Long Swings in Exchange Rates: Are They Really in the Data?

by Franc Klaassen *

Department of Economics
University of Amsterdam

September 25, 2003

Abstract
Several papers report evidence of long swings in U.S. dollar exchange rates by rejecting the random walk in favor of a Markov regime-switching model. We show that this evidence is not robust to an extension of the sample period. One should, however, not conclude that long swings are thus absent, because the tests may have insufficient power. A possible reason is the low data frequency, because existing tests are based on quarterly data and we find that the power can be increased substantially by raising the data frequency. Indeed, weekly data reject the random walk, so that we conclude long swings are in the data. This conclusion is supported by significant superiority of regime-switching over random walk out of sample forecasts of the direction of change.

Key words: data frequency, forecasting, Markov regime-switching, random walk, unidentified parameters.

JEL classification: F31, C52, C53.

*Department of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB, Amsterdam, the Netherlands; tel: +31-20-5254191; fax: +31-20-5254254; E-mail: f.klaassen@uva.nl. I thank Geert Bekaert, Peter Boswijk, Herman Bierens, Harry Huizinga, Henk Jager, Frank de Jong, Frank Kleibergen, Siem Jan Koopman, Bertrand Melenberg, Dirk Veestraten, Kenneth West, an editor and an associate editor and two referees for very useful and constructive comments. An earlier version has appeared as CentER Tilburg University Discussion Paper No. 9908.
1 Introduction

Modeling exchange rates has been a main endeavor. Several structural models of exchange rate determination have been developed, but their empirical validity is often questioned, especially in the short-run (MacDonald and Taylor (1992)). Hence, many researchers have used the random walk, particularly since Meese and Rogoff’s (1983) conclusion that random walk forecasts outperform those from structural exchange rate models.

Nevertheless, the random walk remains unsatisfactory from an economic point of view, as it ignores the impact of economic fundamentals. Moreover, time plots of exchange rates suggest some structural pattern. For instance, the thick lines in the three panels of figure 1 (logarithm of weekly dollar prices of a German mark, Japanese yen and U.K. pound, respectively, from April 1974 to July 2003) indicate persistent appreciation and depreciation periods, that is, long swings.

The swings, however, may also be just a pattern imposed by the eye on the realization of a random walk. To resolve the question whether long swings are a systematic part of exchange rates, one needs a formal test. The aim of this paper is to do such a test using data on the three dollar rates just mentioned.

The outcome of the test will not only help understand the time path of exchange rates, but is also worthwhile for other reasons. For example, swings may originate from economic fundamentals. After all, if economic growth is relevant for exchange rates, then business cycle differences between countries can lead to long swings in exchange rates. Moreover, changes in economic policy may affect the exchange rate generating process (Lucas (1976) critique). Kaminsky (1993), for example, shows theoretically that a change from a contractionary to an expansionary monetary policy regime increases the exchange rate depreciation, resulting in long swings. Hence, knowing whether swings exist could guide future research on exchange rate determination.


¹ In contrast to these papers, which focus on the mean of a series, regime-switching models can also be useful to describe the variance. Persistence of regimes with different unconditional variances can explain part of the conditional heteroskedasticity we often find in high-frequency data. Cai (1994), Hamilton and Susmel (1994), Gray (1996), Bollen, Gray and Whaley (2000) and Klaassen (2002) use such “variance regimes” to model the variance of changes in interest rates, stock indices, interest rates, exchange rates and exchange rates, respectively.
(states) with potentially different mean exchange rate changes, presumably an appreciation and a depreciation regime. If the means are indeed different across both regimes and if these “mean” regimes are persistent, then there exist long swings. If the means are the same across regimes, then the exchange rate follows a random walk (with drift). The test for long swings is therefore a test of the null hypothesis that the means are equal across regimes against the alternative of different means and persistent regimes.

Several authors have already tested for long swings (Engel and Hamilton (1990), Engel (1994)). Using quarterly data up to about 1986, they conclude that long swings exist.

Given that we now have more data, we can redo their tests with quarterly data up to 2003. The results are changed in the sense that there is no longer convincing evidence of long swings. This lack of evidence is confirmed by applying Hansen’s (1992, 1996) testing methodology, which is particularly designed for testing null hypotheses such as the one given above in a Markov switching model.

The lack of evidence, however, is no proof of the absence of long swings; it may be due to insufficient test power. A possible reason is the data frequency: even if swings exist and last for some quarters, sampling at the quarterly frequency may result in too few observations per swing to distinguish the swings from a random walk. We confirm this reasoning by a Monte Carlo simulation study.

To obtain a more powerful test for long swings, we thus increase the data frequency of the three observed dollar rates from quarterly to monthly and then to weekly. Applying Hansen’s test yields our final answer concerning the existence of long swings in exchange rates.

In the next section we describe the regime-switching model and the procedure to test for long swings. Section 3 contains the empirical results. There, we actually test for the existence of long swings and examine the role of the data frequency. That section also presents the estimation results and an out of sample forecasting study that compares long swings to random walk based forecasts. Section 4 concludes.

2 Model and Test for Long Swings

2.1 Regime-Switching Model

Since we want to use a regime-switching model for weekly exchange rates, we have to generalize the basic Hamilton (1989) regime-switching model to incorporate the high-frequency features of exchange rates, in particular conditional heteroskedasticity and leptokurtosis of the innovation. Therefore, we introduce an extended regime-
switching model that includes a generalized autoregressive conditional heteroskedasticity (GARCH) model for the variance (see Bollerslev, Chou and Kroner (1992) for an overview of GARCH) and a t-distribution for the innovation. The model consists of four elements, namely the mean, regime process, variance and distribution. We discuss these elements subsequently and relate our specification to the one in Engel and Hamilton (1990).

Let $S_t$ denote the logarithm of the spot exchange rate at time $t$, that is, the domestic currency price of one unit of foreign currency. We concentrate on the exchange rate change $s_t = 100(S_t - S_{t-1})$, so that $s_t$ is the percentage depreciation of the domestic currency from time $t-1$ to $t$.

