Understanding the behavior of the real exchange rate: Bayesian estimation of nonlinear models

THESIS

presented to the Faculty of Economics and Social Sciences
at the University of Fribourg (Switzerland),
in fulfillment of the requirements for the degree of
Doctor of Economics and Social Sciences

by

Gilles KALTENRIEDER
from Kerzers FR

Accepted by the Faculty of Economics and Social Sciences
on 29 September 2010 at the proposal of
Prof. Dr. Philippe Deschamps (First Advisor) and
Prof. Dr. Martin Wallmeier (Second Advisor)

Neuchâtel, Switzerland 2010

To Mateo and the next ones

Acknowledgements

Writing a Ph.D. thesis is a long way. Despite a lot of moments of solitude when the Markov chains did not converge, I never walked alone on this way. Therefore, it was possible.

First, I am indebted to my supervisor, Prof. Dr. Philippe J. Deschamps, who guided me, encouraged me and simply allowed me to write this thesis. I would also like to thank Prof. Dr. Martin Wallmeier who kindly agreed to read my thesis and who did the job very carefully.

A big thanks to all my friends, but particularly to: David Ardia, thank you for having guided me through the intricacies of Bayesian econometrics and for your jokes at the office; Carlos Ordás Criado, thank you for your help, you have significantly improved the quality of my thesis; Patrick Durrenberger, thank you for the design of the cover, you are a great designer and a true friend; Philippe Jolliet, thank you for your helpful comments and for our "dürüm-times". I am also grateful to Thom and his band.

I am indebted to my family. Thanks to my brother Yves, even though you fell asleep when we were talking about my thesis. I would also like to thank "Monsieur" who still thinks that I have worked on interest rates.

The best for the end. Thanks to you, Johanne, I owe you everything and you give me love everyday. I will always say: you are *all I need*. Thank you for the little Mateo, our sunshine; both of you gave me enough energy to finish this thesis.

Corcelles, Septembre 2010

Contents

List of Figures

List of Tables

Chapter 1

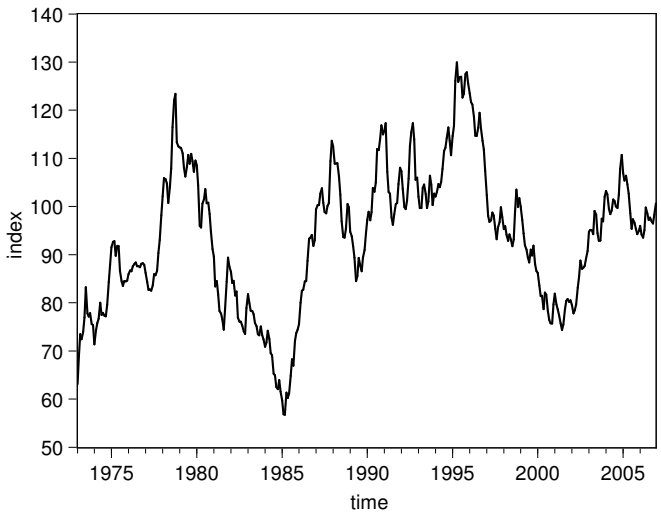
Introduction

1.1 Motivation

Purchasing power parity (PPP) is a theory proposed in the 16th century, which states that once converted to a common currency, national price levels should be equal. The PPP issue is still relevant nowadays, for at least two reasons. Firstly, as Sarno and Taylor [2003] point out, *a discussion of the real exchange rate is tantamount to a discussion of PPP*. Indeed, if PPP is valid, movements in real exchange rate represent deviations from PPP. Therefore, this thesis focuses exclusively on the real exchange rate. Secondly, a strong consensus exists among economists on the fact that the real exchange rate should tend toward PPP in the very long run. The speed of convergence to PPP is extremely slow. Thus, two questions arise: is real exchange rate a multiple regime process? And does the PPP relationship vary over time?

The question as to whether a multiple regime process exists in the real exchange rate series, is linked to the largely debated question of the presence of nonlinearities in real exchange rate dynamics. During the 90s, numerous authors have developed theoretical models of nonlinear real exchange rates. A plausible source of nonlinearities is transaction costs such as transportation costs or tariff and nontariff barriers (see for example Dumas [1992]). Intuitively, these costs create a band for the real exchange rates within which the marginal cost of arbitrage exceeds its marginal benefit. An alternative explanation for nonlinearities in real exchange rates is given by Kilian and Taylor [2003]: there are heterogeneous agents influencing the foreign exchange market, namely economic fundamentalists, chartists and noise traders. Disagreements among these heterogeneous agents endogenously generate threshold behaviors in spot exchange rates and, by extension, in real exchange rates. Sarantis [1999] investigates the existence of these nonlinearities with a smooth transition autoregressive (STAR) model.

Figure 1.1: USD/CHF real exchange rate index



Introduced for the first time by Teräsvirta and Anderson [1992] and Teräsvirta [1994], the STAR model is the starting point for this thesis. The typical STAR model for a time series y_t has the following form:

$$y_t = \phi' x_t + \theta' x_t F(\gamma, c; s_t) + u_t \,, \tag{1.1}$$

where $x_t = (1 \quad y_{t-1} \quad ... \quad y_{t-p})'$ is a vector including lags of the dependent variable and the constant term; $t = 1, ..., T$ and T is the sample size. The terms ϕ and θ are vectors of autoregressive parameters of length $p+1$. The error term u_t is usually assumed to be iid $\mathcal{N}(0, \sigma^2)$. The transition function $F(\gamma, c; s_t)$ is assumed continuous in s_t and bounded between zero and one.

Most articles on STAR models employ a frequentist approach for estimating these models. The most popular frequentist techniques include estimators such as the least squares estimator, the maximum likelihood estimator and the method of moments. The main weakness of the latter approaches is the risk of being trapped into a local extremum of the criterion. Techniques for circumventing this problem exist (see for example Maringer and Meyer [2008]), but recently some authors have successfully chosen to use the Bayesian approach for estimating STAR models (see for example Deschamps [2008] and Lopes and Salazar [2006]). The Bayesian paradigm considers the true parameters of a model as random variables, thus computational methods based on Markov chain Monte Carlo (MCMC) try to simulate their distributions. The frequentist approach assumes instead that the true parameters of a model are fixed but unknown, and that the estimators of these parameters are random variables. Chapter 2 and Chapter 3 compare the advantages of the two approaches.

The second issue dealt with in this thesis is the stability of the PPP relationship over time. According to Engel [2000], instead of assuming temporary deviations from a fixed target level, it is more plausible to consider that real exchange rates return to a target level that changes over time. Hence, Engel and Kim [1999] suggest that real exchange rate movements may be decomposed into both a permanent and a transitory component, where the permanent component represents PPP.

The data used throughout this thesis are the monthly real exchange rate indexes between the United States dollar (USD) and the Swiss franc (CHF). The data cover the post-Bretton Woods period from 1973:01 to 2006:12 and are taken from the Swiss National Bank's *Monthly Statistical Bulletin* database. The monthly real exchange rate index is the average of daily nominal exchange rates deflated by the consumer price index in both countries. Figure 1.1 shows the USD/CHF real exchange rate index. Note that, as we are interested in the behavior of the real exchange rate, our analysis focuses on non-transformed data. This contrasts with other practices for which log-transformed data or other operations involving rescaling are employed.

1.2 Overview

In Chapter 2, we describe the estimation and evaluation of the STAR model with the frequentist approach. After a preliminary analysis of the data - the USD/CHF real exchange rate index - we discuss the specification, estimation and evaluation of the STAR model. In particular, we investigate the performances of information criteria for discriminating between several STAR models. Moreover, a new algorithm is developed to ensure convergence of the maximization of the likelihood function. Finally, misspecification tests such as those developed by Eitrheim and Teräsvirta [1996] as well as a forecast evaluation are applied.

The estimation and evaluation of the STAR model in a Bayesian context are presented in Chapter 3. In particular, we develop the posterior simulator of two approaches proposed respectively by Deschamps [2008] and Lopes and Salazar [2006]. This allows us to highlight the advantages and disadvantages of both approaches. A model comparison and an analysis of dynamic behavior are performed for the MCMC estimates of several STAR models. Model adequacy is evaluated using the specific misspecification tests developed by Eitrheim and Teräsvirta [1996]. Finally, we consider the models' predictive performances.

In Chapter 4, we extend the STAR model in order to allow a transition between two regimes by introducing a third one: the multiple regime smooth transition autoregressive (MRSTAR) model is described and evaluated in a Bayesian context. We illustrate this model with an application to the USD/CHF real exchange rate index. A model comparison and an analysis of dynamic behavior are performed for the MCMC estimates of several MRSTAR models. The predictive performances of the most appropriate MRSTAR model are tested.

Chapter 5 introduces the use of the state space model in the real exchange rate context. We describe the posterior simulator of a general state space model. The Bayesian estimation and evaluation of several unobserved components models are presented. In particular, the decomposition of the USD/CHF real exchange rate index into both a permanent and a transitory component, as in Engel and Kim [1999], is analyzed. Moreover, a comparison of in-sample performances between the models is performed and predictive performances are tested.

We describe the unobserved random walk and smooth transition autoregressive components (RW-STAR) model in Chapter 6 and propose an application to the USD/CHF real exchange rate index. This RW-STAR model is an extension of the unobserved components model analyzed in Chapter 5: the AR process of the transitory component is replaced by a STAR process, which provides an interesting interpretation of the PPP theory: it assumes that the real exchange rate tends to converge towards an intertemporal equilibrium level but that this target may vary over time. Then, a model comparison, an analysis of dynamic behavior and an evaluation of predictive performances are performed for the MCMC estimates of several RW-STAR models.

Finally, Chapter 7 summarizes the main results of this thesis and discusses possible future research.

Chapter 2

Frequentist estimation of the STAR model

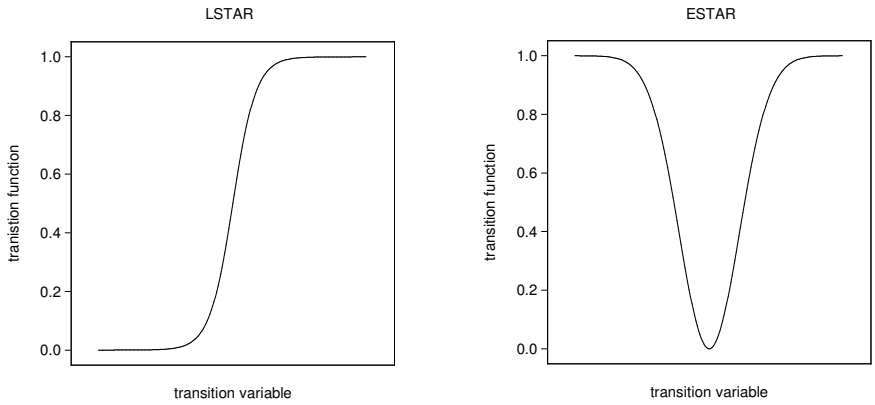
The first question asked in this thesis is whether the real exchange rate is a multiple regime process. Hence we have decided to estimate the smooth transition autoregressive (STAR) model as a starting point. Moreover, we have chosen to employ a frequentist approach to estimate the STAR model. In what follows, we discuss the specification, estimation and evaluation of the STAR model. In particular, we investigate the performances of information criteria for discriminating between several STAR models, misspecification tests such as those developed by Eitrheim and Teräsvirta [1996] are applied and a forecast evaluation is performed.

The outline of this chapter is as follows. We set up the general STAR model in Section 2.1. We perform a preliminary analysis of the data in Section 2.2. In Section 2.3, the three steps of model specification are developed and applied to the USD/CHF real exchange rate index: choice of autoregression length, choice of delay parameter and discrimination between LSTAR and ESTAR models. We treat the specification with information criteria in Section 2.4. The frequentist estimation of ESTAR models with a new algorithm is given in Section 2.5. In Section 2.6, we test the ESTAR models for misspecification. An evaluation of the predictive performances of the ESTAR models is given in Section 2.7 and the main results are summarized in Section 2.8.

2.1 The model

The STAR models were introduced by Teräsvirta and Anderson [1992] and Teräsvirta [1994]. This framework has been used to model various types of economic time series such as industrial production (Teräsvirta and

Figure 2.1: Representation of the transition function of an LSTAR and an ESTAR model versus the value of the transition variable



Anderson [1992]), interest rates (Kapetanios et al. [2003]), unemployment (Skalin and Teräsvirta [2002], van Dijk et al. [2002] and Deschamps [2008]) and exchange rates (Sarantis [1999] and Taylor et al. [2001]).

The typical STAR model has the following form:

$$y_t = \phi' x_t + \theta' x_t F(\gamma, c; s_t) + u_t \tag{2.1}$$

where $x_t = (1 \quad y_{t-1} \quad ... \quad y_{t-p})'$ is a vector including lags of the dependent variable and the constant term; $t = 1, ..., T$ and T is the sample size. The terms ϕ and θ are vectors of autoregressive parameters; their length is $p + 1$. The error term u_t is usually assumed to be iid $\mathcal{N}(0, \sigma^2)$. The transition function $F(\gamma, c; s_t)$ is assumed continuous in s_t and bounded between zero and one. It governs the transition between the two regimes and it depends upon the values of the transition variable s_t. Any variable can be used as a transition variable; nevertheless we use the d^{th} lag of the dependent variable, y_{t-d} here. Teräsvirta [1994] suggests *for notational simplicity* to pick d, the delay parameter, in the range of 1 $\leq d \leq p$, but we do not impose this restriction here. The slope parameter γ measures the smoothness of transitions between regimes while the threshold (or location) parameter c determines the location of the transition.

Teräsvirta and Anderson [1992] introduce two types of STAR models: the logistic (LSTAR) and the exponential (ESTAR) smooth transition autoregressive models. The logistic model is based on the first-order logistic transition function:

$$F(\gamma, c; s_t) = \frac{1}{1 + \exp[-\gamma(s_t - c)]}, \gamma > 0. \tag{2.2}$$

A large or a small value of the transition variable s_t relative to the threshold parameter c defines two different regimes (see the left-hand side of Figure 2.1). Function (2.2) is monotonic increasing in s_t and yields asymmetric adjustment towards equilibrium. The lower the slope parameter γ, the smoother the switch between the two regimes. Note that the LSTAR model becomes linear when the slope parameter $\gamma \to 0$. The LSTAR model can be used when time series are characterized by two unequal dynamic behaviors, e.g. business cycle expansions and contractions.

The second model employs the exponential (transition) function:

$$F(\gamma, c; s_t) = 1 - \exp\Big[-\gamma(s_t - c)^2\Big], \gamma > 0. \tag{2.3}$$

The ESTAR model defines two different regimes in terms of small and large absolute deviations of the values for transition variable s_t from the threshold parameter c. This transition function is symmetrically inverse-bell-shaped around the threshold parameter c, and possesses an "inner" and an "outer" regime (see the right-hand side of Figure 2.1). When the slope parameter of the exponential function (2.3) is such that $\gamma \to \infty$ or $\gamma \to 0$, the nonlinear ESTAR model collapses into a linear model. These properties are attractive for modeling exchange rates. An ESTAR specification of real exchange rates is especially useful when a model includes transaction costs such as transportation costs, tariff and nontariff barriers, as well as any other costs (see Dumas [1992] and Obstfeld and Rogoff [2000]). Intuitively, transaction costs create a band within which no arbitrage profits exist. Therefore, nominal exchange rate deviations from purchasing power parity are not corrected within the band. However, if the real exchange rate moves outside the band, arbitrage brings the real exchange rate back to a value within the band. Moreover, Taylor et al. [2001] and Rapach and Wohar [2006] suggest that time aggregation and non-synchronous adjustment by heterogeneous agents are likely to lead to smooth regime switching, rather than discrete switching. This is especially true for real exchange rates.

From now on, we adopt the following notation for STAR models: let the transition variable be y_{t-d} and p be the number of lags in the vector x_t, the logistic and exponential models will be noted LSTAR(p, d) and ESTAR(p, d) respectively.

2.2 Data and preliminary analysis

The data used throughout this thesis are the monthly real exchange rate index between the United States dollar (USD) and the Swiss franc (CHF). The data cover the post-Bretton Woods period from 1973:01 to 2006:12.

We begin our empirical analysis by verifying the stationarity assumption. In our context, this issue is particularly important, both from a theoretical and an econometric point of view. First, stationarity over time is a necessary condition in order to validate the purchasing power parity (PPP) theory in the long run (see Taylor et al. [2001]). The PPP theory states that national price levels should be equal when expressed in a common currency. When PPP is valid, the real exchange rate should be a constant. Hence, the movements of the real exchange rate represent deviations from PPP. Therefore, as Sarno and Taylor [2003] point out, a discussion of the real exchange rate is tantamount to a discussion of PPP. Note that the PPP theory has been widely debated (see among others Rogoff [1996]). Second, the series has to be stationary and ergodic to ensure the consistency of the nonlinear least square (NLS) estimation below. The necessary and sufficient conditions are stated in Klimko and Nelson [1978].

We test for unit root behavior of the USD/CHF real exchange rate index by computing an augmented Dickey-Fuller (ADF) and a Phillips-Perron (PP) test statistics. The critical values depend on the true underlying dynamic of the series. We can think that the true underlying process contains a drift but no trend, which corresponds to Case 2 in Hamilton [1994]. The Case 2 of the ADF test proceeds from the maximum likelihood estimation of the equation:

$$y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \ldots + \zeta_{p-1} \Delta y_{t-p+1} + \alpha + \rho y_{t-1} + \epsilon_t \qquad (2.4)$$

where ϵ_t is iid $\mathcal{N}(0, \sigma^2)$ and p can be determined by the sequential procedure suggested by Hamilton [1994] p. 530: (2.4) is estimated with p chosen as an upper bound $\bar{p}$. Then the parameter $\zeta_{\bar{p}-1}$ is tested using a t statistic. If the null is not rejected, the joint null hypothesis that both $\zeta_{\bar{p}-1} = 0$ and $\zeta_{\bar{p}-2} = 0$ is tested using an $\mathcal{F}$ statistic. The procedure continues sequentially until a joint null hypothesis that $\zeta_{\bar{p}-1} = 0$, $\zeta_{\bar{p}-2} = 0$, $\ldots$, $\zeta_{\bar{p}-l} = 0$ is rejected for some l. In our case, this procedure indicates $p = 2$, so we use this value for all tests in this section. Returning to the ADF test, the statistic t for the null hypothesis $\rho = 1$ is equal to $(\hat{\rho} - 1)/\hat{\sigma}_{\hat{\rho}}$ and compared to the critical value given in Hamilton [1994]. The same null hypothesis is tested with the Phillips-Perron Z_t statistic (Case 2) but the parameter estimation is based on the equation:

Table 2.1: Unit root tests for the USD/CHF real exchange rate index

Test	Source	Statistic	5% critical value
ADF (t)	Hamilton [1994]	-2.885	-2.87
PP (Z_t)	Hamilton [1994]	-2.705	-2.87
t_{NL}	Kapetanios et al. [2003]	-2.991	-2.93
JLR	Taylor and Sarno [1998]	8.242	3.84

All critical values are taken from Hamilton [1994] for 500 observations, except the JLR which asymptotically follows an $\chi^2(1)$ distribution.

$$y_t = \alpha + \rho y_{t-1} + u_t \tag{2.5}$$

where u_t is serially correlated and possibly heteroskedastic. The t-statistic Z_t described in Hamilton [1994] takes serial correlation into account.

Frömmel [2007] argues that *unit root test for exchange rate should be interpreted cautiously.* Indeed, standard unit root tests such as ADF and PP have low power when non-stationarity is more complex than a standard unit root (see Caporale et al. [2003] for further details). Kapetanios et al. [2003] address this criticism by developing a specific unit root test for nonlinear STAR models. They show that we can check the presence of a unit root by testing the null that $\rho = 0$ against $\rho < 0$ with the t-statistic from the equation:

$$\Delta y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \ldots + \zeta_{p-1} \Delta y_{t-p+1} + \rho y_{t-1}^3 + \epsilon_t \tag{2.6}$$

where ϵ_t is iid $\mathcal{N}(0, \sigma^2)$.

Taylor and Sarno [1998] propose another approach. They show that the power of univariate unit root tests increases strongly when multivariate unit root tests like the Johansen likelihood ratio (JLR) are used. Therefore, we add the real exchange rate between the Great Britain pound (GBP) and the Swiss franc (CHF) for the same period to our data set and compute the JLR statistic[1]. This test is a special case of the likelihood ratio test

[1] Taylor et al. [2001] show that the multivariate augmented Dickey-Fuller (MADF) test has high power in the exchange rate framework as well. We nevertheless don't implement this test here because its power is high even when only one of the series is stationary. They call it the *pitfall* of multivariate unit root tests.

for cointegration of Johansen [1988, 1991]. Consider a vector autoregressive (VAR) model of N series reparametrized into an error correction form:

$$\Delta \mathbf{Q}_t = \mathbf{\Gamma}_1 \Delta \mathbf{Q}_{t-1} + \ldots + \mathbf{\Gamma}_{p-1} \Delta \mathbf{Q}_{t-p+1} + \mathbf{\Gamma}_p \mathbf{Q}_{t-p} + \mu + \omega_t \tag{2.7}$$

where $\Delta \mathbf{Q}_t$ is an (Nx1) vector in first difference, $\mathbf{\Gamma}_i$ are (NxN) matrices of parameters, μ is an (Nx1) vector of constants and ω_t is an (Nx1) vector of white noise errors. The maximum lag p is chosen as in our ADF procedure. Engle and Granger [1987] demonstrate that there can be at most N-1 cointegrating vectors among a system of N I(1) series. If we reject the hypothesis that there are less than N cointegrating vectors, the series has to be stationary. The rank of $\mathbf{\Gamma}_p$ in (2.7) defines the number of distinct cointegrating vectors. Therefore, Taylor and Sarno [1998] express the null hypothesis that at least one of the series is generated by an I(1) process as $H_0 : rank(\mathbf{\Gamma}_p) < N$. The alternative hypothesis that each of the series is stationary is expressed as $H_1 : rank(\mathbf{\Gamma}_p) = N$. The JLR statistic test consists in a likelihood ratio:

$$JLR = -T \ln(1 - \lambda_N) \tag{2.8}$$

where λ_N is the smallest root of the characteristic equation:

$$|\lambda \mathbf{S}_{kk} - \mathbf{S}_{k0} \mathbf{S}_{00}^{-1} \mathbf{S}_{0k}| = 0. \tag{2.9}$$

$\mathbf{S}_{ij}$ is the sum of squared residuals $T^{-1} \sum_{t=1}^{T} \mathbf{R}_{it} \mathbf{R}'_{jt}$ $(i, j = 0, p)$, $\mathbf{R}_{0t}$ and $\mathbf{R}_{pt}$ are the residuals of the regression of $\Delta \mathbf{Q}_t$ and $\mathbf{Q}_{t-p}$ respectively on $\{\imath, \Delta \mathbf{Q}_{t-1}, \Delta \mathbf{Q}_{t-2}, \ldots, \Delta \mathbf{Q}_{t-p+1}\}$, and $\imath$ is the unit vector. The JLR statistic follows an $\chi^2(1)$ asymptotic distribution (see Taylor and Sarno [1998]).

Table 2.1 shows four unit root tests for the USD/CHF real exchange rate index series and their corresponding critical value. The ADF, PP and t_{NL} statistics are really close to the 5% critical value. Therefore, we can barely reject the unit root hypothesis. This indicates the presence of a unit root within the series. But considering that these univariate tests are not powerful, it is preferable to rely on the JLR test. The JLR empirical statistic based on the series of USD/CHF and GBP/CHF real exchange rate indexes is shown in Table 2.1 as well. Its value of 8.242 is much larger than the asymptotic critical value at 5%. Then, we clearly reject the null hypothesis that at least one of the series is I(1). We conclude that the USD/CHF real exchange rate index is stationary with the most powerful test available.

2.3 Specification

A common way of specifying a smooth transition autoregressive model is described in the original article by Teräsvirta [1994]. This author divides the specification process into a sequence of three steps. The first one consists in specifying a linear autoregressive model and choosing the adequate lag length with a standard order selection criterion. Teräsvirta [1994] indicates that any model selection procedure should account for the residual autocorrelation with the portmanteau test of Ljung and Box [1978]. The second step consists in determining whether or not a linear model adequately represents the data generating process. The author applies the linearity test that will be described in Section 2.3.2 for different values of the delay parameter d, and fixes its value by minimizing the p-value of the linearity test. Finally, the choice between the LSTAR and the ESTAR model is made in the last step.

This section applies the Teräsvirta methodology for the specification of models, but, according to the author of this thesis, the results are unreliable and inconsistent. Indeed, the Monte Carlo study in Appendix A shows that the first step of the Teräsvirta procedure can lead to wrong choices. As the following steps of the procedure are governed by the first one, all steps are thus unreliable. Moreover, the Teräsvirta procedure applied below to the USD/CHF real exchange rate shows that the results are inconsistent.

2.3.1 Choice of autoregression length

We start with an iterative procedure which consists in estimating the linear autoregressive models and in determining the lag length by minimizing an information criterion. This lag length is then applied to the nonlinear STAR model. This approach is questionable because if the true real model is nonlinear, then this method may be inadequate. This point of view is shared by Teräsvirta and Anderson [1992]: *if the true model is nonlinear it is possible that a maximum lag greater than the maximum lag in the nonlinear model will be selected.* Note that this consideration has no theoretical foundation. A Monte Carlo study is therefore useful to check to what extent the linear autoregressive models can be used to discriminate between different lags in the nonlinear STAR model. This is the purpose of Appendix A which illustrates with several examples that the first step of the procedure proposed by Teräsvirta [1994] is not reliable when using the AIC (Akaike [1974]) or the BIC (Schwarz [1978]) information criteria. Despite this drawback, we apply these criteria in our case, i.e. the USD/CHF real exchange rate index, but we remain cautious while interpreting the results. Another conclusion of Appendix A is that the BIC selects the right autoregressive parameter p more frequently than does the AIC.

Table 2.2: AIC and BIC of AR models for the USD/CHF real exchange rate index

Models	AIC	BIC
AR(1)	1998.755	2010.782
AR(2)	1951.335	1967.360
AR(3)	1945.734	1965.754
AR(4)	1942.434	1966.443
AR(5)	1939.550	1967.543
AR(6)	1937.261	1969.233
AR(7)	1931.296	1967.242
AR(8)	1920.688	1960.603
AR(9)	1916.229	1960.107
AR(10)	1913.933	1961.771
AR(11)	1907.755	**1959.546**
AR(12)	**1905.883**	1961.623

Table 2.2 shows the value of both the AIC and BIC criteria for the first twelve linear autoregressive models with the USD/CHF real exchange rate index. The minimum Akaike criterion is reached for 12 lags, whereas we reach 11 lags for the BIC. The latter figure is retained for parameter p in the model (2.1).

The lag length selection has to be accompanied by a test for residual autocorrelation. We use the test of Ljung and Box [1978] given by the following statistic:

$$\tilde{Q}(\tilde{r}) \;=\; T(T+2)\sum_{k=1}^{m}\frac{1}{T-k}\cdot\hat{r}_k^2 \tag{2.10}$$

where T is the sample size and $\hat{r}_k^2$ is the value of the autocorrelation function of lag k. Under the null hypothesis that there is no autocorrelation up to order m, $\tilde{Q}(\tilde{r})$ is asymptotically distributed as an χ^2 with m degrees of freedom.

The only linear model whose residuals are autocorrelated is the AR(1) model (the p-value of the test above is under 5% when $m = 1$). In that case, we reject the null hypothesis, whereas we cannot reject this hypothesis for the other models.

2.3.2 Choice of delay parameter

At this step of the specification, Teräsvirta [1994] proposes to choose the value of the delay parameter in using the linearity test described below. Performing this test at a fixed lag but at different delays allows to choose the transition variable y_{t-d} minimizing the p-value. The author justifies this method theoretically as well as numerically, by using Monte Carlo simulations.

It is important to test for linearity before attempting to estimate a nonlinear model. In order to derive a linearity test for the LSTAR model, Teräsvirta [1994] suggests using the artificial regression based on the third-order Taylor approximation of the LSTAR model:

$$y_t = \beta_0 + \beta_1' x_t^* + \beta_2'(x_t^* y_{t-d}) + \beta_3'(x_t^* y_{t-d}^2) + \beta_4'(x_t^* y_{t-d}^3) + v_t \qquad (2.11)$$

where $x_t^* = [y_{t-1}, \ldots, y_{t-p}]'$ and v_t is iid $\mathcal{N}(0, \sigma_v^2)$. The null hypothesis of linearity is $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$.

The test statistic $LM = T(SSR_0 - SSR)/SSR_0$ follows an asymptotic $\chi^2(3p)$ distribution (see Teräsvirta [1994]), where SSR is the sum of squared residuals from the full artificial regression and SSR_0 is the corresponding sum under the null. This linearity test can also be carried out as an $\mathcal{F}$ test. A new test statistic $LM' = \frac{(SSR_0 - SSR)/3p}{SSR/(T-(4p+1))}$ replaces the LM one and it follows a $\mathcal{F}(3p, T-(4p+1))$ distribution under H_0. The $\mathcal{F}$-version of the linearity test has the advantage of having better size properties in small samples than the χ^2-version.

The linearity test for the ESTAR model is based on a first-order Taylor approximation. Therefore, van Dijk et al. [2002] say that the linearity test above has power against ESTAR alternative since all auxiliary regressors in the first-order approximation are nested in (2.11).

Table 2.3 shows the p-values of the linearity $\mathcal{F}$ test for the first twelve lags and delays. We choose to display the p-values of the linearity $\mathcal{F}$ test for all lags and delays in order to show the inconsistency of the procedure. We clearly reject linearity only for the first lag. For all other lags, we cannot reject the linearity hypothesis at a 5% level. These results imply that a STAR is plausible only with a lag of 1. Therefore, it does not make sense to apply the Teräsvirta methodology to other lags. When the value of the lag parameter is one, the smallest p-value is achieved at $d = 2$. Note that we can consider $d = 3$ too. The results are unambiguous, when we consider only the linearity $\mathcal{F}$ test: a STAR(1,2) or a STAR(1,3) model is preferred. Nevertheless, this choice is problematic as it is inconsistent with the first step. This shows that the methodology proposed by Teräsvirta [1994] is unreliable. Note that using the χ^2-version of the linearity test yields very similar results.

Table 2.3: P-values of the linearity test at different lags and delays for the USD/CHF real exchange rate index

	delay											
lag	1	2	3	4	5	6	7	8	9	10	11	12
1	0.548	0.000	0.000	0.005	0.069	0.225	0.222	0.225	0.114	0.039	0.013	0.013
2	0.513	0.130	0.200	0.706	0.946	0.948	0.954	0.973	0.961	0.789	0.879	0.837
3	0.533	0.156	0.184	0.686	0.959	0.974	0.800	0.832	0.675	0.400	0.450	0.483
4	0.703	0.281	0.253	0.518	0.896	0.958	0.825	0.834	0.774	0.601	0.679	0.634
5	0.497	0.355	0.315	0.570	0.828	0.929	0.738	0.702	0.625	0.356	0.474	0.433
6	0.535	0.456	0.460	0.679	0.903	0.910	0.865	0.820	0.758	0.484	0.595	0.475
7	0.627	0.547	0.623	0.823	0.965	0.965	0.864	0.859	0.845	0.573	0.651	0.465
8	0.765	0.674	0.677	0.818	0.942	0.955	0.868	0.826	0.786	0.595	0.621	0.488
9	0.554	0.521	0.597	0.811	0.948	0.968	0.894	0.874	0.835	0.805	0.828	0.676
10	0.392	0.376	0.575	0.807	0.867	0.939	0.922	0.889	0.818	0.740	0.811	0.630
11	0.229	0.271	0.456	0.709	0.790	0.882	0.798	0.894	0.792	0.745	0.788	0.567
12	0.237	0.183	0.380	0.698	0.704	0.846	0.746	0.866	0.806	0.664	0.778	0.427

Table 2.4: P-values of the $\mathcal{F}$ tests at different lags and delays for the USD/CHF real exchange rate index

p	d	p($\mathcal{F}_4$)	p($\mathcal{F}_3$)	p($\mathcal{F}_2$)	first rule[a]	second rule[b]
1	2	0.229	0.001	0.000	LSTAR	ESTAR
1	3	0.790	0.210	0.000	LSTAR	LSTAR

a: the first rule is that of Teräsvirta [1994]; *b*: the second rule is that of Teräsvirta and Anderson [1992].

2.3.3 Discrimination between LSTAR and ESTAR

To discriminate between LSTAR and ESTAR models, Teräsvirta [1994] suggests a simple decision rule based on three $\mathcal{F}$ tests. The three statistics are computed using the artificial regression (2.11) and they are derived from the following hypotheses:

$$\begin{aligned} H_{04} &: \beta_4 = 0 \\ H_{03} &: \beta_3 = 0 \,|\, \beta_4 = 0 \\ H_{02} &: \beta_2 = 0 \,|\, \beta_3 = \beta_4 = 0. \end{aligned}$$

This author proposes to select an ESTAR model if the p-value of $\mathcal{F}_3$ (the test of H_{03}) is the smallest and to choose an LSTAR model otherwise.

Teräsvirta and Anderson [1992] propose another rule to discriminate between the two models: choose the LSTAR model if H_{04} is rejected; select the ESTAR model if H_{03} is rejected but not H_{04}; and finally rejecting H_{02} but neither H_{04} nor H_{03} leads to the LSTAR model. In the extreme case where none of these tests is rejected, the sequence amounts to testing the linearity of data.

Table 2.4 shows the p-values of the three $\mathcal{F}$ tests (noted $p(\mathcal{F}_4)$, $p(\mathcal{F}_3)$ and $p(\mathcal{F}_2)$) corresponding to the hypotheses sequence H_{04}, H_{03} and H_{02} at two pairs of lag and delay. The choice of pairs is based on the results yielded from Table 2.3. The transition function designated by the Teräsvirta rule and the Teräsvirta and Anderson rule is also displayed. The two rules lead to opposite conclusions for the STAR(1,2) model: one points towards the logistic function and the other towards the exponential function. But, as $p(\mathcal{F}_3)$ is close to $p(\mathcal{F}_2)$, an ESTAR(1,2) model seems to be a better choice. On the other hand, for the STAR(1,3) model, both rules indicate the logistic function.

The three modeling steps suggest estimating an ESTAR(1,2) or an LSTAR(1,3) model. But as this analysis has shown some inconsistencies, we are free to explore other specifications. Moreover, we should not give too much importance to the first step because Teräsvirta [1994] indicates that the selection technique is useful for saving time and effort by avoiding the estimation of nonlinear models as long as possible, but that it cannot be applied strictly. Finally, as it is now much faster to estimate a STAR model, it is preferable to estimate several different models and to discriminate between them at the same time through information criteria or misspecification tests.

2.4 Specification with information criteria

Another way to specify a STAR model is to discriminate between further values of parameters p and d for a STAR model at the time of estimation. Indeed, Taylor et al. [2001] say that parameters p and d can be estimated with the other model parameters using a grid search. The minimization of the sum of squared residuals (SSR) for several values of these parameters yields a consistent estimator. First, we can construct two tables of the SSR values of the nonlinear least squares (NLS) estimation for parameters p and d, for both ESTAR and LSTAR models. But a problem appears: this method does not take into account the number of parameters to be estimated. That is why it is preferable to use an information criterion such as the BIC one directly.

Applying this idea, we construct a table of the BIC values of the NLS estimation for parameters p and d in the range of 1 to 12 (Table 2.5). The NLS estimates are performed using a new algorithm developed in the next section. An adaptation of the algorithm is required to ensure that the series has the same length for all estimates. Note that when the algorithm does not converge to a global minimum, no value is given. The analysis of Table 2.5 shows the BIC is minimized when $p = 2$ and $d = 3$. But two others models have very similar values, the ESTAR(2,2) and ESTAR(3,3). Thus, the three models are kept for the following section. All other models are estimated and discarded except the ones for which the slope parameters γ are significantly different from zero. Indeed, if the slope parameter γ of a STAR model is not significantly different from zero, the model is certainly misspecified.

The same analysis is realized for the LSTAR model, but we do not show the results because the algorithm converges to a global minimum for only one couple of p and d values. That may indicate either a misspecification problem of the LSTAR model or the use of a wrong algorithm. So, we do not estimate the LSTAR model in the next section.

Table 2.5: Value of the BIC criterion of the ESTAR model at different lags and delays for the USD/CHF real exchange rate index

	delay											
lag	1	2	3	4	5	6	7	8	9	10	11	12
1	1948	1954	1947	1958	1952	–	–	–	–	1953	1955	1959
2	–	**1913**	**1913**	–	–	–	–	–	–	–	–	–
3	–	–	**1915**	1926	–	–	–	–	–	–	1921	1925
4	–	–	1926	–	–	–	–	–	1934	–	–	–
5	–	–	1931	–	–	–	–	–	1945	–	1942	1946
6	–	–	–	–	–	–	–	–	–	–	–	–
7	–	–	–	–	–	–	–	–	–	–	–	–
8	–	–	–	–	–	–	–	–	–	–	–	–
9	–	–	–	–	–	–	–	–	–	–	–	–
10	–	–	–	1998	–	–	–	1984	–	–	–	–
11	–	–	–	2003	–	–	–	1991	1998	–	–	–
12	–	2004	–	–	–	–	–	–	–	–	–	–

2.5 Estimation

The next step of the modeling cycle is estimating the STAR model. The previous steps allowed us to choose the lag parameter p, delay parameter d and transition function F. As mentioned in Section 2.4, the models ESTAR(2,2), ESTAR(2,3) and ESTAR(3,3) have a minimum BIC. But additional models can be estimated in order to compare the value of their parameters.

Estimating the parameters in the STAR model (2.1) is a straightforward application of nonlinear least squares (NLS), that is, the parameter $\omega = (\theta, \phi, \gamma, c)'$ can be estimated as:

$$\hat{\omega} = \arg\min_{\omega} Q_T(\omega) = \arg\min_{\omega} \sum_{t=1}^{T} (y_t - \phi' x_t - \theta' x_t F(\gamma, c; s_t))^2. \quad (2.12)$$

Under the additional assumption that the u_t errors are normally distributed, NLS is equivalent to maximum likelihood. Otherwise, the NLS estimates can be interpreted as quasi maximum likelihood estimates (see van Dijk et al. [2002] for more details concerning the consistence of NLS estimates).

We know that a joint estimation of all parameters in the STAR models may cause a problem and that the nonlinear optimization may not converge. Teräsvirta [1994] therefore suggests to standardize the transition function by dividing it by $\hat{\sigma}^2(y)$, the sample variance of y_t. Moreover, the choice of starting values for the parameters has great importance in order to achieve convergence. If the estimation does not converge, this author also suggests carrying out the estimation using a grid for parameters γ and c. The idea is as follows: if γ and c are known and fixed, the STAR model becomes linear in the autoregressive parameters ϕ and θ and it can be obtained by ordinary least squares (OLS). Then, the sum of squares function $Q_T(\omega)$ can be concentrated and minimized with respect to the γ and c parameters only. We apply the latter suggestion in the estimations reported in Table 2.6.

To ensure optimization convergence, we propose to use a new algorithm. We start by a function that computes the concentrated sum of squares of the model for given parameters γ and c. Then we minimize it, using the Differential Evolution algorithm initially developed by Rainer Storn (see Price et al. [2005]) and implemented in the statistical software package R (R Development Core Team [2007]) by Ardia [2007]. Once the optimal values of γ and c are found, we estimate the remaining parameters by OLS for the linearized equation. Finally, we estimate the full model by NLS using these starting values. Note that it is possible that the NLS optimization does

not converge due to misspecification. Nevertheless, the careful selection of the starting values ensures that if the NLS optimization converges, the estimation is a global minimum.

This approach is used to estimate ESTAR models for a range of p and d parameters. As mentioned in Section 2.4, the models ESTAR(2,2), ESTAR(2,3) and ESTAR(3,3) have a minimum BIC. The estimates of these models are shown in Table 2.6. Additional models are also shown in Table 2.6 because their slope parameters γ are significantly different from zero.

A first look at Table 2.6 indicates that the estimation of the slope parameter γ is not significantly different from zero for the model ESTAR(2,2). This model is of no interest in the nonlinear framework, because under the null that the slope parameter is zero, the model becomes linear. For the other ESTAR models, γ is significant at 1% or 5%, therefore this result indicates nonlinearity in the USD/CHF real exchange rate index. More generally, the slope parameters of the ESTAR models with delay parameter d equal to 1 or 2 are less significant than with $d = 3$. For the latter, the estimated slope parameters γ are around 1.8-2.2, which indicates a rather smooth switch between the two regimes. And these transitions are around 116, because the estimated threshold parameters c of the models ESTAR(1,3), ESTAR(2,3), ESTAR(3,3) and ESTAR(4,3) are between 115.7 and 116.1.

2.6 Misspecification tests

This section describes the three misspecification tests proposed by Eitrheim and Teräsvirta [1996] and applies them to the ESTAR models estimated in Section 2.5: a test of serial independence of errors, a test to know whether the errors still contain nonlinearities and a test of parameter constancy. Indeed, these tests are specifically developed to measure the adequacy of STAR models in small samples. Thus, they are able to identify the possible misspecification problems in our case.

