\pi_{23t})$	0.2667	-0.1052	0.0463
π_{14t} $(= -\pi_{24t})$	-0.1398	-0.1405	0.0000
π_{16t}	0.1167	0.1595	0.0000
π_{18t}	0.4387	-0.1514	0.0000
π_{19t}	-0.4330	-0.2398	0.0000
π_{111t} $(= -\pi_{211t})$	-0.2811	0.2694	0.0312
π_{112t}	-0.0588	-0.0588	0.0000
π_{20t}	0.1005	0.1005	0.0000
π_{21t}	1.7170	0.0100	0.0135
π_{25t}	0.1127	0.1124	0.0005
π_{26t}	0.1192	0.0765	0.0000
π_{27t}	0.0753	0.0763	0.0008
π_{28t}	-0.2879	0.3023	0.0000
π_{29t}	0.5890	0.3957	0.0000

$\widehat{\sigma}_{\varepsilon} = 1.1968$

Coefficient π_{jk} refers to the coefficient for the kth lag in the jth regime

Table 6e: State Variables for RW-STAR model for Germany

Parameters	Start Value	End Value	$\sigma_{\eta\pi}$
π_{10t}	0.1033	0.0992	0.0018
π_{11t}	-0.3619	-0.3620	0.0003
π_{13t}	0.1720	0.1700	0.0008
π_{14t}	0.1324	0.1391	0.0013
π_{15t}	0.1563	0.1552	0.0006
π_{16t}	0.2282	0.1506	0.0081
π_{110t}	-0.0343	-0.0343	0.0001
π_{111t}	-0.0642	-0.0659	0.0008
π_{20t}	1.9762	1.3597	0.0464
π_{21t}	-0.5374	-0.5374	0.0000
π_{22t}	-0.7379	-0.7379	0.0001
π_{23t}	-0.2538	-0.2538	0.0002
π_{25t}	-0.4426	-0.4426	0.0002
π_{26t}	-0.5371	-0.6719	0.0139
π_{27t}	-0.2809	-0.2809	0.0002
π_{210t}	0.3352	0.3351	0.0006
π_{211t}	0.6504	0.0859	0.0549
π_{212t}	0.4913	0.4617	0.0061

$\widehat{\sigma}_{\varepsilon} =1.4058$

Coefficient π_{jk} refers to the coefficient for the kth lag in the jth regime

Table 6f: State Variables for RW-STAR model for UK

Parameters	Start Value	End Value	$\sigma_{\eta\pi}$
π_{10t}	0.1535	0.1535	0.0000
π_{13t}	0.0634	0.0787	0.0042
π_{14t}	0.1162	0.1162	0.0000
π_{15t}	0.1609	0.1609	0.0000
π_{18t}	0.1646	0.1646	0.0000
π_{21t}	-0.0372	-0.2574	0.0138
π_{22t}	-0.1317	-0.1317	0.0000
π_{23t}	0.2400	-0.1222	0.0406
π_{24t}	-0.2076	-0.2076	0.0000
π_{25t}	-0.1053	-0.0462	0.0256
π_{27t}	-0.0854	-0.0854	0.0000
π_{28t}	-0.3264	-0.3264	0.0000
π_{29t}	-0.1564	-0.1564	0.0000

$\widehat{\sigma}_\varepsilon = 1.2410$

Coefficient π_{jk} refers to the coefficient for the kth lag in the jth regime

Table 6g: State Variables for RW-STAR model for OECD7

Parameters	Start Value	End Value	$\sigma_{\eta\pi}$
π_{10t}	0.1794	0.1207	0.0044
π_{11t}	0.5880	0.4853	0.0149
π_{12t}	0.1897	0.1897	0.0000
π_{18t}	0.1088	0.1088	0.0000
π_{21t}	-0.7187	-0.7187	0.0000
π_{23t}	0.2979	0.2979	0.0000
π_{24t}	0.1728	0.1728	0.0002
π_{210t}	-0.1513	-0.0489	0.0099

$\widehat{\sigma}_\varepsilon = 0.5533$

Coefficient π_{jk} refers to the coefficient for the kth lag in the jth regime

Table 7: Out of Sample One Step Ahead Forecast RMSE
(48 observations from 2000:1 to 2003:12)

	AR	RWAR	LSTAR	TV-STAR	RW-STAR
USA	**0.4318**	*0.4868**	0.4577	0.4449	0.4843
Japan	1.4388	**1.3396**	*1.5605**	1.5403*	1.3423
France	*0.8426*	0.8292	0.8165	**0.7826**	0.7919
Germany	*1.1984*	1.1982	1.1751	**1.1743**	1.1815
UK	**1.0254**	1.0263	1.0763	*1.0886*	1.0544
OECD	0.4877	**0.4842**	0.4980	*0.5333*	0.5043

The lowest RMSE is indicated in bold type while the highest RMSE is indicated in italics. A star indicates that the Diebold Mariano test rejects H_0: The MSE of RW-STAR and starred forecasts are equivalent against the alternative that the RWSTAR forecasts better at the 5% level of significance.