The mean of $s_t$ depends on the (unobservable) regime the process is in. As in Engel and Hamilton (1990), we assume that there are two regimes. Let $r_t \in \{1, 2\}$ denote the regime at time $t$. Within this regime, the mean exchange rate depreciation is $\mu_{rt}$, and we identify the first regime as the low mean regime: $\mu_1 \leq \mu_2$.

We assume that within each regime the mean conditional on previous exchange rate changes is constant over time, though it is possible to allow for autoregressive terms, among other things. Hence the mean equation is

$$s_t = \mu_{rt} + \varepsilon_t,$$

where the expectation of the innovation $\varepsilon_t$ is zero conditional on the current regime $r_t$ and the information set of the data generating process. This set consists of two parts. The first one, $I_{t-1} = (s_{t-1}, s_{t-2}, \ldots)$, denotes the information that is observed by the econometrician; the second part, $\tilde{r}_{t-1}$, is the regime path $(r_{t-1}, r_{t-2}, \ldots)$, which is not observed. We thus get $E\{\varepsilon_t|I_{t-1}, \tilde{r}_t\} = 0$. Using the subscript $t-1$ below an operator (probability, expectation or variance) as short-hand notation for conditioning on $I_{t-1}$, this gives $E_{t-1}\{\varepsilon_t|\tilde{r}_t\} = 0$.

The persistence of both regimes depends on the regime-staying probabilities. Let $p_{t-1}(r_t|\tilde{r}_{t-1})$ denote the probability of going to regime $r_t$ at time $t$ conditional on the information set of the data generating process. As in Engel and Hamilton (1990), we assume that $r_t$ follows a first-order Markov process with constant staying probabilities, so that

$$p_{t-1}(r_t|\tilde{r}_{t-1}) = p(r_t|r_{t-1}) = \begin{cases} p_{11} & \text{if } r_t = r_{t-1} = 1 \\ p_{22} & \text{if } r_t = r_{t-1} = 2. \end{cases}$$

The mean and regime processes just described are able to capture long swings in exchange rates. After all, if $\mu_1 < \mu_2$ and if $p_{11}$ and $p_{22}$ are high, then being in the first regime leads to a probably long period of appreciation (assuming for simplicity $\mu_1 < 0$)
and, after a switch to the second regime, to a probably long period of depreciation (assuming $\mu_2 > 0$). Hence, there are long swings. Note, however, that the model does not impose the existence of swings; we do allow for $\mu_1 = \mu_2$, so that exchange rates can have a constant mean, as in the random walk with drift.

When specifying the conditional variance, $V_{t-1}\{e_t|\tilde{r}_t\}$, Engel and Hamilton (1990) assume that this variance is constant within a regime. This may well be fine for the quarterly data of Engel and Hamilton, but is problematic for our weekly data. After all, if mean regimes are persistent (a few years according to Engel and Hamilton), the variance is also constant for a long time. In particular for high-frequency data this is problematic, as it is well-known that there is conditional heteroskedasticity. Moreover, if the model with constant regime-specific variances is estimated with weekly data, it may well be that the regimes are exploited to capture the strong conditional heteroskedasticity instead of the long swings in which we are interested. This may yield a significant test for the existence of two regimes, even if there are no long swings.

The second feature of the Engel and Hamilton (1990) variance specification is that the variance is different across the two mean regimes. As the authors admit, the perfect dependence between mean and variance can be problematic. For instance, if the appreciation regime is associated with high volatility, a period of unusual volatility can force the process into this appreciation regime, even when the currency is actually depreciating. Moreover, economists are not convinced that there is any relation between the mean and the variance of exchange rates (for instance, see Engle, Ito and Lin (1990)).

As a solution for both issues we disconnect the variance from the regime process, which is thus completely focussed at the long swings. For the variance specification we use the popular GARCH approach.

A direct application of the standard GARCH(1,1) formula in our regime-switching model would define the conditional error variance as

$$V_{t-1}\{e_t|\tilde{r}_t\} = \omega + \alpha e_{t-1}^2 + \beta V_{t-2}\{e_{t-1}|\tilde{r}_{t-1}\}. \tag{3}$$

This specification, however, appears practically infeasible when estimating the model. In building the sample log-likelihood, the econometrician first uses (1) to express the unobserved $e_{t-1}^2$ in terms of the conditioning variables by $e_{t-1}^2 = (s_{t-1} - \mu_{r_{t-1}})^2$. Hence, $V_{t-1}\{e_t|\tilde{r}_t\}$ depends on the unobserved regime $r_{t-1}$. However, it also depends on the lagged variance $V_{t-2}\{e_{t-1}|\tilde{r}_{t-1}\}$, which depends on $r_{t-2}$ and $V_{t-3}\{e_{t-2}|\tilde{r}_{t-2}\}$, where the latter depends on $r_{t-3}$ and $V_{t-4}\{e_{t-3}|\tilde{r}_{t-3}\}$, and so on. Consequently, the conditional variance in (3) depends on the entire sequence of regimes up to time $t-1$. Since the number of possible combinations grows exponentially with $t-1$, this leads to an enormous
number of regime paths to $t-1$. The econometrician, who does not observe regimes, has to integrate out all possible regime paths. This renders estimation intractable.