2.6.1 Serial independence test

Consider the STAR model described in Section 2.1:

$$y_t = \phi' x_t + \theta' x_t F(\gamma, c; s_t) + u_t \tag{2.13}$$

but where:

$$u_t = a' v_t + \epsilon_t, \qquad \epsilon_t \sim iid(0, \sigma_\epsilon^2) \tag{2.14}$$

Table 2.6: Results of frequentist estimation of ESTAR models for the USD/CHF real exchange rate index

Parameter	ESTAR(2,1)	ESTAR(3,1)	ESTAR(1,2)	ESTAR(2,2)
ϕ_0	2068.745	1475.630	-75.768**	6003.471
ϕ_1	-15.384	-11.440	1.682**	14.603
ϕ_2	-1.331	3.942	–	-65.883
ϕ_3	–	-4.208*	–	–
ϕ_4	–	–	–	–
θ_0	-2066.367	-1473.480	77.542**	-6001.001
θ_1	16.663	12.748	-0.699**	-13.322
θ_2	1.027	-4.368	–	65.576
θ_3	–	4.305*	–	–
θ_4	–	–	–	–
γ	1.769*	1.733*	0.085*	4.578
c	117.362**	117.316**	109.507**	115.268**

Parameter	ESTAR(1,3)	ESTAR(2,3)	ESTAR(3,3)	ESTAR(4,3)
ϕ_0	163.941*	303.727	-2830.107**	-2852.191**
ϕ_1	-0.552	-0.202	-1.000	-1.113
ϕ_2	–	-1.518	1.473	1.513
ϕ_3	–	–	24.789**	24.910**
ϕ_4	–	–	–	0.041
θ_0	-162.879*	-302.303	2831.627**	2853.859**
θ_1	1.542*	1.498*	2.320**	2.339**
θ_2	–	1.208	-1.850	-1.904
θ_3	–	–	-24.751**	-24.823**
θ_4	–	–	–	-0.080
γ	2.616**	2.200**	1.819**	1.828**
c	115.712**	115.666**	116.084**	116.086**

** significant at the 1% level. * significant at the 5% level.

where $v_t = (u_{t-1}, \ldots, u_{t-q})'$ and $a = (a_1, \ldots, a_q)'$, $a_q \neq 0$. Assume that the roots of $z^q - \sum_{j=1}^{q} a_j z^{q-j} = 0$ lie inside the unit circle. The test of serial independence uses the first derivative of the transition function with respect to slope and threshold parameters. So, for $s_t = y_{t-d}$ and for an ESTAR model with $F(\gamma, c; y_{t-d}) = F_{t-d} = 1 - \exp\{-\gamma \cdot (y_{t-d} - c)^2\}$, we have :

$$\frac{\partial(\theta' x_t F_{t-d})}{\partial \gamma} = g_\gamma(t) = \exp\{-\gamma(y_{t-d} - c)^2\} \cdot (y_{t-d} - c)^2 \cdot \theta' x_t \qquad (2.15)$$

$$\frac{\partial(\theta' x_t F_{t-d})}{\partial c} = g_c(t) = -2\gamma \cdot \exp\{-\gamma(y_{t-d} - c)^2\} \cdot (y_{t-d} - c) \cdot \theta' x_t. \quad (2.16)$$

The hypothesis of serial independence of errors u_t in (2.13) is $H_0 : a = 0$. Under the null, (2.13) becomes the model estimated in Section 2.5. Eitrheim and Teräsvirta [1996] say that the Lagrange Multiplier (LM) test of H_0 against (2.13) can be performed in four stages as follows.

(i) Estimate the STAR model by NLS (in our case using the algorithm from Section 2.5) under the null of uncorrelated errors and compute the residuals $\hat{u}_t$.

(i') The residuals $\hat{u}_t$ are regressed on x_t, $x_t\hat{F}_{t-d}$, $\hat{g}_\gamma(t)$, and $\hat{g}_c(t)$ to compute new residuals $\tilde{u}_t$, where $\hat{F}_{t-d} = F(\hat{\gamma}, \hat{c}; y_{t-d})$. Then compute the residual sum of squares, $SSR_0 = \sum_{t=1}^{T} \tilde{u}_t^2$.

(ii) Regress $\tilde{u}_t$ on $\hat{v}_t$, x_t, $x_t\hat{F}_{t-d}$, $\hat{g}_\gamma(t)$ and $\hat{g}_c(t)$, then compute the residual sum of squares, SSR.

(iii) Compute the test statistic $\mathcal{F}_{LM} = \frac{(SSR_0 - SSR)/q}{SSR/(T-n-q)}$, where n is the number of parameters in the STAR model.

Stage (i') is added because if the sample size is small and the model difficult to estimate, numerical problems when using the NLS may lead to a solution such that the residual vector $(\hat{u}_1, \ldots, \hat{u}_T)'$ is not precisely orthogonal to the gradient. Hence, stage (i') ensures orthogonality between the residuals $\tilde{u}_t$ and the gradient.

More details are given in Eitrheim and Teräsvirta [1996] about the sufficient conditions ensuring convergence in distribution for this test. In particular, the moment condition $E[u_t^6] < \infty$ must hold.

2.6.2 No remaining nonlinearity test

Consider now an *additive* STAR model:

$$y_t = \phi' x_t + \theta' x_t F_1(\gamma_1, c_1; y_{t-d}) + \pi' x_t F_2(\gamma_2, c_2; y_{t-e}) + u_t \tag{2.17}$$

where u_t is $iid(0, \sigma^2)$ and F_2 is either a logistic or an exponential function with $F_2(0, c_2; y_{t-e}) = 0$. The null hypothesis $H_0 : \gamma_2 = 0$ or $F_2 = 0$ is tested against (2.17). Thus, under the null, the parameters ϕ, θ, γ_1 and c_1 can consistently be estimated by our algorithm. By contrast, the other parameters of the additive STAR model (2.17) are not identified under the null. This problem is circumvented using the third-order Taylor approximation about $\gamma_2 = 0$ on F_2. Indeed, assuming that F_2 is a logistic function, the model (2.17) becomes:

$$y_t = \beta_0' x_t + \theta' x_t F_1(\gamma_1, c_1; y_{t-d}) + \beta_1' x_t^* y_{t-e} + \beta_2' x_t^* y_{t-e}^2 + \beta_3' x_t^* y_{t-e}^3 + \epsilon_t \tag{2.18}$$

where β_0 is ϕ plus the first two elements of the third-order Taylor approximation, $x_t^* = [y_{t-1}, \ldots, y_{t-p}]'$ and ϵ_t is iid $\mathcal{N}(0, \sigma_\epsilon^2)$. Hence, the null hypothesis of no remaining nonlinearity is $H_0' : \beta_1 = \beta_2 = \beta_3 = 0$. Note that the test of H_0' within (2.18) is also powerful against the case where F_2 is an exponential function, and the moment condition requires $E[u_t^8] < 0$.

The test statistic is the same as that of the test of serial independence but with $\hat{v}_t = ((x_t^*)' y_{t-e}, (x_t^*)' y_{t-e}^2, (x_t^*)' y_{t-e}^3)'$. The test may be carried out in four stages, as shown above. The only differences are the new definition of $\hat{v}_t$ at stage (ii) and the degrees of freedom in the $\mathcal{F}$ test, $3p$ and $T - n - 3p$.

2.6.3 Parameter constancy test

The third misspecification test proposed by Eitrheim and Teräsvirta [1996] postulates a parametric alternative to parameter constancy in STAR models which explicitly allows parameters to change smoothly over time. Consider a STAR model for which both ϕ and θ may be subject to change over time:

$$y_t = \phi(t)' x_t + \theta(t)' x_t F(\gamma, c; s_t) + u_t \tag{2.19}$$

where u_t is iid $\mathcal{N}(0, \sigma^2)$. Let the time-varying parameter vectors of length p $\phi(t) = \tilde{\phi} + \lambda_1 H_j(t; \gamma_1, c_0, c_1, c_2)$ and $\theta(t) = \tilde{\theta} + \lambda_2 H_j(t; \gamma_1, c_1)$, where λ_1 and λ_2 are vectors of length p. The null hypothesis of parameter constancy in (2.19) is $H_0 : H_j(t; \gamma_1, c_0, c_1, c_2) = 0$ in all three cases. Three functional forms for $H_j(\cdot)$ are used by the authors:

$$H_1(\cdot) = \Big(1 + \exp\{-\gamma_1(t - c_0)\}\Big)^{-1} - 0.5 \tag{2.20}$$

$$H_2(\cdot) = \Big(1 - \exp\{-\gamma_1(t - c_0)^2\}\Big) \tag{2.21}$$

$$H_3(\cdot) = \Big(1 + \exp\{-\gamma_1(t^3 + c_2t^2 + c_1t + c_0)\}\Big)^{-1} - 0.5 \tag{2.22}$$

where $\gamma_1 > 0$. Hence, the null hypothesis of parameter constancy may be written $H_0 : \gamma_1 = 0$. H_3 is the most flexible function allowing monotonically as well as nonmonotonically changing parameters. H_1 and H_2 are special cases of H_3, thus it is sufficient to describe the test statistic for the latter. Note that the parameters γ, c_0, c_1 and c_2 are identified under the alternative hypothesis but not under the null. To circumvent this identification problem, Eitrheim and Teräsvirta [1996] take the first-order Taylor expansions of H_3 about $\gamma_1 = 0$. After reparametrization, the model (2.19) becomes:

$$\begin{aligned} y_t \;=\; & \beta_0'x_t + \beta_1'tx_t + \beta_2't^2x_t + \beta_3't^3x_t + \\ & \Big(\beta_4'x_t + \beta_5'tx_t + \beta_6't^2x_t + \beta_7't^3x_t\Big) \cdot F(\gamma, c; s_t) + r_t^* \end{aligned} \tag{2.23}$$

where $r_t^* = u_t + R(t; \gamma_1, c_0, c_1, c_2)$. Under H_0, $r_t^* = u_t$. The null hypothesis of parameter constancy is then $H_0' : \beta_j = 0, j = 1, 2, 3, 5, 6, 7$.

The test statistic, called LM_3, is the same as that of the test of serial independence, but with this difference:

$$\hat{v}_t = \Big(tx_t',\, t^2x_t',\, t^3x_t',\, tx_t'F(\hat{\gamma}, \hat{c}; s_t),\, t^2x_t'F(\hat{\gamma}, \hat{c}; s_t),\, t^3x_t'F(\hat{\gamma}, \hat{c}; s_t)\Big)'.$$

The test may be carried out in four stages as shown above. The only differences are the new definition of $\hat{v}_t$ at stage (ii) and the degrees of freedom in the $\mathcal{F}$ test, $6(p + 1)$ and $T - 8(p + 1) - 2$. If the alternative hypothesis is specified as (2.19) with the symmetric nonmonotonic transition function (2.21), the test is based on (2.23) assuming $\beta_3 = 0$ and $\beta_7 = 0$ (test statistic LM_2). If only monotonic change in parameters defined by (2.20) is allowed, equation (2.23) applies with $\beta_2 = \beta_3 = 0$ and $\beta_6 = \beta_7 = 0$ (test statistic LM_1).

2.6.4 Application

Table 2.7 shows the p-values of the three misspecification tests of Eitrheim and Teräsvirta [1996] for several ESTAR models including the best ones designated by the BIC (ESTAR(2,3) and ESTAR(3,3)), and the ESTAR(2,1) and ESTAR(3,1) models. The latter two are chosen because their

Table 2.7: P-values of misspecification tests of four ESTAR models for the USD/CHF real exchange rate index

	ESTAR(2,1)	ESTAR(3,1)	ESTAR(2,3)	ESTAR(3,3)
Serial independence				
q=1	0.603	0.547	0.084	0.046*
q=2	0.266	0.726	0.169	0.042*
q=3	0.427	0.683	0.320	0.023*
q=4	0.462	0.814	0.400	0.044*
q=5	0.604	0.879	0.495	0.071
q=6	0.618	0.923	0.588	0.111
q=7	0.714	0.958	0.701	0.173
q=8	0.647	0.736	0.739	0.193
q=9	0.729	0.797	0.800	0.260
q=10	0.252	0.311	0.206	0.025*
No remaining nonlinearity				
e=1	0.529	0.448	0.941	0.933
e=2	0.142	0.111	0.601	0.678
e=3	0.164	0.150	0.450	0.751
Parameter constancy				
LM3	0.119	0.681	0.210	0.592
LM2	0.026*	0.290	0.051	0.277
LM1	0.040*	0.253	0.448	0.796

** means the p-value is under 1%; * means the p-value is under 5%.

slope parameters are significantly different from zero at a level of 5%. No p-value is less than 5% in the ESTAR(3,1) and ESTAR(2,3) models. So, we are not able to reject the null hypothesis of no misspecification for these models. On the other hand, we reject the null hypothesis of parameter constancy for the ESTAR(2,1) model and the null of serial independence for the ESTAR(3,3) model, both at a level of 5%. These considerations are sufficient to discard these two models.

Before turning to forecast evaluation, it is useful to summarize the results of our analysis of the estimations of STAR models for the USD/CHF real exchange rate index. They indicate that the ESTAR(2,3) model is the best one and that the ESTAR(3,1) model yields very good results too.

2.7 Forecast evaluation

The predictive performance of our best ESTAR models (ESTAR(2,3) and ESTAR(3,1)) is evaluated by comparing their out-of-sample forecasts with those of a linear AR model: the AR(11) model. The comparison with a linear model aims to determine whether the complexity of a nonlinear model estimation is bearing fruit in terms of forecasting. Moreover, this section highlights the problem of choosing the number of forecasts used in forecast evaluation. Indeed, this number can influence the forecast accuracy.

Our analysis is based on 120 one-step ahead forecasts using 120 expanding estimation windows. So, the first estimation window contains 288 observations from 1973:01 to 1996:12 and yields a forecast for 1997:01. The second one contains 289 observations from 1973:01 to 1997:01 and yields a forecast for 1997:02. The last one includes 407 observations from 1973:01 to 2006:11 and yields a forecast for 2006:12. The models are re-estimated with the 120 information sets. Thus, 120 forecasts are available for evaluation. A simple way to proceed would be to use only the maximum of forecasts. Here, we choose to increase the number of forecasts and compare the evaluations in order to show that the choice of the number of forecasts is not insignificant.

The formal evaluations of the predictive performance of our models are analyzed with the Diebold and Mariano [1995] test. According to van Dijk and Franses [2003], the DM test is widely used nowadays for comparing point forecasts from two competing models. Moreover, van Dijk et al. [2003] find that the DM test is more powerful in discriminating linear and nonlinear models than other forecasting techniques.

Table 2.8: Mean absolute prediction errors of three models for the USD/CHF real exchange rate index

Number of forecasts	20	40	60	80	100	120
ESTAR(2,3)	1.622	1.786	1.752	1.716	1.815	1.805
ESTAR(3,1)	**1.524**	**1.689**	**1.678**	**1.658**	**1.776**	**1.759**
AR(11)	1.599	1.803	1.748	1.697	1.800	1.769

The version of the DM test used here is a comparison of the mean absolute prediction errors (MAPE) of two models. The MAPE for a series of S forecasts ($S = 1, ..., 120$) is given by:

$$MAPE_S = \frac{1}{S}\sum_{s=1}^{S}\left|y_{T+1-s} - \hat{y}_{T+1-s,i}\right| \tag{2.24}$$

where $\hat{y}_{T+1-s,i}$ is the one-step ahead forecast of y_{T+1-s} using model i and based on the information set available at time $T - s$, and T is the sample size. The DM statistics are obtained in calculating the loss differential:

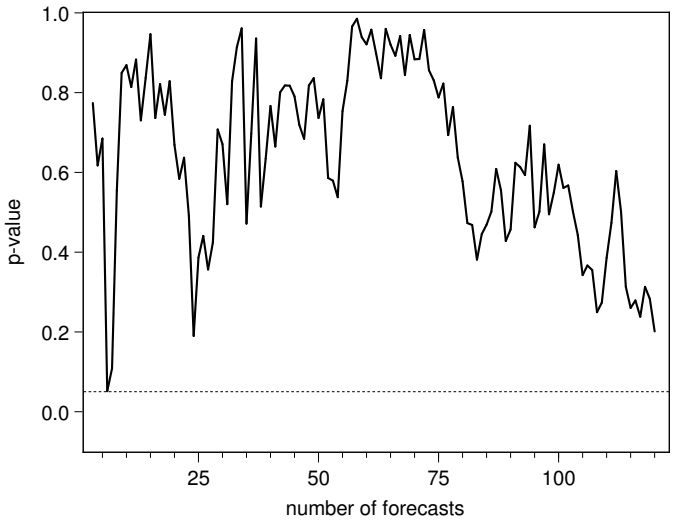
$$D_{s,ij} = \left|y_{T+1-s} - \hat{y}_{T+1-s,i}\right| - \left|y_{T+1-s} - \hat{y}_{T+1-s,j}\right| \tag{2.25}$$

for $s = 1, ..., S$ and then regressing these values on a constant c_{sij} in increasing the number S from $S = 2$ to $S = 120$. The variance of $\hat{c}_{ij}$ is calculated using the Newey-West estimator. The null hypothesis of same $MAPE_S$ is tested using a t statistic. So, rejecting H_0 means that the forecast accuracy of a model is different from another.

Table 2.8 shows the comparison of the $MAPE_S$ of the models ESTAR(2,3), ESTAR(3,1) and AR(11) for an increasing number of forecasts (20, 40, 60, 80, 100 and 120). The $MAPE_S$ of the ESTAR(2,3) and AR(11) models are close to each other. On the other hand, the ESTAR(3,1) model obviously has the smallest value, whatever the number of forecasts may be. However, these results have to be evaluated with the DM test to know whether the $MAPE_S$ of the ESTAR(3,1) model are significantly different from the two others.

In our case, it is more interesting to analyze the DM tests by plotting their p-values in increasing the number of forecasts and look whether the p-values are under 5% or not. Figure 2.2 shows the p-values of DM tests between models ESTAR(2,3) and AR(11) for an increasing number of forecasts. No p-value is under 0.05, thus the DM test does not reject the null of same accuracy of both models, whatever the number of forecasts may be. The analysis of Figure 2.3 indicates that the forecasts of the models

Figure 2.2: P-values of the DM test between the ESTAR(2,3) and AR(11) models for an increasing number of forecasts



ESTAR(3,1) and ESTAR(2,3) do not have the same accuracy. Looking at the values of the $MAPE_S$ of both models, the forecast accuracy of the ESTAR(3,1) is better than that of the ESTAR(2,3) model. Thus, this result implies that the ESTAR(3,1) model is obviously the best nonlinear model in the sense of best forecast accuracy. Moreover, Figure 2.4 shows the p-values of DM tests between the ESTAR(3,1) model and the AR(11) model for an increasing number of forecasts. This figure provides a good comparison between a nonlinear and a linear model. Knowing that the $MAPE_S$ of the ESTAR(3,1) is lower than that of the AR(11), rejecting the null of the DM test implies that the ESTAR(3,1) performs better than the AR(11). Precisely, Figure 2.4 shows that between 35 and 55 forecasts the p-values of DM tests are under 0.05. Hence, if a researcher chooses, for example, 50 forecasts for his analysis, he should conclude that the forecasts of the ESTAR(3,1) model are significantly more accurate than those of the AR(11) model. This would be a mistake because in increasing the number of forecasts, the power of the DM test increases and for more than 55 forecasts, all p-values being over 0.05.

Figure 2.3: P-values of the DM test between the ESTAR(3,1) and ESTAR(2,3) models for an increasing number of forecasts

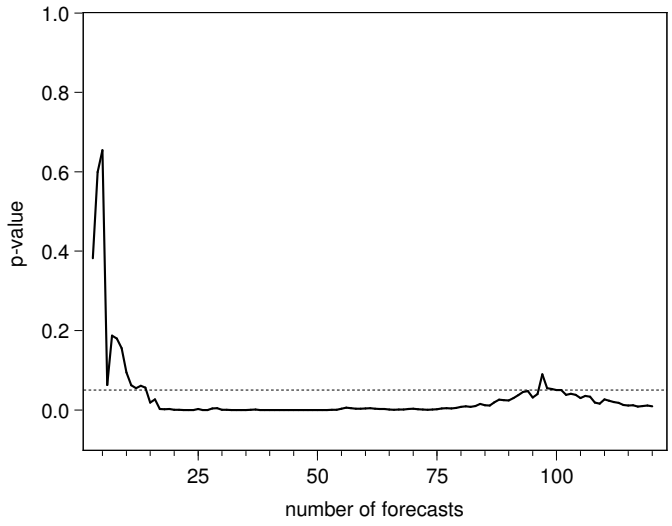
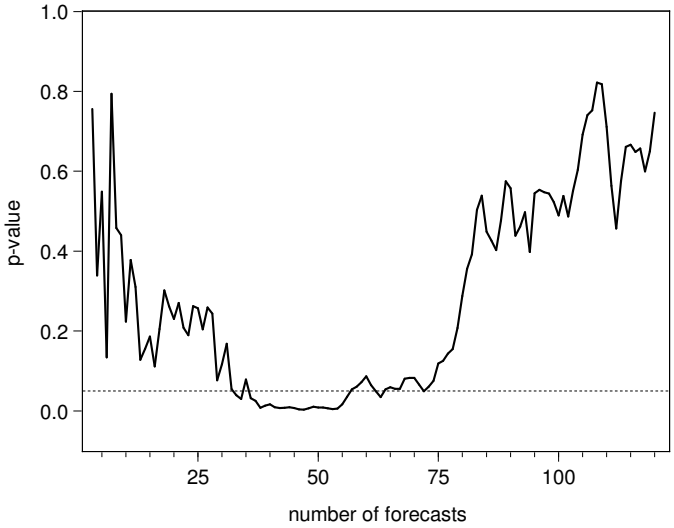


Figure 2.4: P-values of the DM test between the ESTAR(3,1) and AR(11) models for an increasing number of forecasts



2.8 Conclusion

In this chapter, we analyzed the USD/CHF real exchange rate index with unit root tests and found that the series was stationary with the most powerful test available. We discussed specification, estimation and evaluation of the STAR model in a frequentist context. In order to discriminate between STAR models, we first used the methodology proposed by Teräsvirta [1994], but it was, according to the author of this thesis, unreliable and inconsistent. Second, we discriminated between STAR models using information criteria. This methodology was problematic because our new estimation algorithm did not converge for all models. Nevertheless, we could discard some STAR models, including all LSTAR models. Then, the misspecification tests developed by Eitrheim and Teräsvirta [1996] were applied in order to discard more models. Thus, the ESTAR(2,3) model was the best STAR model, in the sense of best fit. Finally, the analysis of predictive performances indicated that the ESTAR(2,3) was beaten by the ESTAR(3,1) model.

Moreover, Haggan and Ozaki [1981] pointed out that the frequentist methods have difficulties in estimating the slope parameter γ. These difficulties, confirmed by Teräsvirta [1994], derive from a lack of information about this parameter. The Bayesian methods allow to solve this problem by introducing subjective information about the parameters. Thus, the following chapter will present the estimation of the STAR model in a Bayesian context.

Chapter 3

Bayesian estimation of the STAR model

In this chapter, the estimation and evaluation of the STAR model in a Bayesian context are presented. We have decided to turn to the Bayesian methods because we have seen in the previous chapter that the frequentist methods are limited. The Bayesian methods allow to compensate for lack of information on some parameters by some *a priori* information. In order to focus our investigations on a few models only, preliminary work was conducted. Some models were then discarded in accordance with two criteria: the inability of the Markov chain Monte Carlo (MCMC) algorithm to converge and the results obtained in Chapter 2 (see Section 2.8). For example, the LSTAR and ESTAR(3,1) models are discarded because of convergence problems. Finally, only four models are compared in what follows: the ESTAR(2,2), ESTAR(2,3), ESTAR(3,3) and ESTAR(4,3) models.

The outline of this chapter is as follows. The MCMC scheme is detailed in Section 3.1. In Section 3.2, some ESTAR models are estimated on the USD/CHF real exchange rate index: a model comparison, a posterior analysis and a sensitivity analysis are performed for the MCMC estimation. In Section 3.3, we analyze the dynamic behavior of the ESTAR models. We test the ESTAR models for misspecification in Section 3.4. An evaluation of the predictive performances of the ESTAR models is given in Section 3.6 and the main results are summarized in Section 3.7.

3.1 Posterior simulator

The MCMC algorithm used in this chapter for the STAR model:

$$y_t = \phi' x_t + \theta' x_t F(\gamma, c; s_t) + u_t \tag{3.1}$$

iterates on the full conditional posteriors of three blocks: the vector:

$$\beta' = (\phi' \quad \theta') \tag{3.2}$$

of length $k = (2p + 2)$, σ^2 and the vector $\vartheta = (\gamma \quad c)'$. The terms $\phi = (\phi_0 \quad \phi_1 \quad ... \quad \phi_p)'$ and $\theta = (\theta_0 \quad \theta_1 \quad ... \quad \theta_p)'$ are vectors of autoregressive parameters of length $(p+1)$, σ^2 is the variance of the errors, and γ and c are parameters of the transition function.

3.1.1 The likelihood function

The first step towards finding the expressions of the full conditional posteriors is to specify the likelihood function of the general STAR model (3.1). The independence and normality assumptions of the errors imply that we can write:

$$p(y|\beta, \sigma^2, \gamma, c) = \frac{1}{(2\pi)^{T/2}} \cdot \frac{1}{\sigma^T} \cdot \exp\left[-\frac{u'u}{2\sigma^2}\right] \tag{3.3}$$

where u is the vector of the u_t implied by (3.1).

3.1.2 The prior

Second, we develop the assumptions on the priors of the three blocks. We assume a multinormal prior on β with mean vector $\underline{\beta}$ and precision matrix $\underline{V}$:

$$p(\beta) = \frac{1}{(2\pi)^{k/2}} \cdot |\underline{V}^{-1}|^{-1/2} \cdot \exp\left[-\frac{1}{2}(\beta - \underline{\beta})' \underline{V} (\beta - \underline{\beta})\right] \tag{3.4}$$

and an independent inverted Gamma prior on σ^2 with shape and scale hyperparameters $\underline{a}$ and $\underline{b}$:

$$p(\sigma^2) = \underline{b}^{\underline{a}} \cdot \frac{1}{G(\underline{a})} \cdot \sigma^{-2(\underline{a}+1)} \cdot \exp\left[-\frac{\underline{b}}{\sigma^2}\right] \tag{3.5}$$

where $G(\cdot)$ is the gamma function.

The assumptions on the priors of the parameters γ and c are crucial in the STAR model. As in Deschamps [2008], we can specify the exponential transition function:

$$F(\gamma, c; s_t) = 1 - \exp\Big[-\gamma(s_t - c)^2\Big], \gamma > 0 \tag{3.6}$$

of the STAR model differently in considering γ^2 instead of γ for the slope parameter. We then define a new transition function in (3.1): $\tilde{F}(\gamma, c; s_t) = 1 - \exp\Big[-\gamma^2(s_t - c)^2\Big]$. Opting for independent normal priors for γ and c:

$$\gamma \sim \mathcal{N}(\underline{\gamma}, \sigma_\gamma^2) \tag{3.7}$$

$$c \sim \mathcal{N}(\underline{c}, \sigma_c^2) \tag{3.8}$$

ensures that the slope parameter is positive.

Another possibility proposed by Lopes and Salazar [2006] is to keep the exponential transition function (3.6) in the STAR model and to assume a Gamma prior on the slope parameter γ. A comparison of the two methods is proposed later.

We take an almost improper prior on β with a null expectation vector, a variance of 10^6 for parameters ϕ_0 and θ_0, and a variance of 10^3 for the other parameters. All covariances are null. Furthermore, to ensure an almost improper prior on σ^2, the $\underline{a}$ and $\underline{b}$ hyperparameters are 10^{-6}.

The MCMC algorithm is sensitive to the choice of priors on parameters γ and c. In particular, if the priors are not sufficiently informative, the algorithm has difficulties converging. Hence, we assume that the hyperparameters $\underline{c}$, σ_c^2, $\underline{\gamma}$ and σ_γ^2 are respectively 115, 1, $\sqrt[4]{2}$ and $1.5 - \sqrt{2}$. Note that this last choice implies $E[\gamma^2] = 1.5$ and $V[\gamma^2] = 0.5$. Indeed, if $x \sim \mathcal{N}(\mu, \sigma^2)$, we have $y = (\frac{1}{\sigma}(x-\mu))^2 \sim \chi_1^2$ and $x^2 = \sigma^2 y + 2\mu x - \mu^2$. Then, $E[x^2] = \sigma^2 + 2\mu^2 - \mu^2$ and $V[x^2] = 2\sigma^4 + 4\mu^2\sigma^2$, because the covariance between y and x is null in this case. Our choices of priors can be criticized because they are not based on existing results. Many attempts have been made before opting for these values. Other choices of priors involve Markov chains with bad properties. More details are given in Section 3.2.4.

3.1.3 The posterior

Finally, we develop the expressions of the full conditional posteriors of each block and then use them to generate random draws. The draws of two blocks are used to generate a draw of the remaining block. The first two blocks are easy to treat. Indeed, the full conditional posteriors of β and σ^2 are analytically known (see among others Koop [2003], Chapter 4). We can then use the Gibbs sampler to produce random draws from the posterior.

More precisely, if γ and c are known, the STAR model (3.1) becomes linear and has the vector form $y = X\beta + u$, where row t of the $T \times k$ matrix X is:

$$(1 \quad y_{t-1} \quad ... \quad y_{t-p} \quad \tilde{F}_t \quad \tilde{F}_t y_{t-1} \quad ... \quad \tilde{F}_t y_{t-p}) \tag{3.9}$$

with $\tilde{F}_t \equiv \tilde{F}(\gamma, c; s_t)$, $k = (2p + 2)$ is the length of β and T is the sample size. Then, the full conditional posterior of β is given by:

$$p(\beta|y, \sigma^2, \gamma, c) \;=\; \frac{1}{(2\pi)^{k/2}} \cdot |\overline{V}^{-1}|^{-1/2} \cdot \exp\left[-\frac{1}{2}(\beta - \overline{\beta})' \, \overline{V} \, (\beta - \overline{\beta})\right] \tag{3.10}$$

with $\overline{\beta} = \overline{V}^{-1}(\frac{1}{\sigma^2} X'y + \underline{V}\underline{\beta})$ and $\overline{V} = \frac{1}{\sigma^2} X'X + \underline{V}$, and the full conditional posterior of σ^2 is given by:

$$p(\sigma^2|y, \beta, \gamma, c) \;=\; \overline{b}^{\overline{a}} \cdot \frac{1}{G(\overline{a})} \cdot \sigma^{-2(\overline{a}+1)} \cdot \exp\left[-\frac{\overline{b}}{\sigma^2}\right] \tag{3.11}$$

with $\overline{a} = \underline{a} + \frac{T}{2}$ and $\overline{b} = \underline{b} + \frac{1}{2}(y - X\beta)'(y - X\beta)$.

The full conditional posterior in the third block is nonstandard, i.e. it does not have a known form. The algorithm for simulating these posteriors is then based on the independence chain Metropolis-Hastings (ICMH) using a multivariate Student distribution as a candidate generating density. Considering the independent normal priors for γ and c, the kernel of the full conditional posterior is:

$$\begin{aligned} \kappa^*(\vartheta) \;=\; \exp\Bigg[& -\frac{(\gamma - \underline{\gamma})^2}{2\sigma_\gamma^2} - \frac{(c - \underline{c})^2}{2\sigma_c^2} \\ & -\frac{1}{2\sigma^2} \sum_{t=1}^{T} \Big(y_t - \phi' x_t - \theta' x_t \tilde{F}(\gamma, c; s_t)\Big)^2 \Bigg]. \end{aligned} \tag{3.12}$$

The parameters of the candidate density of the independence chain Metropolis-Hastings algorithm are usually fixed for each iteration and taken from the ML estimation (see for example Koop [2003], Chapter 4). Deschamps [2008] proposes to estimate these parameters at each iteration, using the following first-order Taylor expansion of model (3.1) with $\tilde{F}(.)$ around (γ^*, c^*):

$$y_t^* \;=\; \gamma x_{1t}^* + c x_{2t}^* + v_t$$

where:

$$y_t^* = y_t - \phi' x_t - \theta' x_t \left(\tilde{F}(\gamma^*, c^*; s_t) - \frac{\partial \tilde{F}_t}{\partial \gamma}\bigg|_{\gamma^*, c^*} \gamma^* - \frac{\partial \tilde{F}_t}{\partial c}\bigg|_{\gamma^*, c^*} c^* \right)$$

$$x_{1t}^* = (\theta' x_t) \frac{\partial \tilde{F}_t}{\partial \gamma}\bigg|_{\gamma^*, c^*}$$

$$x_{2t}^* = (\theta' x_t) \frac{\partial \tilde{F}_t}{\partial c}\bigg|_{\gamma^*, c^*} .$$

The vector $\vartheta^* = (\gamma^* \, c^*)'$ is an approximate solution to the Bayesian update equations:

$$\vartheta^* = S_\vartheta^{-1} \left[\frac{X_*' y_*}{\sigma^2} + \begin{pmatrix} \sigma_\gamma^2 & 0 \\ 0 & \sigma_c^2 \end{pmatrix}^{-1} \begin{pmatrix} \underline{\gamma} \\ \underline{c} \end{pmatrix} \right] \tag{3.13}$$

$$S_\vartheta = \frac{X_*' X_*}{\sigma^2} + \begin{pmatrix} \sigma_\gamma^2 & 0 \\ 0 & \sigma_c^2 \end{pmatrix}^{-1} \tag{3.14}$$

where y_* is the $T \times 1$ vector with elements y_t^* and X_* is the $T \times 2$ matrix with elements x_{1t}^* and x_{2t}^*. Further iterations on (3.13) and (3.14) are used to find the approximate solution. The vector of prior expectation is taken as a starting point. A candidate ϑ is drawn from a multivariate Student density with kernel:

$$\kappa(\vartheta) = \left[1 + \frac{(\vartheta - \vartheta^*)' S_\vartheta (\vartheta - \vartheta^*)}{\nu} \right]^{-\frac{\nu+2}{2}} \tag{3.15}$$

and is accepted with probability:

$$\alpha(\vartheta_{old}, \vartheta) = min \left[\frac{\kappa(\vartheta_{old})}{\kappa(\vartheta)} \frac{\kappa^*(\vartheta)}{\kappa^*(\vartheta_{old})}, 1 \right] \tag{3.16}$$

where ϑ_{old} is the most recently drawn vector. The number ν of degrees of freedom in (3.15) can be chosen by experimentation to ensure a good acceptance rate of the candidate; $\nu = 3$ seems to be a good choice.

The posterior simulators are tested against any analytical or coding errors following Geweke [2004].

Table 3.1: Chain properties of the Bayesian ICMH ESTAR estimation for the USD/CHF real exchange rate index

	ESTAR(2,2)	ESTAR(2,3)	ESTAR(3,3)	ESTAR(4,3)
$\overline{\alpha}(\vartheta_{old}, \vartheta)$	0.638	0.729	0.739	0.735
RNE(γ^2)	0.425	0.443	0.378	0.299
RNE(c)	0.398	0.473	0.282	0.348
corr(γ^2)	0.572	0.670	0.709	0.805
corr(c)	0.667	0.492	0.798	0.794

$\overline{\alpha}(.)$ is the mean acceptance rate, RNE the *relative numerical efficiency* and corr the first lag autocorrelation of the Markov chains.

3.2 MCMC estimation

We run the algorithm for 50'000 iterations and discard the first half as burn-in. Chain convergence of the chains is checked using the convergence diagnostic methodology proposed by Geweke [1992]: the mean of the Markov chain based on the first half of the draws should be close to the mean based on the last half. If these two means are very different, this indicates either that too few draws have been taken, or that too few draws have been discarded as burn-in. In practice, we follow the recommendation of Koop [2003] in Chapter 4, by comparing the 10% first draws to the 40% last draws. So, if the sample draws are drawn from a stationary distribution, the two means are equal and the convergence diagnostic given by:

$$CD = \frac{\hat{\mu}_{S_A} - \hat{\mu}_{S_C}}{\sqrt{\hat{\sigma}_A^2 + \hat{\sigma}_C^2}} \tag{3.17}$$

should follow an asymptotically standard normal distribution. Define S_A and S_C be the 10% first and 40% last draws of the chain, $\hat{\mu}_{S_A}$ and $\hat{\mu}_{S_C}$ are the means of these samples. Moreover, we define $\hat{\sigma}_A^2$ and $\hat{\sigma}_C^2$ as the squares of the numerical standard errors calculated using the spectral methods.

3.2.1 Markov chain properties

The analysis of the results of the MCMC algorithm starts by looking at the properties of the Markov chains. Table 3.1 shows some of these properties and indicates that the Markov chains are well-mixing. First, the mean acceptance rate of the independence chain Metropolis-Hastings algorithm given in (3.16) is about 0.70. Second, the *relative numerical efficiency*

(RNE) developed by Geweke [1989] is the ratio of two measures of the standard error of the mean of a Markov chain. The term in the numerator is the naive standard error calculated by ignoring chain autocorrelation. The term in the denominator is the numerical standard error calculated using spectral methods and also used in the convergence diagnostic above (3.17). The RNE of parameters gives information on the efficiency of the MCMC algorithm. The higher the RNE, the greater the efficiency because the MCMC chain contains little autocorrelation. The RNE of the crucial parameters γ^2 and c is around 0.3-0.4 for the four models. Third, the first lag autocorrelations of the posteriors for those two parameters are all below or equal to 0.8. Obviously, there is no autocorrelation problem with this algorithm.

3.2.2 Model comparison

In this section, we perform the comparison of four models: the ESTAR(2,2), ESTAR(2,3), ESTAR(3,3) and ESTAR(4,3) models. A simple way to discriminate between these models is to compare some information criteria. In the Bayesian context, the likelihood function has to be evaluated at a point estimate. We choose to evaluate the likelihood function and then the information criteria, at the mean of the posterior. First, the AIC by Akaike [1974] is used as a benchmark for the comparison. Second, the BIC by Schwarz [1978] is used because it has the property of favoring parsimonious models. Both are the most used information criteria in econometrics.

Third, we use a criterion recently proposed by Spiegelhalter et al. [2002], the *deviance information criterion* (DIC). This criterion was developed to compare models in which the number of parameters is not clearly defined. Indeed, the idea is to estimate the effective number of the model's parameters given by $p_D = \overline{D} - D(\tilde{\theta})$, where $D(\theta) = -2ln(p(y|\theta))$ is the deviance, $\tilde{\theta} = E[\theta|y]$, $\overline{D} = E[D(\theta)|y]$ and θ the model's parameters. The authors then define the DIC as $DIC = D(\tilde{\theta}) + 2p_D$.

Finally, a commonly used Bayesian criterion for discriminating between models is the marginal likelihood. In a general notation, the marginal likelihood is:

$$p(y) = \int f(y|\theta)p(\theta)d\theta \tag{3.18}$$

where $f(y|\theta)$ is the likelihood and $p(\theta)$ the prior of an unspecified model. Several methods exist to estimate marginal likelihood, but DiCiccio et al. [1997] suggested applying the *bridge sampling* technique to do it. The bridge sampling method, that we know thanks to Meng and Wong [1996], was

introduced (into statistics) to estimate ratios of normalizing constants. In Bayesian econometrics, keeping in mind the Bayes' rule[1], bridge sampling is based on the following identity:

$$1 = \frac{\int p(\theta) f(y|\theta) \alpha(\theta) q(\theta) d\theta}{\int p(y) p(\theta|y) \alpha(\theta) q(\theta) d\theta} \tag{3.19}$$

where $\alpha(\theta)$ is an arbitrary bridge function and $q(\theta)$ is a normalized importance density. As the marginal likelihood does not depend on θ, (3.19) is equal to:

$$p(y) = \frac{\int \left[p(\theta) f(y|\theta) \alpha(\theta)\right] q(\theta) d\theta}{\int \left[\alpha(\theta) q(\theta)\right] p(\theta|y) d\theta} = \frac{E_q \left[p(\theta) f(y|\theta) \alpha(\theta)\right]}{E_p \left[\alpha(\theta) q(\theta)\right]} \tag{3.20}$$

where $E_h[g(\theta)]$ denotes the expectation of $g(\theta)$ with respect to the density $h(\theta)$.

The general bridge sampling estimator $\hat{p}(y)$ of the marginal likelihood is obtained by taking the sample averages in (3.20) where $\tilde{\theta}^l$, $l = 1, ..., L$ are draws from the importance density $q(\theta)$ and θ^m, $m = 1, ..., M$ are draws from the posterior $p(\theta|y)$:

$$\hat{p}(y) = \frac{L^{-1} \sum_{l=1}^{L} p(\tilde{\theta}^l) f(y|\tilde{\theta}^l) \alpha(\tilde{\theta}^l)}{M^{-1} \sum_{m=1}^{M} \alpha(\theta^m) q(\theta^m)} . \tag{3.21}$$

As said in Deschamps [2008], to well choose the function $\alpha(\theta)$ is important for numerical efficiency. Moreover, special choices of $\alpha(\theta)$ lead to common estimators (Frühwirth-Schnatter [2004]). First, if $\alpha(\theta) = 1/q(\theta)$ is used, one obtains the *importance sampling* estimator:

$$\hat{p}_{IS}(y) = \frac{1}{L} \sum_{l=1}^{L} \frac{p(\tilde{\theta}^l) f(y|\tilde{\theta}^l)}{q(\tilde{\theta}^l)} . \tag{3.22}$$

Second, if $\alpha(\theta) = 1/(p(\theta) f(y|\theta))$ is used, one obtains the *reciprocal importance sampling* estimator of Gelfand and Dey [1994]:

$$\hat{p}_{RI}(y) = \left[\frac{1}{M} \sum_{m=1}^{M} \frac{q(\theta^m)}{p(\theta^m) f(y|\theta^m)} \right]^{-1} . \tag{3.23}$$

[1] As Koop [2003] said, the Bayes' rule *lies at the heart of Bayesian econometrics* is written $p(\theta|y)p(y) = f(y|\theta)p(\theta)$.