Table 8: P-Values of Forecasting Performance Tests
(48 observations from 2000:1 to 2003:12)

Model	Ljung-Box Q_6 Statistics			Goodness of Fit Statistics		
	$(z-\overline{z})$	$(z-\overline{z})^2$	$(z-\overline{z})^3$	Location	Scale	Skewness
			USA			
AR	0.6601	0.1119	0.6009	0.0005	0.5637	0.0209
RWAR	0.0032	0.5311	0.0496	0.0039	0.5637	0.5637
LSTAR	0.6928	0.2374	0.8210	0.0005	0.3865	0.0094
TV-STAR	0.2730	0.6645	0.7510	0.1489	1.0000	0.3865
RW-STAR	0.1707	0.4478	0.4868	0.0015	0.2484	0.0094
			JAPAN			
AR	0.0821	0.3024	0.3094	0.0209	0.0015	0.0094
RWAR	0.1383	0.4337	0.4908	0.0833	0.0039	0.0202
LSTAR	0.1662	0.4549	0.5263	0.0209	0.0209	0.0833
TV-STAR	0.1901	0.7677	0.1891	0.2482	0.5637	0.2482
RW-STAR	0.3696	0.1914	0.4587	0.0209	0.0039	0.0833
			FRANCE			
AR	0.4088	0.2320	0.7966	0.0209	0.0209	0.2482
RWAR	0.8205	0.2755	0.8852	0.0433	0.0094	0.2482
LSTAR	0.3914	0.3137	0.7456	0.0039	0.0433	0.1489
TV-STAR	0.4789	0.5492	0.6906	0.0433	0.7729	0.5637
RW-STAR	0.6377	0.1769	0.5690	0.0094	0.0039	0.7729
			GERMANY			
AR	0.5517	0.7171	0.8408	0.0833	0.7729	0.0433
RWAR	0.5468	0.9705	0.9456	0.0433	0.7729	0.5637
LSTAR	0.7394	0.9581	0.9527	0.0833	0.5637	0.2482
TV-STAR	0.9646	0.5953	0.8966	0.1489	0.7729	0.5637
RW-STAR	0.7030	0.7572	0.8478	0.2482	1.0000	0.0209
			UK			
AR	0.2524	0.3305	0.0528	0.0002	1.0000	0.3865
RWAR	0.5105	0.7572	0.4150	0.0001	0.7729	0.0094
LSTAR	0.0770	0.6259	0.0818	0.0005	0.7729	0.1489
TV-STAR	0.0427	0.6065	0.0590	0.1489	0.7729	0.2482
RW-STAR	0.2040	0.4989	0.4850	0.0005	1.0000	0.5637
			OECD			
AR	0.2548	0.3778	0.1440	0.0039	0.0833	0.0209
RWAR	0.3549	0.6066	0.2707	0.0015	0.1489	0.0209
LSTAR	0.2266	0.7298	0.0570	0.0094	0.0833	0.0433
TV-STAR	0.6097	0.3955	0.0439	0.1489	1.0000	0.7729
RW-STAR	0.3549	0.6066	0.2707	0.0039	0.8330	0.2482

The Ljung Box statistics test that the (transformed) forecasts are i.i.d., while the goodness of fit statistics test that the moments of the transformed forecast match those from a U(0,1) distribution.

Figure 1: First Difference of (100 X Log of) Industrial Indices for OECD Countries

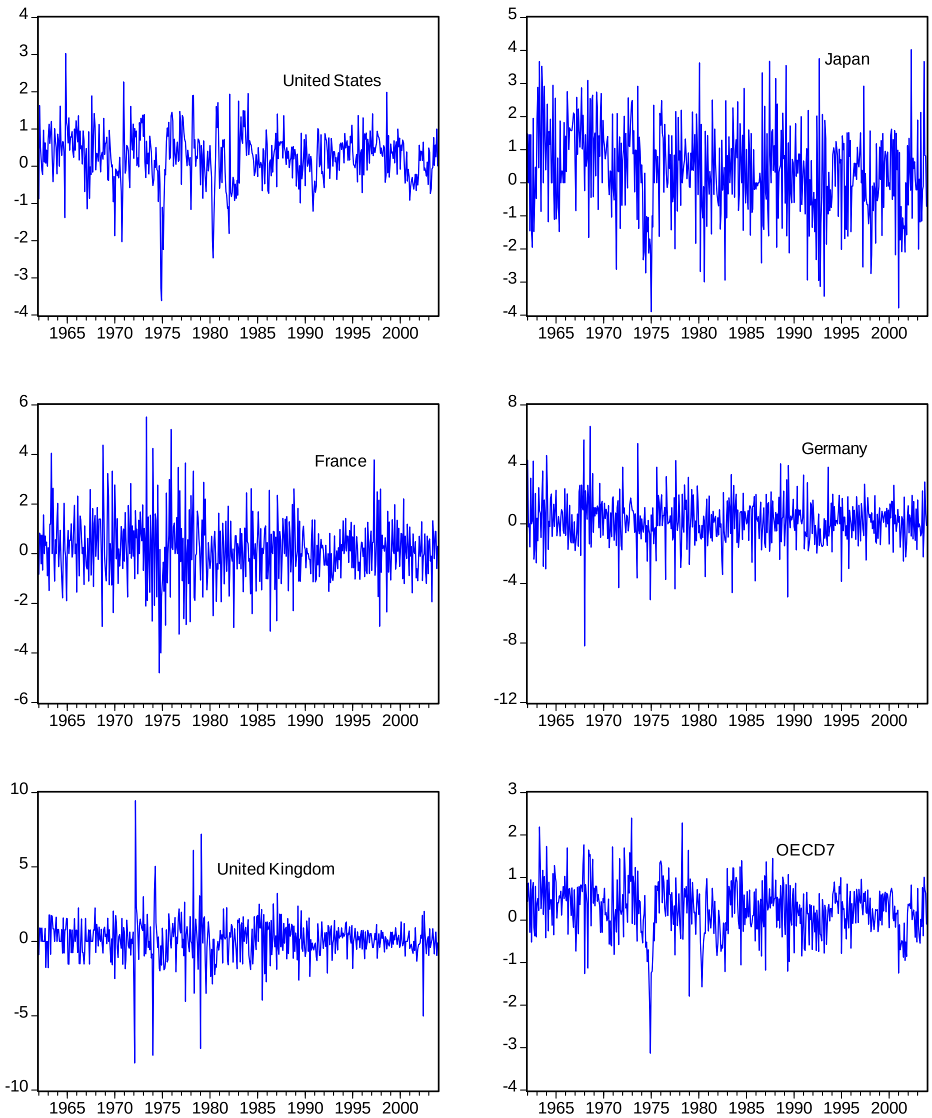

Figure 2: Estimated Intercept and First Lag Coefficients for Linear and Random Walk Autoregressive Model