To avoid this path-dependency problem, it is interesting to realize that the same problem also hampered the application of regime-switching GARCH models, where the conditional variance depends on the variance regime the process is in and where the conditional variance within each variance regime is governed by a GARCH process. For such models, Gray (1996) introduces a way to remove the path-dependence from the likelihood. We apply the basic idea behind that technique to solve the problem also in our regime-switching mean model. That is, we directly average out the regimes $r_{t-1}$ in the source of the path-dependence, $\varepsilon_{t-1}^2 = (s_{t-1} - \mu_{r_{t-1}})^2$, instead of only in the likelihood. This removes the regime-dependence of $V_{t-1}\{\varepsilon_t|\tilde{r}_t}\}$. To average out the regimes, we follow Klaassen’s (2002) approach, as that improves on Gray’s method. That is, we use the observed information $I_{t-1}$, so that $V_{t-1}\{\varepsilon_t|\tilde{r}_t}\} \text{ becomes equal to } V_{t-1}\{\varepsilon_t\}$:

$$V_{t-1}\{\varepsilon_t|\tilde{r}_t\} = V_{t-1}\{\varepsilon_t\} = \omega + \alpha E_{t-1}\{\varepsilon_{t-1}^2\} + \beta V_{t-2}\{\varepsilon_{t-1}\},$$  \hspace{1cm} (4)

with the usual GARCH restrictions $\omega > 0$ and $\alpha, \beta \geq 0$ to ensure $V_{t-1}\{\varepsilon_t\} > 0$ for all $t$. We also assume that $\alpha + \beta < 1$, so that the unconditional variance is $\sigma^2 = \frac{\omega}{1-\alpha-\beta}$.

The fourth and final element of the regime-switching model concerns the conditional distribution of exchange rate changes. Engel and Hamilton (1990) choose a normal distribution for their quarterly series. However, to allow for extra leptokurtosis in our weekly data, we follow other papers by taking a t-distribution (see Bollerslev, Chou and Kroner (1992)). It has $\nu$ degrees of freedom, zero mean, and variance $V_{t-1}\{\varepsilon_t\}$:

$$\varepsilon_t|I_{t-1}, \tilde{r}_t \sim t\left(\nu, 0, V_{t-1}\{\varepsilon_t\}\right).$$  \hspace{1cm} (5)

Equations (1), (2), (4) and (5) describe the complete regime-switching model. We estimate it by maximum likelihood using a sample of $s_t$ for $t = 1, ..., T$. The likelihood function, which has a convenient recursive structure, can be derived in a similar way as in Klaassen (2002). To start up the recursive process, we set the required variables equal to their expectation without conditioning on observable information $I_{t-1}$.

2.2 Testing Procedure for Long Swings

The central feature of the model presented in the previous subsection is the allowance for two persistent regimes with different mean depreciations $\mu_1$ and $\mu_2$, as these are able to capture long swings. Hence, to test for the existence of long swings, we test for the existence of such regimes. More formally, we test the null hypothesis of a
single-regime model \((\mu_1 = \mu_2)\), that is, the random walk with drift, against the regime-switching model of subsection 2.1. In this subsection we set out the testing procedure; the empirical outcomes follow in the next section.

Testing the null restriction \(\mu_1 = \mu_2\) is not straightforward. For instance, under the null there is in fact only one regime that governs the exchange rate, so that the regime-staying probabilities \(p_{11}\) and \(p_{22}\) are not identified; see Hansen (1992) for a discussion of other problems involved. This makes the asymptotic distribution of the usual tests (likelihood ratio, Wald and Lagrange multiplier) no longer \(\chi^2\) (Hansen (1992)).

Garcia (1998) proposes one way to overcome the problems. He derives the asymptotic distribution of the likelihood ratio test for several variants of the regime-switching model. However, the model of subsection 2.1, with the GARCH based variance and \(t\)-distribution for the innovation, is not included in Garcia’s paper.

A generally applicable solution for the testing problems mentioned above is given by Hansen (1992, 1996). His approach can be summarized as follows. The null restriction is equivalent to \(\mu_2 - \mu_1 = 0\). Under this null, \(p_{11}\) and \(p_{22}\) are not identified. The other parameters \((\mu_1, \sigma^2, \alpha, \beta, \nu)\), however, are identified. Hansen proposes to consider a fixed value for \((\mu_2 - \mu_1, p_{11}, p_{22})\). For this point, maximize the log-likelihood across the other five parameters, that is, concentrate out \((\mu_1, \sigma^2, \alpha, \beta, \nu)\). Subtracting the log-likelihood under the null and dividing this difference by its standard deviation yields the standardized likelihood ratio for the \((\mu_2 - \mu_1, p_{11}, p_{22})\) under consideration.

Hansen’s test statistic \(\hat{LR}^*\) is the supremum of these standardized likelihood ratios over all parameter combinations \((\mu_2 - \mu_1, p_{11}, p_{22})\) that are possible under the alternative. In practice Hansen suggests to take the supremum over a finite grid of parameter combinations. The asymptotic p-value of \(\hat{LR}^*\) is not known by itself, but Hansen shows that it is smaller than or equal to the asymptotic p-value using the distribution of another variable. He advises to use this upper bound, though it makes the test conservative (too few rejections of the null). He also explains how the bound can be approximated via simulation. Finally, the p-value bound depends on a bandwidth number \(M\), and Hansen suggests to compute the p-value for different \(M\) (see Hansen (1996) for details).

In this paper we follow Hansen’s method. Details on the implementation are in the empirical section below.

3 Empirical Results

In this section we use the regime-switching model and the testing apparatus developed above to address the central question of this paper, namely whether long swings really
exist. First, we describe the data. Then, we investigate existing tests for long swings. In subsection 3.3 we perform our own test. After that, we analyze the estimates of the regime-switching model. In subsection 3.5 we examine whether taking account of the long swings leads to better out of sample exchange rate forecasts than those generated by the random walk model.

3.1 Data

We use three U.S. dollar exchange rates, namely, the dollar vis-à-vis the German mark, the Japanese yen and the U.K. pound. These exchange rates have been chosen because of their important role on foreign exchange markets and because they behave relatively independently, for instance, compared to several dollar-EMS exchange rates. We have weekly data on exchange rate levels over the post-Bretton-Woods period from April 2, 1974 to June 17, 2003; for the euro period the German mark rates have been derived from the dollar/euro rates and the official mark/euro conversion rate. This gives $T = 1,524$ observations of the percentage dollar depreciations $s_t$. They have been obtained from Datastream.