Table 3.2: Information criteria of the Bayesian ICMH ESTAR estimation for the USD/CHF real exchange rate index

	ESTAR(2,2)	ESTAR(2,3)	ESTAR(3,3)	ESTAR(4,3)
np	9	9	11	13
AIC	1941.175	1927.162	1927.486	**1924.055**
BIC	1977.233	**1963.219**	1971.529	1976.074
p_D	7.386	8.935	9.827	11.790
DIC	1937.946	1927.033	1925.140	**1921.635**
$log(\hat{p}(y))$	-1016.890	**-1011.931**	-1018.747	-1024.418

np is the number of parameters of the model, p_D the effective number of parameters used in the DIC and $log(p(y))$ the logarithmic marginal likelihood.

Finally, Meng and Wong [1996] propose an asymptotically optimal choice of $\alpha(\theta)$ that minimizes the expected relative error of the estimator $\hat{p}(y)$:

$$\alpha(\theta) = \frac{1}{Lq(\theta) + M\frac{p(\theta)f(y|\theta)}{p(y)}} . \tag{3.24}$$

In what follows, we estimate the marginal likelihood using (3.24). This estimator is called *optimal bridge sampling estimator* $\hat{p}_{BS}(y)$ by Frühwirth-Schnatter [2004]. As this choice of $\alpha(\theta)$ depends on the marginal likelihood, an iterative procedure must be applied: we take (3.23) as the starting value and iterate between (3.24) and (3.20) for M = L = 50'000. Note that the three estimators $\hat{p}_{IS}(y)$, $\hat{p}_{RI}(y)$ and $\hat{p}_{BS}(y)$ yield almost the same value for each model, in our case.

We follow the choice of Deschamps [2008] to define the importance density $q(\theta)$: it has the same parametric form as the prior, but with moments that match the empirical posterior ones obtained by MCMC.

Table 3.2 shows the comparison of the four preselected models. The ESTAR(2,3) minimizes the BIC, while the ESTAR(4,3) minimizes the AIC and the DIC. As MCMC estimation is subject to uncertainty, and the AIC of the models ESTAR(2,3) and ESTAR(3,3) are close; the values have been re-estimated ten times. In no case was the order different from the order of Table 3.2. Moreover, the effective number of parameters p_D used in the DIC is lower than the number of parameters np for all models. The difference between these two values explains why the DIC is in favor of the

Table 3.3: Results of the Bayesian ICMH estimation of the four ESTAR models for the USD/CHF real exchange rate index

	ESTAR(2,2)		ESTAR(2,3)		ESTAR(3,3)		ESTAR(4,3)	
	Prior	Posterior	Prior	Posterior	Prior	Posterior	Prior	Posterior
ϕ_0	0	1341.785	0	288.978	0	-489.497	0	38.342
	(10^3)	(556.596)	(10^3)	(149.114)	(10^3)	(645.794)	(10^3)	(688.582)
ϕ_1	0	1.181	0	0.015	0	-0.046	0	-0.679
	$(10^{3/2})$	(2.104)	$(10^{3/2})$	(0.554)	$(10^{3/2})$	(0.530)	$(10^{3/2})$	(0.807)
ϕ_2	0	-11.844	0	-1.595	0	-0.859	0	-4.020
	$(10^{3/2})$	(5.631)	$(10^{3/2})$	(1.234)	$(10^{3/2})$	(1.241)	$(10^{3/2})$	(3.080)
ϕ_3	–	–	–	–	0	6.041	0	7.934
					$(10^{3/2})$	(5.045)	$(10^{3/2})$	(5.922)
ϕ_4	–	–	–	–	–	–	0	-2.684
							$(10^{3/2})$	(2.246)
θ_0	0	-1339.436	0	-287.516	0	490.911	0	-36.797
	(10^3)	(556.539)	(10^3)	(149.110)	(10^3)	(645.803)	(10^3)	(688.593)
θ_1	0	0.112	0	1.281	0	1.363	0	1.999
	$(10^{3/2})$	(2.110)	$(10^{3/2})$	(0.556)	$(10^{3/2})$	(0.533)	$(10^{3/2})$	(0.807)
θ_2	0	11.528	0	1.285	0	0.476	0	3.623
	$(10^{3/2})$	(5.634)	$(10^{3/2})$	(1.235)	$(10^{3/2})$	(1.245)	$(10^{3/2})$	(3.082)
θ_3	–	–	–	–	0	-5.988	0	-7.837
					$(10^{3/2})$	(5.049)	$(10^{3/2})$	(5.927)
θ_4	–	–	–	–	–	–	0	2.648
							$(10^{3/2})$	(2.251)
σ^2	imp^a	6.822	imp^a	6.690	imp^a	6.642	imp^a	6.635
		(0.487)		(0.476)		(0.480)		(0.473)
γ^2	1.5	2.179	1.5	2.086	1.5	2.005	1.5	2.560
	$(\sqrt{0.5})$	(0.607)	$(\sqrt{0.5})$	(0.498)	$(\sqrt{0.5})$	(0.513)	$(\sqrt{0.5})$	(0.602)
c	115	115.322	115	115.648	115	115.775	115	115.681
	(1)	(0.177)	(1)	(0.098)	(1)	(0.159)	(1)	(0.109)

Mean and standard deviation (in brackets) of priors and posteriors; a: almost improper priors because the hyperparameters of the Gamma are 10^{-6} but not nulls.

model ESTAR(4,3). Concerning the Bayesian criterion described above, the ESTAR(2,3) maximizes the logarithmic marginal likelihood. The marginal likelihood is the best way of discriminating between models as long as they have the same prior distribution. It enables us to obtain the Bayes factor (BF), which is the ratio of the marginal likelihood of two models. The Jeffrey's scale is then used to classify the evidence provided by the data in favor of one model against another. The following transformed Bayes factor proposed by Kass and Raftery [1995]: 2 x log(BF) = 2 x (–1011.931 – (–1016.890)) = 9.918 indicates a "strong" evidence for the ESTAR(2,3) model relative to ESTAR(2,2). The transformed BF of ESTAR(2,3) against ESTAR(4,3) is 24.974. So, the Jeffrey's scale indicates a "very strong" evidence in this case. These results indicate an obvious preference for the ESTAR(2,3) model.

Figure 3.1: Histograms of $p(\gamma^2|y)$ ICMH for four ESTAR models

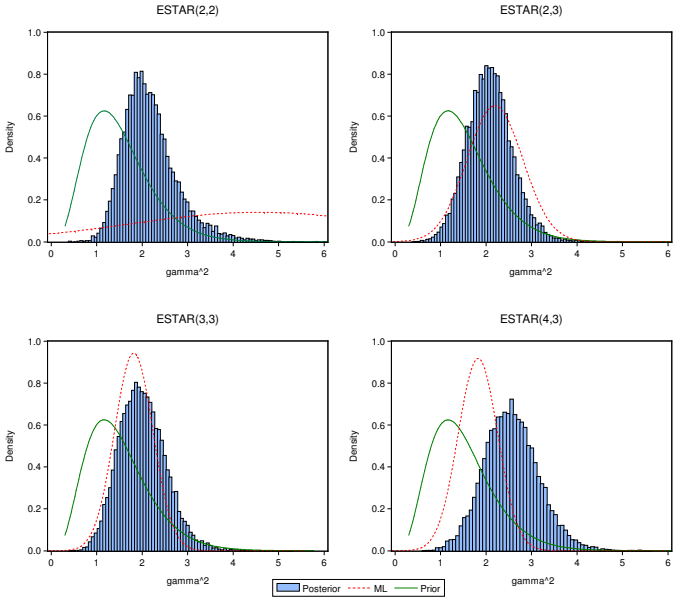


Figure 3.2: Histograms of $p(c|y)$ ICMH for four ESTAR models

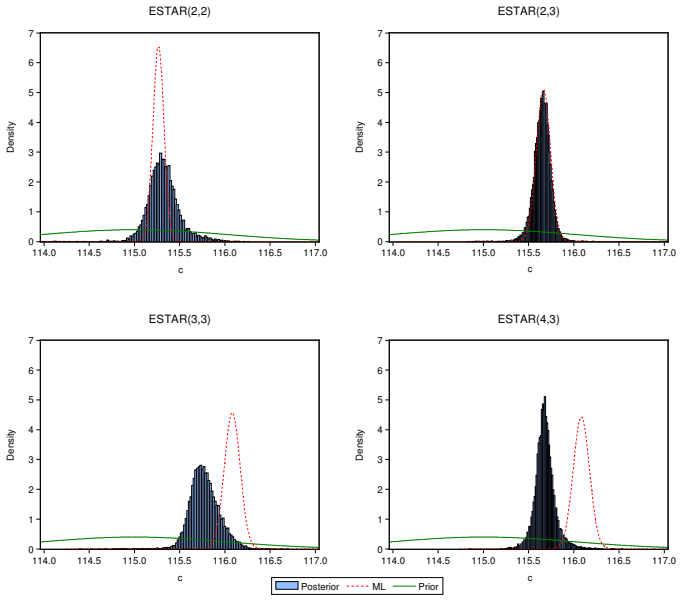


Table 3.4: Chain properties of the Bayesian ICMH ESTAR estimation with looser prior for the USD/CHF real exchange rate index

	ESTAR(2,2)	ESTAR(2,3)	ESTAR(3,3)	ESTAR(4,3)
$\overline{\alpha}(\vartheta_{old}, \vartheta)$	0.553	0.690	0.680	0.668
RNE(γ^2)	0.227	0.102	0.070	0.100
RNE(c)	0.159	0.101	0.066	0.081
corr(γ^2)	0.797	0.879	0.924	0.938
corr(c)	0.897	0.806	0.947	0.936

$\overline{\alpha}(.)$ is the mean acceptance rate, RNE the *relative numerical efficiency* and corr the first lag autocorrelation of the Markov chains.

3.2.3 Posterior analysis

Table 3.3 shows the means and standard deviations of priors and posteriors of the four preselected ESTAR models for the USD/CHF real exchange rate index. The interpretation of the autoregression parameters ϕ and θ has little interest if we look at their mean values only. So, Section 3.3 analyzes the dynamic behavior of the process of the two regimes. The parameters of interest are γ^2 and c, as they indicate the behavior of the transition function between the two regimes. The means of the posteriors of the slope parameter γ^2 are around 2-2.5 for the four models. This indicates a rather smooth switch between the two regimes. The means of the posteriors of the threshold parameter c, interpreted as the purchasing power parity (PPP), are around 115.5 for the four models. In Figures 3.1 and 3.2 we compare the posterior distributions of these parameters for the four ESTAR models. For each model, besides the posterior distributions, the figures present the priors distributions and the asymptotic distributions of the maximum likelihood estimates found in Chapter 2. The posteriors of the ESTAR(2,3) and ESTAR(3,3) models almost match the asymptotic distributions of the likelihood estimates.

3.2.4 Sensitivity analysis

Our estimates are sensitive to the prior specification of the parameters γ and c. Hence, the MCMC algorithm is run using looser priors only for γ given by the hyperparameters $\underline{\gamma} = 1.5$ and $\sigma_\gamma^2 = 0.5$; other priors remain unchanged. These choices imply $E[\gamma^2] = 2.75$ and $V[\gamma^2] = 5$. The purpose of this extension is to highlight the difficulties encountered when some priors are

Figure 3.3: Histograms of $p(\gamma^2|y)$ ICMH with looser prior for four ESTAR models

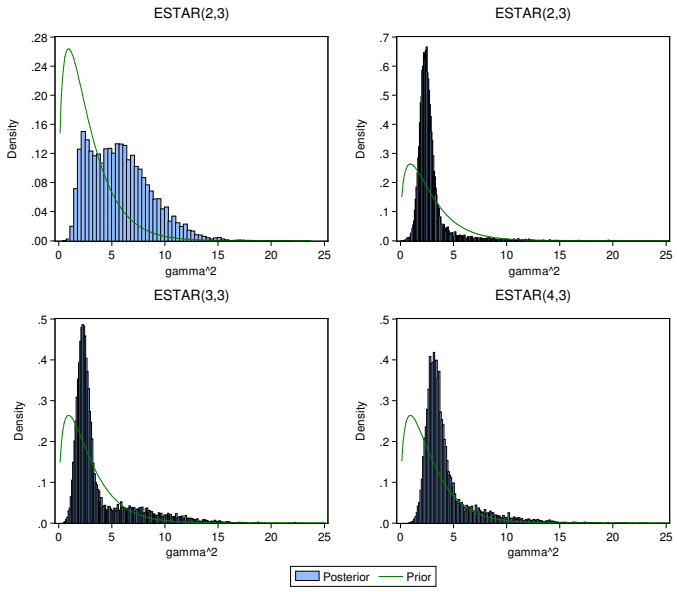


Figure 3.4: Histograms of $p(c|y)$ ICMH with looser prior for four ESTAR models

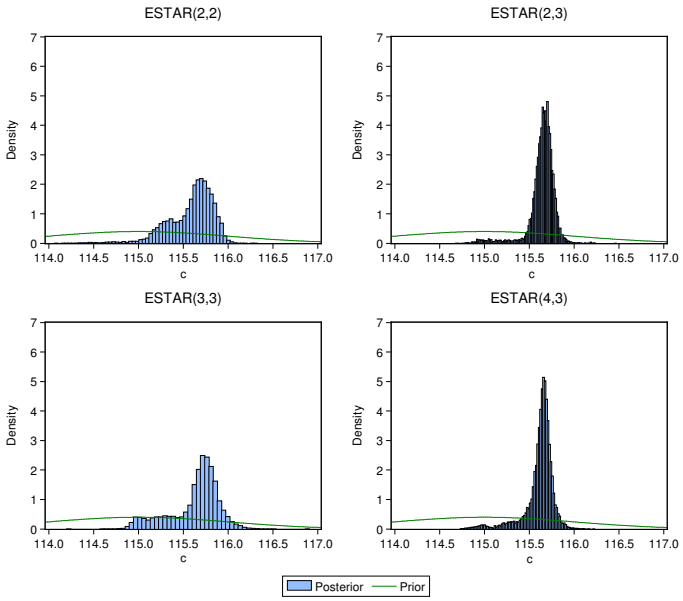


Table 3.5: Dynamic behavior of the ESTAR models for the USD/CHF real exchange rate index

	Regime	Most prominent roots	Modulus	Frequency	Period
ESTAR(2,2)	I	0.590 ± 3.39i	3.442	1.40	4.49
	O	0.964	0.964		
	Stable stationary[a] point: 96.199				
ESTAR(2,3)	I	0.008 ± 1.26i	1.263	1.56	4.02
	O	0.979	0.979		
	Stable stationary[a] point: 103.101				
ESTAR(3,3)	I	-0.848 ± 1.72i	1.913	2.03	3.09
	O	0.981	0.981		
	Stable stationary[a] point: 103.303				
ESTAR(4,3)	I	-1.025 ± 2.32i	2.532	1.99	3.16
	O	0.978	0.978		
	Stable stationary[a] point: 102.885				

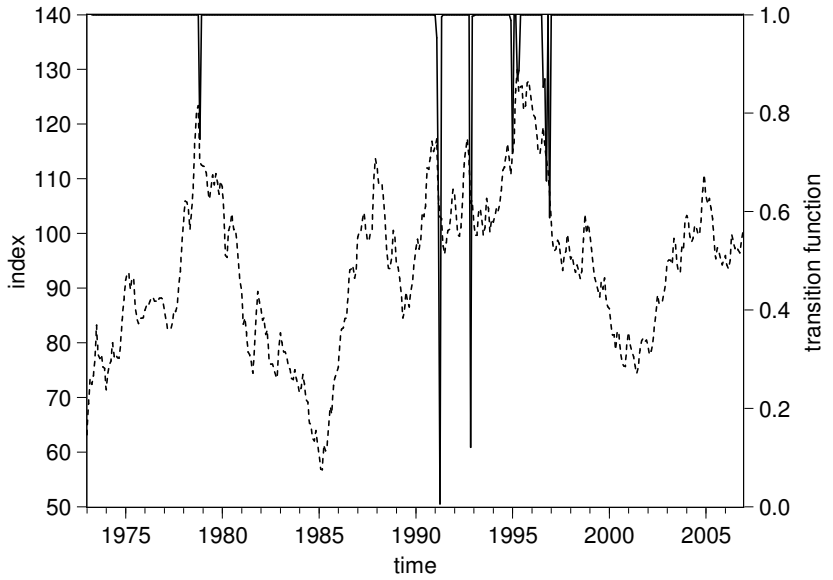
I is the inner regime, O is the outer regime; *a*: value obtained using several starting values.

not sufficiently informative. Table 3.4 shows some properties of the Markov chains obtained with looser priors. Comparing these values to Table 3.1, we notice that they are worse. Indeed, the RNE are smaller and the first lag autocorrelation of the Markov chains are higher for the parameters γ^2 and c. Moreover, Figures 3.3 and 3.4 show the posterior distributions of these parameters. These posteriors do not look like Figures 3.1 and 3.2. They indicate a poor mixing of the Markov chains. Such results are difficult to analyze and confirm the idea that the priors of the parameters of the transition function must be sufficiently informative in our case.

3.3 Dynamic behavior

As described in Sarantis [1999], it is important to investigate the dynamic properties of nonlinear models; in the present context, we want to analyze the economic dynamics of the USD/CHF real exchange rate. Indeed, to understand the behavior of the ESTAR models, we want to know, for example, if the process of the two regimes is stable or not, and why the process switches from one regime to another. To examine the dynamic behavior, we compute the characteristic roots of the extreme regimes of the ESTAR models. We define an *inner* regime when the transition function $\tilde{F}(\gamma, c; s_t) = 0$ in (3.1) and an *outer* regime when $\tilde{F}(\gamma, c; s_t) = 1$. The

Figure 3.5: USD/CHF real exchange rate index and transition function of the Bayesian ICMH ESTAR(2,3) estimation



analysis of the characteristic roots of the polynomial defined by the extreme regimes indicates if the regimes are stable, explosive or have a unit root. In contrast, the long-run properties of the models are found by simulation for $t \to \infty$ (Teräsvirta and Anderson [1992]). The underlying idea is to run the process without noise using several starting values and to observe its evolution over time.

Table 3.5 shows the most prominent roots and moduli of the regimes. The four ESTAR models exhibit cyclical movements in the inner regime with periods of about 3-4 months. Looking at the moduli, we see that these regimes are very explosive. So, when the USD/CHF real exchange rate index gets closer to the inner regime, it quickly moves into the outer regime. On the other hand, the outer regimes of the four ESTAR models are stable. This asymmetric behavior yields an overall stable process. Indeed, the long-run simulations indicate whenever the process converges towards a unique point. The stable stationary point for the four ESTAR models lies below the estimated threshold parameter. According to Michael et al. [1997] the exchange rate should converge towards the threshold parameter interpreted as the PPP of the process. Hence, this observation highlights a contradiction if we consider the threshold parameter as the PPP.

Note that the moduli of the outer regimes are not far from unity, which might explain the difficulties for rejecting the hypothesis of unit root of the series. The Bayesian approach allows to estimate the exact distribution of any function of the model parameters. A simple Monte Carlo exercise yields the distribution of the modulus of the outer regimes. The technique involves calculating the modulus for each iteration of the algorithm in order to estimate the distribution. It is then straightforward to obtain the number of values less than or equal to 1. In our case, the probability that the modulus shown in Table 3.5 is less than or equal to 1 is respectively 99.5%, 93.9%, 93.3% and 94.9%. The probability that the outer regimes have a unit root is low but not negligible.

Figure 3.5 compares, in the same graphics, the USD/CHF real exchange rate index series and the transition function of the Bayesian ICMH ESTAR(2,3) estimation. It is obvious that the series is rarely in the inner regime and never for a long time.

3.4 Misspecification tests: the posterior predictive p-values

Once the comparison between the models has been performed, misspecification tests must be applied to judge their adequacy. We choose some standard misspecification tests and some specific misspecification tests for STAR models.

In Bayesian econometrics, a way to perform those tests is to compute the posterior predictive p-values. This approach was developed by Meng [1994] and is described in Chapter 5 of Koop [2003]. The latter says: *The posterior predictive p-value approach uses the idea that, if the model is a reasonable one, the actual observed data set should be of the type which is commonly generated by the model.* One then computes the posterior predictive p-value of a statistic by evaluating the probability of obtaining a value more extreme than the actual value. First, we define a new vector of residuals $u^\dagger(y^\dagger, \varphi)$, where $y^\dagger$ is the vector of data assumed to be generated by a given model with parameter vector φ. One must distinguish $u^\dagger(y^\dagger, \varphi)$ from $u(y, \varphi)$, where y is the vector of data actually observed. Both are used in the posterior predictive p-value approach. We consider n misspecification tests. Hence, let $s_i^\dagger(y^\dagger, \varphi)$ for $i = 1, ..., n$ be the statistics of interest. The predictive distribution of the statistics $s_i^\dagger(y^\dagger, \varphi)$ can be simulated if a posterior sample from $p(\varphi|y)$ is available; and this is the case here with the Metropolis-Hastings algorithm. Hence, for each replication φ from this sample, one generates $y^\dagger$ by recursively simulating the model, computes the residuals vector $u^\dagger(y^\dagger, \varphi)$ and computes the statistics $s_i^\dagger(y^\dagger, \varphi)$.

Given the definition of the likelihood function (3.3), one recursively defines $p(y_t^\dagger|\varphi)$ as a normal density function of mean $(\phi' x_t^\dagger + \theta' x_t^\dagger F(\gamma, c; y_{t-d}^\dagger))$ and variance σ^2, where $x_t^\dagger = (1 \quad y_{t-1}^\dagger \quad ... \quad y_{t-p}^\dagger)'$. Then, one can write that:

$$u_t^\dagger \; = \; y_t^\dagger - \phi' x_t^\dagger - \theta' x_t^\dagger F(\gamma, c; y_{t-d}^\dagger). \tag{3.25}$$

So, the definition (3.25) is used to simulate the predictive distribution of the statistics $s_i^\dagger(y^\dagger, \varphi)$ and to compare them with their expected value computed with the actual data, $E[s_i(y, \varphi)|y]$. These are simply obtained using the residuals vector $u(y, \varphi)$.

In practice, the following algorithm adapted from Koop [2003] is used to find the posterior predictive p-value of the misspecification tests. We choose to run the algorithm for $S = 25'000$ iterations as the length of the Markov chains.

Step 1: Take a draw, $\varphi^{(s)}$, from the posterior sample.
Step 2: Generate a representative data set, $y^{\dagger(s)}$, recursively, from $p(y_t^{\dagger(s)}|\varphi^{(s)})$ for $t = 1, ..., T$.
Step 3: Set a residuals vector, $u^{\dagger(s)}$, of the representative data set $y^{\dagger(s)}$ using (3.25).
Step 4: Compute the statistic $s_i^{\dagger(s)}$ using the vector $u^{\dagger(s)}$.
Step 5: Set a residuals vector, $u^{(s)}$, of the actually observed data set y.
Step 6: Compute the statistic $s_i^{(s)}$ using the vector $u^{(s)}$.
Step 7: Repeat Steps 1 to 6 S times.
Step 8: Take the average of the S draws $s_i^{(s)}$ to give an estimate of $E[s_i(y, \varphi)|y]$.
Step 9: Calculate the percentage of the S draws $s_i^{\dagger(s)}$ that exceeds the estimation of $E[s_i(y, \varphi)|y]$ from Step 7; if this number is less than 0.5 then it is the estimation of the posterior predictive p-value and otherwise the posterior predictive p-value is one minus this number.

We now describe the statistics s_i used to measure the adequacy of the four ESTAR models estimated in Section 3.2:

1. The skewness statistic given by the expression $SK = \frac{\sqrt{T}\sum u_t^3}{\left(\sum u_t^2\right)^2}$ is a measure of the degree of asymmetry of the errors' distribution.

2. The excess kurtosis statistic given by the expression $EK = \frac{T\sum u_t^4}{\left(\sum u_t^2\right)^2} - 3$ is a measure of the degree of thickness of the tails of the errors' distribution.

Table 3.6: P-values of misspecification tests of the Bayesian ICMH ESTAR estimation for the USD/CHF real exchange rate index

	ESTAR(2,2)	ESTAR(2,3)	ESTAR(3,3)	ESTAR(4,3)
Skewness	0.400	0.120	0.118	0.113
Kurtosis	0.030*	0.123	0.099	0.067
Jarque-Bera	0.089	0.076	0.065	0.042*
AC(12)	0.179	0.170	0.247	0.225
ARCH(12)	0.291	0.442	0.366	0.268
Serial independence				
q=1	0.465	0.091	0.102	0.229
q=2	0.228	0.173	0.085	0.058
q=3	0.384	0.327	0.045*	0.018*
q=4	0.487	0.411	0.079	0.011*
q=5	0.355	0.484	0.124	0.021*
q=6	0.300	0.393	0.181	0.038*
q=7	0.200	0.282	0.266	0.060
q=8	0.404	0.256	0.275	0.067
q=9	0.328	0.194	0.342	0.092
q=10	0.225	0.209	0.034*	0.009**
No remaining nonlinearity				
e=1	0.448	0.063	0.091	0.032*
e=2	0.077	0.410	0.303	0.190
e=3	0.101	0.446	0.237	0.242
Parameter constancy				
LM3	0.312	0.209	0.410	0.323
LM2	0.417	0.092	0.228	0.116
LM1	0.029*	0.461	0.241	0.088

** means the p-value is under 1% and * means the p-value is under 5%.

3. The Jarque-Bera statistic, which combines skewness and kurtosis as $JB = T\left(\frac{SK^2}{6} + \frac{EK^2}{24}\right)$, is used as an indicator of error non-normality.

4. An $\mathcal{F}$ statistic for testing the nullity of autoregression coefficients in an AR(12) model of residuals. This is used to measure autocorrelation in residuals, and is noted AC(12).

5. An $\mathcal{F}$ statistic for testing the nullity of the autoregression coefficients in an AR(12) model of the squared residuals. This is used to measure conditional heteroscedasticity in residuals, and is noted ARCH(12).

6. The test of serial independence developed by Eitrheim and Teräsvirta [1996] is based on an $\mathcal{F}$ statistic for testing the hypothesis of no error autocorrelation of length q in a STAR model.

7. The test of no remaining nonlinearity developed by Eitrheim and Teräsvirta [1996] is based on an $\mathcal{F}$ statistic for testing the hypothesis of no additive STAR (with delay parameter e) component.

8. The test of parameter constancy developed by Eitrheim and Teräsvirta [1996] is based on an $\mathcal{F}$ statistic for testing the hypothesis of the constancy of the autoregressive parameters in the STAR model against three time-varying functional forms of these parameters.

The last three tests are described in detail in Section 2.6.

The estimated p-values of the misspecification tests in Table 3.6 confirm the conclusions of Section 3.2.2: no p-value is under 5% in the model ESTAR(2,3), whereas there is weak evidence of misspecification in the other models. Of course, few p-values are under 5% for the four models because they have been preselected beforehand. Moreover, these results justify the choice of the ESTAR models to describe the real exchange rate process. Finally, note that only the test of serial independence by Eitrheim and Teräsvirta [1996] enables detecting autocorrelation for the ESTAR(4,3) model. Indeed, the standard AC(12) test has a p-value of 0.225. This is an interesting point because the test of serial independence is powerful when the autoregression coefficient of the residuals of the true DGP is high and when the order of autocorrelation in the test is low (see Eitrheim and Teräsvirta [1996]). So, Table 3.7 compares the $\mathcal{F}$ test (AC(q)) with the serial independence test (SI(q)) for the models ESTAR(3,3) and ESTAR(4,3). The standard $\mathcal{F}$ test does not detect an autocorrelation problem because no p-value is under 5% for AC(q). On the other hand, several p-values of the serial independence test are under 5%, in particular for q=3 when this test is powerful. This proves that the serial independence test by Eitrheim and Teräsvirta [1996] is useful in our case.

Table 3.7: P-values of the $\mathcal{F}$ test and serial independence test of the Bayesian ICMH ESTAR estimation for the USD/CHF real exchange rate index

	ESTAR(3,3)		ESTAR(4,3)	
	AC(q)	SI(q)	AC(q)	SI(q)
q=1	0.262	0.102	0.285	0.229
q=2	0.275	0.085	0.316	0.058
q=3	0.431	0.045*	0.334	0.018*
q=4	0.467	0.079	0.359	0.011*
q=5	0.426	0.124	0.466	0.021*
q=6	0.328	0.181	0.414	0.038*
q=7	0.231	0.266	0.319	0.060
q=8	0.245	0.275	0.366	0.067
q=9	0.183	0.342	0.319	0.092
q=10	0.264	0.034*	0.176	0.009**

** means the p-value is under 1% and * means the p-value is under 5%; SI(q) is the serial independence test.

3.5 Other approach

The approach proposed by Deschamps [2008] has several advantages over existing methods. First, the automatic implementation of the MCMC algorithm: Once the tunning parameter ν in (3.15) is chosen, no more adaptation is needed. Moreover, the independence chain Metropolis-Hastings algorithm yields well-mixing chains with high acceptance rates and few autocorrelations. An alternative is the method proposed by Lopes and Salazar [2006].

Lopes and Salazar [2006] use the original exponential transition function:

$$F(\gamma, c; s_t) \;=\; 1 - \exp\Big[-\gamma(s_t - c)^2 \Big] \tag{3.26}$$

specification of the STAR model, with $\gamma > 0$. To ensure that γ is positive, they assume a Gamma prior on this parameter:

$$\gamma \;\sim\; \mathcal{G}(\underline{a_\gamma}\,,\, \underline{b_\gamma}) \tag{3.27}$$

where $\underline{a_\gamma}$ and $\underline{b_\gamma}$ are the *shape* and *rate* parameters. Both are chosen so that the moments of γ match those of γ^2 in the approach by Deschamps [2008]. In our case, these hyperparameters are $\underline{a_\gamma} = 4.5$ and $\underline{b_\gamma} = 3$. So that $E[\gamma] = 1.5$ and $V[\gamma] = 0.5$.

The other priors remain unchanged, but the new specification implies a new kernel of the full conditional posterior of the block $\vartheta = (\gamma \quad c)'$:

$$\kappa^*(\vartheta) \;=\; \exp\left[-\frac{(c-\underline{c})^2}{2\sigma_c^2} - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\Big(y_t - \phi' x_t - \theta' x_t \tilde{F}(\gamma, c; s_t)\Big)^2\right]$$

$$\cdot\, \gamma^{(\underline{a_\gamma}-1)} \cdot \underline{b_\gamma}^{\underline{a_\gamma}} \cdot \exp\Big[- \underline{b_\gamma}\gamma\Big]. \tag{3.28}$$

The algorithm proposed by Lopes and Salazar [2006] for simulating the nonstandard distribution of ϑ is based on a random walk chain Metropolis-Hastings (RWMH). For each iteration, we generate draws from the proposal densities of vector ϑ:

$$\gamma \;\sim\; \mathcal{G}(\gamma_{old}^2/\Delta_\gamma, \gamma_{old}/\Delta_\gamma) \tag{3.29}$$

$$c \;\sim\; \mathcal{N}(c_{old}, \Delta_c) \tag{3.30}$$

where γ_{old} and c_{old} are the most recently drawn parameters. The acceptance rate has the form:

$$\alpha(\vartheta_{old}, \vartheta) \;=\; min\left[\frac{p_{\mathcal{G}}(\gamma_{old}|\gamma^2/\Delta_\gamma, \gamma/\Delta_\gamma)}{p_{\mathcal{G}}(\gamma|\gamma_{old}^2/\Delta_\gamma, \gamma_{old}/\Delta_\gamma)}\,\frac{\kappa^*(\vartheta)}{\kappa^*(\vartheta_{old})}\,,\, 1\right] \tag{3.31}$$

where $p_{\mathcal{G}}$ is the probability density function of gamma. The density of c does not appear in (3.31) because of the symmetry of the normal density. The tuning parameters Δ_γ and Δ_c are chosen to obtain good acceptance rates, or more generally good chain properties. Indeed, the random walk Metropolis-Hastings algorithm has the disadvantage that MCMC chains are not necessarily well-mixing.

Below, we compare the approach by Lopes and Salazar [2006] based on the RWMH algorithm to the approach by Deschamps [2008] based on the ICMH algorithm. Such a comparison allows us to see the advantages and disadvantages of both approaches in a practical exercise. Table 3.8 shows the information criteria when the RWMH algorithm is used. The information criteria yield close values for both approaches. Moreover, Table 3.9 shows the means and standard deviations of priors and posteriors when the RWMH algorithm is used. And Table 3.10 shows the p-values of misspecification tests. Again, all these results are comparable to the previous ones, but the RNE and the first lag autocorrelation of the RWMH indicate a lower chain quality (see Table 3.11). Note that, as above, we run the algorithm for 50'000 iterations and discard the first half as burn-in.

Table 3.8: Information criteria of the Bayesian RWMH ESTAR estimation for the USD/CHF real exchange rate index

	ESTAR(2,2)	ESTAR(2,3)	ESTAR(3,3)	ESTAR(4,3)
AIC	1941.185	1927.186	1927.439	**1924.030**
BIC	1977.242	**1963.243**	1971.482	1976.048
DIC	1937.964	1926.880	1924.508	**1921.645**
$log(\hat{p}(y))$	-1017.226	**-1011.787**	-1018.867	-1024.710

$log(p(y))$ is the logarithmic marginal likelihood.

Table 3.9: Results of the Bayesian RWMH estimation of four ESTAR models for the USD/CHF real exchange rate index

	ESTAR(2,2)		ESTAR(2,3)		ESTAR(3,3)		ESTAR(4,3)	
	Prior	Posterior	Prior	Posterior	Prior	Posterior	Prior	Posterior
ϕ_0	0	1325.496	0	282.910	0	-508.124	0	59.155
	(10^3)	(571.623)	(10^3)	(147.123)	(10^3)	(632.309)	(10^3)	(679.480)
ϕ_1	0	1.180	0	0.044	0	-0.060	0	-0.737
	($10^{3/2}$)	(2.052)	($10^{3/2}$)	(0.539)	($10^{3/2}$)	(0.505)	($10^{3/2}$)	(0.848)
ϕ_2	0	-11.702	0	-1.570	0	-0.838	0	-4.218
	($10^{3/2}$)	(5.705)	($10^{3/2}$)	(1.207)	($10^{3/2}$)	(1.235)	($10^{3/2}$)	(3.275)
ϕ_3	–	–	–	–	0	6.193	0	8.189
					($10^{3/2}$)	(4.927)	($10^{3/2}$)	(6.251)
ϕ_4	–	–	–	–	–	–	0	-2.865
							($10^{3/2}$)	(2.445)
θ_0	0	-1323.143	0	-281.451	0	509.528	0	-57.604
	(10^3)	(571.553)	(10^3)	(147.115)	(10^3)	(632.324)	(10^3)	(679.492)
θ_1	0	0.113	0	1.252	0	1.377	0	2.057
	($10^{3/2}$)	(2.058)	($10^{3/2}$)	(0.542)	($10^{3/2}$)	(0.507)	($10^{3/2}$)	(0.847)
θ_2	0	11.385	0	1.260	0	0.456	0	3.822
	($10^{3/2}$)	(5.708)	($10^{3/2}$)	(1.207)	($10^{3/2}$)	(1.239)	($10^{3/2}$)	(3.276)
θ_3	–	–	–	–	0	-6.640	0	-8.092
					($10^{3/2}$)	(4.931)	($10^{3/2}$)	(6.255)
θ_4	–	–	–	–	–	–	0	2.829
							($10^{3/2}$)	(2.450)
σ^2	imp^a	6.823	imp^a	6.688	imp^a	6.640	imp^a	6.633
		(0.448)		(0.478)		(0.475)		(0.474)
γ	1.5	2.112	1.5	2.033	1.5	2.001	1.5	2.607
	($\sqrt{0.5}$)	(0.583)	($\sqrt{0.5}$)	(0.486)	($\sqrt{0.5}$)	(0.489)	($\sqrt{0.5}$)	(0.631)
c	115	115.320	115	115.649	115	115.780	115	115.677
	(1)	(0.182)	(1)	(0.090)	(1)	(0.146)	(1)	(0.102)

Mean and standard deviation (in brackets) of priors and posteriors; a: almost improper priors.

Table 3.10: P-values of misspecification tests of the Bayesian RWMH ESTAR estimation for the USD/CHF real exchange rate index

	ESTAR(2,2)	ESTAR(2,3)	ESTAR(3,3)	ESTAR(4,3)
Skewness	0.403	0.122	0.119	0.110
Kurtosis	0.030*	0.119	0.101	0.067
Jarque-Bera	0.090	0.073	0.066	0.039*
AC(12)	0.180	0.172	0.246	0.229
ARCH(12)	0.290	0.444	0.355	0.257
Serial independence				
q=1	0.464	0.091	0.102	0.235
q=2	0.229	0.174	0.083	0.059
q=3	0.380	0.331	0.043*	0.018*
q=4	0.490	0.419	0.077	0.009**
q=5	0.357	0.479	0.119	0.018*
q=6	0.297	0.390	0.179	0.037*
q=7	0.198	0.278	0.262	0.062
q=8	0.401	0.249	0.272	0.067
q=9	0.324	0.189	0.338	0.090
q=10	0.225	0.211	0.034*	0.008**
No remaining nonlinearity				
e=1	0.450	0.060	0.087	0.030*
e=2	0.080	0.414	0.291	0.193
e=3	0.103	0.442	0.233	0.241
Parameter constancy				
LM3	0.305	0.213	0.409	0.319
LM2	0.414	0.090	0.232	0.113
LM1	0.026*	0.461	0.249	0.087

** means the p-value is under 1% and * means the p-value is under 5%.

Table 3.11: Chain properties of the Bayesian RWMH ESTAR estimation for the USD/CHF real exchange rate index

	ESTAR(2,2)	ESTAR(2,3)	ESTAR(3,3)	ESTAR(4,3)
$\overline{\alpha}(\vartheta_{old}, \vartheta)$	0.488	0.524	0.497	0.508
RNE(γ)	0.072	0.117	0.072	0.099
RNE(c)	0.119	0.277	0.138	0.164
corr(γ)	0.984	0.973	0.986	0.976
corr(c)	0.876	0.830	0.922	0.931

$\overline{\alpha}(.)$ is the mean acceptance rate, RNE the *relative numerical efficiency* and corr the first lag autocorrelation of the Markov chains.

Running more iterations and selecting one draw out of five improves the RNE and reduces the first lag autocorrelation. For example (not presented in Table 3.11), when we keep one draw on five of 50'000 iterations after another burn-in phase of 50'000 iterations, the MCMC output of the ESTAR(2,3) has the values RNE(γ) = 0.184, RNE(c) = 0.588, corr(γ) = 0.888 and corr(c) = 0.467.

3.6 Forecast evaluation

In order to evaluate the predictive performance of the ESTAR model, we start by simulating the predictive densities. This is the method used in the Bayesian context. Indeed, it is more convenient to make predictions with uncertainty using the Bayesian approach because the frequentist approach involves the use of asymptotic distributions. Hence, for each posterior draw $(\beta^{(s)},\sigma^{(s)},\gamma^{2(s)},c^{(s)})$ and for $h = 1,...,H$, we generate draws $y^{(s)}_{T+h|T}$ from normal distributions with expectations $y^*_{T+h}(y^*_{T+h-1},...,y^*_{T+h-p},y^*_{T+h-d},\beta^{(s)},\sigma^{(s)},c^{(s)})$ and standard error $\sigma^{(s)}$, where y^*_{T+h-j} is the observed value y_{T+h-j} if $j \geq h$, and the previously simulated value $y^{(s)}_{T+h-j|T}$ otherwise.

The predictive performance of the ESTAR model is evaluated by comparing its out-of-sample forecasts with those of a linear AR model and with a simple random walk model. The comparison with a linear model aims at highlighting the advantages of the nonlinear ESTAR model, while the random walk model provides a benchmark in the context of exchange rates. The details of the estimation of the linear and random walk models are given in Appendix B.

3.6.1 Graphical illustrations

The obtained predictive densities are illustrated with two information sets. The author's choices are as follows. The first information set (S_1) contains observations from 1973:01 to 1999:06, and the second information set (S_2) adds observations through 2002:03. As the last observation of our sample is 2006:12, the lengths of the sets are respectively 90 and 57. The models have been re-estimated with the two information sets. Figure 3.6 shows the predictive medians compared to the observed values and surrounded by the bounds of 95% highest prediction density (HPD) intervals obtained with the information sets. If the predictive median of the ESTAR(2,3) model follows the USD/CHF real exchange rate index rather well, we also see that the HPD bounds quickly grow around it.

Figure 3.6: ESTAR(2,3) prediction intervals using information sets S_1 and S_2

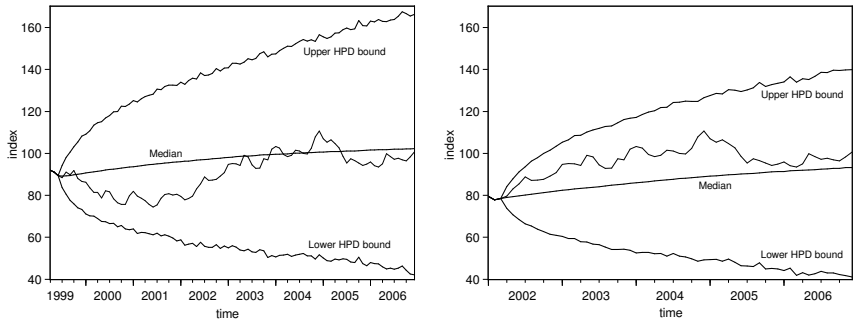


Table 3.12: Mean absolute prediction error of three models for the USD/CHF real exchange rate index

	Horizons					
	1	2	3	4	5	6
ESTAR(2,3)	1.818	2.949	3.730	4.133	4.527	4.887
AR(2)	**1.796**	2.919	3.631	**3.945**	**4.326**	**4.692**
RW	1.868	**2.872**	**3.613**	4.052	4.444	4.739

3.6.2 Forecast evaluation tests

The formal evaluations of the predictive performance of our models are analyzed with two tests. The first is a Diebold and Mariano [1995] test and the second is a West and McCracken [1998] test. Both use the median of the predictive densities at horizons $h = 1, ..., H$ for 112 information sets. All information sets begin with the observation 1973:01. The first information set ends with the observation 1997:02.