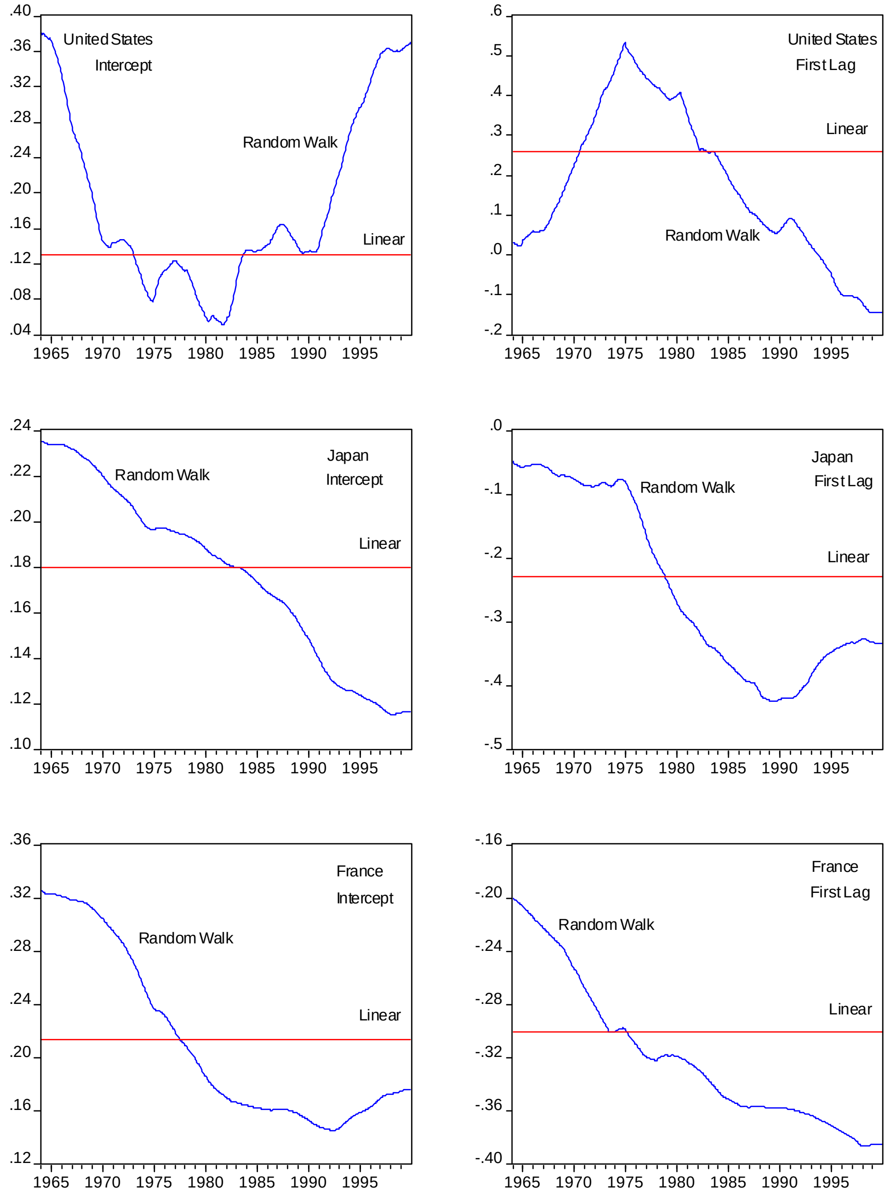

Figure 3: Transition Functions for LSTAR models of the First Differences of (100 X Log of) Industrial Production Indices for OECD Countries