The thick lines in the three panels in figure 1 show the behavior of the logarithm of the three exchange rates $S_t$ over the sample period. At first sight, exchange rates indeed seem to be characterized by long swings. However, this graphical suggestion is of course no formal evidence, as the long swings may be only a pattern imposed by the eye on the realization of a random walk. Hence, we need a test.

3.2 Existing Tests for Long Swings

As explained in subsection 2.2, testing for long swings is not straightforward. For instance, the regime-staying probabilities are not identified under the random walk null hypothesis (a single-regime model), which makes the asymptotic distribution of conventional tests no longer $\chi^2$. Several authors, such as Engel and Hamilton (1990) and Engel (1994), circumvent the testing problems by introducing an auxiliary assumption to nevertheless be able to test for long swings. This subsection describes their approach and investigates the robustness of their results to an extension of their sample period. Following their work, we here use quarterly data, obtained by elimination from the weekly level series described above.

The Engel and Hamilton (1990) approach, also followed by Engel (1994), is embedded in a Markov regime-switching model where the exchange rate change is normally distributed with mean $\mu_1$ and variance $\sigma_1^2$ in the first regime and normally distributed with $\mu_2$ and $\sigma_2^2$ in the second regime, where $(\mu_1, \sigma_1^2) \neq (\mu_2, \sigma_2^2)$. The model thus always
has two regimes. This is the auxiliary assumption that the authors make to circumvent the testing problems mentioned above, which originated from the allowance for one regime.

The authors employ two tests for the existence of long swings. The first one is based on the fact that there are no long swings if the two regimes have the same mean. The test therefore concerns the null hypothesis $\mu_1 = \mu_2$ against the alternative $\mu_1 \neq \mu_2$.

The second Engel and Hamilton test investigates whether the regimes are persistent in the sense that the probability of being in a regime depends on the previous regime. If this is not the case, then the probability of staying in the first regime equals the one of switching from the second to the first regime, so that there are no long swings. Hence, they test $p_{11} = 1 - p_{22}$ versus $p_{11} \neq 1 - p_{22}$.

For both tests the authors use the conventional likelihood ratio test with 5% critical value 3.84. We use the same significance level throughout the paper. Because the existence of long swings implies that both $\mu_1 \neq \mu_2$ and $p_{11} \neq 1 - p_{22}$, only a rejection of both tests is evidence in favor of long swings.

We first compute both tests using similar data as in Engel and Hamilton (1990) and Engel (1994), that is, the 1974:II through 1986:IV subsample of our data set. The likelihood ratios the for the $\mu_1 = \mu_2$ and $p_{11} = 1 - p_{22}$ tests, respectively, are (3.15, 3.07) for Germany, (6.86, 4.35) for Japan, and (9.43 , 8.01) for the U.K. The results point at the existence of long swings and thereby corroborate the conclusion of Engel and Hamilton (1990) and Engel (1994).

Now we extend the data range by including 1987:I through 2003:II. The likelihood ratios become (3.23, 1.04), (7.95, 2.73) and (3.31, 17.99), so that there is no evidence of long swings any more, except perhaps some indication for the U.K. For instance, the estimates for Japan, where $\mu_1 = \mu_2$ is rejected but not $p_{11} = 1 - p_{22}$, reveal that the estimated regime-staying probability $p_{11}$ is 0.441, implying that the low mean regime is short-lived instead of persistent, so that the estimates do not point at long swings.

We conclude that existing evidence of long swings depends on the sample period. Prolonging the data range no longer yields convincing evidence of long swings.

### 3.3 Long Swings in Exchange Rates: Are They Really in the Data?

As indicated in the introduction, the lack of evidence of long swings from quarterly data may be caused by the low data frequency. Therefore, in this subsection we analyze the effect of raising the data frequency for the test outcome. To take account of the potential conditional heteroskedasticity and leptokurtosis, we return to the model of subsection 2.1. We use Hansen’s (1992, 1996) testing methodology described in 2.2,
so that we do not have to make the auxiliary assumption that there are two regimes as in the previous subsection. To get further insight into the test, we first apply it to the same quarterly data as used in the previous subsection. After that we increase the frequency, so as to derive our final answer to the question whether long swings exist.

3.3.1 Quarterly Data

As usual for quarterly data, we find no evidence of conditional heteroskedasticity and non-normality. Hence, for the sake of parsimony, in this subsection we consider the model of subsection 2.1 restricted by $\alpha = \beta = 0$ and $\nu = \infty$. We test the null hypothesis $\mu_1 = \mu_2$, so the null model is the random walk with drift.

Since Hansen’s test is computed in practice by maximization of the standardized likelihood ratio over a grid of $(\mu_2 - \mu_1, p_{11}, p_{22})$, we first have to choose such a grid. This requires a specification of the parameter space for these coefficients under the alternative hypothesis of long swings. Hence, what values for $(\mu_2 - \mu_1, p_{11}, p_{22})$ are reasonable in case of long swings? We think that for the countries under consideration reasonable long appreciation and depreciation swings are not more than about 10 percentage points different on a quarterly basis, so that $\mu_2 - \mu_1$ should be at most 10. We take the grid $\mu_2 - \mu_1 \in \{0.5, 1, 2, ..., 10\}$.

The grid for the regime-staying probabilities $p_{11}$ and $p_{22}$ is based on what degree of persistence in swings is considered “long”. One usually thinks in terms of quarters or even years. Therefore, we assume that swings are called long if they last for at least, say, two quarters. Because the expected duration of a regime $r$ is $(1 - p_{rr})^{-1}$ (see Hamilton (1989)), this condition implies that $p_{rr} \geq 0.5$. However, if both $p_{11}$ and $p_{22}$ are small, $p_{11} + p_{22}$ is close to one, which implies that the current regime is almost independent of the previous one. This is in contrast with the idea of long swings. We therefore impose the extra restriction that $p_{11} + p_{22}$ is at least 1.2, say. The grid is $p_{11}, p_{22} \in \{0.50, 0.55, 0.60, ..., 0.95\}$ such that $p_{11} + p_{22} \geq 1.2$. The total number of grid points $(\mu_2 - \mu_1, p_{11}, p_{22})$ is 990.²

Table 1 presents the test results. The first column for each country reports them for the series from 1974:II through 1986:IV, similar to the data used by Engel and Hamilton (1990) and Engel (1994). The Hansen statistic $\hat{LR}^*$ is 1.95 for Germany, 2.28 for Japan and 2.69 for the U.K. The p-value upper bounds are presented for different bandwidth parameters $M$; they are based on 10,000 simulations. The bounds are above the level of 5%. Extending the sample period by including data from 1987:I through

²The precise grid choice is inevitably arbitrary. For many Hansen tests in the paper we have therefore experimented with other grids. The results from those grids are essentially the same.
2003:II raises the bounds even further. Hence, we again find no evidence of long swings in quarterly data.