The version of the Diebold and Mariano [1995] test used in this work is a comparison of the mean absolute prediction errors (MAPE) of two models. The MAPE is defined as $\frac{1}{112}\sum_{t=1}^{112}|y_t - \hat{y}_{t,h,i}|$, where $\hat{y}_{t,h,i}$ is the median h-step forecast of y_t using model i and based on the information set available at time $t-h$. An absolute loss function of prediction errors is used, but the traditional quadratic loss function yields almost identical results. There is a simple way to calculate the statistics: the loss differential:

$$D_{t,ijh} = |y_t - \hat{y}_{t,h,i}| - |y_t - \hat{y}_{t,h,j}| \tag{3.32}$$

is regressed on a constant c_{ijt} and the variance of $\hat{c}_{ijh}$ is calculated using a heteroscedasticity and autocorrelation consistent (HAC) method, here the

Table 3.13: Diebold-Mariano tests of three models for the USD/CHF real exchange rate index

		Horizons					
		1	2	3	4	5	6
S.[a]	ESTAR(2,3)-AR(2)	1.684	1.832	1.619	1.303	1.276	1.146
	ESTAR(2,3)-RW	1.107	1.108	1.264	0.469	0.388	0.577
	AR(2)-RW	1.568	0.612	0.130	0.561	0.450	0.144
P.[b]	ESTAR(2,3)-AR(2)	0.095	0.070	0.108	0.195	0.205	0.254
	ESTAR(2,3)-RW	0.271	0.270	0.209	0.621	0.699	0.565
	AR(2)-RW	0.120	0.542	0.896	0.576	0.653	0.885

a: test statistics; *b*: test p-values.

Newey-West estimator. The null hypothesis that has the same MAPE is tested using a t statistic. So, rejecting H_0 means that the forecast accuracy of a model is better than that of others.

The second test, called *efficiency test* by West and McCracken [1998], consists in regressing the series y_t on a constant and the median h-step forecast of all models. This is repeated for each value of h. Under the null hypothesis of forecast efficiency of the model i, all regression coefficients are zero except for model i. The null is tested using an $\mathcal{F}$ statistic of $y_t = \hat{y}_{t,h,i} + \epsilon_t$ against the unrestricted model. The variance of parameters is once more obtained using the Newey-West estimator.

Table 3.12 shows the comparison of the MAPEs of the ESTAR(2,3), AR(2) and RW models for six h-step forecasts. The nonlinear model is beaten in each case by one of the simple linear models. Indeed, the AR(2) and the RW models have the smallest MAPE in four, respectively two cases. But those results have to be evaluated with the DM test because the MAPE are not necessarily significantly different from each other. Indeed, the analysis of Table 3.13 indicates that none of the DM statistics rejects the null that has the same MAPEs.

Table 3.14: Efficiency tests of three models for the USD/CHF real exchange rate index

		Horizons					
		1	2	3	4	5	6
Statistics	ESTAR(2,3)	2.059	4.908	6.352	4.986	4.199	5.054
	AR(2)	1.408	4.367	5.338	4.140	2.962	3.936
	RW	6.547	2.444	1.134	1.685	1.220	1.778
P-values	ESTAR(2,3)	0.091	0.001	0.000	0.001	0.003	0.001
	AR(2)	0.244	0.006	0.002	0.008	0.035	0.010
	RW	0.000	0.068	0.339	0.175	0.306	0.156

Moreover, Table 3.14 shows the efficiency test statistics and p-values for the ESTAR(2,3), AR(2) and RW models. The ESTAR(2,3) model rejects the hypothesis of efficiency compared to other models for all horizons except $h = 1$. In this case, ESTAR(2,3) and AR(2) are both efficient, even though the p-value for AR(2) is higher than the p-value for ESTAR(2,3). For $h = 2, ..., 6$, the simple RW model is clearly the only efficient one. All these considerations tend to prove that the ESTAR model is not the best choice for forecasting the USD/CHF real exchange rate. And this is confirmed by the fact that the same results are obtained using the ESTAR(3,3) model instead of ESTAR(2,3). Nevertheless, looking at Figure 3.5, we note that the series does not approach the inner regime inside the forecast sets. This is an important fact, because forecasting using a regime-switching model such as the STAR model does not make sense if it remains in the same regime.

3.7 Conclusion

In this chapter, we studied the ESTAR model in a Bayesian context. We demonstrated the advantages of using Bayesian rather than frequentist methods to estimate and evaluate such models. Indeed, in Chapter 2, we concluded that frequentist methods were not appropriate. The Bayesian approach compensated for the lack of information on slope parameter γ in specifying a close prior distribution of this parameter. Introducing some a priori information is essential in our case. Moreover, the Bayesian methods offered some tools such as marginal likelihood to discriminate between models, or posterior predictive p-values to judge model adequacy. Two MCMC algorithms have been compared. Obviously, the ICMH algorithm proposed by Deschamps [2008] was better than the RWMH algorithm proposed by Lopes and Salazar [2006]. Based on preliminary work, we chose to focus only

on four ESTAR models for the application to the USD/CHF real exchange rate index. The model comparison and misspecification tests allowed us to designate the ESTAR(2,3) model as the model having the best fit. Concerning predictive performance, our analysis first indicated that the efficiency test was more powerful than the DM test and second that the ESTAR(2,3) model was not the best choice for forecasting the real exchange rate. Indeed, this model did not pass the efficiency test for all horizons, certainly because the process spent most of the time in one regime only. The analysis also reveals that the serial independence test by Eitrheim and Teräsvirta [1996] was useful for detecting the problem of autocorrelation in our models.

From an economic point of view, our results indicated that the USD/CHF real exchange rate was mean-reverting and rarely reached PPP. It even seemed that it converged towards a value below PPP. This contradiction motivated further investigations by extending the ESTAR model or seeking out other models. The ESTAR model will be extended in the following chapter by adding a third regime that describes the behavior of the real exchange rate around the convergence point of the current process. Another possibility would be to consider a time-varying ESTAR model - where only the threshold parameter varies over time - but the results were not encouraging.

Chapter 4

Bayesian estimation of the Multiple Regime STAR model

This chapter is an extension of the previous chapter because we describe, estimate and evaluate the multiple regime exponential smooth transition autoregressive (MRESTAR) model in a Bayesian context. Indeed, in analyzing the dynamic behavior of ESTAR models in Section 3.3, we concluded that each of the processes converged towards a stable stationary point lying below the estimated threshold parameter. Our interpretation was that there was a third regime around the stable stationary point. We therefore ask the following question: would the introduction of a third regime improve the performances of the ESTAR model in the real exchange rate context? To answer this question, we compare the information criteria and predictive performances of the ESTAR and MRESTAR models.

The outline of this chapter is as follows. We set up the general MRSTAR model in Section 4.1. The MCMC scheme is detailed in Section 4.2. In Section 4.3, the MRESTAR model is applied to the USD/CHF real exchange rate index. In Section 4.4, we analyze the dynamic behavior of the MRESTAR models. An evaluation of the predictive performances of the MRESTAR models is given in Section 4.5. We summarize the main results in Section 4.6.

4.1 The model

The basic STAR model cannot accommodate more than two regimes, no matter which transition function is chosen. But the results of Chapter 3 indicate that a third regime could improve the ESTAR model (see Section

3.7). So, van Dijk et al. [2002] propose two ways of describing a STAR model with more than two regimes, depending on whether the regimes can be characterized by a single or by several transition variables.

The easiest way of obtaining a three-regime STAR model with a single transition variable is to add a second nonlinear component to the basic STAR model:

$$y_t = \phi' x_t + \theta' x_t F(\gamma_1, c_1; y_{t-d}) + \pi' x_t F(\gamma_2, c_2; y_{t-d}) + u_t \qquad (4.1)$$

where $x_t = (1 \quad y_{t-1} \quad ... \quad y_{t-p})'$ is a vector including lags of the dependent variable and the constant term. The terms $\phi = (\phi_0 \quad \phi_1 \quad ... \quad \phi_p)'$, $\theta = (\theta_0 \quad \theta_1 \quad ... \quad \theta_p)'$ and $\pi = (\pi_0 \quad \pi_1 \quad ... \quad \pi_p)'$ are vectors of autoregressive parameters of length $(p+1)$. The error term u_t is assumed to be iid $\mathcal{N}(0, \sigma^2)$. $F(\gamma_j, c_j; y_{t-d})$, for $j = 1, 2$, is the transition function and is assumed continuous in y_{t-d} and bounded between zero and one. The terms $F(\gamma_1, c_1; y_{t-d})$ and $F(\gamma_2, c_2; y_{t-d})$ govern the transition between two regimes and they depend on the values of a single transition variable y_{t-d}. The slope parameter γ_j measures the smoothness of transitions between regimes while the threshold parameter c_j determines the location of the transition.

Given the Multiple Regime STAR (MRSTAR)[1] model (4.1), an exponential transition function:

$$F(\gamma_j, c_j; y_{t-d}) = 1 - \exp\Big[-\gamma_j (y_{t-d} - c_j)^2 \Big] \qquad (4.2)$$

where $\gamma_j > 0$, $j = 1, 2$ and $c_1 \neq c_2$ is specified.

The three extreme regimes of this MRESTAR model are defined as follows. If y_{t-d} is equal to the c_1 value, $F(\gamma_1, c_1; y_{t-d}) = 0$ and the process is an AR model with parameters $(\phi + \pi)$, because $F(\gamma_2, c_2; y_{t-d}) = 1$; we call it the *first inner regime*. On the other hand, if y_{t-d} is equal to the c_2 value, $F(\gamma_2, c_2; y_{t-d}) = 0$ and the process is an AR model with parameters $(\phi + \theta)$, because $F(\gamma_1, c_1; y_{t-d}) = 1$; this is the *second inner regime*. Finally, if y_{t-d} is sufficiently far away from c_1 and c_2 simultaneously to imply $F(\gamma_1, c_1; y_{t-d}) = F(\gamma_2, c_2; y_{t-d}) = 1$, the process is an AR model with parameters $(\phi + \theta + \pi)$; we call it the *outer regime*. Of course, the transitions between these extreme regimes are smooth.

[1] van Dijk and Franses [1999] propose another way to describe a STAR model with more than two regimes and call it the Multiple Regime STAR (MRSTAR) model. Here, we use the name MRSTAR to talk about model (4.1) although this amounts to misusing the language.

4.2 Posterior simulator

The MCMC algorithm used in this chapter for the MRESTAR model iterates on the full conditional posteriors of four blocks: the vector:

$$\beta' = (\phi' \quad \theta' \quad \pi') \tag{4.3}$$

of length $k = (3p + 3)$, the scalar σ^2, the vectors $\vartheta_1 = (\gamma_1 \quad c_1)'$ and $\vartheta_2 = (\gamma_2 \quad c_2)'$.

4.2.1 The likelihood function

The first step towards finding the expressions of the full conditional posteriors is to specify the likelihood function of the general MRSTAR model (4.1). The independence and normality assumptions of the errors imply that we can write:

$$p(y|\beta, \sigma^2, \gamma_1, \gamma_2, c_1, c_2) \;=\; \frac{1}{(2\pi)^{T/2}} \cdot \frac{1}{\sigma^T} \cdot \exp\left[-\frac{u'u}{2\sigma^2}\right] \tag{4.4}$$

where T is the sample size and u is the vector of the u_t implied by (4.1).

4.2.2 The prior

Second, we develop the assumptions on the priors of the four blocks. We assume a multinormal prior on β with mean vector $\underline{\beta}$ and precision matrix $\underline{V}$:

$$\beta \;\sim\; \mathcal{N}(\underline{\beta}\,,\, \underline{V}^{-1}) \tag{4.5}$$

and an independent inverted Gamma prior on σ^2 with shape and scale hyperparameters $\underline{a}$ and $\underline{b}$:

$$\sigma^2 \;\sim\; \mathcal{IG}(\underline{a}\,,\, \underline{b})\,. \tag{4.6}$$

The posterior simulators of two approaches are compared in Chapter 3. In this chapter, only the approach proposed by Deschamps [2008] is adapted to the MRSTAR model. So, the exponential transition function (4.2) of the MRSTAR model (4.1) is specified differently in considering γ_j^2 instead of γ_j as the slope parameter. We then define the transition function in (4.1) as:

$$\tilde{F}(\gamma_j, c_j; y_{t-d}) \;=\; 1 - \exp\Big[-\gamma_j^2(y_{t-d} - c_j)^2\Big] \tag{4.7}$$

where $j = 1, 2$ and $c_1 \neq c_2$. Opting for independent normal priors for γ_j and c_j:

$$\gamma_j \sim \mathcal{N}(\underline{\gamma_j}, \sigma^2_{\gamma_j}) \tag{4.8}$$

$$c_j \sim \mathcal{N}(\underline{c_j}, \sigma^2_{c_j}) \tag{4.9}$$

ensures that the slope parameter is positive.

We take an almost improper prior on β with a null expectation vector, a variance of 10^6 for the parameters ϕ_0, θ_0 and π_0, and a variance of 10^3 for the other parameters. All covariances are null. Furthermore, to ensure an almost improper prior on σ^2, the hyperparameters $\underline{a}$ and $\underline{b}$ are 10^{-6}.

The posterior of the parameters γ_1 and c_1 of the first transition function of the MRSTAR model (4.1) are supposed to be close to the posteriors of the parameters γ and c of the transition function of the STAR model (3.1) in Chapter 3. Hence, the same priors are used. The hyperparameters $\underline{c_1}$, $\underline{\sigma^2_{c_1}}$, $\underline{\gamma_1}$ and $\underline{\sigma^2_{\gamma_1}}$ are respectively 115, 1, $\sqrt[4]{2}$ and $1.5 - \sqrt{2}$.

The second transition function of the MRSTAR model (4.1) is assumed to describe the behavior of the series around the convergence point of the process of the STAR model (3.1) (see Section 3.3). In a Bayesian approach, this implies that we have information on the parameter c_2. Moreover, the prior of the parameter γ_2 is supposed to be the same as γ_1. Hence, the hyperparameters $\underline{c_2}$, $\underline{\sigma^2_{c_2}}$, $\underline{\gamma_2}$ and $\underline{\sigma^2_{\gamma_2}}$ are respectively 102, 0.5, $\sqrt[4]{2}$ and $1.5-\sqrt{2}$.

4.2.3 The posterior

Finally, we develop the expressions of the full conditional posteriors of each block and then use the Gibbs sampler to generate random draws. The full conditional posteriors of β and σ^2 are easy to treat because they are analytically known and are described in Section 3.1.3.

More precisely, if γ_1, c_1, γ_2 and c_2 are known, the MRSTAR model (4.1) becomes linear and has the vector form $y = X\beta + u$, where row t of the $T \times k$ matrix X is:

$$(1 \; y_{t-1} \; ... \; y_{t-p} \; \tilde{F}_{1t} \; \tilde{F}_{1t}y_{t-1} \; ... \; \tilde{F}_{1t}y_{t-p} \; \tilde{F}_{2t} \; \tilde{F}_{2t}y_{t-1} \; ... \; \tilde{F}_{2t}y_{t-p}) \tag{4.10}$$

with $\tilde{F}_{jt} \equiv \tilde{F}(\gamma_j, c_j; y_{t-d})$, $k = (3p+3)$ is the length of β and T is the sample size. Then, the full conditional posterior of β is given by:

$$\beta|y, \sigma^2, \gamma_j, c_j \sim \mathcal{N}(\overline{\beta}, \overline{V}^{-1}) \tag{4.11}$$

with $\overline{\beta} = \overline{V}^{-1}(\frac{1}{\sigma^2}X'y + \underline{V}\underline{\beta})$ and $\overline{V} = \frac{1}{\sigma^2}X'X + \underline{V}$, and the full conditional posterior of σ^2 is given by:

$$\sigma^2|y, \beta, \gamma_j, c_j \sim \mathcal{IG}(\overline{a}, \overline{b}) \tag{4.12}$$

with $\overline{a} = \underline{a} + \frac{T}{2}$ and $\overline{b} = \underline{b} + \frac{1}{2}(y - X\beta)'(y - X\beta)$.

The full conditional posteriors in the third and the fourth blocks are non-standard. The algorithm for simulating these posteriors is then based on the independence chain Metropolis-Hastings (ICMH) using a multivariate Student distribution as a candidate generating density. If γ_2 and c_2 are known, the MRSTAR model (4.1) becomes a simple STAR model with two regimes as studied in Chapter 3, but with $(p+1)$ more autoregressive parameters. Hence, if β and σ^2 are known too, and considering the independent normal priors for γ_1 and c_1, the kernel of the full conditional posterior of the third block is:

$$\kappa^*(\vartheta_1) = \exp\left[-\frac{(\gamma_1 - \underline{\gamma_1})^2}{2\sigma^2_{\gamma_1}} - \frac{(c_1 - \underline{c_1})^2}{2\sigma^2_{c_1}} \right.$$

$$\left. -\frac{1}{2\sigma^2}\sum_{t=1}^{T}\Big(y_t - \phi' x_t - \pi' x_t \tilde{F}_{2t} - \theta' x_t \tilde{F}(\gamma_1, c_1; y_{t-d})\Big)^2 \right]. \tag{4.13}$$

As proposed by Deschamps [2008], the parameters of the candidate density of the ICMH algorithm are estimated at each iteration using a linearization of the model. In our case, this estimation is based on the following first-order Taylor expansion of the model (4.1) around (γ_1^*, c_1^*) considering the parameters ϕ, θ, π, σ^2, γ_2 and c_2 as known:

$$y_{1t}^* = \gamma_1 x_{11t}^* + c_1 x_{12t}^* + v_t$$

where:

$$y_{1t}^* = y_t - \phi' x_t - \pi' x_t \tilde{F}_{2t} - \theta' x_t \left(\tilde{F}(\gamma_1^*, c_1^*; y_{t-d}) - \frac{\partial \tilde{F}_{1t}}{\partial \gamma_1}\bigg|_{\gamma_1^*, c_1^*} \gamma_1^* - \frac{\partial \tilde{F}_{1t}}{\partial c_1}\bigg|_{\gamma_1^*, c_1^*} c_1^* \right)$$

$$x_{11t}^* = (\theta' x_t) \frac{\partial \tilde{F}_{1t}}{\partial \gamma_1}\bigg|_{\gamma_1^*, c_1^*}$$

$$x_{12t}^* = (\theta' x_t) \frac{\partial \tilde{F}_{1t}}{\partial c_1}\bigg|_{\gamma_1^*, c_1^*}.$$

The vector $\vartheta_1^* = (\gamma_1^* \quad c_1^*)'$ is an approximate solution to the Bayesian update equations:

$$\vartheta_1^* = S_{\vartheta_1}^{-1} \left[\frac{X_{1*}' y_{1*}}{\sigma^2} + \begin{pmatrix} \sigma_{\gamma_1}^2 & 0 \\ 0 & \sigma_{c_1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \underline{\gamma_1} \\ \underline{c_1} \end{pmatrix} \right] \tag{4.14}$$

$$S_{\vartheta_1} = \frac{X_{1*}' X_{1*}}{\sigma^2} + \begin{pmatrix} \sigma_{\gamma_1}^2 & 0 \\ 0 & \sigma_{c_1}^2 \end{pmatrix}^{-1} \tag{4.15}$$

where y_{1*} is the $T \times 1$ vector with elements y_{1t}^* and X_{1*} is the $T \times 2$ matrix with elements x_{11t}^* and x_{12t}^*. Further iterations on (4.14) and (4.15) are used to find the approximate solution. The vector of prior expectation is taken as a starting point. A candidate ϑ_1 is drawn from a multivariate Student density with kernel:

$$\kappa(\vartheta_1) = \left[1 + \frac{(\vartheta_1 - \vartheta_1^*)' S_{\vartheta_1} (\vartheta_1 - \vartheta_1^*)}{\nu_1} \right]^{-\frac{\nu_1 + 2}{2}} \tag{4.16}$$

and is accepted with probability:

$$\alpha(\vartheta_{1old}, \vartheta_1) = min \left[\frac{\kappa(\vartheta_{1old})}{\kappa(\vartheta_1)} \frac{\kappa^*(\vartheta_1)}{\kappa^*(\vartheta_{1old})}, 1 \right] \tag{4.17}$$

where ϑ_{1old} is the most recently drawn vector. The number ν_1 of degrees of freedom in (4.16) can be chosen by experimentation to ensure a good acceptance rate of the candidate; $\nu_1 = 3$ seems to be a good choice.

The algorithm for simulating the full conditional posteriors of the fourth block is obtained by switching the indices 1 and 2 in the algorithm above. So, if γ_1 and c_1 are known, the kernel of the full conditional posterior of the fourth block is:

$$\kappa^*(\vartheta_2) = \exp \left[-\frac{(\gamma_2 - \underline{\gamma_2})^2}{2\sigma_{\gamma_2}^2} - \frac{(c_2 - \underline{c_2})^2}{2\sigma_{c_2}^2} \right.$$

$$\left. -\frac{1}{2\sigma^2} \sum_{t=1}^{T} \left(y_t - \phi' x_t - \theta' x_t \tilde{F}_{1t} - \pi' x_t \tilde{F}(\gamma_2, c_2; y_{t-d}) \right)^2 \right]. \tag{4.18}$$

Table 4.1: Chain properties of the Bayesian ICMH MRESTAR estimations for the USD/CHF real exchange rate index

	MRESTAR(2,3)	MRESTAR(3,3)	MRESTAR(4,3)
$\overline{\alpha}(\vartheta_{1old}, \vartheta_1)$	0.730	0.745	0.736
$\overline{\alpha}(\vartheta_{1old}, \vartheta_2)$	0.704	0.738	0.723
RNE(γ_1^2)	0.414	0.351	0.254
RNE(c_1)	0.530	0.295	0.305
corr(γ_1^2)	0.666	0.707	0.803
corr(c_1)	0.495	0.795	0.794
RNE(γ_2^2)	0.467	0.512	0.425
RNE(c_2)	0.520	0.301	0.303
corr(γ_2^2)	0.536	0.556	0.641
corr(c_2)	0.623	0.752	0.798

$\overline{\alpha}(.)$ is the mean acceptance rate, RNE the *relative numerical efficiency* and corr the first lag autocorrelation of the Markov chains.

4.3 MCMC estimation

We run the algorithm for 50'000 iterations and discard the first half as burn-in. The convergence diagnostic test proposed by Geweke [1992] is used to test the convergence of the chains.

4.3.1 Markov chain properties

The analysis of the results of the MCMC algorithm starts by looking at the properties of the Markov chains. Table 4.1 shows some of these properties and indicates that the Markov chains are well-mixing. First, the mean acceptance rate of the ICMH algorithm is about 0.75 for both transition functions. This is a characteristic of the approach proposed by Deschamps [2008]: it ensures a high acceptance rate. Second, the RNE developed by Geweke [1989], which gives information on the efficiency of the MCMC algorithm, is calculated for the crucial parameters of both transition functions. The RNE of γ_1^2 and c_1 are around 0.3-0.5 for the three models, and are even higher for γ_2^2 and c_2. Third, the first lag autocorrelations of the posteriors for the parameters γ_1^2, c_1, γ_2^2 and c_2 are all under or equal to 0.8. Obviously, there is no autocorrelation problem with this algorithm.

Table 4.2: Information criteria of the Bayesian ICMH MRESTAR estimations for the USD/CHF real exchange rate index

	MRESTAR(2,3)	MRESTAR(3,3)	MRESTAR(4,3)
np	14	17	20
AIC	1929.024	1932.517	**1928.131**
BIC	**1985.113**	2000.583	2008.159
p_D	12.055	13.588	15.880
DIC	1925.135	1925.693	**1919.891**
$log(\hat{p}(y))$	**-1024.478**	-1034.647	-1043.309

np is the number of parameters of the model, p_D the effective number of parameters used in the DIC and $log(p(y))$ the logarithmic marginal likelihood.

4.3.2 Model comparison

In this section, we compare three models: the MRESTAR(2,3), MRESTAR(3,3) and MRESTAR(4,3) models. In Chapter 3, a fourth combination of parameters p and d ($p = 2$ and $d = 2$) for the ESTAR model was estimated but was clearly the worst. This result implies that the transition variable y_{t-3} is the best choice for the transition function. Hence, there is no reason to estimate the MRESTAR(2,2) model in this chapter. Then, a simple way to discriminate between the three models above is to compare some information criteria. In the Bayesian context, the likelihood function has to be evaluated at a point estimate. We choose to evaluate the likelihood function and then the information criteria, at the mean of the posterior. First, the AIC of Akaike [1974] is used as a benchmark for the comparison. Second, the BIC of Schwarz [1978] is used because it has the property of favoring parsimonious models. Both are the most used information criteria in econometrics. Third, we use a criterion recently proposed by Spiegelhalter et al. [2002], the DIC. Finally, the bridge sampling estimator of the marginal likelihood proposed by Meng and Wong [1996] is used as a Bayesian criterion. Marginal likelihood allows us to calculate the Bayes factors (BF). Then, we use the Jeffrey's scale to classify the evidence provided by the data in favor of one model against another. More details were given in Section 3.2.2.

Table 4.2 shows the comparison of the three preselected models. The MRESTAR(2,3) minimizes the BIC, whereas the MRESTAR(4,3) minimizes the AIC and DIC. The difference between the effective number of parameters p_D used in the DIC and the number of parameters np is the

highest for the MRESTAR(4,3) model. This explains why the DIC is in favor of the MRESTAR(4,3) model. Concerning the Bayesian criterion, the MRESTAR(2,3) model maximizes the logarithmic marginal likelihood. The following transformed Bayes factor proposed by Kass and Raftery [1995]: 2 x log(BF) = 2 x (–1024.478 – (–1034.647)) = 20.338 indicates a "very strong" evidence for the MRESTAR(2,3) model relative to MRESTAR(3,3). The transformed BF of MRESTAR(2,3) against MRESTAR(4,3) is 37.662. So, the Jeffrey's scale indicates a "very strong" evidence in this case. These results indicate an obvious preference for the MRESTAR(2,3) model.

The analysis of Table 4.2 and Table 3.2 allows to discriminate between the MRESTAR models estimated in this chapter and the ESTAR models estimated in Chapter 3. This comparison indicates divergent results: the ESTAR(4,3) minimizes the AIC, the ESTAR(2,3) minimizes the BIC and the MRESTAR(4,3) minimizes the DIC. Note that the marginal likelihood must not be used because all models do not have the same prior distribution. As Kass and Raftery [1995] say, the BIC gives an asymptotic approximation of the marginal likelihood. Hence, the BIC is our favorite criterion for discriminating between these models. Obviously, the BICs of the MRESTAR models are worse than the ones of the ESTAR models: the introduction of a third regime does not improve the fit of the ESTAR models.

4.3.3 Posterior analysis

Table 4.3 shows the means and standard deviations of priors and posteriors of the three preselected MRESTAR models for the USD/CHF real exchange rate index. The autoregression parameters ϕ, θ and π allow to analyze the dynamic behavior of the process of the three regimes. The parameters γ_1^2 and c_1 describe the behavior of the first transition function. The posteriors of these parameters are really close to those found in Section 3.2.3, namely in the ESTAR model with only one transition function. So, it is more interesting to look at the parameters γ_2^2 and c_2 of the second transition function. The means of the posteriors of the slope parameter γ_2^2 are around 1.5 for the three models. These values are smaller than those found for the parameter γ_1^2 even though the priors are the same. This indicates a smoother switch for the second transition function than for the first. The means of the posteriors of the threshold parameter c_2 are slightly under 102 for the three models.

Table 4.3: Results of the Bayesian ICMH estimations of the three MRESTAR models for the USD/CHF real exchange rate index

	MRESTAR(2,3)		MRESTAR(3,3)		MRESTAR(4,3)	
	Prior	Posterior	Prior	Posterior	Prior	Posterior
ϕ_0	0	371.437	0	-513.489	0	-19.557
	(10^3)	(152.362)	(10^3)	(655.293)	(10^3)	(707.397)
ϕ_1	0	0.514	0	0.431	0	-0.235
	$(10^{3/2})$	(0.706)	$(10^{3/2})$	(0.717)	$(10^{3/2})$	(0.908)
ϕ_2	0	-2.900	0	-2.1212	0	-4.945
	$(10^{3/2})$	(1.404)	$(10^{3/2})$	(1.429)	$(10^{3/2})$	(3.135)
ϕ_3	–	–	0	6.990	0	8.017
			$(10^{3/2})$	(5.226)	$(10^{3/2})$	(6.171)
ϕ_4	–	–	–	–	0	-1.677
					$(10^{3/2})$	(2.315)
θ_0	0	-287.318	0	553.313	0	-57.807
	(10^3)	(148.375)	(10^3)	(649.736)	(10^3)	(707.045)
θ_1	0	1.288	0	1.361	0	2.000
	$(10^{3/2})$	(0.549)	$(10^{3/2})$	(0.521)	$(10^{3/2})$	(0.798)
θ_2	0	1.277	0	0.493	0	3.665
	$(10^{3/2})$	(1.226)	$(10^{3/2})$	(1.218)	$(10^{3/2})$	(3.051)
θ_3	–	–	0	-6.542	0	-7.607
			$(10^{3/2})$	(5.073)	$(10^{3/2})$	(5.915)
θ_4	–	–	–	–	0	2.557
					$(10^{3/2})$	(2.209)
π_0	0	-82.688	0	-38.428	0	78.981
	(10^3)	(36.545)	(10^3)	(237.687)	(10^3)	(270.320)
π_1	0	-0.511	0	-0.487	0	-0.459
	$(10^{3/2})$	(0.450)	$(10^{3/2})$	(0.493)	$(10^{3/2})$	(0.441)
π_2	0	1.318	0	1.287	0	0.916
	$(10^{3/2})$	(0.702)	$(10^{3/2})$	(0.767)	$(10^{3/2})$	(0.778)
π_3	–	–	0	-0.426	0	-0.305
			$(10^{3/2})$	(2.256)	$(10^{3/2})$	(2.463)
π_4	–	–	–	–	0	-0.944
					$(10^{3/2})$	(0.729)
σ^2	imp^a	6.611	imp^a	6.589	imp^a	6.536
		(0.476)		(0.472)		(0.474)
γ_1^2	1.5	2.100	1.5	1.980	1.5	2.554
	$(\sqrt{0.5})$	(0.491)	$(\sqrt{0.5})$	(0.500)	$(\sqrt{0.5})$	(0.589)
c_1	115	115.648	115	115.793	115	115.685
	(1)	(0.096)	(1)	(0.155)	(1)	(0.107)
γ_2^2	1.5	1.437	1.5	1.725	1.5	1.767
	$(\sqrt{0.5})$	(0.697)	$(\sqrt{0.5})$	(0.754)	$(\sqrt{0.5})$	(0.760)
c_2	102	101.910	102	101.947	102	101.610
	(0.5)	(0.457)	(0.5)	(0.445)	(0.5)	(0.395)

Mean and standard deviation (in brackets) of priors and posteriors; *a*: almost improper priors because the hyperparameters of the Gamma are 10^{-6} but not nulls.

Table 4.4: Dynamic behavior of MRESTAR models for the USD/CHF real exchange rate index

	Regime	Most prominent roots	Modulus	Frequency	Period
MRESTAR(2,3)	I_1	0.002 ± 1.26i	1.258	1.57	4.00
	I_2	0.901 ± 0.90i	1.274	0.79	8.00
	O	0.980	0.980		
	Stable stationary[a] interval: 102.266 - 102.823				
MRESTAR(3,3)	I_1	-0.882 ± 1.75i	1.961	2.04	3.08
	I_2	0.681 ± 0.760	1.021	0.84	7.48
	O	0.980	0.980		
	Stable stationary[a] point: 102.809				
MRESTAR(4,3)	I_1	-1.019 ± 2.31i	2.522	1.99	3.16
	I_2	0.449 ± 1.01i	1.105	1.15	5.45
	O	0.975	0.975		
	Stable stationary[a] point: 103.204				

I_1 and I_2 are the first and second inner regimes, O is the outer regime.
a: value obtained using several starting values.

4.4 Dynamic behavior

The importance of analyzing the dynamic properties of a nonlinear model was shown in Section 3.3. In particular, we have seen that the estimates of ESTAR models for the USD/CHF real exchange rate index converge towards a value below the supposed PPP. Hence, the analysis of the dynamic behavior of our MRESTAR models can yield information on this convergence point. Moreover, we want to know if any of the three regimes is stable or not.

To examine dynamic behavior, we compute the characteristic roots of the three extreme regimes of the MRESTAR models. Knowing that the two transition functions cannot be simultaneously null, we define a *first inner* regime when the transition function $\tilde{F}(\gamma_1, c_1; y_{t-d}) = 0$ in (4.1), a *second inner* regime when the transition function $\tilde{F}(\gamma_2, c_2; y_{t-d}) = 0$ and an *outer* regime when $\tilde{F}(\gamma_1, c_1; y_{t-d}) = 1$ and $\tilde{F}(\gamma_2, c_2; y_{t-d}) = 1$. The analysis of the characteristic roots of the polynomial defined by the extreme regimes indicates if the regimes are stable, explosive or have a unit root. In contrast, the long-run properties of the models are found by simulations for $t \to \infty$ (see Teräsvirta and Anderson [1992]). The underlying idea is to run the process without noise using several starting values and to observe its evolution over time.

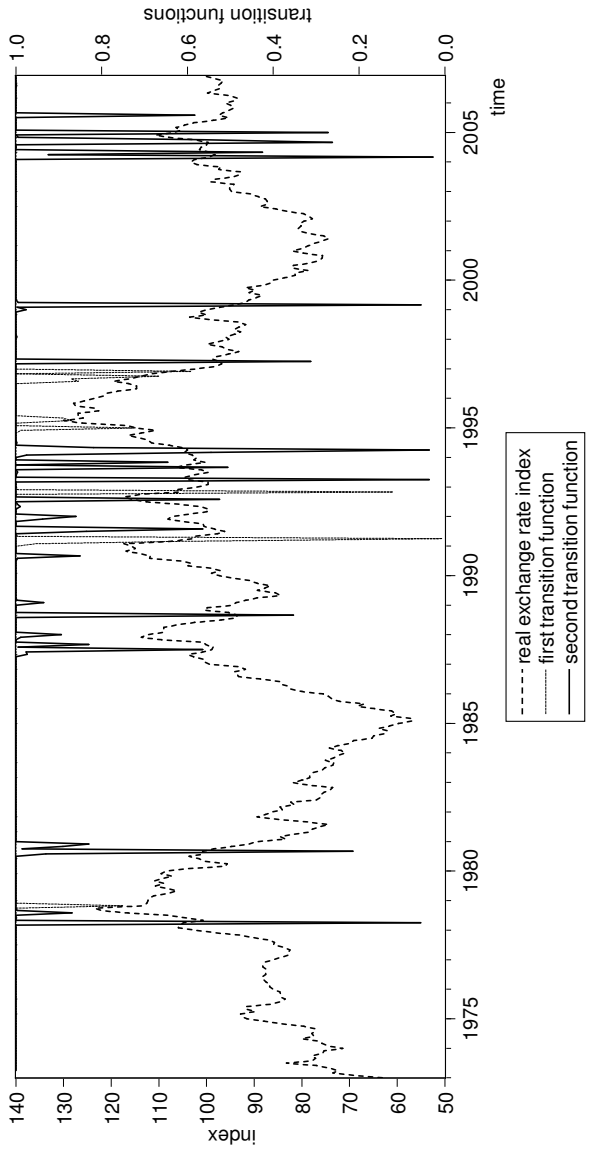


Figure 4.1: USD/CHF real exchange rate index and transition functions of the Bayesian estimation of the MRESTAR(2,3) model

Table 4.4 shows the most prominent roots and moduli of the regimes. The three MRESTAR models exhibit cyclical movements in both inner regimes with periods of about 3-4 months for the first inner regime and 5-8 months for the second one. Looking at the moduli, we see that these regimes are explosive. So, when the USD/CHF real exchange rate index gets closer to one inner regime, it quickly moves into the outer regime. On the other hand, the outer regimes of the three MRESTAR models are stable. This asymmetric behavior yields an overall stable process. Indeed, the long-run simulations indicate that the process converges towards a unique point for the MRESTAR(3,3) and MRESTAR(4,3) models, or towards an interval for MRESTAR(2,3). Specifically, the overall process of this model does not converge towards a unique point, but moves within the interval. The stable stationary point (or interval) for the three MRESTAR models lies between the two estimated threshold parameters, but very near the mean of the posteriors of c_2 (see Table 4.3). According to Michael et al. [1997], the real exchange rate should converge towards the threshold parameter interpreted as the PPP of the process. Hence, the threshold parameter of the second inner regime can be interpreted as the PPP of the USD/CHF real exchange rate index. This interpretation is questionable particularly because it provides no explanation as to the first inner regime. Theoretically, there is no reason for having two PPPs between two countries. So, the threshold parameter of the first inner regime must be something else than a PPP.

Figure 4.1 shows the USD/CHF real exchange rate index series superimposed on the two transition functions of the Bayesian estimation of the MRESTAR(2,3) model. The series is more often in the second inner regime than in the first one, but in both cases the series does not spend much time there. Hence, the series spends most of the time in the outer regime because both inner regimes are explosive.

4.5 Forecast evaluation

The simulation of the predictive densities is used to evaluate the predictive performance of the MRESTAR(2,3) model selected above. More details about the predictive densities were given in Section 3.6. The predictive performance of the MRESTAR(2,3) model is evaluated by comparing its out-of-sample forecasts with those of the ESTAR(2,3) model, the linear AR(2) model and a simple random walk (RW) model.

The performance of the MRESTAR(2,3) model is analyzed with two tests. The first is the Diebold and Mariano [1995] test and the second is the West and McCracken [1998] test. Both use the median of the predictive densities

Table 4.5: Mean absolute prediction error of four models for the USD/CHF real exchange rate index

	Horizons					
	1	2	3	4	5	6
MRESTAR(2,3)	1.810	2.914	3.701	4.182	4.642	5.094
ESTAR(2,3)	1.818	2.949	3.730	4.133	4.527	4.887
AR(2)	**1.796**	2.919	3.631	**3.945**	**4.326**	**4.692**
RW	1.868	**2.872**	**3.613**	4.052	4.444	4.739

Table 4.6: P-values of the Diebold-Mariano tests between the MRESTAR(2,3) model and three models for the USD/CHF real exchange rate index

	Horizons					
MRESTAR(2,3)	1	2	3	4	5	6
-ESTAR(2,3)	0.715	0.505	0.689	0.558	0.255	0.087
-AR(2)	0.558	0.939	0.564	0.178	0.096	0.127
-RW	0.304	0.582	0.397	0.468	0.387	0.209

at horizons $h = 1, ..., H$ for 112 information sets. All information sets begin with the observation 1973:01. The first information set ends with the observation 1997:02.

The version of the Diebold and Mariano [1995] test used in this work is a comparison of the mean absolute prediction error (MAPE) of two models. The MAPE is defined as $\frac{1}{112}\sum_{t=1}^{112}|y_t - \hat{y}_{t,h,i}|$, where $\hat{y}_{t,h,i}$ is the median h-step forecast of y_t using model i and based on the information set available at time $t-h$. A simple way to calculate the statistics: the loss differential:

$$D_{t,ijh} = |y_t - \hat{y}_{t,h,i}| - |y_t - \hat{y}_{t,h,j}| \tag{4.19}$$

is regressed on a constant c_{ijt} and the variance of $\hat{c}_{ijh}$ is calculated using a heteroscedasticity and autocorrelation consistent (HAC) method, here the Newey-West estimator. The null hypothesis that has the same MAPE is tested using a t statistic. So, rejecting H_0 means that the forecast accuracy of a model is better than others.

The second test, called the *efficiency test* by West and McCracken [1998], consists in regressing the series y_t on a constant and the median h-step forecast of all models. This is repeated for each value of h. Under the null hypothesis of forecast efficiency of the model i, all regression coefficients are zero except for model i. The null is tested using an $\mathcal{F}$ statistic of

Table 4.7: Mean squared prediction errors of four models for the USD/CHF real exchange rate index

	Horizons					
	1	2	3	4	5	6
MRESTAR(2,3)	**4.920**	12.806	20.214	26.154	31.389	37.597
ESTAR(2,3)	4.996	12.877	20.192	25.401	30.020	34.638
AR(2)	4.930	**12.516**	19.316	**24.008**	**27.873**	**31.857**
RW	5.202	12.693	**19.206**	24.267	28.762	33.115

$y_t = \hat{y}_{t,h,i} + \epsilon_t$ against the unrestricted model. The variance of parameters is one more obtained using the Newey-West estimator.

Table 4.5 shows the comparison of the MAPE of the models MRESTAR(2,3), ESTAR(2,3), AR(2) and RW for six h-step forecasts. In each case, the nonlinear model is beaten by one of the simple linear models. Indeed, the AR(2) and the RW models have the smallest MAPE in four cases, respectively two cases. But those results have to be evaluated with the DM test because the MAPE are not necessarily significantly different from each other. Indeed, the analysis of Table 4.6 indicates that none of the DM statistics rejects the null that has the same MAPE.

In our case, it is interesting to look at the traditional mean squared prediction errors (MSPE). Table 4.7 shows the comparison of the models MSPE of the MRESTAR(2,3), ESTAR(2,3), AR(2) and RW for six h-step forecasts. The MSPE is defined as $\frac{1}{112}\sum_{t=1}^{112}(y_t - \hat{y}_{t,h,i})^2$, where $\hat{y}_{t,h,i}$ is the median h-step forecast of y_t using model i and based on the information set available at time $t-h$. The MSPE of the MRESTAR(2,3) model is the minimum of the four models at horizon $h = 1$ with a value of 4.920. Nevertheless, a straightforward adaptation of the DM test of the quadratic loss function is used but not presented here, to say that this value is not significantly different of the MSPE of the other models at horizon $h = 1$.