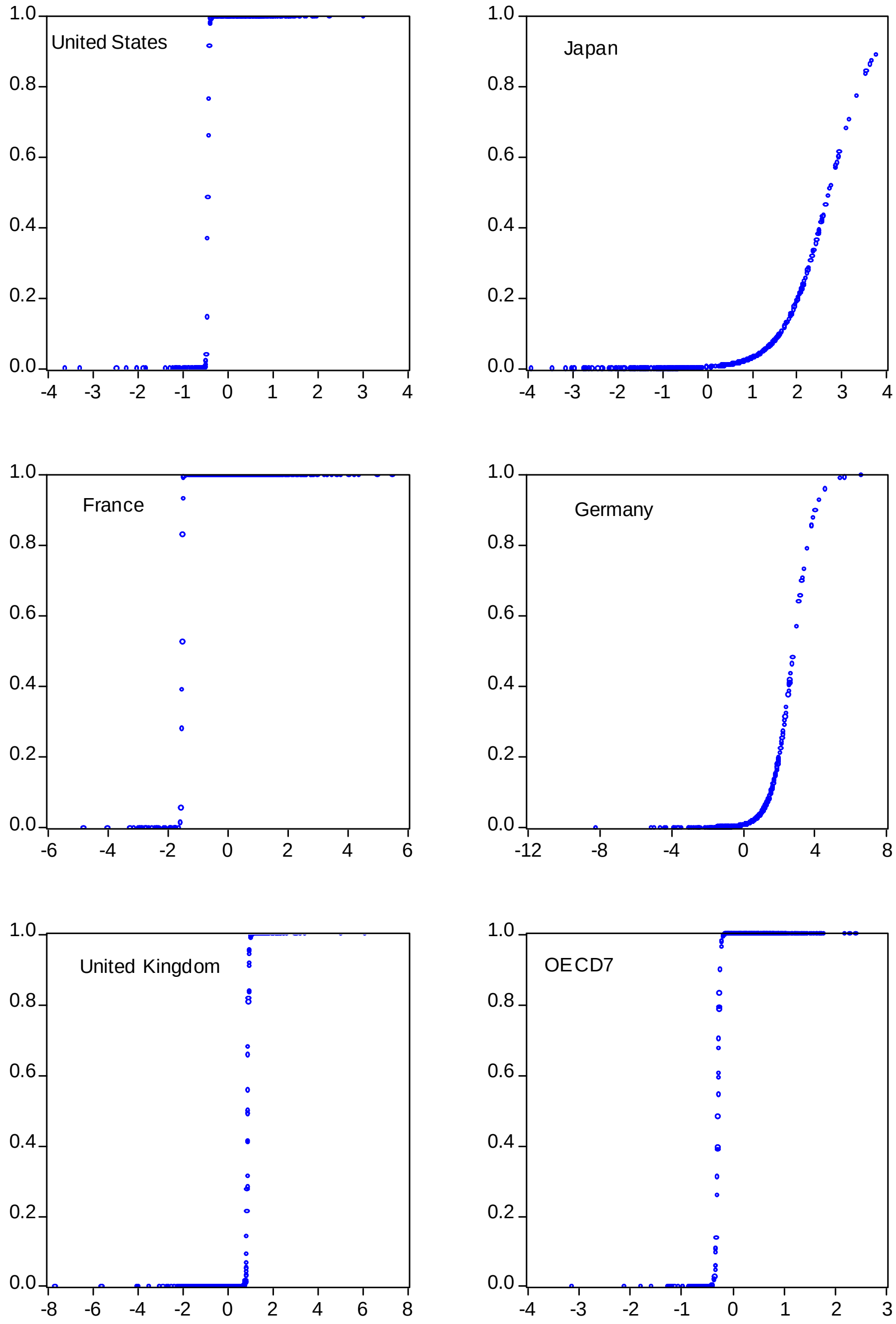

Figure 4: Transition Functions for TV-STAR Models of the First Diffeences of (100 X Log of) Industrial Production Indices for OECD Countries