3.3.2 The Relevance of Data Frequency for Test Power

The inability to reject the random walk may be caused by the absence of long swings. However, it may also be due to a lack of power of the tests in a situation where swings are present. One possible reason is that the data frequency is too low. After all, even if swings are part of the exchange rate generating process and last for some quarters, a sample of quarterly data may result in too few observations per swing to distinguish the swings from a random walk. Although this argument may sound intuitively appealing, there is no direct support for it yet. Therefore, we explore the idea in this subsection.

The most closely related paper is Coe (2002). He does a Monte Carlo experiment to compute the power of Garcia’s (1998) test for the presence of regimes in a homoskedastic normal model. He first takes estimates based on annual GDP data over 95 years as true parameter values for the Monte Carlo simulations of the regime-switching model. Then he repeats the study using estimates from quarterly GDP data over 37 years. Though Coe’s focus is not on the relevance of data frequency for test power, he does report an interesting result for our paper, that is, he finds a power of 0.48 for data sets containing 100 observations based on the annual estimates and a power of 0.89 for data sets of 400 observations generated from the quarterly estimates. This gain in power may come from an increase in data frequency for a fixed time span. However, because the underlying annual and quarterly estimates are essentially different, so that the simulated annual series are not simply the time aggregated versions of the simulated quarterly series, it is not clear what the contribution of the increase in data frequency to the gain in power really is. In addition, the impact of conditional heteroskedasticity and excess kurtosis cannot be inferred.

In the literature on unit root tests the relevance of data frequency for test power has received considerably more attention. The common idea is that a longer span of the data is more important than a higher frequency to raise the power, but that a higher frequency for a given time span can also provide significant improvements in the finite sample power (Maddala and Kim (1998, p.130)). The latter part, which is the relevant one for this paper since we have a fixed time span, is based on homoskedastic normal processes, as used in Choi and Cheung (1995), among others.

There is also recent evidence that for conditionally heteroskedastic and leptokurtic processes the power of unit root tests increases substantially if one explicitly accounts for both features (Seo (1999)). Because they can usually be captured with
high-frequency but not with low-frequency data, the result suggests a second reason why increasing the data frequency can lead to improvements in the power of unit root tests.

Because the current paper examines exchange rates, which have conditional heteroskedasticity and excess kurtosis in high-frequency data, and because our testing approach can account for that, the results from the unit root literature suggest that raising the data frequency may indeed improve the power of long swings tests, as argued above.

To further support that claim, we perform a Monte Carlo experiment. We generate a sample of 1,000 exchange rate changes $s_t$ that exhibit long swings. As data generating process (DGP) we use the GARCH-t regime-switching model of subsection 2.1 with parameters similar to the estimates for weekly data discussed below ($\mu_1 = -0.2, \mu_2 = 0.2, p_{11} = 0.98, p_{22} = 0.98, \sigma^2 = 1, \alpha = 0.1, \beta = 0.85, \nu = 5$); the DGP is initialized by regime $r_1 = 1$ with probability 0.5 and by the variance $V_0\{\varepsilon_1\} = \sigma^2$. We call this the high-frequency sample. We carry out Hansen’s test for $\mu_1 = \mu_2$ taking account of the GARCH-t features in the data and using the grid \{0.1, 0.2, ..., 0.8\} for $\mu_2 - \mu_1$ and \{0.96, 0.97, 0.98, 0.99\} for $p_{11}$ and $p_{22}$.3 Next, we derive the low-frequency sample by dividing the high-frequency sample in groups of 10 consecutive exchange rate changes and adding up the changes, so that the low-frequency sample contains 100 observations. Since there appears to be no evidence of GARCH-t effects in this sample, we apply Hansen’s test under homoskedasticity and normality. The grid for $\mu_2 - \mu_1$ is \{1, 2, ..., 8\} and the one for $p_{11}$ and $p_{22}$ is \{0.6, 0.7, 0.8, 0.9\}. The procedure is repeated 500 times. The power is estimated by the relative frequency of rejections at the 5% nominal size using the bandwidth parameter $M = 10$ for Hansen’s test.

The estimated power using low-frequency data turns out to be 0.27. For the high-frequency data it is 0.91. Hence, the power increases considerably when raising the frequency. This conclusion appears robust when experimenting with other Monte Carlo designs (various parameter values and grids), though the magnitude of the power improvement depends on the specific set-up.

The results in the unit root literature as discussed above suggest that the gain in power originates from two sources, namely the increase in frequency for a homoskedastic normal DGP and the fact that only with high-frequency data one can account for conditional heteroskedasticity and leptokurtosis in the DGP. To examine to what extent this is true for long swings tests, we repeat the simulation study but now without the

---

3The number of grid points, 128, is much less than that for the real data analyses, because we want to limit the computer time for the Monte Carlo (for 128 points it already takes several days per parameter combination). We have experimented with denser grids, but the results are similar.
GARCH-t effects, that is, with $\alpha = \beta = 0$ and $\nu = \infty$. The low-frequency power becomes 0.36 and the high-frequency one 0.62. Hence, from the total gain of 0.64 we attribute 0.26 to the increase in frequency for homoskedastic normal processes and 0.38 to the fact that high-frequency data allows one to capture conditional heteroskedasticity and excess kurtosis.