Finally, Table 4.8 shows the efficiency test statistics and p-values for the models MRESTAR(2,3), ESTAR(2,3), AR(2) and RW. The model MRESTAR(2,3) does not reject the hypothesis of efficiency compared to other models for the horizon $h = 1$ and 2. For $h = 1$, only the MRESTAR(2,3) model is efficient. For $h = 2$, both MRESTAR(2,3) and RW are efficient. For longer horizons, both nonlinear models reject the hypothesis of efficiency. Moreover, as stated by Deschamps [2008] on forecasting of smooth transition and Markow switching autoregressive models: *studying predictions at horizons greater than one month is of somewhat marginal additional value.* This author explains this by *the*

Table 4.8: Efficiency tests of four models for the USD/CHF real exchange rate index

		Horizons					
		1	2	3	4	5	6
Statistics	MRESTAR(2,3)	1.319	2.235	4.400	3.592	3.241	3.426
	ESTAR(2,3)	3.173	4.013	5.377	3.396	3.220	4.603
	AR(2)	2.582	3.433	4.538	2.886	2.256	3.430
	RW	4.969	1.975	0.856	1.078	0.994	1.534
P-values	MRESTAR(2,3)	0.268	0.070	0.002	0.009	0.015	0.011
	ESTAR(2,3)	0.017	0.005	0.001	0.012	0.015	0.002
	AR(2)	0.041	0.011	0.002	0.026	0.068	0.011
	RW	0.001	0.104	0.493	0.371	0.414	0.198

high persistence of the process that is investigated, which leads to very high correlations between successive forecasts at horizons larger than one, resulting in loss of power. The results of the efficiency test imply that the MRESTAR(2,3) model is a good choice for constructing one-step ahead forecasts of the real USD/CHF exchange rate at the horizon at which the test is the most powerful.

4.6 Conclusion

In this chapter, we analyzed the performances of the MRESTAR model, an extension of the ESTAR model, in a Bayesian context. A comparison between the information criteria of three preselected MRESTAR models allowed us to designate the MRESTAR(2,3) model as the best one, in the sense of best fit. Nevertheless, comparing the MRESTAR models' BIC estimated in this chapter and the ESTAR models' BIC estimated in Chapter 3, we noted higher values when a third regime was assumed. According to this in-sample criterion, our extension was not an improvement of the ESTAR model. The analysis of the predictive performances of the MRESTAR(2,3) model confirmed the lack of power of the DM test. On the other hand, the efficiency test indicated that the MRESTAR(2,3) model was an efficient model at one and two-month horizons. Comparing with the ESTAR(2,2), AR(2) and RW models, the MRESTAR(2,3) model even was the only efficient model for one-step ahead forecasts. Hence, our three-regime model improved forecast accuracy.

Our investigations confirmed a theoretical economic result: the USD/CHF real exchange rate was mean-reverting and converged toward the PPP represented by the threshold parameter of the second inner regime of the MRESTAR model. On the other hand, we provided no economic interpretation of the threshold parameter of the first inner regime of this model. It remains an open issue.

Chapter 5

Bayesian estimation of the Unobserved Components model

The second question asked in this thesis is whether the PPP relationship varies over time. To answer this question, Engel and Kim [1999] proposed to decompose the real exchange rate into both a permanent and a transitory component using an unobserved components (UC) model, which can be written as a state space model. Based on the idea by Engel and Kim [1999], we discuss the specification, estimation and evaluation of several specific UC models. In particular, a comparison of in-sample performance between models is performed and predictive performances are tested.

The outline of this chapter is as follows. The features of the state space models are presented in Section 5.1. The MCMC scheme of the general UC model is detailed in Section 5.2. In Section 5.3, we set up a specific UC model with its specific MCMC scheme. We apply the UC model to the USD/CHF real exchange rate index in Section 5.4. An evaluation of the predictive performances of the UC model is given in Section 5.5. We summarize the main results in Section 5.6.

5.1 Features of the State Space models

The state space models are models commonly used with time series data. On the one hand, they offer a different way of writing existing models, such as the autoregressive moving average ARMA models, and on the other hand, they especially act as a general tool for analyzing time-varying parameters models or unobserved components models. To illustrate the features of the state space model, first, we write it in the general form proposed by Koop [2003]:

$$y_t = X_t\beta + Z_t\alpha_t + \epsilon_t \tag{5.1}$$

and

$$\alpha_{t+1} = T_t\alpha_t + u_t. \tag{5.2}$$

Equation (5.1) is called the *observation equation*. y_t is the observed series, X_t and Z_t are vectors of length k and p containing explanatory variables, and β is a vector of k parameters. The error term ϵ_t is usually assumed to be iid $\mathcal{N}(0, \sigma^2)$. Equation (5.2) is called the *state equation*. In that form, there are in fact p state equations because α_t is a vector of length p. T_t is then a $p \times p$ matrix of known constants or unknown parameters. We assume that the errors vector u_t is iid $\mathcal{N}(0, H^{-1})$. As the state equation describes the behavior of α_t, α_t is called the *state vector*. Moreover, an assumption must be made for the initial value α_1: $\alpha_1 = a_1 + u_0$ and $u_0 \sim \mathcal{N}(0, P_1)$, where a_1 and P_1 are known. In (5.1), t runs from 1 through N while in (5.2), it runs from 1 through $N - 1$.

Considering the general state space model above, if $Z_t = 0$, the equation (5.1) reduces itself to a linear regression model and if X_t contains lagged endogenous variables, the model is an autoregressive one. Moreover, if X_t is a matrix containing a column of ones and the variable y_{t-1}, if $Z_t = \rho$, $T_t = 0$, $\epsilon_t = u_t$ for all t and $p = 1$, then the observation and state equations become:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \rho\alpha_t + u_t$$

and

$$\alpha_{t+1} = u_t.$$

By rewriting the model in a single-equation form, we obtain:

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t + \rho u_{t-1}$$

which is an ARMA(1,1) model. Note that this result can easily be generalized by increasing k and p.

A second illustration of the possibilities provided by the state space model is to write the time-varying parameters model in a state space form. For example, an autoregressive model of order $p = 1$ with time-varying parameter is obtained when $X_t = 0$ and $Z_t = y_{t-1}$ in (5.1). In this case, the simplest form of the state equation (5.2) is a random walk. Indeed, if $T_t = 1$, the time-varying parameter α_{t+1} follows a random walk. So, we have the model:

$$y_t = \alpha_t y_{t-1} + \epsilon_t$$
$$\alpha_{t+1} = \alpha_t + u_t.$$

This kind of models is useful when structural changes occur over time, such as with the macroeconomic time series.

5.2 Posterior simulator of unobserved components models

This section describes the Bayesian inference of a simplification of the general state space model (5.1)-(5.2) above called the *unobserved components* (UC) model. An unobserved components model is obtained assuming $X_t = 0$ and $\sigma^2 = 0$ in the model (5.1)-(5.2). So, the observation equation (5.1) becomes an identity without any error term:

$$y_t = Z_t \alpha_t \tag{5.3}$$

and the state equation remains the same:

$$\alpha_{t+1} = T_t \alpha_t + u_t \tag{5.4}$$

where the errors vector u_t is iid $\mathcal{N}(0, H^{-1})$. The MCMC algorithm usually used is the Gibbs sampler, because the full conditional posteriors of blocks of parameters can easily be analytically calculated. Considering the case where T_t is a matrix of known constants, the algorithm iterates on the full conditional posteriors of two blocks: the $p \times p$ matrix H, and the vectors α_t for $t = 1, ..., N$ of length p each. Indeed, if α_t for $t = 1, ..., N$ was known, the state equation (5.4) would be a Seemingly Unrelated Regression (SUR) model. The Gibbs sampler involves a method for taking random draws from $p(\alpha_1, ..., \alpha_N | y, H)$ called the *simulation smoother*.

5.2.1 The likelihood function

The first step towards finding the expressions of the full conditional posteriors is to specify the likelihood function of the UC model (5.3)-(5.4). As proposed by Durbin and Koopman [2001], one can represent this model as a linear regression model:

$$y = c + \omega, \qquad \omega \sim \mathcal{N}(0, \Omega) \tag{5.5}$$

where:

$$y = (y_1, ..., y_N)'$$

$$c = Z \left[I_p, T_1, T_2 T_1, ..., \prod_{t=N-1}^{1} T_t \right]' a_1$$

$$\omega = \begin{bmatrix} Z_1 u_0 & + & 0 & + & 0 \\ \prod_{t=1}^{1} Z_2 T_t u_0 & + & Z_2 u_1 & + & 0 \\ \prod_{t=2}^{1} Z_3 T_t u_0 & + & \prod_{t=2}^{2} Z_3 T_t u_1 & + & Z_3 u_2 \\ & & \vdots & & \vdots \\ \sum_{i=0}^{N-2} \prod_{t=N-1}^{i+1} Z_N T_t u_i & & & + & Z_N u_{N-1} \end{bmatrix}$$

and where $Z = diag(Z_1, ..., Z_N)$ and Ω represents the covariance structure.

Then, the likelihood function of the model (5.3)-(5.4) can be written:

$$p(y|\alpha_1, ..., \alpha_N, T_1, ..., T_N, H) = \frac{1}{(2\pi)^{N/2}} \cdot |\Omega|^{-1/2} \cdot \exp\left[-\frac{1}{2}(y-c)'\Omega^{-1}(y-c) \right]. \tag{5.6}$$

5.2.2 The prior

Second, we develop the assumptions on the prior of the two blocks. One assumes a Wishart prior on H with the scalar degrees of freedom hyperparameter $\underline{\nu}$ and the positive define matrix $\underline{H}$:

$$p(H) = \frac{1}{c_W} \cdot |H|^{\frac{\underline{\nu}-N-1}{2}} \cdot |\underline{H}|^{-\frac{\underline{\nu}}{2}} \cdot \exp\left[-\frac{1}{2} tr(\underline{H}^{-1} H) \right] \tag{5.7}$$

where:

$$c_W = 2^{\frac{\underline{\nu} N}{2}} \cdot \pi^{\frac{N(N-1)}{4}} \cdot \prod_{i=1}^{N} G\left(\frac{\underline{\nu} + 1 - i}{2} \right).$$

For the elements of the state vector, one treats (5.4) as a hierarchical prior:

$$p(\alpha_1, ..., \alpha_T) = p(\alpha_1|H) p(\alpha_2|\alpha_1, H) ... p(\alpha_N|\alpha_{N-1}, H)$$

with:

$$p(\alpha_{t+1}|\alpha_t, H) \sim \mathcal{N}(\alpha_{t+1}|T_t\alpha_t\,,\, H) \tag{5.8}$$

where $t = 1, ..., N-1$ and:

$$p(\alpha_1|H) \sim \mathcal{N}(\alpha_1|0\,,\, H). \tag{5.9}$$

5.2.3 Seemingly Unrelated Regression

Third, the full conditional posterior of the matrix H is developed. So, considering the UC model (5.3)-(5.4), if α_t for $t = 1, ..., N$ is known, the state equation (5.4) is a Seemingly Unrelated Regressions (SUR) model. The SUR model is a special case of the simultaneous equations model whose variables are not related but whose error terms are. So, in the system of equations (5.4), the precision matrix H is not necessarily diagonal. Then, if T_t is a matrix of known constant and if the prior on H is given by the equation (5.7), the full conditional posterior of H is given by:

$$p(H|y, \alpha_1, ..., \alpha_N) = \frac{1}{c_W} \cdot \left| H \right|^{\frac{\overline{\nu}-N-1}{2}} \cdot \left| \overline{H} \right|^{-\frac{\overline{\nu}}{2}} \cdot \exp\left[-\frac{1}{2} tr(\overline{H}^{-1} H) \right] \quad (5.10)$$

with:

$$\overline{\nu} = N + \underline{\nu}$$

and:

$$\overline{H} = \left[\underline{H}^{-1} + \sum_{t=0}^{N-1} (\alpha_{t+1} - T_t \alpha_t)(\alpha_{t+1} - T_t \alpha_t)' \right]^{-1}.$$

5.2.4 Simulation smoother

Finally, the simulation smoother is described in order to take random draws from $p(\alpha_1, ..., \alpha_N|y, H)$. A simulation smoother is an algorithm for drawing samples from the conditional distribution of the states given the observations, states being the unobserved variables of the state equation. A simulation smoother widely used in many applications is described in DeJong and Shephard [1995]. Their idea is to draw the disturbances recursively, by running the *Kalman filter* and a disturbance smoother and then to construct the states from the simulated disturbances.

Later, Durbin and Koopman [2002] have developed a new simulation smoother which is, according to them, simple and computationally efficient relative to that of DeJong and Shephard [1995]. After using the two smoothers, one can say that that of Durbin and Koopman [2002] is actually simpler and more efficient in our case. First they consider the following general state space model:

$$\begin{array}{lcll} y_t & = & Z_t \alpha_t + \epsilon_t, & \epsilon_t \sim \mathcal{N}(0, H_t) \\ \alpha_{t+1} & = & T_t \alpha_t + R_t \eta_t, & \eta_t \sim \mathcal{N}(0, Q_t) \end{array} \quad (5.11)$$

where $t = 1, ..., N$, y_t is an $m \times 1$ vector of observations, α_t is a $p \times 1$ state vector and ϵ_t and η_t are vectors of disturbances. Matrices Z_t, T_t, R_t, H_t and Q_t are assumed to be known. Moreover, they assume that $\alpha_1 \sim \mathcal{N}(a_1, P_1)$, where a_1 and P_1 are known. The idea of this new simulation smoother is to draw random vectors from the conditional distribution $p(\alpha|y)$ where α is the stacking of the state vectors α_t of the model (5.11) for $t = 1, ..., N + 1$: $\alpha = (\alpha_1', ..., \alpha_{N+1}')'$. For that, the Kalman filter, a disturbance smoother and a state smoother are used. More details about the algorithm of this simulation smoother are given in Appendix C.

5.3 Two specific unobserved components models

The purpose of this section is to connect the UC models to the real exchange rate. Hence, two UC models are described in Section 5.3.1 and 5.3.2. Both are specific cases of the general UC model presented in the previous section. The first one, suggested by Engel and Kim [1999], is a model with two unobserved components: a random walk process and an autoregressive process. The second one, proposed by Kleijn and van Dijk [2001], replaces the simple random walk by an integrated random walk process of order I(2). Before describing these models, we give some economic arguments justifying our choices. As discussing the real exchange rate is tantamount to discussing PPP, it is important to study the literature on the tests of the PPP hypothesis. If PPP is valid, the real exchange rate would be stationary. So, many unit root tests have been used by researchers. At first, the tests rejected the hypothesis of real exchange rate stationarity. Then, as researchers increased the data to work on long spans of 100 years or more, the unit root tests did not reject stationarity anymore. These results should show that PPP is valid, i.e. the real exchange rate is mean-reverting. However, Engel [2000] argues, using a simulation study, that PPP is not valid for two reasons: the unit root tests have serious size biases and the stationarity tests have very low power. Therefore, this author suggests that instead of temporary deviations from a fixed target level, it is more plausible to assume that the real exchange rate returns to a target level that changes over time. The real exchange rate then exhibits a unit root component and a stationary component. The theoretical explanation of this decomposition assumes that the domestic and foreign prices that are included in the calculation of the real exchange rate are weighted averages of traded and non-traded goods. According to the theory of international economics, the price of traded goods must be stationary. So, if the real exchange rate is not stationary, this must be due to the non-stationarity of the price of non-traded goods.

From the above considerations, Engel and Kim [1999] suggested decomposing the real exchange rate between the United States dollar (USD) and the Great Britain pound (GBP) into both a permanent and a transitory component. The aim of this decomposition is to model the behavior of the real exchange rate in the long-run with the permanent component and the behavior of the real exchange rate when the series move away from the long-run component with the transitory component. They use monthly data from January 1885 to November 1995. The observation equation is given by:

$$y_t = p_t + c_t \tag{5.12}$$

where y_t is the log of the real exchange rate, p_t the permanent component and c_t the transitory component. Then, p_t follows a random walk process:

$$p_t = p_{t-1} + w_{1,t}, \qquad w_{1,t} \sim \mathcal{N}(0, \sigma^2_{1,t}) \tag{5.13}$$

and c_t is assumed to follow an AR(2) process:

$$c_t = \phi_1 c_{t-1} + \phi_2 c_{t-2} + w_{2,t}, \qquad w_{2,t} \sim \mathcal{N}(0, \sigma^2_{2,t}) \tag{5.14}$$

where $w_{2,t}$ is serially uncorrelated.

This long time series includes several exchange regimes, so Engel and Kim [1999] assume that variances $\sigma^2_{1,t}$ and $\sigma^2_{2,t}$ of the state equations are heteroskedastic and switch between various values. Specifically, the authors assume a Markov switching heteroskedasticity. More details are given in Engel and Kim [1999] and in Frühwirth-Schnatter [2001].

Engel and Kim [1999] assume that there is a permanent component in series with over one hundred years of real exchange rates, despite the fact that the literature rejects the null hypothesis of a unit root in the data. Two reasons motivate them to make this assumption. First, as discussed above, Engel [2000] has shed light on the fact that unit root tests have a large size bias and stationarity tests have very low power when a series contains both a permanent and a transitory component. Second, it is clear that structural shifts occurred during the last century. For example, the passage from a fixed exchange rate regime to a floating exchange rate regime must involve a structural shift in the data generating process. Moreover, the authors say that *the literature has not established how robust unit root tests are to changes in regimes.*

5.3.1 RW and AR(r) processes

The model developed in this section is a model with two unobserved components: a random walk process (noted RW) and an autoregressive process of order r (noted AR(r)).

$$\begin{array}{rcl} y_t & = & p_t + c_t \\ p_{t+1} & = & p_t + w_{1,t} \\ c_{t+1} & = & \phi_1 \cdot c_t + \phi_2 \cdot c_{t-1} + ... + \phi_r \cdot c_{t-r+1} + w_{2,t} \end{array} \tag{5.15}$$

where $w_{1,t}$ follows an iid $\mathcal{N}(0, \sigma_1^2)$ and $w_{2,t}$ an iid $\mathcal{N}(0, \sigma_2^2)$. As in Engel and Kim [1999], we suppose that $w_{1,t}$ and $w_{2,t}$ are independent.

The model (5.15) can be written in the general notation as follows:

$$\begin{array}{rcl} y_t & = & Z_t \alpha_t + \epsilon_t \\ \alpha_{t+1} & = & T_t \alpha_t + u_t \end{array} \tag{5.16}$$

where ϵ_t follows an iid $\mathcal{N}(0, \sigma^2)$, but where its variance is null, i.e. $\sigma^2 = 0$. Moreover, α_t is given by the following vector of length $p = r + 1$:

$$\alpha_t = \begin{bmatrix} p_t \\ c_t \\ \alpha_{3,t} \\ \vdots \\ \alpha_{r+1,t} \end{bmatrix}$$

and the following row vector Z_t of length p and the $p \times p$ matrix T_t are constant across time:

$$Z_t = Z = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \end{bmatrix}$$

$$T_t = T = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \phi_1 & 1 & & \vdots \\ \vdots & \vdots & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & & 1 \\ 0 & \phi_r & 0 & \cdots & 0 \end{bmatrix}.$$

Finally, the vector u_t of length p follows an iid $\mathcal{N}(0, H^{-1})$ and contains the values $w_{1,t}$ and $w_{2,t}$:

$$u_t = \begin{bmatrix} w_{1,t} \\ w_{2,t} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{and} \quad H^{-1} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & \cdots & 0 \\ 0 & \sigma_2^2 & & & \vdots \\ \vdots & & 0 & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}.$$

Some explanations are given here to understand the link between the expressions (5.15) and (5.16). First, it is obvious that the observation equation is identical in both expressions. Second, knowing the definition of α_t, T_t and u_t, the state equations in (5.16) can be written as follows:

$$\begin{array}{lcl} p_{t+1} & = & p_t + \omega_{1,t} \\ c_{t+1} & = & \phi_1 \cdot c_t + \alpha_{3,t} + \omega_{2,t} \\ \alpha_{3,t+1} & = & \phi_2 \cdot c_t + \alpha_{4,t} \\ \alpha_{4,t+1} & = & \phi_3 \cdot c_t + \alpha_{5,t} \\ \vdots & & \vdots \\ \alpha_{r,t+1} & = & \phi_{r-1} \cdot c_t + \alpha_{r+1,t} \\ \alpha_{r+1,t+1} & = & \phi_r \cdot c_t. \end{array}$$

The first equation above is that of (5.15). Then, by inserting the last equation above into the previous one, and so on until the equation of c_{t+1} is reached, we find the third equation of (5.15) again:

$$\begin{array}{rcl} c_{t+1} & = & \phi_1 \cdot c_t + \alpha_{3,t} + \omega_{2,t} \\ & = & \phi_1 \cdot c_t + \phi_2 \cdot c_{t-1} + \alpha_{4,t-1} + \omega_{2,t} \\ & = & \phi_1 \cdot c_t + \phi_2 \cdot c_{t-1} + \phi_3 \cdot c_{t-2} + \alpha_{5,t-2} + \omega_{2,t} \\ & & \vdots \\ & = & \phi_1 \cdot c_t + \phi_2 \cdot c_{t-1} + \phi_3 \cdot c_{t-2} + ... + \phi_r \cdot c_{t-r+1} + \omega_{2,t}. \end{array}$$

The Gibbs sampler of the model (5.16) iterates on the full conditional posteriors of three blocks: the scalars σ_1^2 and σ_2^2, the parameters $\phi_1, ..., \phi_r$ of the matrix T and the vectors α_t for $t = 1, ..., N$.

The hypothesis of independence between $\omega_{1,t}$ and $\omega_{2,t}$ and the structure of the state equations of the model (5.15) imply that we do not treat it as a SUR model. Indeed, the RW process of the permanent component p_t and the AR(r) process of the transitory component c_t are treated independently above. Therefore, we assume an inverted Gamma prior on σ_i^2 with shape and scale hyperparameters $\underline{a_i}$ and $\underline{b_i}$:

$$p(\sigma_i^2) = \underline{b_i}^{\underline{a_i}} \cdot \frac{1}{G(\underline{a_i})} \cdot \sigma_i^{-2(\underline{a_i}+1)} \cdot \exp\left[-\frac{\underline{b_i}}{\sigma_i^2}\right], \qquad \text{for } i = 1, 2. \tag{5.17}$$

In the case of the AR(r) process, defining $\phi = (\phi_1, ..., \phi_r)'$, we assume an independent multinormal prior on ϕ with mean vector $\underline{\phi}$ and precision matrix $\underline{V_\phi}$:

$$p(\phi) = \frac{1}{(2\pi)^{r/2}} \cdot \left|\underline{V_\phi}^{-1}\right|^{-1/2} \cdot \exp\left[-\frac{1}{2}(\phi - \underline{\phi})' \, \underline{V_\phi} \, (\phi - \underline{\phi})\right]. \tag{5.18}$$

The third block is composed of the state vectors α_t. To sample the state vectors, we use the simulation smoother described in Appendix C. The initialization of the simulation smoother is adapted from that of Frühwirth-Schnatter [2001]. So, our prior is $\alpha_1 \sim \mathcal{N}(a_1, P_1)$, where $a_1 = (y_1\, 0 \cdots 0)'$ and P_1 is an identity matrix multiplied by the scalar 10^4. Moreover, choices have to be made on the prior parameter values. These values introduce little prior information into the algorithm. Indeed, we suppose that $\underline{a_1} = 10^{-3}$, $\underline{b_1} = 10^{-3}$, $\underline{a_2} = 10^{-3}$, $\underline{b_2} = 10^{-3}$. For the autoregressive parameters, $\underline{\phi}$ is a null vector and $\underline{V_\phi}$ is an identity matrix multiplied by the scalar 10^{-2}.

Then, the full conditional posteriors of each block are developed. The first step of the Gibbs sampler is to run the simulation smoother to draw a sample from the full conditional posteriors of the state vectors α_t for $t = 1, ..., N, N+1$, knowing the parameters σ_i^2 and ϕ. Moreover, the full conditional posteriors of the parameters σ_i^2 are given by:

$$p(\sigma_i^2|y, \phi, \alpha_1, ..., \alpha_{N+1}) = \overline{b_i}^{\overline{a_i}} \cdot \frac{1}{G(\overline{a_i})} \cdot \sigma^{-2(\overline{a_i}+1)} \cdot \exp\left[-\frac{\overline{b_i}}{\sigma_i^2}\right] \tag{5.19}$$

with:

$$\overline{a_i} = \underline{a_i} + \frac{N}{2}, \qquad \text{for } i = 1, 2$$

$$\overline{b_1} = \underline{b_1} + \frac{1}{2}\sum_{t=1}^{N}(p_{t+1} - p_t)^2$$

and:

$$\overline{b_2} = \underline{b_2} + \frac{1}{2}\sum_{t=r}^{N}(c_{t+1} - \phi_1 \cdot c_t - \phi_2 \cdot c_{t-1} - ... - \phi_r \cdot c_{t-r+1})^2 .$$

For a better understanding of the MCMC algorithm, the equation of the transitory component of the model (5.15) is rewritten in vector form $y_c = X_c\phi + \omega_2$, where row t of the vector y_c and the matrix X_c are, respectively:

$$(c_{t+1}) \qquad \text{and} \qquad (c_t \quad c_{t-1} \quad ... \quad c_{t-r+1}).$$

Knowing the parameters σ_2^2 and $\alpha_1, ..., \alpha_N$, the full conditional posterior of the vector ϕ is given by:

$$p(\phi|y, \sigma_2^2, \alpha_1, ..., \alpha_{N+1}) = \frac{1}{(2\pi)^{r/2}} \cdot \left|\overline{V_\phi}^{-1}\right|^{-1/2} \cdot \exp\left[-\frac{1}{2}(\phi - \overline{\phi})'\,\overline{V_\phi}\,(\phi - \overline{\phi})\right] \tag{5.20}$$

with:

$$\overline{\phi} = \left(\overline{V_\phi}\right)^{-1} \left(\frac{1}{\sigma_2^2} X_c' y_c + \underline{V_\phi}\,\underline{\phi}\right)$$

and:

$$\overline{V_\phi} = \frac{1}{\sigma_2^2} X_c' X_c + \underline{V_\phi}\,.$$

As proposed by Frühwirth-Schnatter [2001], we use a technique to force the AR(r) process to be stationary. Indeed, an underlying hypothesis of the model (5.15) is that the process of the transitory component is stationary. So, within an iteration of the Gibbs sampler, we repeat the sampling of ϕ until the stationarity condition is fulfilled. The stationarity condition is verified by ensuring that the modulus of the most prominent root of the AR(r) process is inferior to one.

More details are given here about the sample size. First, the time series y contains N observations. Then, the simulation smoother provides $N+1$ state vectors α_t. Finally, for the sampling of the parameters of the AR(r) process, r values of the state vectors are used to initialize the process, then the sample size is $N+1-r$.

5.3.2 RWI(2) and AR(r) processes

With the purpose of allowing more smoothness in the RW-AR(r) model presented above, we replace the simple random walk process with an integrated random walk of order I(2). Indeed, Kleijn and van Dijk [2001] argue that the random walk component has an erratic behavior and is non-smooth. So, a random walk tends to pick up the short-term (and medium-term) dynamics, whereas the permanent component of the series should pick up only the long-term dynamics. Moreover, they argue that the estimation of the variance of an integrated random walk of order I(2) is more precise than the estimation of the variance of a simple I(1) random walk. Therefore, the model developed in this section is an unobserved components model with an integrated random walk process of order I(2) (noted RWI(2)) and an autoregressive process of order r (noted AR(r)):

$$\begin{array}{rcl} y_t & = & d_t + p_t + c_t \\ d_{t+1} & = & d_t + p_t \\ p_{t+1} & = & p_t + w_{1,t} \\ c_{t+1} & = & \phi_1 \cdot c_t + \phi_2 \cdot c_{t-1} + ... + \phi_r \cdot c_{t-r+1} + w_{2,t} \end{array} \tag{5.21}$$

where $w_{1,t}$ follows an iid $\mathcal{N}(0, \sigma_1^2)$ and $w_{2,t}$ an iid $\mathcal{N}(0, \sigma_2^2)$. As above, we suppose that $w_{1,t}$ and $w_{2,t}$ are independent.

The model (5.21) can be written in a more general notation as follows:

$$\begin{aligned} y_t &= Z_t \alpha_t + \epsilon_t \\ \alpha_{t+1} &= T_t \alpha_t + u_t \end{aligned} \tag{5.22}$$

where ϵ_t follows an iid $\mathcal{N}(0, \sigma^2)$, but where its variance is null, i.e. $\sigma^2 = 0$. Moreover, α_t is given by the following vector of length $p = r + 2$:

$$\alpha_t = \begin{bmatrix} d_t \\ p_t \\ c_t \\ \alpha_{4,t} \\ \vdots \\ \alpha_{r+2,t} \end{bmatrix}$$

and the following row vector Z_t of length p and the $p \times p$ matrix T_t are constant across time:

$$Z_t = Z = \begin{bmatrix} 1 & 1 & 1 & 0 & \cdots & 0 \end{bmatrix}$$

$$T_t = T = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & 0 & \phi_1 & 1 & & \vdots \\ \vdots & \vdots & \vdots & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & \vdots & & 1 \\ 0 & 0 & \phi_r & 0 & \cdots & 0 \end{bmatrix}.$$

Finally, the vector u_t of length p follows an iid $\mathcal{N}(0, H^{-1})$ and contains the values $w_{1,t}$ and $w_{2,t}$:

$$u_t = \begin{bmatrix} 0 \\ w_{1,t} \\ w_{2,t} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{and} \quad H^{-1} = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \sigma_1^2 & & & & \vdots \\ \vdots & & \sigma_2^2 & & & \vdots \\ \vdots & & & 0 & & \vdots \\ \vdots & & & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{bmatrix}.$$

The link between expressions (5.21) and (5.21) is detailed here. First, it is obvious that the observation equation is identical in both expressions. Second, knowing the definition of α_t, T_t and u_t, the state equations in (5.15) can be written as follows:

$$\begin{aligned}
d_{t+1} &= d_t + p_t \\
p_{t+1} &= p_t + \omega_{1,t} \\
c_{t+1} &= \phi_1 \cdot c_t + \alpha_{4,t} + \omega_{2,t} \\
\alpha_{4,t+1} &= \phi_2 \cdot c_t + \alpha_{5,t} \\
\alpha_{5,t+1} &= \phi_3 \cdot c_t + \alpha_{6,t} \\
\vdots & \quad \vdots \\
\alpha_{r+1,t+1} &= \phi_{r-1} \cdot c_t + \alpha_{r+1,t} \\
\alpha_{r+2,t+1} &= \phi_r \cdot c_t.
\end{aligned}$$

The two first equations above are those of (5.15). Then, by inserting the last equation above into the previous one, and so on until the equation of c_{t+1} is reached, we find the third equation of (5.15) again:

$$\begin{aligned}
c_{t+1} &= \phi_1 \cdot c_t + \alpha_{4,t} + \omega_{2,t} \\
&= \phi_1 \cdot c_t + \phi_2 \cdot c_{t-1} + \alpha_{5,t-1} + \omega_{2,t} \\
&= \phi_1 \cdot c_t + \phi_2 \cdot c_{t-1} + \phi_3 \cdot c_{t-2} + \alpha_{6,t-2} + \omega_{2,t} \\
&\quad \vdots \\
&= \phi_1 \cdot c_t + \phi_2 \cdot c_{t-1} + \phi_3 \cdot c_{t-2} + ... + \phi_r \cdot c_{t-r+1} + \omega_{2,t}.
\end{aligned}$$

The Gibbs sampler of the RWI(2)-AR(r) model (5.22) is similar to that of the RW-AR(r) model (5.16). The Gibbs sampler iterates on the full conditional posteriors of three blocks: the scalars σ_1^2 and σ_2^2, the parameters $\phi_1, ..., \phi_r$ of the matrix T and the vectors α_t for $t = 1, ..., N$.

Our choice of notation for the RWI(2)-AR(r) model (5.21) implies that we can use the same equations developed for the RW-AR(r) model. Therefore, the full conditional posteriors of the first two blocks are given by the equations (5.19) and (5.20), but where the state variable p_t is replaced by d_t. For the third block, the simulation smoother described in Appendix C yields a sample of the state vectors with distribution $p(\alpha_t|y, \sigma_1^2, \sigma_2^2, \phi)$ for $t = 1, ..., N, N + 1$.

5.4 MCMC estimation

We run the algorithm for 50'000 iterations and discard the first fifth as burn-in. For the posteriors distributions of the parameters, we take every 20th iteration to reduce serial correlation (see Section 5.4.1 for more details on this problem). The analysis of the posteriors is then based on

Table 5.1: Chain properties of the Bayesian estimations of Unobserved Component models for the USD/CHF real exchange rate index

	RW-AR(2)	RW-AR(3)	RWI(2)-AR(2)	RWI(2)-AR(3)
RNE(σ_1^2)	0.195	0.181	0.192	0.201
RNE(σ_2^2)	0.275	0.196	0.918	0.842
corr(σ_1^2)	0.892	0.931	0.866	0.873
corr(σ_2^2)	0.472	0.773	0.039	0.051

RNE is the *relative numerical efficiency* and corr the first lag autocorrelation of the Markov chains.

two thousand iterations. The convergence of the chains is checked using the convergence diagnostic methodology proposed by Geweke [1992] (see Chapter 3, Section 3.2).

5.4.1 Markov chain properties

We start the analysis of the results of the MCMC algorithm by looking at the properties of Markov chains. Table 5.1 shows some of these properties. A possible problem with the Bayesian estimation of the unobserved components models is the autocorrelation of the Markov chain. Indeed, within an iteration of the Gibbs sampler, there is a great dependence between the draw of the simulation smoother and the draw of the variance of the permanent component. So, the chain of the parameter σ_1^2 in our model may have too much autocorrelation, especially if its prior is not sufficiently informative. Too much autocorrelation of one of the replication series indicates a poor mixing of the chain and the whole chain cannot be used for posterior inference. A careful analysis of the Markov chain of the parameter σ_1^2, and also of the parameter σ_2^2, is then useful. Therefore, the first lag autocorrelations of the posteriors for those two parameters are calculated in Table 5.1. The values of Table 5.1 are based on the reduced chains of length 2'000. The first lag autocorrelations for σ_2^2 are under 0.8 for the four models and imply a well-mixing of those chains. On the other hand, the first lag autocorrelation for σ_1^2 are all over 0.8, even over 0.9 for the RW-AR(3) model. These values are high, but this does not affect the posteriors distributions. Indeed, the length of the Markov chains was increased, but no significant difference was noted at the time of the analysis of the posteriors.

A second interesting property of the Markov chains is the *relative numerical efficiency* developed by Geweke [1989]. It is the ratio of two measures of the standard error of the mean of a Markov chain. The term in the numerator is the naive standard error calculated ignoring the chain's autocorrelation. The term in the denominator is the numerical standard error calculated using the spectral methods. The RNE of the parameters gives information on the efficiency of the MCMC algorithm. The higher the RNE, the greater the MCMC algorithm efficiency because the MCMC chain contains little autocorrelation. The RNE of the parameter σ_1^2 are under 0.2 and the RNE of the parameters σ_2^2 are around 0.2-0.9. In conclusion, there are some autocorrelation problems with this algorithm, but they do not affect the posterior inference.

5.4.2 Model comparison

We perform the comparison of four models: the RW-AR(2), RW-AR(3), RWI(2)-AR(2) and RWI(2)-AR(3) models. Indeed, we would like to find out if an integrated random walk of order I(2) performs better than a simple random walk to model the permanent component. On the other hand, we would like to find the best length of the autoregressive process of the transitory component. A simple way to discriminate between these models is to compare some information criteria. The information criteria are all based on an evaluation of the likelihood function. In the Bayesian context, the likelihood function has to be evaluated at a point estimate. We choose to evaluate the likelihood function, and then the information criteria, at the mean of the posterior. Therefore, the means of the posterior of the parameters σ_1^2, σ_2^2 and ϕ are used.

The first and basic idea for evaluating the likelihood function of our models (5.15) and (5.21) is to solve the following general expression of our models recursively and then to find the exact likelihood function in terms of the parameters σ_1^2, σ_2^2 and ϕ:

$$\begin{aligned} y_t &= Z\alpha_t \\ \alpha_{t+1} &= T\alpha_t + u_t \end{aligned} \tag{5.23}$$

where the matrices Z and T, and the vectors α_t and u_t for $t = 1, ..., N$ are defined in Sections 5.3.1 and 5.3.2 for the models (5.15) and (5.21) respectively.

This technique yields:

$$\begin{array}{rcccccccccc}
y_1 & = & ZT\alpha_0 & + & Zu_0 & & & & & & \\
y_2 & = & ZT^2\alpha_0 & + & ZTu_0 & + & Zu_1 & & & & \\
y_3 & = & ZT^3\alpha_0 & + & ZT^2u_0 & + & ZTu_1 & + & Zu_2 & & \\
\vdots & & & & & & & & & & \\
y_N & = & ZT^N\alpha_0 & + & ZT^{N-1}u_0 & + & ZT^{N-2}u_1 & + & ... & + & Zu_{N-1}
\end{array}$$

Therefore, the general model (5.23) can be rewritten as follows:

$$y = c + \omega \qquad \omega \sim \mathcal{N}(0, \Omega), \tag{5.24}$$

where the vectors c and ω of length N are:

$$c = \begin{bmatrix} ZT \\ ZT^2 \\ \vdots \\ ZT^n \end{bmatrix} \alpha_0$$

$$\text{and} \quad \omega = \begin{bmatrix} Zu_0 & & & \\ ZTu_0 & + & Zu_1 & \\ \vdots & & & \\ ZT^{N-1}u_0 & + & ZT^{N-2}u_1 & +... + ZTu_{N-2} + Zu_{N-1} \end{bmatrix}.$$

The hypothesis of independence between error terms u_t, for $t = 0, 1, ..., N$, implies that we can define the precision matrix H of the vector u_t for the RW-AR(r) and RWI(2)-AR(r) models respectively:

$$H^{-1}_{RW-AR(r)} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & \cdots & 0 \\ 0 & \sigma_2^2 & & & \vdots \\ \vdots & & 0 & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}$$

$$\text{and} \quad H^{-1}_{RWI(2)-AR(r)} = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \sigma_1^2 & & & & \vdots \\ \vdots & & \sigma_2^2 & & & \vdots \\ \vdots & & & 0 & & \vdots \\ \vdots & & & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{bmatrix}.$$

The covariance matrix Ω is given by the following $N \times N$ symmetric matrix:

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1N} \\ \omega_{12} & \omega_{22} & \cdots & \omega_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{1N} & \omega_{2N} & \cdots & \omega_{NN} \end{bmatrix}$$

where $\omega_{11} = ZH^{-1}Z'$, $\omega_{12} = ZH^{-1}T'Z'$, $\omega_{1N} = ZH^{-1}T'^{N-1}Z'$, $\omega_{22} = Z(TH^{-1}T' + H^{-1})Z'$, $\omega_{2N} = Z(TH^{-1}T'^{N-1} + H^{-1}T'^{N-2})Z'$ and $\omega_{NN} = Z(T^{N-1}H^{-1}T'^{N-1} + ... + H^{-1})Z'$.

Finally, the likelihood function is given by:

$$p(y|\beta, \sigma_1^2, \sigma_2^2) = \frac{1}{(2\pi)^{N/2}} \cdot |\Omega|^{-1/2} \cdot \exp\left[-\frac{1}{2}(y-c)'\Omega^{-1}(y-c) \right]. \quad (5.25)$$

Evaluating the likelihood function using the latter expression is obviously not efficient. This evaluation implies, for example, the inversion of a $N \times N$ matrix. Therefore, it is more efficient to evaluate the likelihood function using the *prediction error decomposition* given by Schweppe [1965] and described by Francke et al. [2008]. This decomposition is based on the idea that:

$$p(y|\cdot) = p(y_1|\cdot) \cdot \prod_{t=2}^{N} p(y_t|Y_{t-1}, \cdot)$$

where $Y_t = (y_1, ..., y_t)'$. The Kalman filter (C.2) yields the prediction error $v_t = y_t - E[y_t|Y_{t-1}]$ and the covariance matrix $F_t = V[y_t|Y_{t-1}]$. Then, $F_t = V[v_t]$ and the equation (5.25) can be rewritten as:

Table 5.2: Information criteria of the Bayesian estimations of Unobserved Component models for the USD/CHF real exchange rate index

	RW-AR(2)	RW-AR(3)	RWI(2)-AR(2)	RWI(2)-AR(3)
np	4	5	4	5
AIC	**1980.196**	1985.441	1998.296	2003.190
BIC	**2020.286**	2035.554	2038.387	2053.302
p_D	2.886	3.778	1.893	2.573
DIC	**1977.968**	1982.997	1994.083	1998.337
$log(\hat{p}(y))$	**-19.096**	-25.210	-28.269	-34.852

np is the number of parameters of the model, p_D the effective number of parameters used in the DIC and $log(p(y))$ the logarithmic marginal likelihood.

$$p(y|\cdot) = \frac{1}{(2\pi)^{N/2}} \cdot \Big(\prod_{t=1}^{N} |F_t| \Big)^{-\frac{1}{2}} \cdot \exp \left[-\frac{1}{2} \sum_{t=1}^{N} v_t' F_t^{-1} v_t \right].$$

Table 5.2 shows the four information criteria, already used in previous chapters of this thesis, to perform the model comparison. First, the AIC by Akaike [1974] is used as a benchmark for the comparison. Second, the BIC by Schwarz [1978] is used because it has the property of favoring parsimonious models. Third, we use the DIC proposed by Spiegelhalter et al. [2002]. This criterion was developed to compare models in which the number of parameters is not clearly defined. Finally, a commonly used Bayesian criterion for discriminating between models is the marginal likelihood $p(y) = \int f(y|\theta)p(\theta)d\theta$, where $f(y|\theta)$ is the likelihood and $p(\theta)$ the prior of the model. In what follows, we estimate the marginal likelihood using the *optimal bridge sampling estimator* $\hat{p}_{BS}(y)$ (3.21) based on the bridge function $\alpha(\theta)$ (3.24) proposed by Meng and Wong [1996]. It is an iterative procedure for which we choose to take 10'000 iterations and take the *importance sampling* $\hat{p}_{IS}(y)$ (3.22) estimator as starting value. Moreover, we follow the choice of Deschamps [2008] to define the importance density $q(\theta)$: it has the same parametric form as the prior, but with moments that match the empirical posterior moments obtained by MCMC. More details about the four information criteria are given in Section 3.2.2.