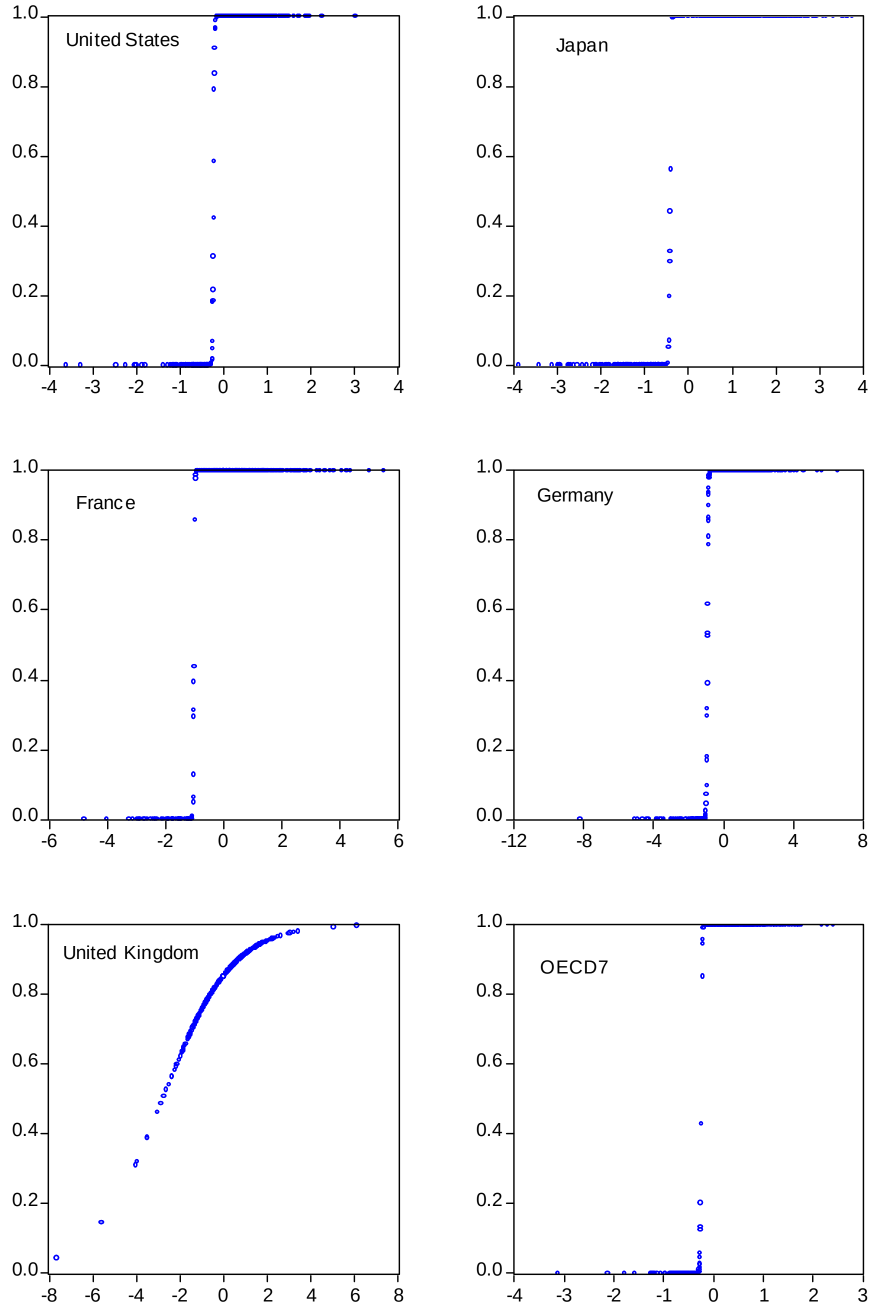

Figure 5: Transition Functions in Time for TV-STAR Models of the First Differences of (100 X Log of) Industrial Production Indices for OECD Countries

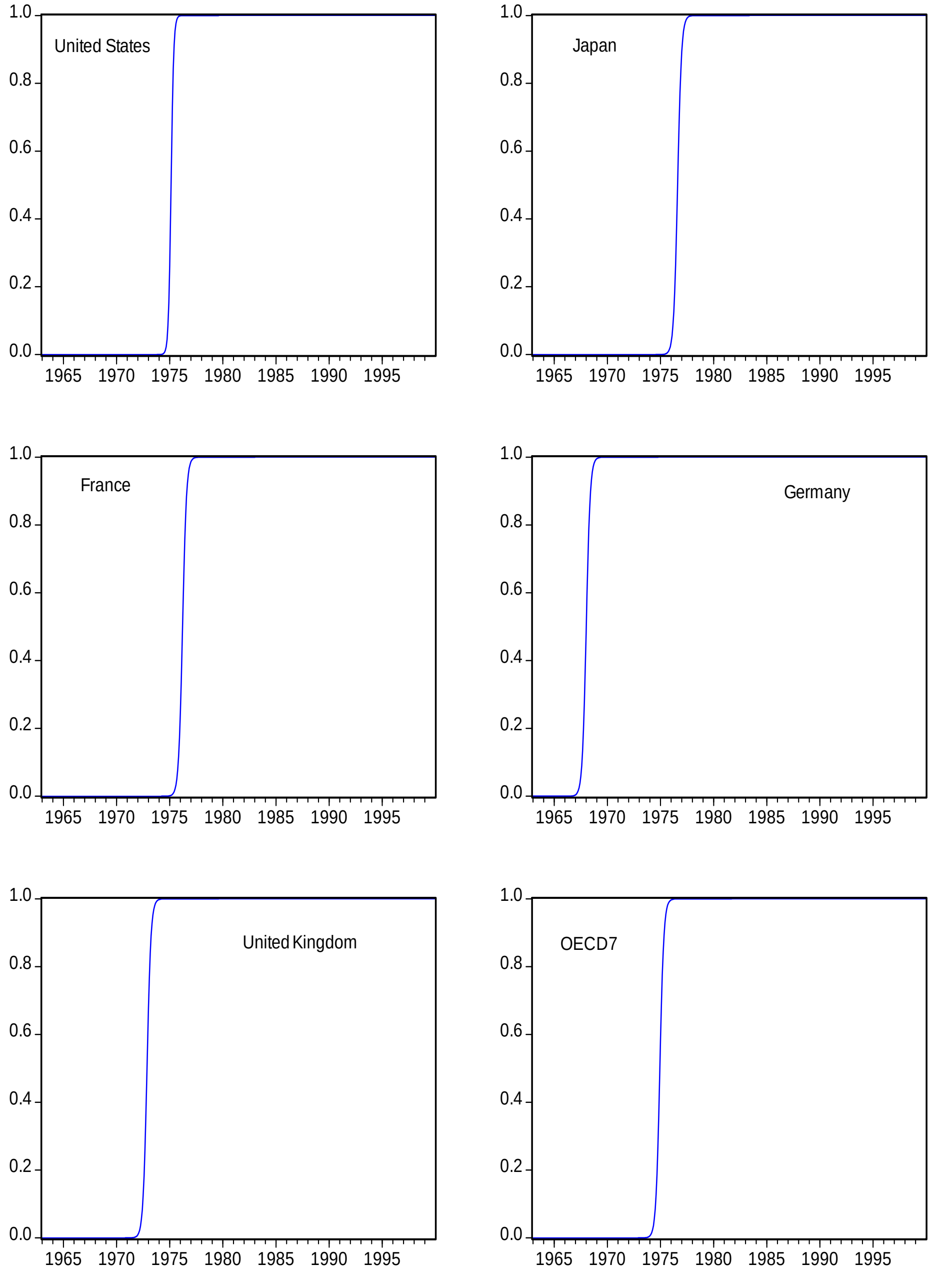

Figure 6: Transition Functions for the RW-STAR Models of the First Differences of (100X Log of) Industrial Production Indices for OECD Countries