The gain in power may reflect that the actual size of the test using high-frequency data is much higher than for low-frequency data. To study this, we redo the original Monte Carlo but now with $\mu_1 = \mu_2 = 0$ for the DGP. The size for the low-frequency data is estimated at 0.014 and for the high-frequency at 0.006. Hence, the concern just mentioned is not supported empirically.

The fact that both sizes are lower than the nominal size of 0.05 reflects the conservativeness of Hansen’s bound-based procedure. The extent of conservativeness here may seem to contradict Hansen’s (1992) conclusion that the conservativeness is small for his application. However, concentrating on the high frequency, if we take a similar set-up as in Hansen (1992) by using a homoskedastic normal DGP ($\alpha = \beta = 0$ and $\nu = \infty$) and widening the grid for $p_{11}$ and $p_{22}$ to $\{0.1, 0.3, 0.5, 0.7, 0.9\}$, keeping the other elements the same, then the size becomes 0.04, in line with Hansen’s finding. Hence, the lower size in our original set-up is caused by the characteristics of that set-up. Further analysis reveals that both the GARCH effects, the t-distribution and the narrower grid are relevant, because adding either one of them to the Hansen-type set-up already reduces the size.

### 3.3.3 Weekly Data

The positive effect of data frequency on test power suggests to redo the long swings test for the three dollar exchange rates using a higher than quarterly frequency. We first increase the data frequency to monthly, but then the tests are still insignificant.

Then we use weekly data. The grid choice is based on similar arguments as in subsection 3.3.1, which leads to $\mu_2 - \mu_1 \in \{0.01, 0.10, 0.20, \ldots, 1.00\}$ and $p_{11}, p_{22} \in \{0.950, 0.955, \ldots, 0.995, 0.999\}$, in total 1331 grid points. As table 1 shows, the p-value upper bounds now yield evidence of the existence of long swings, particularly if one realizes that the test is conservative, so that the true p-values are smaller.

Our final conclusion is thus that the data really suggest that long swings exist. The previous inability to find long swings using quarterly data is not caused by the absence of swings but is rather due to statistical reasons: the low data frequency leads to too few observations per swing to significantly distinguish long swings from a single-regime process. Weekly data give the test enough power.
3.4 Estimation Results

We now present the estimation results for the regime-switching model and, for comparison, the results for the random walk, all for the weekly data. For Germany we add a first-order autoregressive term to mean equation (1), so that the mean equation for Germany is \( s_t = \mu_r + \theta (s_{t-1} - \mu_r) + \varepsilon_t \) instead of (1).

As table 2 shows, all three exchange rates are characterized by an appreciation \( (\mu_1 < 0) \) and a depreciation swing \( (\mu_2 > 0) \). The strong persistence of these swings is represented by the large regime-staying probabilities \( p_{11} \) and \( p_{22} \). Conditional homoskedasticity and conditional normality are strongly rejected. Finally, the autocorrelations of the normalized residuals and their squares (not reported) support the model specification.

To get a better idea about the degree of regime persistence that the staying probabilities imply, we first compute the expected duration of a regime \( r \), which is \( (1 - p_{rr})^{-1} \). The average estimates of \( p_{11} \) and \( p_{22} \) imply an expected duration of about a year for both regimes. These are comparable to Engel and Hamilton (1990), despite the fact that our estimates are based on weekly instead of quarterly data.

A second way to examine the persistence of regimes is by inspecting smoothed regime probabilities. The smoothed probability of a particular regime at time \( t \) is the probability that the process was in that regime at time \( t \) using the complete data set \( I_T \) (see Klaassen (2002) for its computation). The smoothed regime probabilities are plotted in figure 1, and they confirm that regimes are persistent.

3.5 Forecasting Performance

Given the existence of long swings, a natural question is whether this can be exploited to forecast exchange rates. Therefore, this subsection compares the out of sample forecast quality of the regime-switching models estimated in subsection 3.4 to the random walk.

The forecast under consideration is \( \hat{E}_{t-1}\{S_t\} \), the expected exchange rate at some future time \( \tau \) using information at time \( t - 1 \). The forecasts for the regime-switching model are computed using similar formulas as in Klaassen (2002); for the random walk the forecasts are the previous exchange rate \( S_{t-1} \) plus a drift term. We examine three horizons \( \tau - (t - 1) \), namely one week, one month (4 weeks) and one quarter (13 weeks). There are 1,524 out of sample forecasts, resulting from the procedure described in the note below table 3.

We evaluate the forecasts in two ways. First, we examine how well a forecast \( \hat{E}_{t-1}\{S_t\} \) predicts the level of the future exchange rate \( S_\tau \). This is point prediction. As usual, we take the mean squared error (MSE) criterion \( \frac{1}{T} \sum_{t=1}^{T} (S_\tau - \hat{E}_{t-1}\{S_t\})^2 \). In
the left panel of table 3 we report the MSE of the regime-switching model (left column for each country) and the MSE difference with respect to the random walk (right column). Their standard errors are heteroskedasticity and autocorrelation consistent.4

Table 3 shows that the regime-switching model does not outperform the random walk in terms of MSE. This corroborates Engel (1994), but it contrasts with the conclusion of Engel and Hamilton (1990) that the regime-switching forecasts are better. However, the latter paper uses a different out of sample period, that is, the first quarter of 1984 through the first quarter of 1988. If we take the same period (but still weekly data), then we also find that the regime-switching model does better. Apparently, the difference in out of sample periods explains the difference in results.

One may ask whether a comparison of point predictions is a good way to evaluate forecasts for our data. First, exchange rates are very volatile, so that relating point forecasts to realizations introduces a great amount of noise. This makes it difficult to find a forecast improvement of using swings, even if swings exist.