The comparison of the four information criteria presented in Table 5.2 could not be more obvious: the RW-AR(2) model minimizes the AIC, the BIC and the DIC, and maximizes the logarithmic marginal likelihood. This unanimity implies that the RW-AR(2) performs better than the other using in-sample criteria. Specifically, on the one hand, Table 5.2 indicates that a simple random walk process performs better than an integrated random walk of order I(2) for modeling the permanent component. On the other hand, the best length of the stationary autoregressive process for modeling the transitory component is two.

5.4.3 Estimation results

Table 5.3 shows the means and standard deviations of priors and posteriors of the four Unobserved Component models for the USD/CHF real exchange rate index. The parameters of interest are the variances σ_1^2 and σ_2^2. Indeed, we would like to compare the weight of the permanent component with that of the transitory component. The means of the posteriors of σ_1^2 are 0.337 for the RW-AR(2) model and 0.553 for the RW-AR(3). These values are almost a hundred times smaller for the RWI(2)-AR(r) models: 0.005 for the RWI(2)-AR(2) and 0.006 for the RWI(2)-AR(3). This great difference is due to the structure of the permanent component. We now consider the means of the posteriors of σ_2^2. Despite the fact that those values are smaller than these of σ_1^2 for the RW-AR(r) models, we can say that they are relatively close for the four models and that they are obviously higher than the means of the posteriors of σ_1^2. So, the movement of the USD/CHF real exchange rate is principally dictated by the transitory component. Moreover, the means and standard deviations of the distributions of the most prominent root of the autoregressive process are shown in Table 5.3. Indeed, within each iteration, the most prominent root is calculated. For the four models, the mean of the most prominent root is clearly inferior to one. Of course, these values cannot exceed one because of the stationarity condition introduced in the MCMC algorithm.

As we are particularly interested in comparing the representation of the permanent component, the analysis is completed by Figure 5.1 and Figure 5.2. They plot the posterior distributions of the parameters of the Bayesian estimation of the RW-AR(2) and RWI(2)-AR(2) models. The histograms of both parameters σ_1^2 do not have a Gaussian form.

An important byproduct of the MCMC algorithm used in this chapter is the sample of the smoothed components. So, the means of all draws of the unknown components p_t, d_t and c_t for $t = 1, ..., N, N + 1$ are computed and yield the Bayesian equivalent of the state smoother. Figure 5.3 and

Figure 5.1: Histograms of $p(\sigma_1^2|y)$, $p(\sigma_2^2|y)$, $p(\phi_1|y)$ and $p(\phi_2|y)$ of the RW-AR(2) model

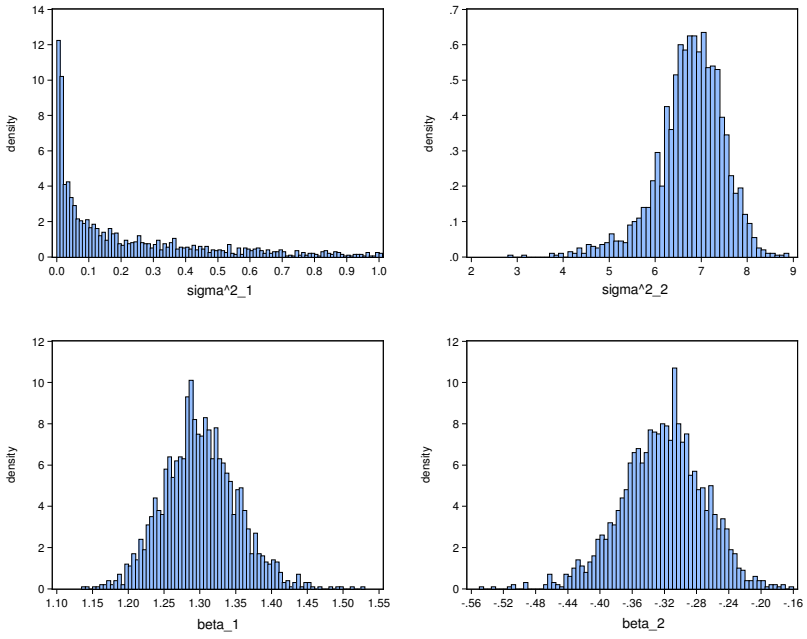


Figure 5.2: Histograms of $p(\sigma_1^2|y)$, $p(\sigma_2^2|y)$, $p(\phi_1|y)$ and $p(\phi_2|y)$ of the RWI(2)-AR(2) model

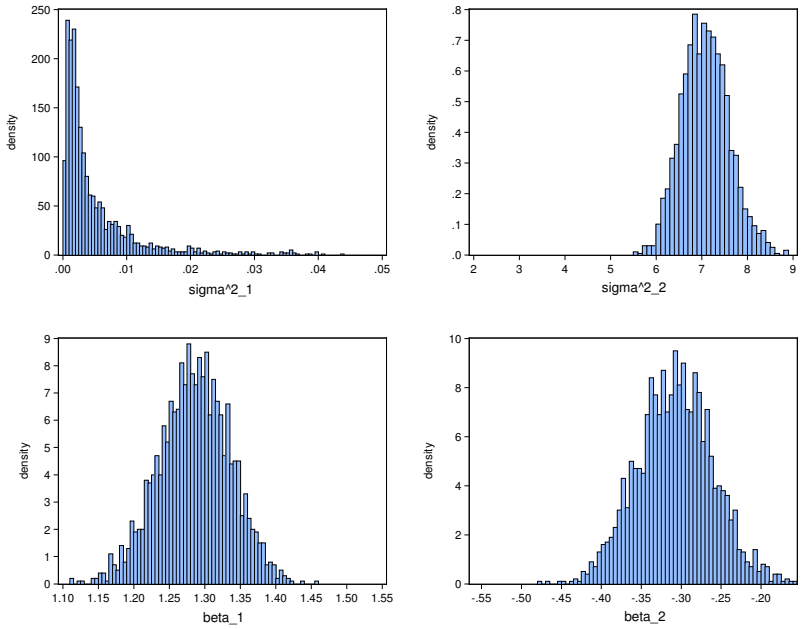


Figure 5.3: Posterior means of p_t and c_t of the RW-AR(2) model

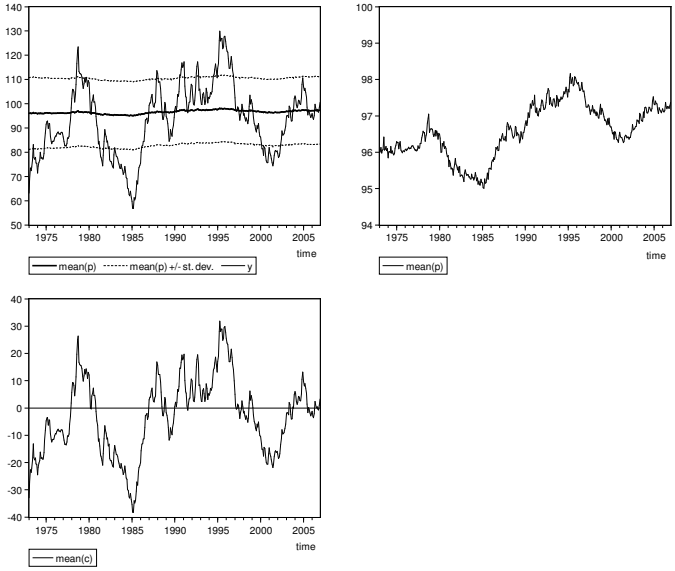


Figure 5.4: Posterior means of p_t, d_t and c_t of the RWI(2)-AR(2) model

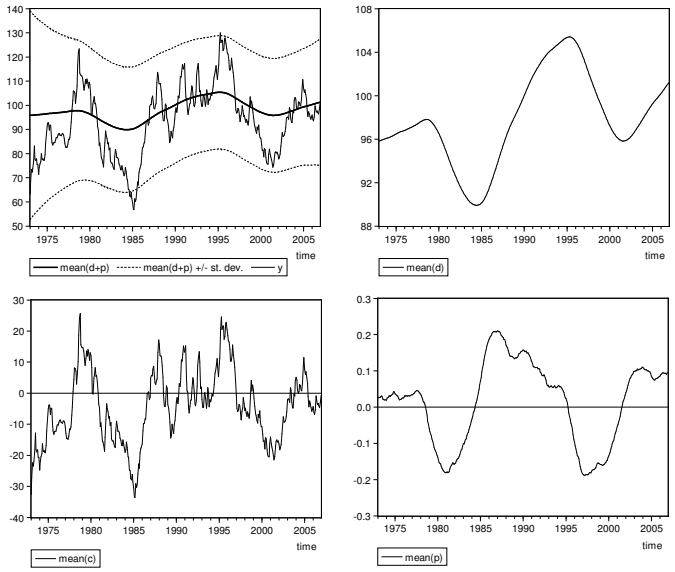


Table 5.3: Results of the Bayesian estimations of Unobserved Component models for the USD/CHF real exchange rate index

	RW-AR(2)		RW-AR(3)		RWI(2)-AR(2)		RWI(2)-AR(3)	
	Prior	Posterior	Prior	Posterior	Prior	Posterior	Prior	Posterior
σ_1^2	imp^a	0.337	imp^a	0.553	imp^a	0.005	imp^a	0.006
		(0.504)		(0.935)		(0.007)		(0.010)
σ_2^2	imp^a	6.762	imp^a	6.469	imp^a	7.078	imp^a	7.057
		(0.731)		(1.168)		(0.519)		(0.512)
ϕ_1	0	1.301	0	1.352	0	1.286	0	1.312
	(10)	(0.051)	(10)	(0.092)	(10)	(0.050)	(10)	(0.052)
ϕ_2	0	-0.322	0	-0.480	0	-0.308	0	-0.426
	(10)	(0.051)	(10)	(0.141)	(10)	(0.047)	(10)	(0.080)
ϕ_3	–	–	0	0.109	–	–	0	0.096
			(10)	(0.065)			(10)	(0.050)
root^b		0.967		0.974		0.967		0.975
		(0.017)		(0.015)		(0.027)		(0.022)

Mean and standard deviation (in brackets) of priors and posteriors; *a*: almost improper priors because the hyperparameters of the Gamma are 10^{-3} but not nulls; b: most prominent root.

Figure 5.4 show the smoothed values of the unknown components of the RW-AR(2) and RWI(2)-AR(2) models. The upper-left part of these figures compares the trend of the models, i.e. the smoothed values of the permanent component, with the USD/CHF real exchange rate index series. The trends of both models follow the same movements, but the one of the RWI(2)-AR(2) model is wider and smoother. Recall that the trends are interpreted as PPP in our unobserved components models. Hence, PPP increases over time with a *wave* behavior. That is particularly obvious with the RWI(2)-AR(2) model as its trend pick up only the long-term dynamics. Finally, the smoothed values of the transitory components of both models are almost the same.

5.5 Forecast evaluation

The purpose of this section is to evaluate the predictive performance of the unobserved components models described above. In Bayesian context, the uncertainty on future unobserved data is captured by predictive densities. As for other models, the predictive densities of the unobserved components models can be obtained by simulations. Given the simplified general notation of our unobserved components models (5.16) and (5.22):

$$\begin{aligned} y_t &= Z\alpha_t \\ \alpha_{t+1} &= T\alpha_t + u_t \end{aligned}$$

where u_t is iid $\mathcal{N}(0, H^{-1})$, for each posterior draw $\left(H^{(s)}, T^{(s)}, \alpha_N^{(s)}\right)$ and for $h = 1, ..., 6$, we generate draws of the vector $\alpha_{N+h|N}^{(s)}$ from normal distributions with expectations $\alpha_{N+h}^*(T^{(s)}, \alpha_{N+h-1}^*)$ and precision matrix $H^{(s)}$, where α_{N+h-1}^* is the draw $\alpha_N^{(s)}$ if $h = 1$, and otherwise the previously simulated value $\alpha_{N+h-1|N}^{(s)}$. The draws $y_{N+h|N}^{(s)}$ are given by the following adaptation of the observation equation: $y_{N+h|N}^{(s)} = Z\alpha_{N+h|N}^{(s)}$.

For example, the predictive densities of the RW-AR(1) model (5.15) are obtained as follows: within one iteration s, we initialize p_N^* and c_N^* by drawing $p_N^{(s)}$ and $c_N^{(s)}$ respectively from their posterior distribution. Then, p_{N+h}^* is drawn from the distribution $\mathcal{N}(p_{N+h-1}^*, \sigma_1^{2(s)})$ and c_{N+h}^* from $\mathcal{N}(\phi_1^{(s)} \cdot c_{N+h-1}^*, \sigma_2^{2(s)})$, for $h = 1, ..., 6$. Finally, $y_{N+h|N}^{(s)}$ is computed by adding up p_{N+h}^* and c_{N+h}^*. Adapting this algorithm to any unobserved components model is easily done.

Moreover, the state space models, and hence the unobserved components models, have the advantage of providing the one-step ahead predictive

Table 5.4: Efficiency tests of four models for the USD/CHF real exchange rate index

		Horizons					
		1^a	2	3	4	5	6
Statistics	RW-AR(2)	2.058	2.359	5.118	3.746	3.210	3.406
	RWI(2)-AR(2)	3.992	4.333	7.707	5.850	5.281	6.311
	MRESTAR(2,3)	1.279	1.881	5.021	3.810	3.711	4.153
	ESTAR(2,3)	2.307	2.625	6.141	3.935	3.527	3.824
	AR(2)	1.797	2.183	5.171	3.323	2.653	2.751
	RW	3.821	3.594	1.856	2.126	1.789	2.108
P-values	RW-AR(2)	0.064	0.035	0.000	0.002	0.006	0.004
	RWI(2)-AR(2)	0.001	0.001	0.000	0.000	0.000	0.000
	MRESTAR(2,3)	0.273	0.091	0.000	0.002	0.002	0.001
	ESTAR(2,3)	0.039	0.021	0.000	0.001	0.003	0.002
	AR(2)	0.107	0.050	0.000	0.005	0.019	0.016
	RW	0.002	0.003	0.095	0.056	0.108	0.058

a: Analytical calculations for the UC models.

density analytically. Indeed, the full condition posteriors of the state vectors α_t are available until $t = N + 1$. Thus, the MCMC algorithm provides the posterior distribution $p(\alpha_{N+1}|y)$. Multiplying each draw of this posterior by the matrix Z, we obtain the draws $y^{(s)}_{N+1|N}$. These draws is used as the predictive density at horizon $h = 1$, $y^{(s)}_{N+1|N}$. Note that, in our case, the analytical version of the predictive density at horizon $h = 1$ yields more accurate results than the simulated one.

The predictive performance of the RW-AR(2) and RWI(2)-AR(2) models is evaluated by comparing their out-of-sample forecasts with those of the MRESTAR(2,3), ESTAR(2,3), AR(2) and simple RW models. The details of the Bayesian estimation of AR(2) and RW are given in Appendix B.

The predictive performance of our models is solely analyzed with the *efficiency test* by West and McCracken [1998]. This test was described in Section 3.6.2 and uses, for 112 information sets, the median predictive densities at horizons $h = 2, ..., 6$ and the median function of the posterior distribution $p(Z\alpha_{N+1}|y)$ instead of the predictive density at horizon $h = 1$. All information sets begin with the observation 1973:01. The first information set ends with the observation 1997:02 and the last one ends with the observation 2006:12. Note that the DM test by Diebold and Mariano [1995] was used and analyzed in the previous chapters, but is not used in this chapter because of its lack of power.

Table 5.5: BIC of the Bayesian estimates of a selection of models for the USD/CHF real exchange rate index

Models	BIC
ESTAR(2,2)	1977.233
ESTAR(2,3)	**1963.219**
ESTAR(3,3)	1971.529
ESTAR(4,3)	1976.074
MRESTAR(2,3)	1985.113
MRESTAR(3,3)	2000.583
MRESTAR(4,3)	2008.159
RW-AR(2)	2020.286
RW-AR(3)	2035.554
RWI(2)-AR(2)	2038.387
RWI(2)-AR(3)	2053.302

Table 5.4 shows the efficiency test statistics and p-values for the RW-AR(2), RWI(2)-AR(2), MRESTAR(2,3), ESTAR(2,3), AR(2) and RW models. The UC models are then compared to a selection of the models estimated in the previous chapters. The RWI(2)-AR(2) obviously rejects the hypothesis of efficiency for all horizons. On the other hand, the RW-AR(2) rejects the null for all horizons, except $h = 1$. In this case, the RW-AR(2) is efficient because the p-value of the efficiency test is 0.064. This model is not the only efficient model for $h = 1$, the MRESTAR(2,3) and AR(2) models are efficient too. At horizon $h = 2$, the MRESTAR(2,3) and AR(2) models are the only efficient ones because their p-values are 0.091 and 0.050, respectively. For $h = 3, 4, 5, 6$, all models reject the hypothesis of efficiency except the RW model. Note that the ESTAR(2,3) model rejects the null for all horizons. Hence, this model may be discarded for forecasting the USD/CHF real exchange rate. These considerations tend to prove that a particular UC model is as efficient as a three-regime ESTAR and a linear autoregressive model for one-step ahead forecasts.

5.6 Conclusion

In this chapter, we analyzed a particular type of state space model from a Bayesian point of view: the UC model. Our UC models contained both a permanent and a transitory component. The permanent component was described by two kinds of random walk processes: a simple random walk and an integrated random walk of order I(2). The transitory component was described by autoregressive processes of order r. Our results pointed to the fact that the movements of the USD/CHF real exchange rate were

principally dictated by the transitory component. Using in-sample criteria, the model comparison indicated that the RW-AR(2) performed better than the other UC models analyzed in this chapter.

On the other hand, a comparison between the information criteria of the estimates of the ESTAR, MRESTAR and UC models allows us to evaluate the in-sample fit of our UC models. Note that the marginal likelihood must not be used because all models do not have the same prior distribution. Moreover, between AIC, BIC and DIC, we choose to use BIC because it gives an asymptotic approximation of marginal likelihood (see Kass and Raftery [1995]). Hence, Table 5.5 summarizes the BIC values shown in Tables 3.2, 4.2 and 5.2. Obviously, the UC models perform worse than the MRESTAR models which in turn perform worse than the ESTAR models. This result confirms the well-known idea that BIC favors the more parsimonious models.

Concerning predictive performance, we showed that the RW-AR(2) model was as efficient as the MRESTAR(2,3) and AR(2) models for one-step ahead forecasts. Finally, the results of this chapter indicated that the USD/CHF real exchange rate could be successfully decomposed into a long-run component and a transitory component. The long-run component, interpreted as PPP, is characterized by a stochastic trend. Hence, PPP may vary over time and may have its own dynamics. These results are very encouraging and lead us to investigate this way of modeling the real exchange rate, i.e. the decomposition into a permanent component and a transitory component. So, the next chapter will extend the UC model in describing the transitory component with an ESTAR process instead of an AR process.

Chapter 6

Bayesian estimation of the Unobserved RW-STAR Components model

This chapter is an extension of the previous one. Indeed, we describe, estimate and evaluate the unobserved random walk and exponential smooth transition autoregressive components (RW-ESTAR) model in a Bayesian context. This is an extension because the AR process of the transition component of the UC models described in Chapter 5 is replaced by an ESTAR process. Hence, our interpretation of this new link between UC and ESTAR models is as follows: the PPP relationship varies over time following the permanent component when the transitory component is a two-regime process. We would like to know whether, in the real exchange rate context, this extension improves the performances of the models estimated in previous chapters. To answer this question, we compare the information criteria and predictive performances of all models estimated in this thesis, including the RW-ESTAR.

The outline of this chapter is as follows. We set up the model in Section 6.1. The MCMC scheme is detailed in Section 6.2. In Section 6.3, the RW-ESTAR model is applied to the USD/CHF real exchange rate index: the Empirical Bayes method, a model comparison and an analysis of the dynamic behavior are performed for the MCMC estimations. An evaluation of the predictive performances of the RW-ESTAR model is given in Section 6.4 and the main results are summarized in Section 6.5.

6.1 The model

The unobserved RW-STAR components model is an unobserved components model with a random walk (RW) process for the permanent component and a smooth transition autoregressive (STAR) process for the transitory component. Remember that the STAR model is a nonlinear model where the transition between two regimes is given by a continuous transition function bounded between zero and one. In this chapter we focus on those specifications (i) which possess an exponential transition function characterized by the d^{th} lag of the unobserved component and (ii) for which both regimes are governed by an autoregressive process of length r (AR(r)). Hence, our transitory component is an ESTAR(r,d) process.

The unobserved components model with a RW and an ESTAR(r,d) process is given by:

$$\begin{array}{rcl} y_t & = & p_t + c_t \\ p_{t+1} & = & p_t + w_{1,t} \\ c_{t+1} & = & (\phi_1 \cdot c_t + \phi_2 \cdot c_{t-1} + ... + \phi_r \cdot c_{t-r+1}) + \\ & & (\theta_1 \cdot c_t + \theta_2 \cdot c_{t-1} + ... + \theta_r \cdot c_{t-r+1}) \cdot \\ & & \Big(1 - \exp\Big[- \gamma^2 \cdot (c_{t-d+1} - 0)^2 \Big]\Big) + w_{2,t} \end{array} \tag{6.1}$$

where $w_{1,t}$ follows an iid $\mathcal{N}(0, \sigma_1^2)$ and $w_{2,t}$ an iid $\mathcal{N}(0, \sigma_2^2)$. We suppose that $w_{1,t}$ and $w_{2,t}$ are independent. The particular choice of the exponential transition function implies that the transitory component possesses an inner and an outer regime. Thus, when the variable c_{t-d+1} approaches zero, the transitory component is in the inner regime and the series y_t is close to the long-run trend, i.e. the permanent component p_t. When the variable c_{t-d+1} moves away from zero, the transitory component is in the outer regime and the series y_t is far from its long-run trend, because the transitory component is far from zero. The slope parameter γ^2 measures the smoothness of the transitions between the two regimes. The lower the slope parameter, the smoother the switch between the two regimes. We choose the squared form γ^2 as in Deschamps [2008] to simplify the MCMC algorithm.

6.2 Posterior simulator

This section describes the posterior simulator of a particular unobserved components model: the two unobserved components model with a RW process and an ESTAR(2,1) process:

$$\begin{array}{rcl} y_t & = & p_t + c_t \\ p_{t+1} & = & p_t + w_{1,t} \\ c_{t+1} & = & (\phi_1 \cdot c_t + \phi_2 \cdot c_{t-1}) \\ & & + (\theta_1 \cdot c_t + \theta_2 \cdot c_{t-1}) \cdot \Big(1 - \exp\Big[-\gamma^2 \cdot c_t^2\Big]\Big) + w_{2,t} \end{array} \tag{6.2}$$

where $w_{1,t}$ follows an iid $\mathcal{N}(0, \sigma_1^2)$ and $w_{2,t}$ an iid $\mathcal{N}(0, \sigma_2^2)$ and we assume that the latter two terms are independent from each other. The model (6.2) is a state space model, so it can be written in the following general notation to simplify the use of the MCMC algorithm:

$$\begin{array}{rcl} y_t & = & Z_t \alpha_t + \epsilon_t \\ \alpha_{t+1} & = & T(\alpha_t) + u_t \end{array} \tag{6.3}$$

where ϵ_t follows an iid $\mathcal{N}(0, \sigma^2)$, but where its variance is null, i.e. $\sigma^2 = 0$. The state α_t is the following vector of length $p = 3$:

$$\alpha_t = \begin{bmatrix} p_t \\ c_t \\ \alpha_{3,t} \end{bmatrix}$$

and the state equation is a nonlinear function where:

$$T(\alpha_t) = \begin{bmatrix} p_t \\ (\phi_1 \cdot c_t + \phi_2 \cdot \alpha_{3,t}) + (\theta_1 \cdot c_t + \theta_2 \cdot \alpha_{3,t}) \cdot \Big(1 - \exp[-\gamma^2 \cdot c_t^2]\Big) \\ c_t \end{bmatrix}.$$

Moreover, the following row vector Z_t of length p remains constant over time:

$$Z_t = Z = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}.$$

Finally, the vector u_t of length p follows an iid $\mathcal{N}(0, H)$ and contains the values $w_{1,t}$ and $w_{2,t}$:

$$u_t = \begin{bmatrix} w_{1,t} \\ w_{2,t} \\ 0 \end{bmatrix} \quad \text{and} \quad H = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The development of the state equation of the model (6.3) implies that it can be written as follows:

$$\begin{aligned} y_t &= p_t + c_t \\ p_{t+1} &= p_t + w_{1,t} \\ c_{t+1} &= (\phi_1 \cdot c_t + \phi_2 \cdot \alpha_{3,t}) \\ &\quad + (\theta_1 \cdot c_t + \theta_2 \cdot \alpha_{3,t}) \cdot \left(1 - \exp\left[-\gamma^2 \cdot c_t^2\right]\right) + w_{2,t} \\ \alpha_{3,t+1} &= c_t. \end{aligned} \tag{6.4}$$

6.2.1 The prior

The MCMC algorithm used to estimate the model (6.3) is a Gibbs sampler iterating on the full conditional posteriors of four blocks: the scalars σ_1^2 and σ_2^2, the vector $\beta' = (\phi_1 \quad \phi_2 \quad \theta_1 \quad \theta_2)$, the parameter γ and the vectors α_t for $t = 1, ..., N$.

Once the assumptions on the priors of the four blocks are formulated, the posteriors are developed. The hypothesis of independence between $\omega_{1,t}$ and $\omega_{2,t}$ and the structure of the state equations of the model (6.3) imply that the RW process of the permanent component p_t and the ESTAR(2,1) process of the transitory component c_t are treated independently. Therefore, we assume an inverted Gamma prior on σ_i^2 with shape and scale hyperparameters $\underline{a_i}$ and $\underline{b_i}$:

$$p(\sigma_i^2) = \underline{b_i}^{\underline{a_i}} \cdot \frac{1}{G(\underline{a_i})} \cdot \sigma_i^{-2(\underline{a_i}+1)} \cdot \exp\left[-\frac{\underline{b_i}}{\sigma_i^2}\right], \qquad \text{for } i = 1, 2. \tag{6.5}$$

In the case of the ESTAR(2,1) process, we assume an independent multinormal prior for the vector $\beta' = (\phi_1 \quad \phi_2 \quad \theta_1 \quad \theta_2)$ with mean vector $\underline{\beta}$ and precision matrix $\underline{V_\beta}$:

$$p(\beta) = \frac{1}{(2\pi)^{4/2}} \cdot \left|\underline{V_\beta}^{-1}\right|^{-1/2} \cdot \exp\left[-\frac{1}{2}(\beta - \underline{\beta})' \, \underline{V_\beta} \, (\beta - \underline{\beta})\right]. \tag{6.6}$$

Moreover, we suppose an independent normal prior for parameter γ with mean $\underline{\gamma}$ and variance σ_γ^2:

$$p(\gamma) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma_\gamma} \cdot \exp\left[-\frac{(\gamma - \underline{\gamma})^2}{2\sigma_\gamma^2}\right]. \tag{6.7}$$

The hyperparameters $\underline{\gamma}$ and σ_γ^2 are chosen depending on the desired values for $E[\gamma^2]$ and $V[\gamma^2]$.

The fourth block is composed of the state vectors α_t. As the state equations of the model (6.3) are nonlinear, we use a special simulation smoother to sample them. This is an adaptation of the simulation smoother of Durbin and Koopman [2002] which employs the extended Kalman filter (EKF) described by Koopman and Lee [2008]. The simulation smoother for nonlinear state space models is described in Appendix D. The EKF provides approximate estimates of the state by applying the standard Kalman filter to the Taylor approximation of the nonlinear functions expanded around the estimated state from the filter. The first order approximation to the state equation is given by:

$$\alpha_{t+1} \approx T(a_{t|t}) + \tilde{T}_t \cdot (\alpha_t - a_{t|t}) + u_t.$$

In our case, the partial derivative of the function $T(\alpha_t)$ in model (6.3) is derived analytically:

$$\tilde{T}_t = \left.\frac{\partial T(\alpha_t)}{\partial \alpha_t'}\right|_{\alpha_t = a_{t|t}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & T_{22} & T_{23} \\ 0 & 1 & 0 \end{bmatrix}$$

where:

$$T_{22} = \phi_1 + \theta_1 \cdot \left(1 - \exp[-\gamma^2 \cdot c_{t|t}^2]\right) + (\theta_1 \cdot c_{t|t} + \theta_2 \cdot \alpha_{3,t|t}) \cdot 2 \cdot \gamma^2 \cdot c_{t|t} \cdot \exp[-\gamma^2 \cdot c_{t|t}^2]$$

and:

$$T_{23} = \phi_2 + \theta_2 \cdot \left(1 - \exp[-\gamma^2 \cdot c_{t|t}^2]\right).$$

The initialization of the simulation smoother is as follows: $\alpha_1 \sim \mathcal{N}(a_1, P_1)$, where $a_1 = (y_1\, 0 \cdots 0)'$ and P_1 is an identity matrix multiplied by the scalar 10^4. The hyperparameters used in the MCMC algorithm of model (6.2) are chosen so that little prior information is introduced into the algorithm, except for the parameter γ. We set $\underline{a_1} = 10^{-3}$, $\underline{b_1} = 10^{-3}$, $\underline{a_2} = 10^{-3}$, $\underline{b_2} = 10^{-3}$. For the autoregressive parameters, $\underline{\beta}$ is a null vector and $\underline{V_\beta}$ is an identity matrix multiplied by the scalar 10^{-2}. The parameter γ is crucial in STAR models, therefore the hyperparameters $\underline{\gamma}$ and σ_γ^2 must be chosen carefully. Our data are of little information in that respect. Hence, the prior must compensate for this lack of information. We use a purely Bayesian technique called *empirical Bayes method* to address that issue (see Section 6.3.1 for further details).

6.2.2 The posterior

We now develop the expressions of the full conditional posteriors of each block and then use them to generate random draws. The first step of the Gibbs sampler is to run the simulation smoother for nonlinear state space models in order to draw a sample of the full conditional posteriors of the state vectors α_t for $t = 1, ..., N, N+1$, knowing the parameters σ_1^2, σ_2^2, β and γ. Then, the full conditional posteriors of the parameters σ_i^2 are given by:

$$p(\sigma_i^2|y, \beta, \gamma, \alpha_1, ..., \alpha_{N+1}) = \overline{b_i}^{\overline{a_i}} \cdot \frac{1}{G(\overline{a_i})} \cdot \sigma^{-2(\overline{a_i}+1)} \cdot \exp\left[-\frac{\overline{b_i}}{\sigma_i^2}\right] \qquad (6.8)$$

where:

$$\overline{a_i} = \underline{a_i} + \frac{N}{2}, \qquad \text{for } i = 1, 2$$

$$\overline{b_1} = \underline{b_1} + \frac{1}{2}\sum_{t=1}^{N}(p_{t+1} - p_t)^2$$

and:

$$\overline{b_2} = \underline{b_2} + \frac{1}{2}\sum_{t=1}^{N}\Big(c_{t+1} - (\phi_1 \cdot c_t - \phi_2 \cdot \alpha_{3,t})$$

$$- (\beta_1 \cdot c_t - \beta_2 \cdot \alpha_{3,t}) \cdot \Big(1 - \exp[-\gamma^2 \cdot c_t^2]\Big)\Big)^2.$$

If γ and α_t for $t = 1, ..., N, N+1$ are known, the ESTAR(2,1) process of model (6.4) becomes linear and has the form $c_{t+1} = x_t'\beta + w_{2,t}$, where the vector x_t is:

$$x_t = \begin{bmatrix} c_t \\ \alpha_{3,t} \\ \tilde{F}_t c_t \\ \tilde{F}_t \alpha_{3,t} \end{bmatrix}$$

with $\tilde{F}_t \equiv 1 - \exp[-\gamma^2 \cdot c_t^2]$. Then, the full conditional posterior of β is given by:

$$p(\beta|y, \gamma, \sigma_2^2, \alpha_1, ..., \alpha_{N+1}) = \frac{1}{(2\pi)^{4/2}} \cdot |\overline{V}^{-1}|^{-1/2} \cdot \exp\left[-\frac{1}{2}(\beta - \overline{\beta})' \overline{V} (\beta - \overline{\beta})\right] \qquad (6.9)$$

where:

$$\overline{\beta} = \overline{V}^{-1}\left(\frac{1}{\sigma_2^2}\sum_{t=1}^{N} x_t' c_{t+1} + \underline{V}\underline{\beta}\right)$$

and:

$$\overline{V} \;=\; \frac{1}{\sigma_2^2}\sum_{t=1}^{N} x_t x_t' + \underline{V}\,.$$

As the full conditional posterior of γ is nonstandard, we use the independence chain Metropolis-Hastings (ICMH) algorithm as in Chapter 3. The candidate generating density is a multivariate Student distribution. Considering the normal prior for γ, the kernel of the full conditional posterior is:

$$\kappa^*(\vartheta) \;=\; \exp\left[-\frac{(\gamma-\underline{\gamma})^2}{2\sigma_\gamma^2} - \frac{1}{2\sigma_2^2}\sum_{t=1}^{N}\Big(c_{t+1} - (\phi_1\cdot c_t - \phi_2\cdot\alpha_{3,t}) - (\beta_1\cdot c_t - \beta_2\cdot\alpha_{3,t})\cdot\tilde{F}_t\Big)^2\right]. \tag{6.10}$$

The parameters for the candidate density of the independence chain Metropolis-Hastings algorithm are estimated for each iteration as proposed by Deschamps [2008]. For that purpose, we use the following first-order Taylor expansion of the ESTAR(2,1) process of model (6.4) with $\tilde{F}_t$ around γ^*:

$$y_t^* \;=\; \gamma x_t^* + v_t$$

where:

$$y_t^* \;=\; c_{t+1} - (\phi_1\cdot c_t + \phi_2\cdot\alpha_{3,t}) - (\theta_1\cdot c_t + \theta_2\cdot\alpha_{3,t})\left(\tilde{F}_t - \frac{d\tilde{F}_t}{d\gamma}\bigg|_{\gamma^*}\cdot\gamma^*\right)$$

$$x_t^* \;=\; (\theta_1\cdot c_t + \theta_2\cdot\alpha_{3,t})\frac{d\tilde{F}_t}{d\gamma}\bigg|_{\gamma^*}.$$

The parameter γ^* is an approximate solution to the Bayesian update equations:

$$\gamma^* = S_\gamma^{-1}\left[\frac{x'_* y_*}{\sigma_2^2} + \left(\frac{\underline{\gamma}}{\sigma_\gamma^2}\right)\right] \qquad \text{and} \qquad S_\gamma = \frac{x'_* x_*}{\sigma_2^2} + \frac{1}{\sigma_\gamma^2} \tag{6.11}$$

where y_* is the $N \times 1$ vector with elements y_t^* and x_* is the $N \times 1$ vector with elements x_t^*. Further iterations on (6.11) are used to find the approximate solution. The vector of prior expectation is taken as a starting point. A candidate γ_{new} is drawn from a univariate Student density with kernel:

$$\kappa(\gamma_{new}) = \left[1 + \frac{S_\gamma}{\nu} \cdot (\gamma_{new} - \gamma^*)^2\right]^{-\frac{\nu+1}{2}} \tag{6.12}$$

and is accepted with probability:

$$\alpha(\gamma_{old}, \gamma_{new}) = min\left[\frac{\kappa(\gamma_{old})}{\kappa(\gamma_{new})}\frac{\kappa^*(\gamma_{new})}{\kappa^*(\gamma_{old})}, 1\right] \tag{6.13}$$

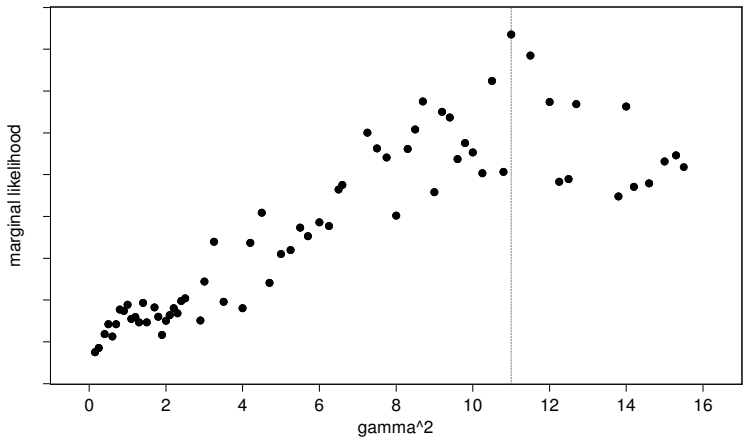
where γ_{old} is the most recently drawn vector. The parameter ν, which corresponds to the degree of freedom in (6.12), can be chosen by experimentation to ensure a good acceptance rate of the candidate. In our case, $\nu = 3$ seems to be a good choice.

In Chapter 5, a stationary constraint was applied to the transitory component following a proposition of Frühwirth-Schnatter [2001]. Here, a similar technique is used to force the ESTAR(2,1) process of model (6.2) to be globally stationary. Indeed, the outer regime of the ESTAR(2,1) process must be stationary, even though the inner regime can be explosive or have a unit root. Thus, within an iteration of the Gibbs sampler, we repeat the sampling of β until the stationarity condition is fulfilled. The stationarity condition is verified by ensuring that the modulus of the most prominent root of the outer regime of the ESTAR(2,1) process is inferior to one. Finally, before any estimation, our MCMC algorithm is tested against any analytical or coding errors following Geweke [2004].

6.3 MCMC estimation

The MCMC estimation is performed by running the algorithm for 50'000 iterations and discarding the first fifth as burn-in. The posterior distributions of the parameters are based on every 20^{th} iteration to avoid serial correlation across iterations. Therefore the analysis of the posteriors relies on two thousand iterations. Chain convergence is checked using the convergence diagnostic methodology proposed by Geweke [1992] (see Chapter 3, Section 3.2).

Figure 6.1: Marginal likelihood of the RW-ESTAR(2,1) model for a grid of values of $E[\gamma^2]$



6.3.1 Empirical Bayes method

As Koop [2003] emphasizes, empirical Bayes methods *involve estimating prior hyperparameters from the data, rather than choosing values for them or setting them to noninformative values.* They allow to choose the prior hyperparameters which maximize marginal likelihood. We apply these methods to the two unobserved components RW-ESTAR(2,1) model in this section, in particular for estimating the expectation $E[\gamma^2]$ of prior distribution (6.7). The variance $V[\gamma^2]$ of this prior is fixed to 0.01 to ensure algorithm convergence. Knowing the values $E[\gamma^2]$ and $V[\gamma^2]$, the hyperparameters $\underline{\gamma}$ and $\sigma^2_{\underline{\gamma}}$ are easily calculated. The other parameters take the value chosen in Section 6.2.1. Then, it is straightforward to adapt the techniques described in Section 3.2.2 for estimating the marginal likelihood of our model (6.3) for several values of $E[\gamma^2]$. The marginal likelihood is defined as $p(y) = \int f(y|\theta)p(\theta)d\theta$, where $f(y|\theta)$ is the likelihood and $p(\theta)$ the prior of the model. We choose to use the *importance sampling* $\hat{p}_{IS}(y)$ (3.22) with 10'000 iterations to estimate marginal likelihood because this estimator converges faster than the *optimal bridge sampling estimator* $\hat{p}_{BS}(y)$ (3.21).

Figure 6.1 displays the marginal likelihood of the RW-ESTAR(2,1) model over the range $E[\gamma^2] \in (0.15,\ 15.50)$. Note that the MCMC algorithm does not converge when $E[\gamma^2] > 15.50$. The analysis of Figure 6.1 shows that marginal likelihood increases when $E[\gamma^2]$ grows until a maximum is reached around 11.00. Therefore, the empirical Bayes method for the RW-ESTAR(2,1) model finds $E[\gamma^2] = 11.00$. This approach is used to estimate $E[\gamma^2]$ in the other models considered.

Table 6.1: Chain properties of the estimations of six RW-ESTAR models for the USD/CHF real exchange rate index

	RNE(σ_1^2)	RNE(σ_2^2)	corr(σ_1^2)	corr(σ_2^2)
RW-ESTAR(2,1)	0.242	0.909	0.810	0.004
RW-ESTAR(2,2)	0.251	1.023	0.840	0.009
RW-ESTAR(2,3)	0.645	1.142	0.216	-0.011
RW-ESTAR(3,1)	0.274	1.016	0.728	-0.003
RW-ESTAR(3,2)	0.249	0.773	0.762	0.000
RW-ESTAR(3,3)	0.747	0.996	0.147	0.000

RNE is the *relative numerical efficiency* and corr the first lag autocorrelation of Markov chains.