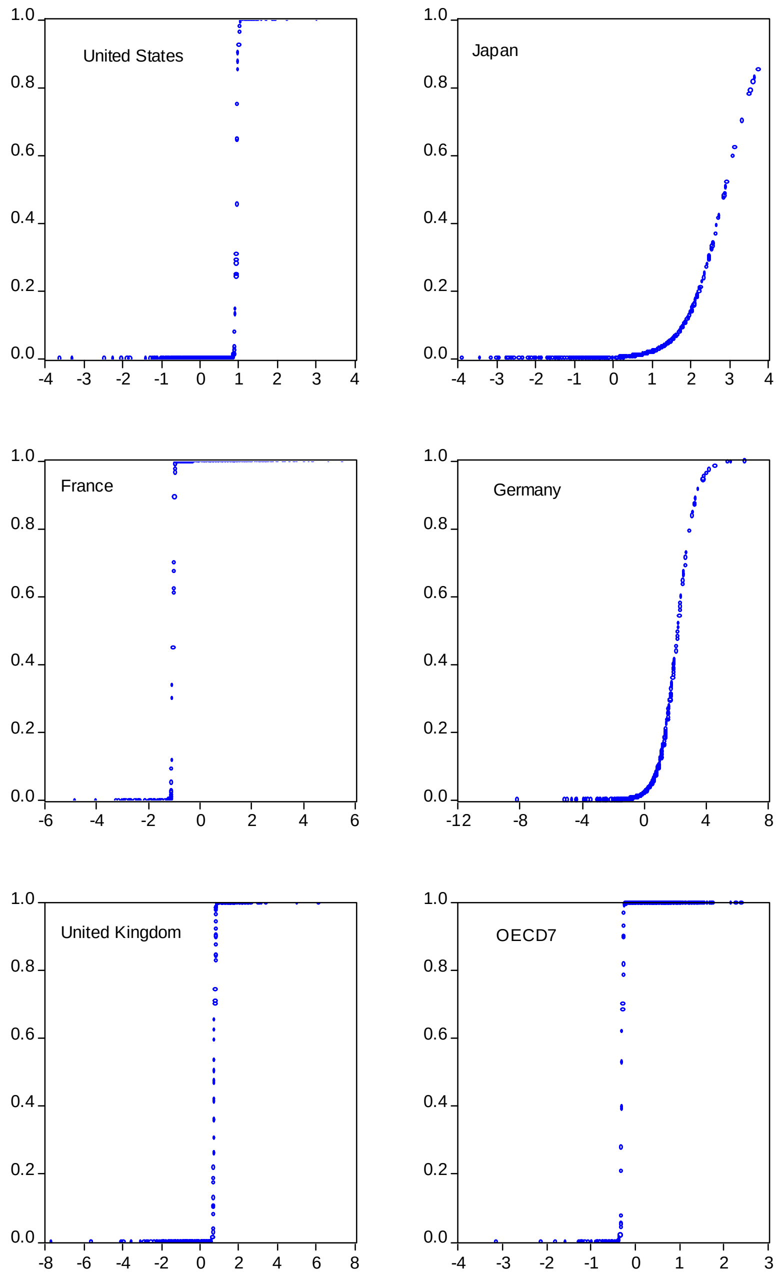

Figure 7: Selected Random Walk Coefficients from the RW-STAR Models

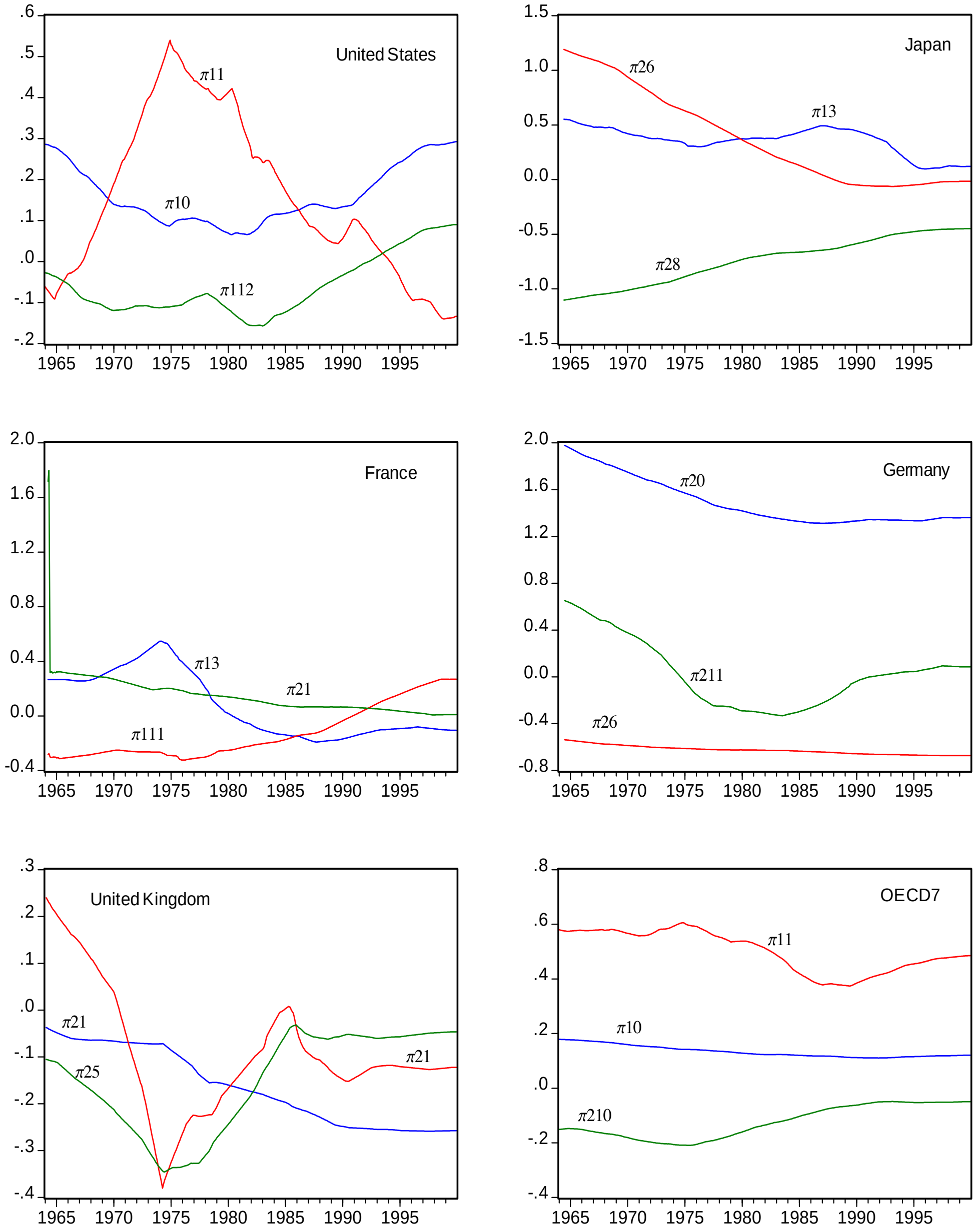

Figure 8a: Empirical vs Theoretical Densities of Transformed Forecasts for the United States (with 95% confidence bands)