A second potential objection to the use of point predictions is that the point prediction errors are only weakly related to trading profits (Leitch and Tanner (1991)). The authors show that directional accuracy of the forecasts, on the other hand, is highly correlated with profits. Forecasting the direction instead of the exact magnitude of exchange rate changes may also be of interest to central banks when deciding whether to intervene (Engel (1994)).

Our second forecast evaluation is thus based on the direction of change. That is, we compute how often the sign of \( \hat{E}_{t-1}\{S_t\} - S_{t-1} \) equals the sign of \( S_t - S_{t-1} \). The percentage of correctly forecasted exchange rate changes is presented in the right panel of table 3. The left column for each country gives the percentage for the regime-switching model; the right column gives the percentage after subtracting the one for the random walk.

Table 3 reveals that the regime-switching model predicts the correct direction in more than 50% of the cases, particularly at the one-week and one-month horizons, where five of the six outperformances are significant. Regime-switching forecasts are also better than those from the random walk, and four of the six improvements at the one-week and one-month horizons are significant. For the one-quarter horizon the outperformance is less clear. Further investigation shows that the improvement is robust across the four out of sample quarters. Hence, taking account of long swings

\[4\text{The standard errors are based on Newey and West (1987). They result from a regression of, for instance, } (S_t - \hat{E}_{t-1}\{S_t\})^2 \text{ on a constant and following West and Cho (1995) by taking Bartlett weights and using the same data-dependent automatic lag selection rule. This rule, introduced by Newey and West (1994), has certain asymptotic optimality properties.} \]
helps predict the direction of exchange rate changes for a few weeks. Other papers point in the same direction, although our evidence is stronger. For instance, Engel (1994) concludes that there is weak evidence that the regime-switching model outperforms the random walk in predicting the direction of change, and that this evidence is smaller for longer horizons. Though this outperformance is interesting in itself, it also strengthens our conclusion from the Hansen (1992, 1996) test in subsection 3.3 that the regime-switching model improves on the random walk.

4 Conclusion

This paper tests the null hypothesis of a random walk against the alternative of long swings in exchange rates, where the long swings are modeled by a Markov regime-switching model. Existing papers conclude from quarterly data through 1986 that long swings exist. However, we show that there is no longer evidence when the quarterly sample is extended to 2003.

We then try to explain the lack of evidence by investigating the power of the test. In particular, the role of the data frequency is examined. Because high-frequency exchange rates exhibit conditional heteroskedasticity and leptokurtosis, we extend the basic Hamilton (1989) regime-switching model by including a GARCH variance and a t-distribution for the innovations. Indeed, a Monte Carlo study based on Hansen’s (1992, 1996) testing methodology demonstrates a substantial gain in power when raising the data frequency. Apparently, one needs enough observations per swing to distinguish them from a random walk.

Hence, we redo the long swings test with weekly instead of quarterly data. Now Hansen’s test rejects the random walk in favor of the long swings model. An out of sample forecasting study also supports the existence of swings in the sense that the regime-switching model yields better forecasts for the direction of exchange rate change than the random walk. Hence, we conclude that there is really evidence of long swings in the data. More specifically, we find an appreciation and a depreciation regime, with expected durations of about a year.

Our results are useful in several other ways. First, the long swings show that there is some systematic component in exchange rates. This may originate from the relevance of economic fundamentals for exchange rates. Hence, the finding of long swings may guide future research on exchange rate determination.

Second, our conclusion provides a basis for papers that already use the existence of long swings as a starting point to examine other economic issues, such as Engel and Hamilton (1990) on uncovered interest parity (UIP), Kaminsky (1993) on peso

Third, the relevance of data frequency for the power of long swings tests may serve as a motivation to reconsider tests for other long-run phenomena when the variables involved can be observed on a high-frequency basis. For instance, redoing unit root tests using high-frequency data may change the outcome from non-rejection to rejection because of a gain in power.

Finally, the rejection of the random walk shows that there is some structural long-run serial dependence in exchange rates and that the regime-switching model is able to make that visible. However, this does not imply that the regime-switching model is the best model. Maybe there is a better model, perhaps an extension of our model. There are a few suggestions in this respect. First, other variables, such as forward rates, can be included in the mean equation. Second, the assumption of time-constant regime-switching probabilities can be relaxed. For instance, one can let them depend on indicators of economic policy and directly analyze the role of policy changes for switches in the exchange rate regime. In this respect, monetary policy announcements may have an effect, as in Kaminsky (1993). Moreover, one can test whether foreign exchange interventions affect the regime-switching probabilities, as such interventions may signal changes in future monetary policy. These suggestions are left for future research.
References


Table 1: Tests for existence of long swings in exchange rates

<table>
<thead>
<tr>
<th>Data frequency</th>
<th>GERMANY</th>
<th>JAPAN</th>
<th>U.K.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subperiod</td>
<td>quarterly</td>
<td>weekly</td>
<td>quarterly</td>
</tr>
<tr>
<td>-.86 ——</td>
<td>1.95</td>
<td>3.04</td>
<td>2.28</td>
</tr>
<tr>
<td>Hansen’s $\hat{LR}^*$ test</td>
<td>0.21</td>
<td>0.43</td>
<td>0.11</td>
</tr>
<tr>
<td>$M=1$</td>
<td>0.19</td>
<td>0.40</td>
<td>0.13</td>
</tr>
<tr>
<td>$M=2$</td>
<td>0.18</td>
<td>0.38</td>
<td>0.12</td>
</tr>
<tr>
<td>$M=5$</td>
<td>0.14</td>
<td>0.38</td>
<td>0.10</td>
</tr>
<tr>
<td>$M=10$</td>
<td>0.12</td>
<td>0.36</td>
<td>0.08</td>
</tr>
</tbody>
</table>