6.3.2 Markov chain properties

Table 6.1 shows the efficiency properties and autocorrelation measures for the Bayesian estimates of six RW-ESTAR models. Chapter 5 explains why a careful analysis of the Markov chain of the parameters σ_1^2 and σ_2^2 is important in this context. On the one hand, the first lag autocorrelations for σ_2^2 are close to zero for the six models, which implies well-mixing chains. On the other hand, the first lag autocorrelations for parameter σ_1^2 are quite large (between 0.728 and 0.840) in all specifications, except in the RW-ESTAR(2,3) and RW-ESTAR(3,3) models. For these models, autocorrelation is small, 0.216 and 0.147 respectively. Thus, there is no autocorrelation problem in the Markov chains. This analysis is confirmed in the RNE analysis. The RNEs for parameter σ_2^2 are clearly higher than the ones for σ_1^2 in all models except for the RW-ESTAR(2,3) and RW-ESTAR(3,3) ones. Both models have high values of RNE for σ_1^2 and σ_2^2. Hence, in both cases, the Markov chains are particularly well-mixing.

6.3.3 Model comparison

In this section, we perform a comparison between six preselected models: the RW-ESTAR(r,d) models, resulting from the combination of r=2,3 and d=1,2,3. In a Bayesian context, a good way of discriminating between these models is to compare their marginal likelihood. The latter combines prior information and data information, and the importance sampling evaluation $\hat{p}_{IS}(y)$ of the marginal likelihood has already been implemented in the empirical Bayes method in Section 6.3.1. Once the marginal likelihood is known, we can find the optimal delay of the transition variable of the ESTAR process and the optimal length of the autoregressive process of the transitory component.

The evaluation of marginal likelihood involves evaluating the likelihood function. Thus, the likelihood function itself is evaluated using the *prediction error decomposition* given by Schweppe [1965] and described by Francke et al. [2008]. This decomposition is based on the expression:

$$p(y|\cdot) = p(y_1|\cdot) \cdot \prod_{t=2}^{N} p(y_t|Y_{t-1}, \cdot)$$

where $Y_t = (y_1, ..., y_t)'$. The extended Kalman filter (D.2) yields the prediction error $v_t = y_t - E[y_t|Y_{t-1}]$ and the covariance matrix $F_t = V[y_t|Y_{t-1}]$. Then, $F_t = V[v_t]$ and the likelihood function can be written as:

$$p(y|\cdot) = \frac{1}{(2\pi)^{N/2}} \cdot \Big(\prod_{t=1}^{N} |F_t| \Big)^{-\frac{1}{2}} \cdot \exp \left[-\frac{1}{2} \sum_{t=1}^{N} v_t' F_t^{-1} v_t \right].$$

Table 6.2 shows the estimates for $E[\gamma^2]$ given by the empirical Bayes method described in Section 6.3.1, the BIC and the marginal likelihood of the RW-ESTAR(r,d) models. Our comparison is based only on marginal likelihood because it allows to calculate the Bayes factors (BF). Then, we use the Jeffrey's scale to classify the evidence provided by the data in favor of a model against another. More details were given in Section 3.2.2. The RW-ESTAR(r=2,d) models with d=1,2,3 perform better compared to their r=3 counterparts because their $log(\hat{p}(y))$ are higher. Moreover, RW-ESTAR(r,d=1) dominates the competing specifications RW-ESTAR(r,d=2,3) in terms of logarithmic marginal likelihood. Globally, the RW-ESTAR(2,1) maximizes the in-sample information criterion with a value of $log(\hat{p}(y)) = -25.213$, but the RW-ESTAR(2,3) exhibits a very close result. Indeed, the following transformed BF proposed by Kass and Raftery [1995]: 2 x log(BF) = 2 x (–25.213 – (–25.700)) = 0.974 indicates a "not worth more than a bare mention" for the RW-ESTAR(2,1) model relative to RW-ESTAR(2,3).

Table 6.2: Logarithmic marginal likelihood of the estimations of six RW-ESTAR models for the USD/CHF real exchange rate index

	$E[\gamma^2]$	BIC	$log(\hat{p}(y))$
RW-ESTAR(2,1)	11.00	**2061.577**	**-25.213**
RW-ESTAR(2,2)	11.00	2085.449	-26.315
RW-ESTAR(2,3)	7.00	2117.312	-25.700
RW-ESTAR(3,1)	5.00	2138.031	-29.985
RW-ESTAR(3,2)	11.00	2162.840	-33.025
RW-ESTAR(3,3)	5.00	2242.698	-31.607

$E[\gamma^2]$ is estimated using the empirical Bayes method.

6.3.4 Estimation results

Table 6.3 shows the means and standard deviations of priors and posteriors in the RW-ESTAR(2,1) and RW-ESTAR(2,3) models for the USD/CHF real exchange rate index. The posterior mean for σ_1^2 is 0.637 in the RW-ESTAR(2,1) case and 1.310 for the RW-ESTAR(2,3) model. Thus, the residuals' variance of the permanent component of the RW-ESTAR(2,3) model is almost twice as high as that of the RW-ESTAR(2,1). This result is confirmed in Figures 6.2 and 6.3. The posteriors mean of σ_2^2 are obviously higher than those of σ_1^2. As described in Chapter 5, the movements of the USD/CHF real exchange rate are principally dictated by the transitory component. The autoregressive parameters ϕ_1, ϕ_2, θ_1 and θ_2 are analyzed in Section 6.3.5. The posterior distributions of γ^2 are very similar to the prior distributions for both models. This reflects the limited information contained in our data for this parameter in the RW-ESTAR models.

The analysis is completed by Figures 6.4 and 6.5. They plot the posterior distributions of σ_1^2, σ_2^2 and γ in the RW-ESTAR(2,1) and RW-ESTAR(2,3) models. The histograms of parameter σ_1^2 are not symmetrical, in particular for the RW-ESTAR(2,1) model where the posterior of σ_1^2 is shifted to the left compared to the RW-ESTAR(2,3) model and seems to be truncated at zero.

The means of all draws of the unknown components p_t and c_t for $t = 1, ..., N, N + 1$ yield the so-called smoothed component. Thus, Figure 6.2 and Figure 6.3 show the smoothed values of the unknown components in the RW-ESTAR(2,1) and RW-ESTAR(2,3) models respectively. The upper-left part of these figures compares the trend of the models, i.e. the smoothed values of the permanent component, with the USD/CHF real exchange rate index series. The trends of both models follow the same

Table 6.3: Results of the estimations of two RW-ESTAR models for the USD/CHF real exchange rate index

	RW-ESTAR(2,1)		RW-ESTAR(2,3)	
	Prior	Posterior	Prior	Posterior
σ_1^2	imp^a	0.637	imp^a	1.310
		(0.573)		(1.457)
σ_2^2	imp^a	7.157	imp^a	9.544
		(3.869)		(7.498)
ϕ_1	0	0.746	0	1.105
	(10)	(6.851)	(10)	(4.311)
ϕ_2	0	-0.362	0	-0.442
	(10)	(3.015)	(10)	(4.500)
θ_1	0	0.552	0	-0.013
	(10)	(6.851)	(10)	(4.315)
θ_2	0	0.048	0	0.339
	(10)	(3.019)	(10)	(4.504)
$\phi_1 + \theta_1$	0	1.299	0	1.092
	$(\sqrt{200})$	(0.062)	$(\sqrt{200})$	(0.105)
$\phi_2 + \theta_2$	0	-0.314	0	-0.103
	$(\sqrt{200})$	(0.061)	$(\sqrt{200})$	(0.105)
γ^2	11.00	11.001	7.00	7.004
	(0.1)	(0.098)	(0.1)	(0.098)

Mean and standard deviation (in brackets) of priors and posteriors; a: almost improper priors because the hyperparameters of the Gamma are 10^{-3} but not nulls.

Figure 6.2: Posterior means of p_t and c_t of the RW-ESTAR(2,1) model

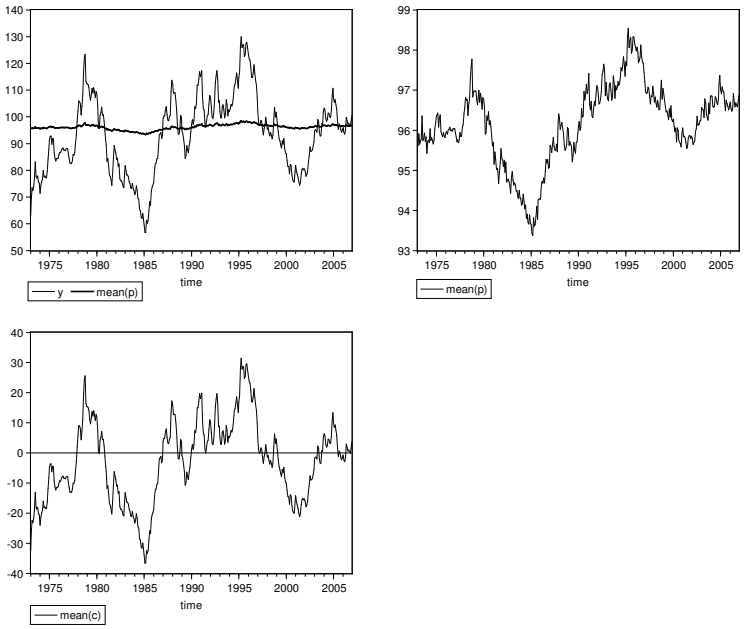


Figure 6.3: Posterior means of p_t and c_t of the RW-ESTAR(2,3) model

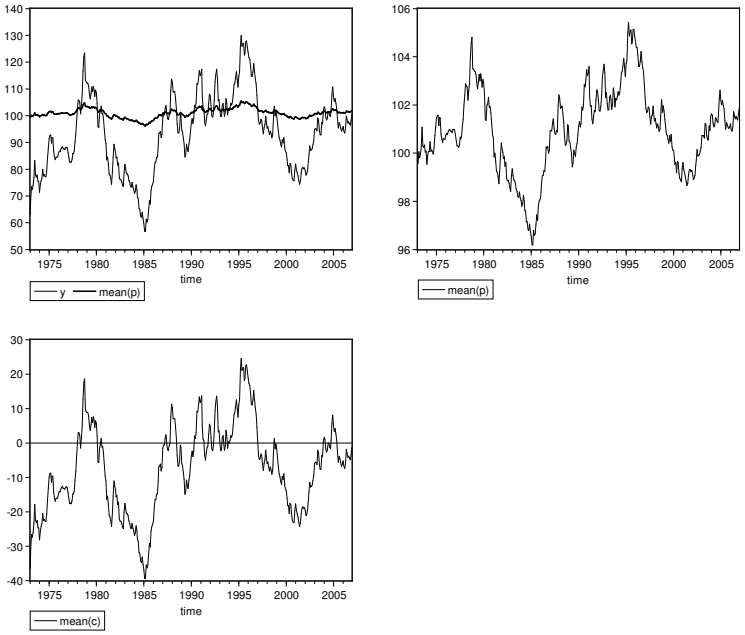


Figure 6.4: Histograms of $p(\sigma_1^2|y)$, $p(\sigma_2^2|y)$ and $p(\gamma^2|y)$ of the RW-ESTAR(2,1) model

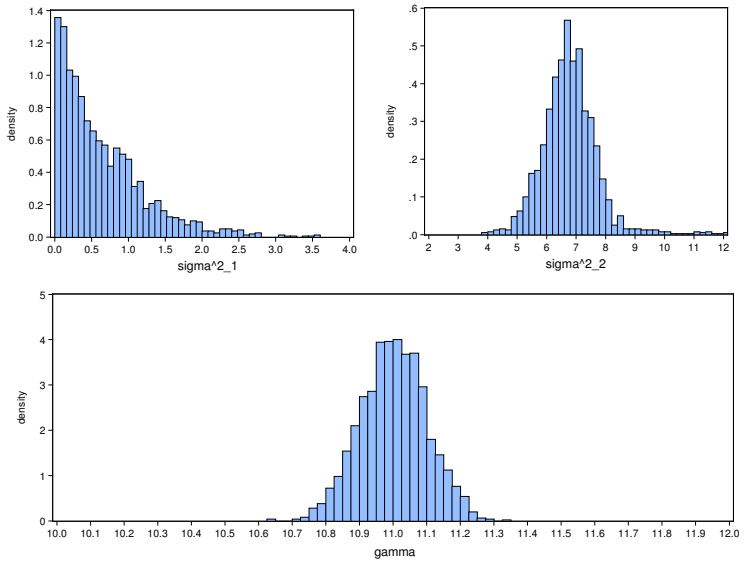


Figure 6.5: Histograms of $p(\sigma_1^2|y)$, $p(\sigma_2^2|y)$ and $p(\gamma^2|y)$ of the RW-ESTAR(2,3) model

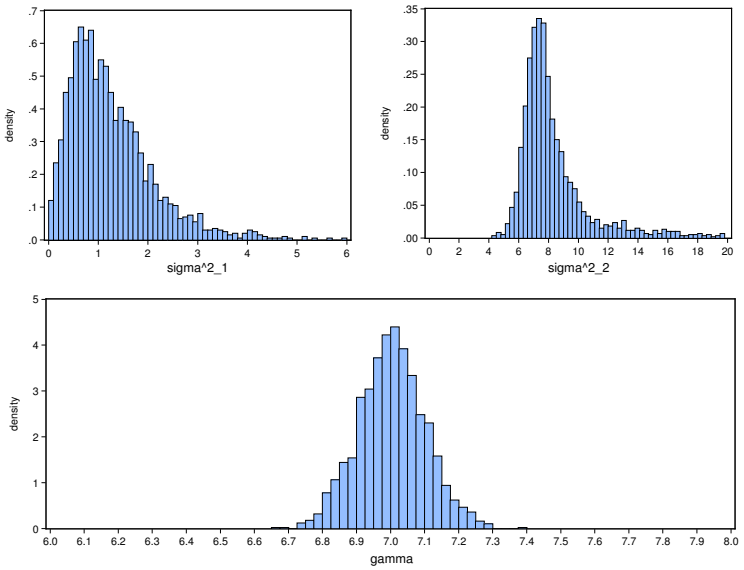


Figure 6.6: Transition function of the RW-ESTAR(2,1) model

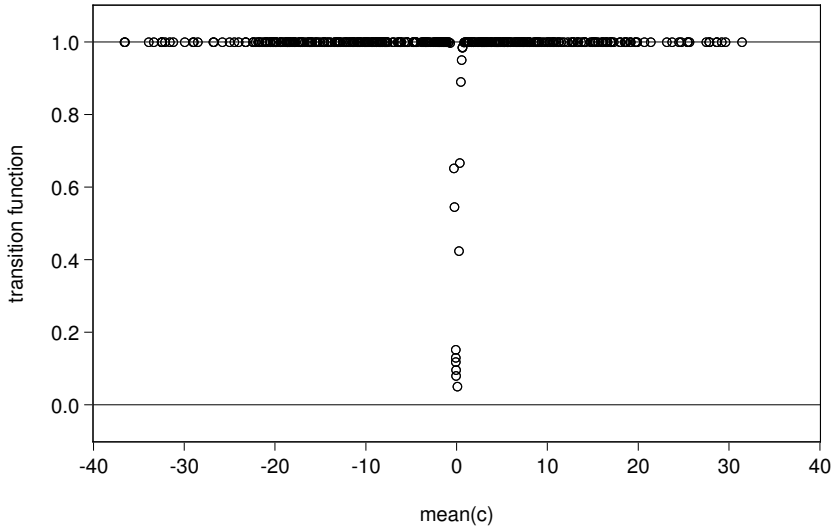


Figure 6.7: Transition function of the RW-ESTAR(2,3) model

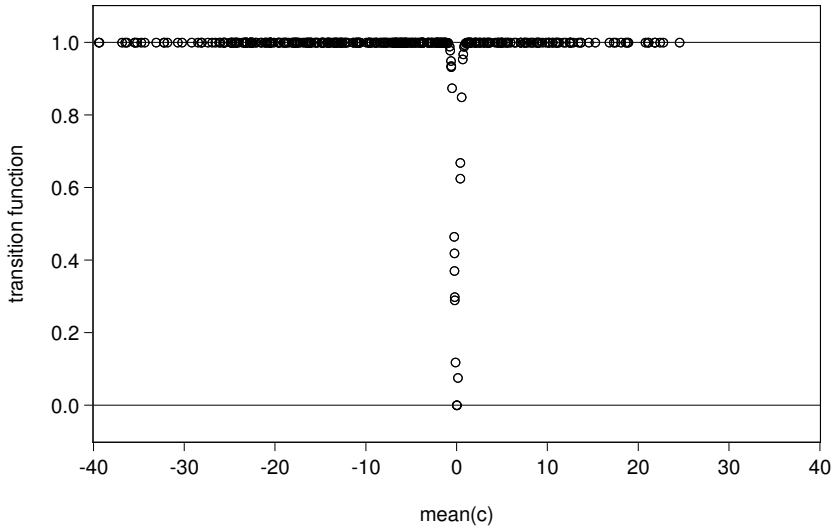


Table 6.4: Dynamic behavior of two RW-ESTAR models for the USD/CHF real exchange rate index

	Regime	Most prominent roots	Modulus	Frequency	Period
RW-ESTAR(2,1)	I	0.373 ± 0.47i	0.601	0.90	6.97
	O	0.977	0.977		
		(0.016)			
RW-ESTAR(2,3)	I	0.553 ± 0.37i	0.665	0.59	10.66
	O	0.987	0.987		
		(0.012)			

I is the inner regime, O is the outer regime; the standard deviations of the most prominent roots are in brackets.

movement, but that of the RW-ESTAR(2,3) model has a higher amplitude. This is most obvious in the upper-right part of these figures which plot only the trend. The locations of the two trends are different. Indeed, the trend of the RW-ESTAR(2,1) evolves between 93 and 99, whereas that of the RW-ESTAR(2,3) evolves between 96 and 106. That makes a difference because these trends are interpreted here as the equilibrium level of the real exchange rate index. Indeed, it is interesting to note that the value of PPP found when estimating the MRESTAR(2,3) model in Chapter 4, i.e. 101.910, is in the range of the trend of the RW-ESTAR(2,3) model.

Figure 6.6 and Figure 6.7 show the transition function of the ESTAR process compared to the posterior mean of the transitory component c_t for the RW-ESTAR(2,1) model and the RW-ESTAR(2,3) model respectively. Both describe the typically inverse-bell shape of an ESTAR process, but with a high value for slope parameter γ^2.

6.3.5 Dynamic behavior

The analysis of the dynamic behavior of the transitory component is very important in order to understand our Unobserved RW-STAR Components model. The transitory component is a nonlinear process describing two extreme regimes: an *inner* regime when the permanent component is exactly equal to the series and an *outer* regime when the permanent component is sufficiently far from the series. Thus, the characteristic roots of the polynomial defined by the extreme regimes indicate if the regimes are stable, explosive or have a unit root.

Figure 6.8: USD/CHF real exchange rate index and transition function of the RW-ESTAR(2,1) estimation

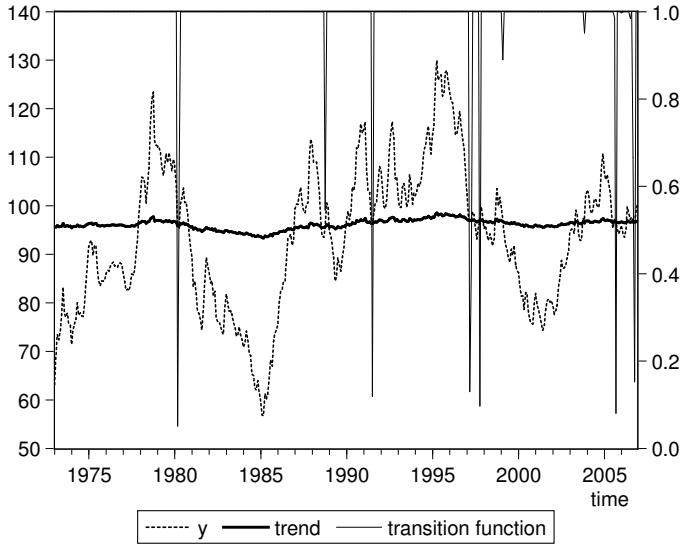


Figure 6.9: USD/CHF real exchange rate index and transition function of the RW-ESTAR(2,3) estimation

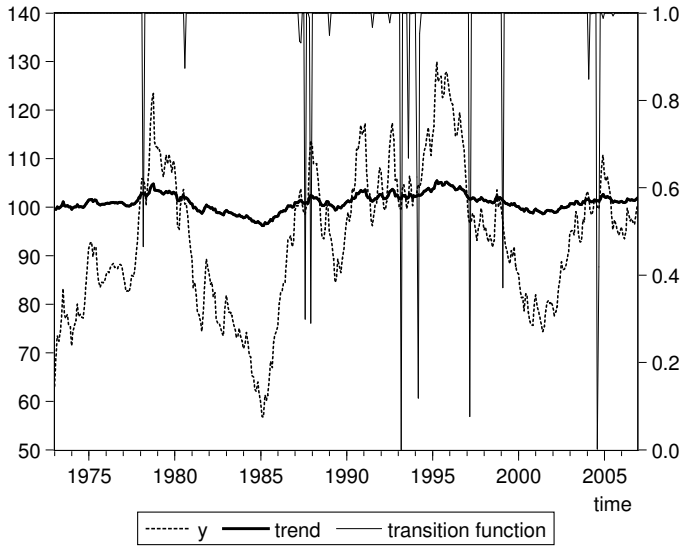


Table 6.4 shows the most prominent roots and moduli of the inner and outer regimes of the transitory components. On the one hand, the RW-ESTAR(2,1) model exhibits a cyclical movement in the inner regime with periods of about 7 months. The modulus indicates that this regime is not explosive. On the other hand, the outer regime of this model is stable because its most prominent root is strictly under 1. Note that this stability is imposed by the stationarity condition introduced in the MCMC algorithm, but that the stationarity condition implies no value for the convergence speed of the transitory process. So, when the USD/CHF real exchange rate index is far from the permanent component, it converges slowly towards it; and when the series get closer to this long-run trend, it moves cyclically around it. Then, the series may diverge from the trend only if the latter moves suddenly. As we can see in Figures 6.6 and 6.8, this often occurs. Indeed, the USD/CHF real exchange rate index is far from the long-run trend most of the time. The RW-ESTAR(2,3) model exhibits the same movements in the inner and outer regimes compared to the RW-ESTAR(2,1). The only significant difference is the cyclical movement of the process in the inner regime, which has longer periods (between 10 and 11 months).

6.4 Forecast evaluation

This section compares the predictive performances of our RW-ESTAR models with a selection of each type of model estimated throughout this thesis.

In a Bayesian context, the uncertainty on future unobserved data is captured by predictive densities. The predictive densities are obtained by simulations. Let us give the simplified general notation of the unobserved RW-STAR components model adapted from (6.3):

$$\begin{aligned} y_t &= Z\alpha_t \\ \alpha_{t+1} &= T(\alpha_t) + u_t \end{aligned}$$

where u_t is iid $\mathcal{N}(0, H)$. For each posterior draw $\left(H^{(s)}, T(\cdot)^{(s)}, \alpha_N^{(s)}\right)$ and for $h = 1, ..., 6$, a draw of the vector $\alpha_{N+h|N}^{(s)}$ is generated from a normal distribution with expectation α_{N+h}^* depending on $T(\alpha_{N+h-1}^*)^{(s)}$ and with covariance matrix $H^{(s)}$, where α_{N+h-1}^* is the draw $\alpha_N^{(s)}$ if $h = 1$, and the previously simulated value $\alpha_{N+h-1|N}^{(s)}$ otherwise. The draws $y_{N+h|N}^{(s)}$ are given by the following adaptation of the observation equation $y_{N+h|N}^{(s)} = Z\,\alpha_{N+h|N}^{(s)}$. Note that one can analytically calculate the

Table 6.5: Efficiency tests of four UC models for the USD/CHF real exchange rate index

		Horizons					
		1^a	2	3	4	5	6
Statistics	RW-ESTAR(2,3)	1.042	1.828	4.790	2.679	2.961	2.885
	RW-ESTAR(2,1)	3.214	4.378	2.617	1.587	2.648	1.983
	RW-AR(2)	1.093	1.947	3.488	3.125	3.881	2.753
	RWI(2)-AR(2)	2.699	4.274	4.334	4.923	5.214	5.166
P-values	RW-ESTAR(2,3)	0.389	0.129	0.001	0.036	0.023	0.026
	RW-ESTAR(2,1)	0.016	0.003	0.039	0.183	0.037	0.102
	RW-AR(2)	0.364	0.108	0.010	0.018	0.006	0.032
	RWI(2)-AR(2)	0.035	0.003	0.003	0.001	0.001	0.001

a: Analytical calculations.

one-step ahead predictive density of the RW-ESTAR models. As the full conditional posteriors of the state vectors α_t are available until $t = N + 1$, the MCMC algorithm provides the posterior distribution $p(\alpha_{N+1}|y)$. Thus, the predictive density at horizon $h = 1$ is replaced by the application $p(Z\alpha_{N+1}|y)$.

In order to compare all models estimated in this thesis, the predictive performances are evaluated solely using the *efficiency test* of West and McCracken [1998] described in Section 3.6.2. This test uses, for 112 information sets, the median of the predictive densities at horizons $h = 2, ..., 6$ and the median of the posterior distribution $p(Z\alpha_{N+1}|y)$ instead of the predictive density at horizon $h = 1$. All information sets begin with the observation 1973:01. The first information set ends with the observation 1997:02 and the last information set ends with the observation 2006:12.

Our analysis begins with a comparison between a selection of our UC models. Indeed, it is interesting to determine which UC models are efficient compared to the others. Thus, Table 6.5 shows the efficiency test statistics and p-values for four UC models estimated in this chapter and in Chapter 5: the RW-ESTAR(2,1), RW-ESTAR(2,3), RW-AR(2) and RWI(2)-AR(2) models. On the one hand, the RW-ESTAR(2,1) and RW-AR(2) do not reject the hypothesis of efficiency at horizons $h = 1$ and $h = 2$, but reject it otherwise. On the other hand, the RW-ESTAR(2,3) is the only efficient model for $h = 4$ and $h = 6$. Moreover, the RWI(2)-AR(2) rejects the null at all horizons. Hence, compared to the UC models estimated in Chapter 5, the new UC models developed in this chapter either have the same predictive performance (for $h = 1$ and $h = 2$) or a better predictive performance (for $h = 4$ and $h = 6$).

Table 6.6: Efficiency tests of six models for the USD/CHF real exchange rate index

		Horizons					
		1^a	2	3	4	5	6
Statistics	RW-ESTAR(2,1)	3.103	2.617	6.243	4.986	4.015	4.805
	RW-ESTAR(2,3)	3.893	2.043	3.224	3.685	4.069	3.773
	RW-AR(2)	2.886	2.531	6.436	6.173	4.838	4.687
	MRESTAR(2,3)	2.572	1.946	6.401	5.988	4.486	4.328
	AR(2)	2.977	2.701	6.740	5.771	4.341	4.126
	RW	8.678	2.896	1.235	1.903	2.226	2.177
P-values	RW-ESTAR(2,1)	0.008	0.021	0.000	0.000	0.001	0.000
	RW-ESTAR(2,3)	0.002	0.066	0.006	0.002	0.001	0.002
	RW-AR(2)	0.012	0.025	0.000	0.000	0.000	0.000
	MRESTAR(2,3)	0.023	0.080	0.000	0.000	0.000	0.001
	AR(2)	0.010	0.018	0.000	0.000	0.001	0.001
	RW	0.000	0.012	0.294	0.087	0.046	0.051

a: Analytical calculations for the UC models.

It is interesting now to compare the predictive performances of the models estimated throughout this thesis. Table 6.6 shows the efficiency test statistics and p-values for a selection of six models: the RW-ESTAR(2,1), RW-ESTAR(2,3), RW-AR(2), MRESTAR(2,3), AR(2) and RW models. We select the three UC models because of the results shown in Table 6.5. The analysis carried out in Chapter 4 shows that the MRESTAR(2,3) model has good predictive performances. Moreover, both linear models are considered as benchmarks. At horizon $h = 1$, all models reject the null hypothesis of efficiency at the 5% level. Nevertheless, in looking at p-values more precisely, three groups can be distinguished: first, the RW-ESTAR(2,3) and RW models obviously reject the null; second, the RW-ESTAR(2,1), RW-AR(2) and AR(2) models have p-values around 1%; third, the MRESTAR(2,3) has the highest p-value (2.3%). Hence, only the MRESTAR(2,3) model does not reject the hypothesis of efficiency at the 1% level. This is tenuous evidence for the good predictive performance of the MRESTAR(2,3) model. The results are more obvious at horizon $h = 2$: the RW-ESTAR(2,3) and MRESTAR(2,3) models are both efficient at the 5% level. So, for forecasting the USD/CHF real exchange rate on a short horizon, the MRESTAR(2,3) model seems to be the better choice. For $h = 3, ..., 6$, the p-values of the RW model are the only values exceeding or approaching the 5% level.

Table 6.7: BIC of the Bayesian estimates of a selection of models for the USD/CHF real exchange rate index

Models	BIC
AR(2)	1967.387
ESTAR(2,3)	**1963.219**
MRESTAR(2,3)	1985.113
RW-AR(2)	2020.286
RW-ESTAR(2,1)	2061.577
RW-ESTAR(2,3)	2117.312

6.5 Conclusion

The RW-ESTAR model studied in this chapter was an UC model specifically designed for the real exchange rate context: the real exchange rate tends to converge toward an intertemporal equilibrium level but this target level may vary over time. We studied the application of the RW-ESTAR model to the USD/CHF real exchange rate index in a Bayesian context. The model comparison indicated that the RW-ESTAR(2,1) and RW-ESTAR(2,3) performed better than the other UC models preselected for our analysis using the marginal likelihood criterion, but that the Bayes factor could not discriminate obviously between these two models.

A comparison between the information criteria of all models estimated throughout this thesis allows us to evaluate the in-sample fit of our models. Note that the marginal likelihood must not be used because all models do not have the same prior distribution. Moreover, between AIC, BIC and DIC, we choose to use BIC because it gives an asymptotic approximation of marginal likelihood (see Kass and Raftery [1995]). Hence, Table 6.7 summarizes the values of BIC showed in Tables 3.2, 4.2, 5.2 and B.1. The choice of the models shown in Table 6.7 is based on marginal likelihood. The ESTAR(2,3) model performs better than the other models. This result confirms the well-known idea that BIC favors the more parsimonious models.

The behaviors of the trend of the UC models estimated in this thesis were compared in order to appreciate the improvement provided by the RW-STAR models. The behaviors of the trend of the RW-AR(2) model estimated in Chapter 5 and the RW-ESTAR(2,1) model estimated in this chapter were very similar (see Figure 5.3 and Figure 6.8). On the other hand, the RW-ESTAR(2,3) (Figure 6.9) exhibited a wider trend movement.

Concerning predictive performance, we showed that the MRESTAR(2,3) model may be considered as efficient for one-step ahead forecasts compared to a selection of the best models estimated throughout this thesis. This was tenuous evidence for the good predictive performance of the MRESTAR(2,3) model. For two-step ahead forecast, the RW-ESTAR(2,3) and MRESTAR(2,3) were obviously both efficient and for longer-step forecasts, the RW was the only efficient model. Hence, the extension described in this chapter improved, in the real exchange rate context, the performances of the models estimated in previous chapters.

Chapter 7

Conclusion

The econometric modeling of real exchange rates is a broad topic. Hence, a thesis on the behavior of the real exchange rate necessarily involves focusing solely on parts of the topic. Following our own personal interests, we focused on modeling the real exchange rate by two methods: (i) we studied the smooth transition autoregressive (STAR) models to investigate whether the real exchange rate was a multiple regime process and (ii) we analyzed the unobserved components models to investigate whether the purchasing power parity (PPP) relationship varied over time. Note that other models were investigated. For example, a Markov switching autoregressive (MS-AR) model was implemented following Engel and Hamilton [1990], but the MS-AR model yielded results that were not very encouraging.

In Chapter 2, we analyzed the USD/CHF real exchange rate index with unit root tests and found that the series was stationary with the most powerful test available. We discussed specification, estimation and evaluation of the STAR model in a frequentist context. In order to discriminate between STAR models, we first used the methodology proposed by Teräsvirta [1994], but, according to the author of this thesis, it was unreliable and inconsistent. Second, we discriminated between STAR models using information criteria. This methodology was problematic because our new estimation algorithm did not converge for all models. Nevertheless, we could discard some STAR models, including some exponential STAR (ESTAR) models and all logistic STAR (LSTAR) models. Then, the misspecification tests developed by Eitrheim and Teräsvirta [1996] were applied in order to discard more models. Thus, the ESTAR(2,3) model was the best STAR model, in the sense of best fit. Finally, the analysis of predictive performances indicated that the ESTAR(2,3) was beaten by the ESTAR(3,1) model. The predictive performances also highlighted that the number of forecasts influences the Diebold and Mariano [1995] test.

We studied the ESTAR model in a Bayesian context in Chapter 3. The reader has been able to appreciate the advantages of using the Bayesian method to estimate and evaluate such models. In particular, the Bayesian method offered some tools, such as marginal likelihood, for discriminating between models or posterior predictive p-values for judging model adequacy. Moreover, despite convergence difficulties when the priors were not sufficiently informative, the Bayesian method did not encounter problems like the frequentist method did when maximizing the likelihood function. Based on preliminary work, we chose to focus only on four ESTAR models for the application to the USD/CHF real exchange rate index. The model comparison and misspecification tests allowed to designate the ESTAR(2,3) model as the model having the best fit. Concerning predictive performances, our analysis indicated first that the efficiency test was more powerful than the DM test and, second, that the ESTAR(2,3) model was not the best choice for forecasting the real exchange rate. Indeed, this model did not pass the efficiency test for all horizons certainly because the process spent most of its time in one regime only. The analysis also revealed that the serial independence test by Eitrheim and Teräsvirta [1996] was useful for detecting the autocorrelation problem in our models. From an economic point of view, the results obtained in Chapter 3 indicated that the real exchange rate between Switzerland and the United States was mean-reverting and that it rarely reached PPP. It even seemed that it converged to a value below PPP. This contradiction motivated further investigations through improving the ESTAR model by adding a third regime that described the behavior of the real exchange rate around the convergence point of the current process.

In Chapter 4, we studied the performances of the multiple regime exponential smooth transition autoregressive (MRESTAR) model, an extension of the ESTAR model, in a Bayesian context. A comparison between the information criteria of three preselected MRESTAR models allowed to designate the MRESTAR(2,3) model as the best one, in the sense of best fit. Nevertheless, comparing the MRESTAR models' BIC estimated in this chapter and the ESTAR models' BIC estimated in Chapter 3, we noted higher values when a third regime was assumed. According to this in-sample criterion, our extension was not an improvement on the ESTAR model. The analysis of the predictive performances of the MRESTAR(2,3) model confirmed the lack of power of the DM test. On the other hand, the efficiency test indicated that the MRESTAR(2,3) model was efficient at one and two-month horizons. Compared to the ESTAR(2,2), AR(2) and RW models, the MRESTAR(2,3) model was even the only efficient model for one-step ahead forecasts. Hence, our three-regime model improved forecast accuracy. Our investigations confirmed a theoretical economic result: the USD/CHF real exchange rate was mean-reverting and converged toward

the PPP represented by the threshold parameter of the second inner regime of the MRESTAR model. On the other hand, we provided no economic interpretation for the threshold parameter of the first inner regime of this model. It remained an open issue.

In Chapter 5, we analyzed a particular type of state space model from a Bayesian point of view: the unobserved components (UC) models. Our UC models contained both a permanent and a transitory component. The permanent component was described by two kinds of random walk processes: a simple random walk and an integrated random walk of order I(2). The transitory component was described by autoregressive processes of order r. Our results pointed to the fact that the movements of the USD/CHF real exchange rate were principally dictated by the transitory component. Using in-sample criteria, the model comparison indicated that the RW-AR(2) performed better than the other UC models analyzed in this chapter. On the other hand, a comparison between BICs of the ESTAR, MRESTAR and UC models indicated that UC models performed worse than MRESTAR models which in turn performed worse than ESTAR models. Concerning predictive performance, we showed that the RW-AR(2) model was as efficient as the MRESTAR(2,3) and AR(2) models for one-step ahead forecasts. Finally, the results of this chapter indicated that the USD/CHF real exchange rate could be successfully decomposed into both a long-run and a transitory component. The long-run component, interpreted as PPP, was characterized by a stochastic trend. Hence, PPP could vary over time and could have its own dynamics. These results were very encouraging and led us to investigate this way of modeling the real exchange rate, i.e. the decomposition into both a permanent and a transitory component. The next step was describing the transitory component with an ESTAR process instead of with an AR process.

The unobserved random walk and exponential smooth transition autoregressive components (RW-ESTAR) model studied in Chapter 6 was an UC model specifically designed for the real exchange rate context: the real exchange rate tended to converge toward an intertemporal equilibrium level but this target level could vary over time. Hence, we studied the application of the RW-ESTAR model to the USD/CHF real exchange rate index in a Bayesian context. The model comparison indicated that the RW-ESTAR(2,1) and the RW-ESTAR(2,3) performed better than the other UC models preselected for our analysis using the marginal likelihood criterion, but the Bayes factor could not discriminate obviously between these two models. On the other hand, a comparison between the BICs of the models estimated throughout this thesis allowed us to evaluate the in-sample fit of our models: the ESTAR(2,3) model performed better than the other models. Concerning predictive performance, we showed that

the MRESTAR(2,3) model could be considered as efficient for one-step ahead forecasts compared to a selection of the best models estimated throughout this thesis. This was tenuous evidence for the good predictive performance of the MRESTAR(2,3) model. For two-step ahead forecasts, the RW-ESTAR(2,3) and MRESTAR(2,3) were obviously both efficient and for longer-step forecasts, the RW was the only efficient model. Hence, the extension described in this last chapter improved, in the real exchange rate context, the performances of the models estimated in previous chapters.

Suggestions for further work

First, from an economic point of view, the interpretation of the MRESTAR model in Chapter 4 is questionable because it provides no explanation about the first inner regime. Hence, explanations about the first inner regime should be found using theoretical works.

Second, the error terms for the unobserved components models described in Chapters 5 and 6 are supposed to be independent from each other. It might be interesting to relax this assumption.

Finally, in terms of data, we chose to use only the monthly USD/CHF real exchange rate index in all our applications. In order to provide more general results, all models could be applied to other currency couples as in Sarantis [1999].

Appendix A

Monte Carlo study of the STAR models

The purpose of this appendix is to show that the first step of the specification procedure proposed by Teräsvirta [1994] is not reliable and can lead to wrong choices. The first step consists in using the minimization of an information criterion on linear autoregressive models to select the value of parameter p in the STAR(p, d) model. We therefore construct a Monte Carlo study in generating six STAR models with various parameters and applying the mentioned method to select parameter p. We can then know the frequency of right choices and wrong choices for each specified model.

We consider the following data generating process (DGP) based on a general STAR(p, d) model:

$$y_t = \phi' x_t + \theta' x_t F(\gamma, c; y_{t-d}) + u_t \tag{A.1}$$

$$\text{and} \quad F(\gamma, c; y_{t-d}) = (exp\{-\gamma(y_{t-d} - c)\})^{-1} \tag{A.2}$$

$$\text{or} \quad F(\gamma, c; y_{t-d}) = 1 - exp\{-\gamma(y_{t-d} - c)^2\} \tag{A.3}$$

where $x_t = [1, y_{t-1}, ..., y_{t-p}]'$, ϕ and θ are vectors of autoregressive parameters of length $p + 1$, and $u_t \sim \mathcal{N}(0, \sigma^2)$. Six models are generated with different values of parameters ϕ, θ, γ, c, p and d. These models, described in Table A.1, are based (with sometimes a few adjustments) on existing literature models or on our own created models. For each model 5'000 replications are simulated with $\sigma = 0.02$, 0.2 or 1, and the sample size T is 100 or 400. Moreover, the first 200 observations are discarded in order to avoid initialization effects. For each replication two criteria are applied to choose the lag of the linear autoregression of the STAR model. We use the most common information criteria: the AIC (Akaike [1974]) and BIC (Schwarz [1978]). Note that the value of parameter d is 3 for all six models, but the results of this Monte Carlo study are not affected by this parameter.

Table A.1: Description of the six models used as DGP for the Monte Carlo study of Tables A.2 and A.3

	A	B	C	D	E	F
ϕ_0	-0.19	0.0	-0.047	163.941	0.02	-2830.11
ϕ_1	0.9	1.8	0.651	-0.552	0.9	-1.00
ϕ_2	–	-1.06	0.291	–	-0.265	1.47
ϕ_3	–	–	0.046	–	–	24.79
θ_0	0.38	0.02	-0.050	-162.879	7.02	2831.63
θ_1	0.0	-0.9	0.577	1.542	-1.6	2.32
θ_2	–	0.795	-0.323	–	0.415	-1.85
θ_3	–	–	-0.279	–	–	-24.75
γ	10	100	3.153	0.616	3	1.82
c	0	0.02	-3	115.712	4	166.08
transition	LSTAR	LSTAR	LSTAR	ESTAR	ESTAR	ESTAR
p	1	2	3	1	2	3
d	3	3	3	3	3	3
source	Lund.[a]	Teräs.[b]	Desch.[c]	own	Teräs.[b]	own

a: Lundbergh and Teräsvirta [2002]; b: Teräsvirta [1994]; c: Deschamps [2008].