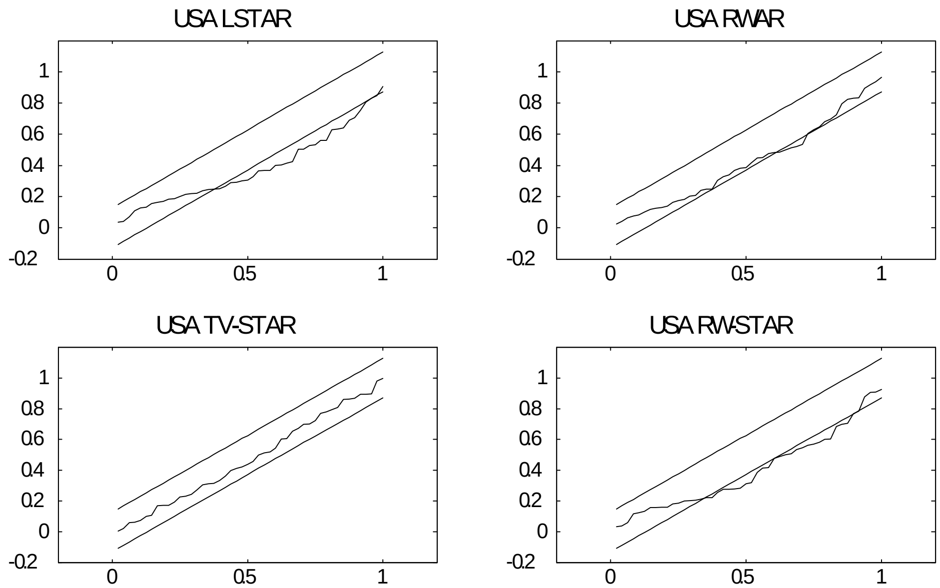

Figure 8b: Empirical vs Theoretical Densities of Transformed Forecasts for Japan (with 95% confidence bands)

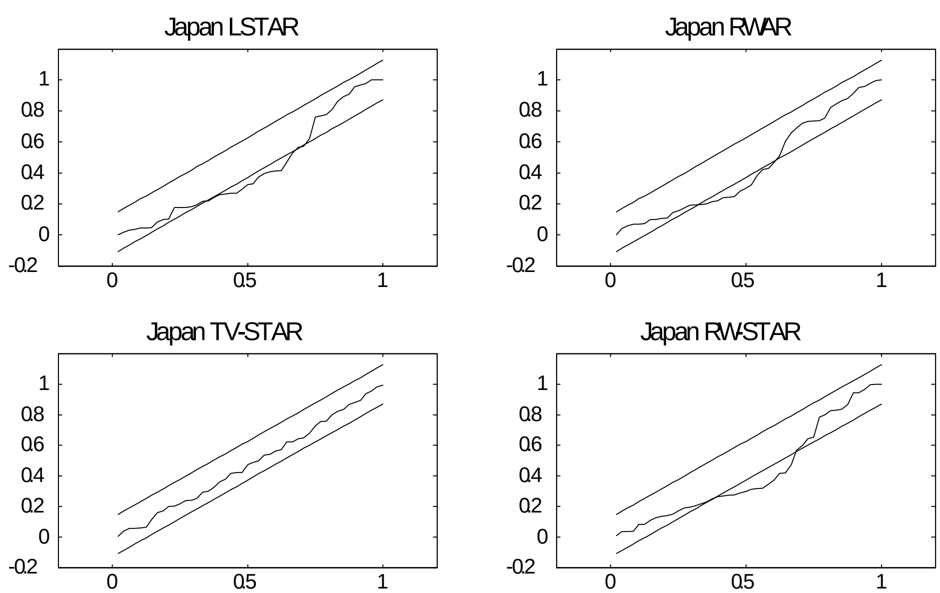

Figure 8c: Empirical vs Theoretical Densities of Transformed Forecasts for France (with 95% confidence bands)