The results are shown in Tables A.2 and A.3. First, Table A.2 refers to LSTAR models. Model A is an LSTAR(1,3) model, therefore the method can not choose a value for p less than the right choice. As a general rule, we see that right choice frequency is relatively high when σ is 0.02 or 0.2. But when σ increases, right choice frequency falls. This is particularly true for models A and B. On the other hand, model C yields a low right choice frequency with both criteria. As an example, when N=100 and σ=0.02, the AIC criterion selects the right lag in only 12.5% of cases and the BIC criterion in only 12.1% of cases. Second, Table A.3 refers to ESTAR models. Model D is an ESTAR(1,3) model, we note again that right choice frequency falls when σ increases. But we do not observe this with models E and F, because their right choice frequency is low in all cases.

Finally, it appears clearly that the BIC criterion selects the parameter p of the DGP more frequently than the AIC in all cases. However, we cannot determine whether this method tends to select a lag greater or smaller than the right lag. We can therefore say, as a conclusion to this appendix, that the assertion above seems suitable: the use of information criteria to determine the lag of the STAR model is not reliable, in particular if the standard deviation of the error is high.

Table A.2: Selection frequency of autoregression lag in the Monte Carlo study for several LSTAR models

			AIC			BIC		
model	N	σ	$<^a$	$=^b$	$>^c$	$<$	$=$	$>$
A	100	0.02	–	85.5	14.5	–	86.3	13.7
		0.2	–	63.0	37.0	–	82.4	17.6
		1	–	0.0	100.0	–	44.4	55.6
	400	0.02	–	84.6	15.4	–	85.4	14.6
		0.2	–	64.2	35.8	–	85.0	15.0
		1	–	0.0	100.0	–	62.4	37.6
B	100	0.02	7.3	83.2	9.5	8.0	84.1	7.9
		0.2	0.0	59.1	40.9	0.1	85.6	14.4
		1	0.0	0.0	100.0	0.0	35.4	64.6
	400	0.02	0.0	88.3	11.7	0.0	90.1	9.9
		0.2	0.0	53.4	46.6	0.0	85.3	14.7
		1	0.0	0.0	100.0	0.0	66.1	33.9
C	100	0.02	81.7	12.5	5.8	82.5	12.1	5.4
		0.2	58.8	16.0	25.2	80.0	12.3	7.7
		1	0.0	0.0	100.0	39.2	13.2	47.6
	400	0.02	50.5	42.7	6.8	51.5	42.3	6.2
		0.2	44.6	22.3	33.1	67.8	21.3	10.9
		1	0.0	0.0	100.0	49.9	23.5	26.6

a: frequency of selecting a lag smaller than the right lag ; b: frequency of selecting the right lag ; c: frequency of selecting a lag greater than the right lag.

Table A.3: Selection frequency of autoregression lag in the Monte Carlo study for several ESTAR models

			AIC			BIC		
model	N	σ	$<^a$	$=^b$	$>^c$	$<$	$=$	$>$
D	100	0.02	–	77.8	22.2	–	78.5	21.5
		0.2	–	59.9	40.1	–	78.4	21.6
		1	–	0.0	100.0	–	39.3	60.7
	400	0.02	–	78.8	21.2	–	79.4	20.6
		0.2	–	62.3	37.7	–	80.2	19.8
		1	–	0.0	100.0	–	62.4	37.6
E	100	0.02	0.0	0.0	100.0	0.6	0.1	99.3
		0.2	0.0	0.0	100.0	6.5	1.4	92.1
		1	0.0	0.0	100.0	10.9	10.9	78.2
	400	0.02	0.0	0.0	100.0	0.0	0.0	100.0
		0.2	0.0	0.0	100.0	0.0	0.0	100.0
		1	0.0	0.0	100.0	6.8	28.5	64.6
F	100	0.02	87.0	7.0	6.0	88.0	6.3	5.7
		0.2	62.4	14.0	23.7	84.7	8.2	7.0
		1	0.0	0.0	100.0	44.2	12.1	43.7
	400	0.02	84.7	8.8	6.5	85.8	7.9	6.2
		0.2	60.7	17.6	21.7	85.7	8.3	6.0
		1	0.0	0.0	100.0	69.9	16.6	13.5

a: frequency of selecting a lag smaller than the right lag ; b: frequency of selecting the right lag ; c: frequency of selecting a lag greater than the right lag.

Appendix B

Bayesian estimation of linear models

The purpose of this appendix is to describe the MCMC estimation of an autoregressive model (AR) and a simple random walk model (RW).

B.1 Autoregressive model

The autoregressive model is:

$$y_t = \phi' x_t + u_t \qquad u_t \sim \mathcal{N}(0\,,\,\sigma^2) \tag{B.1}$$

where $x_t = [1, y_{t-1}, \ldots, y_{t-p}]'$ is a vector of the lags of the dependent variable including the constant term. Given this model, the likelihood is:

$$p(y|\phi, \sigma^2) = \frac{1}{(2\pi)^{T/2}} \cdot \frac{1}{\sigma^T} \cdot exp\left[-\frac{u'u}{2\sigma^2}\right] \tag{B.2}$$

where u is the vector of the $u_t(y)$ and T is the sample size.

We assume a multinormal prior on ϕ with expectation vector $\underline{\phi}$ and precision matrix $\underline{V}$:

$$\phi \sim \mathcal{N}(\underline{\phi}\,,\,\underline{V}^{-1}) \tag{B.3}$$

$$p(\phi) = \frac{1}{(2\pi)^{(p+1)/2}} \cdot |\underline{V}^{-1}|^{-1/2} \cdot exp\left[-\frac{1}{2}(\phi - \underline{\phi})'\,\underline{V}\,(\phi - \underline{\phi})\right] \tag{B.4}$$

and an independent inverted Gamma prior on σ^2 with shape and scale parameters $\underline{a}$ and $\underline{b}$:

$$p(\sigma^2) = \underline{b}^{\underline{a}} \cdot \frac{1}{G(\underline{a})} \cdot \sigma^{-2(\underline{a}+1)} \cdot exp\left[-\frac{\underline{b}}{\sigma^2}\right] \tag{B.5}$$

where $G(\cdot)$ is the gamma function.

Table B.1: Information criteria of the Bayesian estimates of AR models for the USD/CHF real exchange rate index

Models	AIC	BIC	DIC	$log(\hat{p}(y))$
AR(1)	2000.099	2012.126	1999.252	-1020.820
AR(2)	1951.362	1967.387	1951.333	**-1017.729**
AR(3)	1945.776	1965.796	1945.857	-1023.845
AR(4)	1942.494	1966.503	1942.405	-1031.095
AR(5)	1939.627	1967.620	1939.625	-1038.563
AR(6)	1937.362	1969.334	1937.322	-1046.293
AR(7)	1931.424	1967.360	1931.393	-1052.240
AR(8)	1920.843	1960.758	1920.871	-1055.922
AR(9)	1916.407	1960.285	1916.324	-1062.593
AR(10)	1914.153	1961.991	1914.047	-1070.304
AR(11)	1908.011	**1959.802**	1908.083	-1076.156
AR(12)	**1906.171**	1961.911	**1906.141**	-1084.088

$log(p(y))$ is the logarithmic marginal likelihood.

The full conditional posteriors of ϕ and σ^2 are analytically known. We can then use the Gibbs sampler to produce random draws of the posterior.

The full conditional posterior of ϕ is given by:

$$p(\phi|y,\sigma^2,\gamma,c) \;=\; \frac{1}{(2\pi)^{(p+1)/2}}\cdot|\overline{V}^{-1}|^{-1/2}\cdot exp\left[-\frac{1}{2}(\phi-\overline{\phi})'\,\overline{V}\,(\phi-\overline{\phi})\right] \quad \text{(B.6)}$$

with $\overline{\phi}=\overline{V}^{-1}(\frac{1}{\sigma^2}X'y+\underline{V}\underline{\phi})$ and $\overline{V} \;=\; \frac{1}{\sigma^2}X'X+\underline{V}$,

and the full conditional posterior of σ^2 is given by:

$$p(\sigma^2|y,\phi,\gamma,c) \;=\; \overline{b}^{\overline{a}}\cdot\frac{1}{G(\overline{a})}\cdot\sigma^{-2(\overline{a}+1)}\cdot exp\left[-\frac{\overline{b}}{\sigma^2}\right] \quad \text{(B.7)}$$

with $\overline{a} \;=\; \underline{a}+\frac{T}{2}$ and $\overline{b} \;=\; \underline{b}+\frac{1}{2}(y-X\phi)'(y-X\phi)$.

Table B.1 compares the standard information criteria and the logarithmic marginal likelihood of autoregressive models with $p=1,...,12$. The latter is obtained using the bridge sampling estimator of Meng and Wong [1996]. The AR(12) minimizes the AIC and the DIC, while the AR(11) minimizes the BIC and the AR(2) maximizes the logarithmic marginal likelihood. Despite divergent results, the AR(2) model is considered as the best because logarithm marginal likelihood is the highest. Indeed the marginal likelihood

is the best way for discriminating between models if these ones have the same prior distribution. It allows to obtain the Bayes factor (BF), which is the ratio of the marginal likelihood of two models. The Jeffrey's scale is then used to classify the evidence provided by the data in favor of a model against another. The following transformed Bayes factor proposed by Kass and Raftery [1995]: 2 x log(BF) = 2 x (–1017.729 – (–1076.156)) = 116.854 indicates a "very strong" evidence for the AR(2) model relative to AR(11). The transformed BF of AR(2) against AR(12) is even greater with a value of 132.718.

B.2 Random walk model

The simplest random walk model can be written as follow:

$$y_t = y_{t-1} + u_t \qquad u_t \sim \mathcal{N}(0\,,\,\sigma^2) \tag{B.8}$$

or

$$\Delta y_t \equiv y_t - y_{t-1} = u_t. \tag{B.9}$$

Given this model, the likelihood is:

$$p(\Delta y|\sigma^2) = \frac{1}{(2\pi)^{T/2}} \cdot \frac{1}{\sigma^T} \cdot exp\left\{-\frac{\Delta y'\Delta y}{2\sigma^2}\right\} \tag{B.10}$$

where Δy is the vector of the Δy_t.

We assume an inverted Gamma prior on the only parameter of the model σ^2 with shape and scale parameters $\underline{a}$ and $\underline{b}$:

$$p(\sigma^2) = \underline{b}^{\underline{a}} \cdot \frac{1}{G(\underline{a})} \cdot \sigma^{-2(\underline{a}+1)} \cdot exp\left\{-\frac{\underline{b}}{\sigma^2}\right\} \tag{B.11}$$

where $G(\cdot)$ is the gamma function.

Then, it is straightforward to obtain the posterior of σ^2:

$$p(\sigma^2|\Delta y) \propto \sigma^{-2(\underline{a}+1)} \cdot \sigma^{-T} \cdot exp\left\{-\frac{\underline{b}}{\sigma^2} - \frac{\Delta y'\Delta y}{2\pi^2}\right\} \tag{B.12}$$

and defining $\overline{a} = \underline{a} + \frac{T}{2}$ and $\overline{b} = \underline{b} + \frac{\Delta y'\Delta y}{2}$, the distribution:

$$\sigma^2|\Delta y \sim \mathcal{IG}(\overline{a}\,,\,\overline{b}). \tag{B.13}$$

Finally, the draws of $\sigma^2|\Delta y$ are simulated by Monte Carlo methods.

Appendix C

Algorithm of the simulation smoother for linear state space models

In this appendix, some details are given about the algorithm of the simulation smoother used in the Bayesian estimation of Unobserved Component Models of Chapter 5.

Consider the following general state space model:

$$\begin{array}{lcll} y_t & = & Z_t\alpha_t + \epsilon_t, & \epsilon_t \sim \mathcal{N}(0, H_t) \\ \alpha_{t+1} & = & T_t\alpha_t + R_t\eta_t, & \eta_t \sim \mathcal{N}(0, Q_t) \end{array} \tag{C.1}$$

where $t = 1, ..., N$, y_t is an $m \times 1$ vector of observations, α_t is a $p \times 1$ state vector and ϵ_t and η_t are vectors of disturbances. Matrices Z_t, T_t, R_t, H_t and Q_t are assumed to be known. Moreover, one assumes that $\alpha_1 \sim \mathcal{N}(a_1, P_1)$, where a_1 and P_1 are known.

The aim of the algorithm developed by Durbin and Koopman [2002] is to draw random vectors $\tilde{\alpha}$ from the conditional distribution $p(\alpha|y)$, where α is the state vector defined above. As the distribution $p(\alpha|y)$ is unknown, the basic idea is to obtain $\tilde{\alpha}$ by drawing vectors from the multivariate normal $\mathcal{N}(0, A)$ and adding a known vector $\hat{\alpha}$ defined below. An important property of the matrix $A = V[\alpha|y]$ is to not depend on y. The proof of this is given by *Theorem 2.5.1* in Anderson [2003] which states that the covariance matrix of the conditional distribution of the subvector $X^{(1)}$ given $X^{(2)} = x^{(2)}$ is $\Sigma^{(12)} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, where $X^{(1)}$ and $X^{(2)}$ are the subvectors of a given vector X following a multivariate normal $\mathcal{N}(\mu, \Sigma)$, and where Σ is divided into four submatrix Σ_{11}, Σ_{12}, Σ_{21} and Σ_{22}. So, the covariance matrix Σ_{12} does not depend on the vector $x^{(2)}$.

Define $\tilde{\alpha} = \hat{\alpha} + \alpha^+ - \hat{\alpha}^+$, where $\hat{\alpha}^+ = E[\alpha^+|y^+]$. α^+ and y^+ are compute using the model (C.1) and the vector of disturbance w^+ drawn from the distribution $p(w)$, where $w = (\epsilon_1', \eta_1', ..., \epsilon_N', \eta_N')'$. The conditional expectation of the vector $\tilde{\alpha}$ given y is:

$$\begin{aligned} E[\tilde{\alpha}|\, y] &= E[\hat{\alpha} + \alpha^+ - \hat{\alpha}^+|y] \\ &= E[\alpha^+ - \hat{\alpha}^+|y] + \hat{\alpha} \\ &= \hat{\alpha}. \end{aligned}$$

As $(\alpha^+ - \hat{\alpha}^+)$ is independent of y, $E[\alpha^+ - \hat{\alpha}^+|y] = 0$. The variance of $\tilde{\alpha}$ given y is:

$$\begin{aligned} V[\tilde{\alpha}|\, y] &= V[\hat{\alpha} + \alpha^+ - \hat{\alpha}^+|y] \\ &= V\Big[E[\alpha|y] + \alpha^+ - E[\alpha^+|y^+]\,|y\Big] \\ &= V[\alpha^+|y]. \end{aligned}$$

Moreover, the independence between A and y implies that $V[\alpha^+|y] = V[\alpha^+]$, and as α^+ is drawn from $p(w)$, we have $V[\alpha^+] = V[\alpha]$. Considering all these calculations we can say that $V[\tilde{\alpha}|y] = A$.

Finally, as $E[\tilde{\alpha}|y] = \hat{\alpha}$ and $V[\tilde{\alpha}|y] = A$, drawing from $\tilde{\alpha}$ is equivalent to drawing a vector from a multivariate normal $\mathcal{N}(0, A)$ and add $\hat{\alpha}$. This is also equivalent to draw a vector form $p(\alpha|y)$.

Knowing that the vector $\tilde{\alpha}$ can be drawn using the vectors $\hat{\alpha}$, $\hat{\alpha}^+$ and α^+, the question is how to compute these three vectors. On the one hand, the vector α^+ is obtained by means of its definition above. This is easily achieve drawing a vector w^+ from the distribution $p(w)$ and generating the elements of the vector α^+ using recursively the model (C.1). On the other hand, the vectors $\hat{\alpha}$ and $\hat{\alpha}^+$ are obtain by smoothing the vectors α and α^+. Durbin and Koopman [2002] use in their algorithm the state smoother of Koopman [1993]. The method of state smoothing consists of three steps. First, application of the Kalman filter, given by the following equations for the model (C.1):

$$\begin{gathered} v_t = y_t - Z_t a_t, \qquad F_t = Z_t P_t Z_t' + H_t, \qquad K_t = T_t P_t Z_t' F_t^{-1} \\ L_t = T_t - K_t Z_t, \quad a_{t+1} = T_t a_t + K_t v_t \\ P_{t+1} = T_t P_t L_t' + R_t Q_t R_t' \end{gathered} \tag{C.2}$$

for $t = 1, ..., N$ with a_1 and P_1 as the mean vector and variance matrix of the initial state vector α_1. Second, the following disturbance smoother is used to compute the vector:

$$\hat{w}_t = \begin{bmatrix} H_t F_t^{-1} & -H_t K_t' \\ 0 & Q_t R_t' \end{bmatrix} \begin{pmatrix} v_t \\ r_t \end{pmatrix}, \quad \hat{w} = (\hat{w}_1', ..., \hat{w}_N')' \tag{C.3}$$

where r_t is evaluated by the backwards recursion:

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t \tag{C.4}$$

for $t = N, N-1, ..., 1$ with $r_N = 0$. Third, the state smoother is given by:

$$\hat{\alpha}_{t+1} = T_t \hat{\alpha}_t + R_t Q_t R_t' r_t \tag{C.5}$$

with the initialization $\hat{\alpha}_1 = a_1 + P_1 r_0$.

Consequently, the three steps of the algorithm are as follows:

Step 1. Obtain a random vector w^+ from the density $p(w)$. Then, generate α^+ and y^+ by means of recursion of the model (C.1) with w^+ is used instead of w. The recursion is initialized by the draw $\alpha_1^+ \sim \mathcal{N}(a_1, P_1)$.

Step 2. Compute $\hat{\alpha}$ and $\hat{\alpha}^+$ by mean of the Kalman filter (C.2) forward, the disturbance smoother (C.3) and (C.4) backward and the state smoother (C.5) forward.

Step 3. Calculate $\tilde{\alpha} = \hat{\alpha} - \hat{\alpha}^+ + \alpha^+$.

As Durbin and Koopman [2002] say, when a single draw $\tilde{\alpha}$ is required, it is computationally more efficient to compute $\tilde{\alpha}$ by constructing the artificial observations $y^* = y - y^+$ and using $\tilde{\alpha} = \hat{\alpha}^* + \alpha^+$ where $\hat{\alpha}^* = E[\alpha | y^*]$.

Appendix D

Algorithm of the simulation smoother for nonlinear state space models

A description of the algorithm of the simulation smoother used in the Bayesian estimation of Unobserved RW and STAR Components Models of Chapter 6 is developed in this appendix.

A general nonlinear state space model can be defined by the equations:

$$\begin{array}{lcll} y_t & = & Z_t(\alpha_t) + \epsilon_t, & \epsilon_t \sim \mathcal{N}(0, H_t) \\ \alpha_{t+1} & = & T_t(\alpha_t) + R_t\eta_t, & \eta_t \sim \mathcal{N}(0, Q_t) \end{array} \tag{D.1}$$

where $t = 1, ..., N$, y_t is an $m \times 1$ vector of observations, α_t is a $p \times 1$ state vector and ϵ_t and η_t are vectors of disturbances. Matrices R_t, H_t and Q_t are assumed to be known. Moreover, one assumes that $\alpha_1 \sim \mathcal{N}(a_1, P_1)$, where a_1 and P_1 are known. The function $Z_t(.)$ is supposed to be linear here and, on the other hand, the function $T_t(.)$ is supposed to be nonlinear.

The current algorithm of the simulation smoother is adapted from Durbin and Koopman [2002] by using the extended Kalman filter (EKF) described by Koopman and Lee [2008]. The simulation of Durbin and Koopman [2002] is treated in Appendix C. Thus, only the EKF is developed here. The EKF provides approximation estimates of the state by applying the standard Kalman filter to the Taylor approximation of the nonlinear functions expanded around the estimated state from the filter. The first order approximation to the state equation is given by:

$$\alpha_{t+1} \approx T(a_{t|t}) + \tilde{T}_t \cdot (\alpha_t - a_{t|t}) + u_t \qquad \text{where} \quad \tilde{T}_t = \left.\frac{\partial T(\alpha_t)}{\partial \alpha_t'}\right|_{\alpha_t = a_{t|t}}.$$

The equations of the extended Kalman filter are given by the following equations for the model (D.1):

$$\begin{aligned} v_t = y_t - Z_t a_t, \qquad F_t = Z_t P_t Z_t' + H_t, \qquad K_t = \tilde{T}_t P_t Z_t' F_t^{-1} \\ L_t = \tilde{T}_t - K_t Z_t, \qquad a_{t+1} = T_t(a_t) + K_t v_t \\ P_{t+1} = \tilde{T}_t P_t L_t' + R_t Q_t R_t'. \end{aligned} \tag{D.2}$$

Moreover, the state smoother is given by:

$$r_{t-1} = Z_t F_t^{-1} v_t + L_t' r_t \qquad \text{and} \qquad \hat{\alpha}_{t+1} = a_t + P_t r_{t-1} \tag{D.3}$$

for $t = N, N-1, ..., 1$ with $r_N = 0$.

The adaptation of the three steps algorithm described in Appendix C is presented above. The notations are defined in Appendix C.

Step 1. Obtain a random vector w^+ from the density $p(w)$. Then, generate α^+ and y^+ by means of recursion of the model (D.1) with w^+ is used instead of w. The recursion is initialized by the draw $\alpha_1^+ \sim \mathcal{N}(a_1, P_1)$.

Step 2. Compute $\hat{\alpha}$ and $\hat{\alpha}^+$ by mean of the extended Kalman filter (D.2) and the state smoother (D.3).

Step 3. Calculate $\tilde{\alpha} = \hat{\alpha} - \hat{\alpha}^+ + \alpha^+$.

D.1 Application to the RW-ESTAR(2,1) model

The matrix $\tilde{T}_t$ depends on the form of the nonlinear function $T_t(.)$. Below, the analytic calculi are detailed for six unobserved RW and STAR components models. In the case of the RW-ESTAR(2,1) model, we have:

$$\tilde{T}_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & T_{22} & T_{23} \\ 0 & 1 & 0 \end{bmatrix}$$

with:

$$T_{22} = \phi_1 + \theta_1 \cdot \Big(1 - \exp[-\gamma^2 \cdot c_{t|t}^2]\Big) + (\theta_1 \cdot c_{t|t} + \theta_2 \cdot \alpha_{3,t|t}) \cdot 2 \cdot \gamma^2 \cdot c_{t|t} \cdot \exp[-\gamma^2 \cdot c_{t|t}^2]$$

and:

$$T_{23} = \phi_2 + \theta_2 \cdot \Big(1 - \exp[-\gamma^2 \cdot c_{t|t}^2]\Big).$$

D.2 Application to the RW-ESTAR(2,2) model

The RW-ESTAR(2,2) model can be written as follows:

$$\begin{array}{rcl} y_t & = & p_t + c_t \\ p_{t+1} & = & p_t + w_{1,t} \\ c_{t+1} & = & (\phi_1 \cdot c_t + \phi_2 \cdot \alpha_{3,t}) \\ & & + (\theta_1 \cdot c_t + \theta_2 \cdot \alpha_{3,t}) \cdot \Big(1 - \exp\Big[-\gamma^2 \cdot \alpha_{3,t}^2\Big]\Big) + w_{2,t} \\ \alpha_{3,t+1} & = & c_t. \end{array} \tag{D.4}$$

Thus, we have:

$$\tilde{T}_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & T_{22} & T_{23} \\ 0 & 1 & 0 \end{bmatrix}$$

with:

$$T_{22} = \phi_1 + \theta_1 \cdot \Big(1 - \exp[-\gamma^2 \cdot \alpha_{3,t|t}^2]\Big)$$

and:

$$\begin{array}{c} T_{23} = \phi_2 + \theta_2 \cdot \Big(1 - \exp[-\gamma^2 \cdot \alpha_{3,t|t}^2]\Big) \\ +(\theta_1 \cdot c_{t|t} + \theta_2 \cdot \alpha_{3,t|t}) \cdot 2 \cdot \gamma^2 \cdot \alpha_{3,t|t} \cdot \exp[-\gamma^2 \cdot \alpha_{3,t|t}^2]. \end{array}$$

D.3 Application to the RW-ESTAR(2,3) model

The RW-ESTAR(2,3) model can be written as follows:

$$\begin{array}{rcl} y_t & = & p_t + c_t \\ p_{t+1} & = & p_t + w_{1,t} \\ c_{t+1} & = & (\phi_1 \cdot c_t + \phi_2 \cdot \alpha_{4,t}) \\ & & + (\theta_1 \cdot c_t + \theta_2 \cdot \alpha_{4,t}) \cdot \Big(1 - \exp\Big[-\gamma^2 \cdot \alpha_{3,t}^2\Big]\Big) + w_{2,t} \\ \alpha_{3,t+1} & = & \alpha_{4,t} \\ \alpha_{4,t+1} & = & c_t. \end{array} \tag{D.5}$$

Thus, we have:

$$\tilde{T}_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & T_{22} & T_{23} & T_{24} \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

with:

$$T_{22} = \phi_1 + \theta_1 \cdot \Big(1 - \exp[-\gamma^2 \cdot \alpha_{3,t|t}^2]\Big)$$

$$T_{23} = (\theta_1 \cdot c_{t|t} + \theta_2 \cdot \alpha_{4,t|t}) \cdot 2 \cdot \gamma^2 \cdot \alpha_{3,t|t} \cdot \exp[-\gamma^2 \cdot \alpha_{3,t|t}^2]$$

and:

$$T_{24} = \phi_2 + \theta_2 \cdot \Big(1 - \exp[-\gamma^2 \cdot \alpha_{3,t|t}^2]\Big).$$

D.4 Application to the RW-ESTAR(3,1) model

The RW-ESTAR(3,1) model can be written as follows:

$$\begin{array}{rcl} y_t & = & p_t + c_t \\ p_{t+1} & = & p_t + w_{1,t} \\ c_{t+1} & = & (\phi_1 \cdot c_t + \phi_2 \cdot \alpha_{4,t} + \phi_3 \cdot \alpha_{3,t}) \\ & & + (\theta_1 \cdot c_t + \theta_2 \cdot \alpha_{4,t} + \theta_3 \cdot \alpha_{3,t}) \\ & & \cdot \Big(1 - \exp\Big[-\gamma^2 \cdot c_t^2\Big]\Big) + w_{2,t} \\ \alpha_{3,t+1} & = & \alpha_{4,t} \\ \alpha_{4,t+1} & = & c_t. \end{array} \tag{D.6}$$

Thus, we have:

$$\tilde{T}_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & T_{22} & T_{23} & T_{24} \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

with:

$$T_{22} = \phi_1 + \theta_1 \cdot \Big(1 - \exp[-\gamma^2 \cdot c_{t|t}^2]\Big)$$
$$+(\theta_1 \cdot c_{t|t} + \theta_2 \cdot \alpha_{4,t|t} + \theta_3 \cdot \alpha_{3,t|t}) \cdot 2 \cdot \gamma^2 \cdot c_{t|t} \cdot \exp[-\gamma^2 \cdot c_{t|t}^2]$$

$$T_{23} = \phi_3 + \theta_3 \cdot \Big(1 - \exp[-\gamma^2 \cdot c_{t|t}^2]\Big)$$

and:

$$T_{24} = \phi_2 + \theta_2 \cdot \Big(1 - \exp[-\gamma^2 \cdot c_{t|t}^2]\Big).$$

D.5 Application to the RW-ESTAR(3,2) model

The RW-ESTAR(3,2) model can be written as follows:

$$\begin{array}{rcl} y_t & = & p_t + c_t \\ p_{t+1} & = & p_t + w_{1,t} \\ c_{t+1} & = & (\phi_1 \cdot c_t + \phi_2 \cdot \alpha_{4,t} + \phi_3 \cdot \alpha_{3,t}) \\ & & + (\theta_1 \cdot c_t + \theta_2 \cdot \alpha_{4,t} + \theta_3 \cdot \alpha_{3,t}) \\ & & \cdot \Big(1 - \exp\Big[-\gamma^2 \cdot \alpha_{4,t}^2\Big]\Big) + w_{2,t} \\ \alpha_{3,t+1} & = & \alpha_{4,t} \\ \alpha_{4,t+1} & = & c_t. \end{array} \tag{D.7}$$

Thus, we have:

$$\tilde{T}_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & T_{22} & T_{23} & T_{24} \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

with:

$$T_{22} = \phi_1 + \theta_1 \cdot \Big(1 - \exp[-\gamma^2 \cdot \alpha_{4,t|t}^2]\Big)$$

$$T_{23} = \phi_3 + \theta_3 \cdot \Big(1 - \exp[-\gamma^2 \cdot \alpha_{4,t|t}^2]\Big)$$

and:

$$\begin{array}{c} T_{24} = \phi_2 + \theta_2 \cdot \Big(1 - \exp[-\gamma^2 \cdot \alpha_{4,t|t}^2]\Big) \\ +(\theta_1 \cdot c_{t|t} + \theta_2 \cdot \alpha_{4,t|t} + \theta_3 \cdot \alpha_{3,t|t}) \cdot 2 \cdot \gamma^2 \cdot \alpha_{4,t|t} \cdot \exp[-\gamma^2 \cdot \alpha_{4,t|t}^2]. \end{array}$$

D.6 Application to the RW-ESTAR(3,3) model

The RW-ESTAR(3,3) model can be written as follows:

$$\begin{array}{rcl} y_t & = & p_t + c_t \\ p_{t+1} & = & p_t + w_{1,t} \\ c_{t+1} & = & (\phi_1 \cdot c_t + \phi_2 \cdot \alpha_{4,t} + \phi_3 \cdot \alpha_{3,t}) \\ & & + (\theta_1 \cdot c_t + \theta_2 \cdot \alpha_{4,t} + \theta_3 \cdot \alpha_{3,t}) \\ & & \cdot \Big(1 - \exp\Big[-\gamma^2 \cdot \alpha_{3,t}^2\Big]\Big) + w_{2,t} \\ \alpha_{3,t+1} & = & \alpha_{4,t} \\ \alpha_{4,t+1} & = & c_t. \end{array} \tag{D.8}$$

Thus, we have:

$$\tilde{T_t} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & T_{22} & T_{23} & T_{24} \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

with:

$$T_{22} = \phi_1 + \theta_1 \cdot \Big(1 - \exp[-\gamma^2 \cdot \alpha^2_{3,t|t}]\Big)$$

$$T_{23} = \phi_3 + \theta_3 \cdot \Big(1 - \exp[-\gamma^2 \cdot \alpha^2_{3,t|t}]\Big)$$
$$+(\theta_1 \cdot c_{t|t} + \theta_2 \cdot \alpha_{4,t|t} + \theta_3 \cdot \alpha_{3,t|t}) \cdot 2 \cdot \gamma^2 \cdot \alpha_{3,t|t} \cdot \exp[-\gamma^2 \cdot \alpha^2_{3,t|t}]$$

and:

$$T_{24} = \phi_2 + \theta_2 \cdot \Big(1 - \exp[-\gamma^2 \cdot \alpha^2_{3,t|t}]\Big).$$

Computational Details

The algorithms have been written in the R language, version 2.6.2 (see R Development Core Team [2007]). Most algorithms of the Bayesian estimation have been written solely using the R package *base* 2.11, but the R packages *MCMCpack* 1.0-6 and *tseries* 0.10-22 have also been used to perform specific tasks. The R program itself and packages are available from CRAN at `http://CRAN.R-project.org`.

A new algorithm was used for the frequentist estimation of the STAR models in Chapter 2. The optimal starting values of the NLS estimation was found using the Differential Evolution algorithm initially developed by Rainer Storn (see Price et al. [2005]) and implemented in the R package *DEoptim* 2.0-5 by Ardia [2007]. More details were given in Section 2.5.

Finally, all graphics were obtained using the EViews 6 software. Practice had shown that this software was user-friendly and yielded high quality graphics.

Abbreviations

AC	Autocorrelation
ADF	Augmented Dickey-Fuller
AIC	Akaike information criterion
AR	Autoregressive
ARCH	Autoregressive conditional heteroscedasticity
ARMA	Autoregressive moving average
BF	Bayes factor
BIC	Schwarz (or Bayesian) information criterion
BS	Bridge sampling
CHF	Swiss franc
DGP	Data generating process
DIC	Deviance information criterion
DM	Diebold and Mariano
EK	Excess kurtosis
EKF	Extended Kalman filter
ESTAR	Exponential STAR
GBP	Great Britain pound
HAC	Heteroscedasticity and autocorrelation consistent
HPD	Highest prediction density
ICMH	Independence chain Metropolis-Hastings
iid	Independent and identically distributed
IS	Importance sampling
JB	Jarque-Bera
JLR	Johansen likelihood ratio
LM	Lagrange multiplier
LSTAR	Logistic STAR
MADF	Multivariate augmented Dickey-Fuller
MAPE	Mean absolute prediction errors
MCMC	Markov chain Monte Carlo
MRSTAR	Multiple regime STAR
MS-AR	Markov switching autoregressive
MSPE	Mean squared prediction errors

(continued)

NL	Nonlinear
NLS	Nonlinear least squares
OLS	Ordinary least squares
PP	Phillips-Perron
PPP	Purchasing power parity
RI	Reciprocal importance sampling
RNE	Relative numerical efficiency
RW	Random walk
RWI(2)	Integrated random walk process of order I(2)
RWMH	Random walk chain Metropolis-Hastings
SI	Serial independence
SK	Skewness
SSR	Sum of squared residuals
STAR	Smooth transition autoregressive
SUR	Seemingly unrelated regression
UC	Unobserved components
USD	United States dollar
VAR	Vector autoregressive

Bibliography

H. Akaike. A new look at the statistical model identification. *IEEE Transaction in Automatic Control*, AC-19:716–723, 1974.

T.W. Anderson. *An introduction to multivariate statistical analysis.* Wiley-Interscience, 3rd edition, 2003.

D. Ardia. *DEoptim package : Differential Evolution Optimization in R. R package version 1.2-1*, 2007.

G.M. Caporale, N. Pittis, and P. Sakellis. Testing for ppp : The erratic behaviour of unit root tests. *Economics Letters*, 80:277–284, 2003.

P. DeJong and N. Shephard. The simulation smoother for time series models. *Biometrika*, 82(2):339–350, 1995.

P.J. Deschamps. Comparing smooth transition and markov switching autoregressive models of us unemployment. *Journal of Applied Econometrics*, 23:435–462, 2008.

T.J. DiCiccio, R.E. Kass, A. Raftery, and L. Wasserman. Computing bayes factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association*, 92(439):903–915, 1997.

F.X. Diebold and R.S. Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263, 1995.

B. Dumas. Dynamic equilibrium and the real exchange rate in a spatially separated world. *The Review of Financial Studies*, 5(2):153–180, 1992.

J. Durbin and S.J. Koopman. A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–615, 2002.

J. Durbin and S.J. Koopman. *Times Series Analysis by State Space Methods.* Oxford University Press, 2001.

O. Eitrheim and T. Teräsvirta. Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics*, 74:59–72, 1996.

C. Engel and J.D. Hamilton. Long swings in the dollar : Are they in the data and do markets know it? *American Economic Review*, 80(4):689–713, 1990.

C.M. Engel. Long-run ppp may not hold after all. *Journal of International Economics*, 51(2):243–273, 2000.

C.M. Engel and C.-J. Kim. The long-run u.s.\u.k. real exchange rate. *Journal of Money, Credit and Banking*, 31(3):335–356, 1999.

R.F. Engle and C.W.J. Granger. Co-integration and error correction : Representation, estimation and testing. *Econometrica*, 55(2):251–276, 1987.

M.K. Francke, S.J. Koopman, and A.F. de Vos. Likelihood functions for state space models with diffuse initial conditions. Tinbergen Institute Discussion Papers 08-040/4, Tinbergen Institute, 2008.

M. Frömmel. The dollar exchange rate, adjustment to ppp, and the interest rate differential. Discussion paper, Ghent University, 2007.

S. Frühwirth-Schnatter. Fully bayesian analysis of switching gaussian state space models. *Annals of the Institute of Statistical Mathematics*, 53(1): 31–49, 2001.

S. Frühwirth-Schnatter. Estimating marginal likelihoods for mixture and markov switching models using bridge sampling techniques. *Econometrics Journal*, 7:143–167, 2004.

A.E. Gelfand and D.K. Dey. Bayesian model choice : Asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B*, 56 (3):501–514, 1994.

J. Geweke. Getting it right : Joint distribution tests of posterior simulators. *Journal of the American Statistical Association*, 99:799–804, 2004.

J. Geweke. Baysian inference in econometric models using monte carlo integration. *Econometrica*, 57(6):1317–1339, 1989.

J. Geweke. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J.M. Bernardo, J.O. Dawid, and A.P. Smith, editors, *Bayesian Statistics 4*, pages 169–193. Oxford University Press, 1992.

V. Haggan and T. Ozaki. Modelling nonlinear random vibrations using an amplitude-dependent autoregressive times series model. *Biometrika*, 68 (1):189–196, 1981.

J.D. Hamilton. *Time series analysis*. Princeton University Press, 1994.

S. Johansen. Statistical analysis of cointegrating vectors. *Journal of Economic Dynamics and Control*, 12(2-3):231–254, 1988.

S. Johansen. Estimation and hypothesis testing of cointegrating vectors in gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580, 1991.

G. Kapetanios, Y. Shin, and A. Snell. Testing for a unit root in the nonlinear star framework. *Journal of Econometrics*, 112(2):359–379, 2003.

R.E. Kass and A.E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

L. Kilian and M.P. Taylor. Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics*, 60(1): 85–107, 2003.

R. Kleijn and H.K. van Dijk. A bayesian analysis of the ppp puzzle using an unobserved components model. Tinbergen Institute Discussion Papers 01-105/4, Tinbergen Institute, 2001.

L.A. Klimko and P.I. Nelson. On conditional least squares estimation for stochastic processes. *Annals of Statistics*, 6(3):629–642, 1978.

G. Koop. *Bayesian Econometrics*. Wiley, 2003.

S.J. Koopman. Disturbance smoother for state space models. *Biometrika*, 80(1):117–126, 1993.

S.J. Koopman and K.M. Lee. Seasonality with trend and cycle interactions in unobserved components models. Tinbergen Institute Discussion Papers 08-028/4, Tinbergen Institute, 2008.

G.M. Ljung and G.E.P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65:297–303, 1978.

H.F. Lopes and E. Salazar. Bayesian model uncertainty in smooth transition autoregressions. *Journal of Time Series Analysis*, 27(1):99–117, 2006.

S. Lundbergh and T. Teräsvirta. Forecasting with smooth transition autoregressive models. In M.P. Clements and D.F. Hendry, editors, *A Companion to Economic Forecasting*, chapter 21, pages 485–509. Blackwell, 2002.

D.G. Maringer and M. Meyer. Smooth transition autoregressive models – new approaches to the model selection problem. *Studies in Nonlinear Dynamics & Econometrics*, 12(1), 2008.

X.-L. Meng. Posterior predictive p-values. *The Annals of Statistics*, 22(3): 1142–1160, 1994.

X.-L. Meng and W.H. Wong. Simulating ratios of normalizing constants via a simple identity : A theoretical exploration. *Statistica Sinica*, 6:831–860, 1996.

P. Michael, A.R. Nobay, and D.A. Peel. Transactions costs and nonlinear adjustment in real exchange rates : An empirical investigation. *The Journal of Political Economy*, 105(4):862–879, 1997.

M. Obstfeld and K. Rogoff. The six major puzzles in international macroeconomics : Is there a common cause? *NBER Macroeconomics Annual*, 15:339–390, 2000.

K.V. Price, R.M. Storn, and J.A. Lampinen. *Differential evolution - A practical approach to global optimization*. Springer, 2005.

R Development Core Team. *R : A Language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2007. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

D.E. Rapach and M.E. Wohar. The out-of-sample forecasting performance of nonlinear models of real exchange rate behavior. *International Journal of Forecasting*, 22(2):341–361, 2006.

K. Rogoff. The purchasing power parity puzzle. *Journal of Economic Literature*, 34(2):647–668, 1996.

N. Sarantis. Modeling non-linearities in real effective exchange rates. *Journal of International Money and Finance*, 18(1):27–45, 1999.

L. Sarno and P. Taylor. *The economics of exchange rates*. Cambridge University Press, 2003.

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

F.C. Schweppe. Evaluation of likelihood functions for gaussian signals. *IEEE Transactions on Information Theory*, 11:61–70, 1965.

J. Skalin and T. Teräsvirta. Modeling asymmetries and moving equilibria in unemployment rates. *Macroeconomic Dynamics*, 6(2):202–241, 2002.

D.J. Spiegelhalter, N.G. Best, B.P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit (with discussion). *Journal of The Royal Statistical Society: Series B*, 64(4):583–639, 2002.

M.P. Taylor and L. Sarno. The behavior of real exchange rates during the post-bretton woods period. *Journal of International Economics*, 46:281–312, 1998.

M.P. Taylor, D.A. Peel, and L. Sarno. Nonlinear mean-reversion in real exchange rates : Towards a solution to the purchasing power parity puzzles. *International Economic Review*, 42(4):1015–1042, 2001.

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

T. Teräsvirta and H.M. Anderson. Characterizing nonlinearities in business cycle using smooth transition autoregressive models. *Journal of Applied Econometrics*, 7:S119–S136, 1992.

D. van Dijk and P.H. Franses. Modeling multiple regimes in the business cycle. *Macroeconomic Dynamics*, 3(3):311–340, 1999.

D. van Dijk and P.H. Franses. Selecting a nonlinear time series model using weighted tests of equal forecast accuracy. *Oxford Bulletin of Economics and Statistics*, 65(s1):727–744, 2003.

D. van Dijk, T. Teräsvirta, and P.H. Franses. Smooth transition autoregressive models - a survey of recent developments. *Econometric Reviews*, 21 (1):1–47, 2002.

D. van Dijk, P.H. Franses, M.P. Clements, and J. Smith. On setar nonlinearity and forecasting. *Journal of Forecasting*, 22:359–375, 2003.

K.D. West and M.W. McCracken. Regression-based tests of predictive ability. *International Economic Review*, 39(4):817–840, 1998.