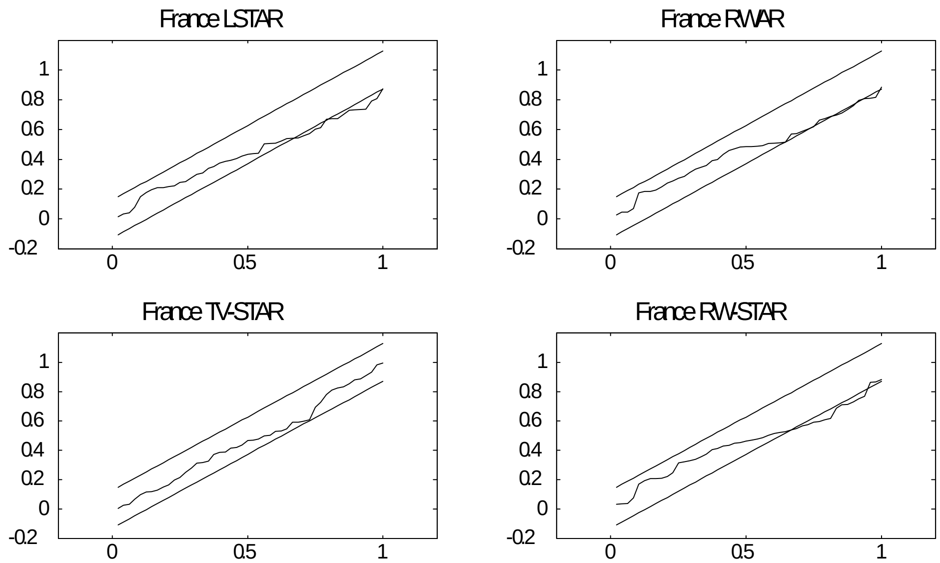

Figure 8d: Empirical vs Theoretical Densities of Transformed Forecasts for Germany (with 95% confidence bands)

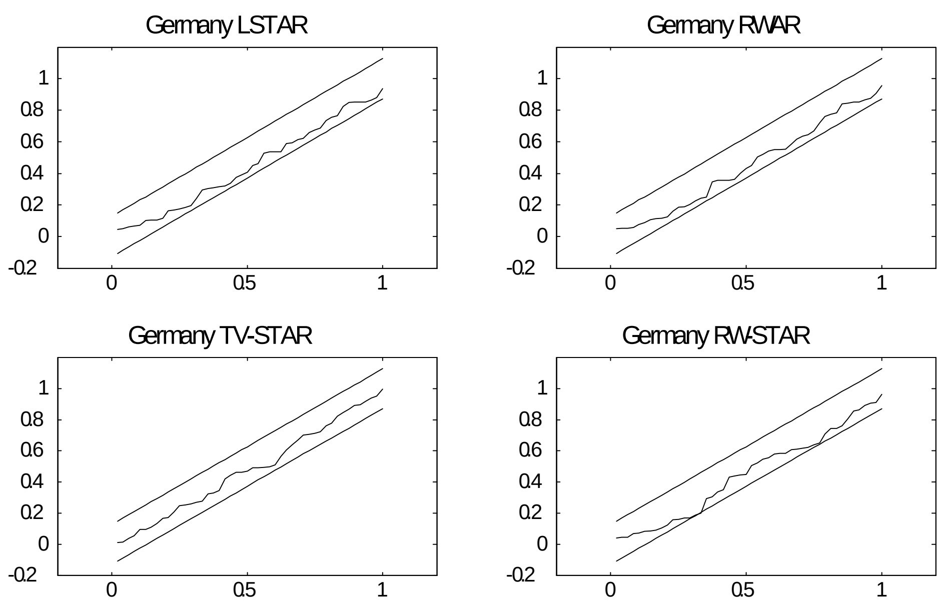

Figure 8e: Empirical vs Theoretical Densities of Transformed Forecasts for the United Kingdom (with 95% confidence bands)

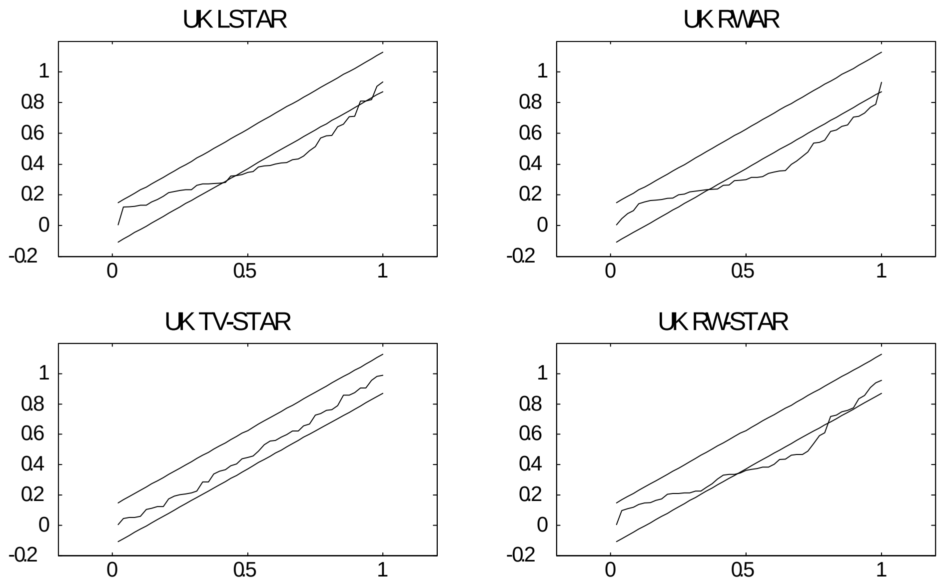

Figure 8f: Empirical vs Theoretical Densities of Transformed Forecasts for the OECD (with 95% confidence bands)