The test for long swings is a test of the null hypothesis of the random walk against the alternative of the regime-switching model of subsection 2.1, which contains the null as the special case $\mu_1 = \mu_2$. For the quarterly frequency both models have been restricted by imposing conditional homoskedasticity ($\alpha=\beta=0$) and normality ($\nu^{-1}=0$). The left column for the quarterly frequency uses only data through 1986:IV, that is, similar data as in Engel and Hamilton (1990) and Engel (1994). The $\hat{LR}^*$ statistic is based on Hansen’s (1992, 1996) method. Hence the upper bounds for the p-values may depend on a bandwidth parameter $M$; the bounds are based on 10,000 simulations. For the quarterly data, the grid is $\mu_2 - \mu_1 \in \{0.5, 1.0, 2.0, \ldots, 10.0\}$ and $p_{11}, p_{22} \in \{0.50, 0.55, 0.60, \ldots, 0.95\}$ such that $p_{11} + p_{22} \geq 1.2$, in total 990 points. For the weekly data, the grid is $\mu_2 - \mu_1 \in \{0.01, 0.10, 0.20, \ldots, 1.00\}$ and $p_{11}, p_{22} \in \{0.950, 0.955, \ldots, 0.995, 0.999\}$, in total 1331 points. The required CPU times on a Pentium 733 computer are 0.2, 0.5 and 15 hours for the three columns per country.
<table>
<thead>
<tr>
<th></th>
<th>GERMANY</th>
<th></th>
<th>JAPAN</th>
<th></th>
<th>U.K.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RW</td>
<td>RS</td>
<td>RW</td>
<td>RS</td>
<td>RW</td>
<td>RS</td>
</tr>
<tr>
<td>Mean of regime</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \mu_1 )</td>
<td>0.03</td>
<td>-0.21</td>
<td>-0.00</td>
<td>-0.30</td>
<td>0.01</td>
<td>-0.25</td>
</tr>
<tr>
<td>(0.03)</td>
<td>(0.08)</td>
<td>(0.03)</td>
<td>(0.15)</td>
<td>(0.03)</td>
<td>(0.15)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>( \mu_2 )</td>
<td></td>
<td></td>
<td>0.21</td>
<td>0.18</td>
<td></td>
<td>0.11</td>
</tr>
<tr>
<td>(0.08)</td>
<td>(0.09)</td>
<td>(0.09)</td>
<td>(0.07)</td>
<td>(0.07)</td>
<td>(0.07)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Regime stay prob</td>
<td>( p_{11} )</td>
<td>0.989</td>
<td>0.956</td>
<td>0.984</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.011)</td>
<td>(0.032)</td>
<td>(0.032)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>( p_{22} )</td>
<td>0.990</td>
<td>0.962</td>
<td>0.994</td>
<td>0.994</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.012)</td>
<td>(0.029)</td>
<td>(0.029)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Uncond. ( \sigma^2 ) variance</td>
<td>2.47</td>
<td>2.52</td>
<td>1.70</td>
<td>1.37</td>
<td>2.30</td>
<td>2.28</td>
</tr>
<tr>
<td>(0.57)</td>
<td>(0.66)</td>
<td>(0.80)</td>
<td>(0.76)</td>
<td>(0.67)</td>
<td>(0.67)</td>
<td>(0.67)</td>
</tr>
<tr>
<td>ARCH ( \alpha )</td>
<td>0.12</td>
<td>0.12</td>
<td>0.06</td>
<td>0.06</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>GARCH ( \beta )</td>
<td>0.85</td>
<td>0.84</td>
<td>0.93</td>
<td>0.93</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>T-distrib. ( \nu^{-1} )</td>
<td>0.09</td>
<td>0.12</td>
<td>0.17</td>
<td>0.19</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-2660</td>
<td>-2651</td>
<td>-2625</td>
<td>-2616</td>
<td>-2535</td>
<td>-2530</td>
</tr>
</tbody>
</table>

Standard errors in parentheses.

“RW” denotes the random walk, “RS” the regime-switching model of subsection 2.1. For Germany, however, the regime-switching model is extended by a first-order autoregressive term to correct for some short-term serial correlation; the estimate for the autoregressive parameter is 0.05 (0.03).

We report the inverse of the degrees of freedom of the t-distribution, because testing for conditional normality then boils down to simply testing whether \( \nu^{-1} \) differs from zero.
Table 3: Out-of-sample forecasting statistics

<table>
<thead>
<tr>
<th>Forecast horizon</th>
<th>Mean squared error (MSE)</th>
<th>Percentage correct direction of change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Germany</td>
<td>Japan</td>
</tr>
<tr>
<td></td>
<td>RS</td>
<td>-RW</td>
</tr>
<tr>
<td>Week</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.08</td>
<td>-0.00</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Month</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.64</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>(0.68)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Quarter</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>35.46</td>
<td>1.28</td>
</tr>
<tr>
<td></td>
<td>(2.87)</td>
<td>(1.28)</td>
</tr>
</tbody>
</table>

Standard errors in parentheses; they are heteroskedasticity and autocorrelation consistent (see note 4). “Percentage correct direction of change” denotes the percentage of forecasts that imply the same direction of exchange rate change as observed in the data. The columns headed by “RS” contain the forecasting statistic of the regime-switching model, the columns headed by “-RW” give the forecasting statistic of the regime-switching model minus that of the random walk. A * reflects that the regime-switching model is significantly better than the benchmark at the 5% level (one-sided tests), where the benchmark for the RS columns in the direction of change panel is 50% and for the -RW columns in both panels it is the random walk.

Each statistic is based on 1,524 out of sample comparisons (for the one-week horizon; for longer horizons somewhat fewer), generated as follows. One commonly splits the sample into two parts, re-estimates a model using only the first part (estimation period) and generates forecasts for the second part (out of sample period). The estimation period must be long enough, particularly for a regime-switching model on exchange rate data, as exchange rate swings are quite long and we need sufficiently many swings to get accurate estimates. Suppose we take the first three quarters of the data. This leaves only one quarter (381 observations) for the out of sample analysis. This may result in a lack of power of the tests. Therefore, we also take the three other combinations, that is, quarters 2,3,4 for estimation with 1 for forecasting, 1,3,4 with 2, and 1,2,4 with 3. This gives 1,524 out of sample observations.
Figure 1: Logarithm of exchange rates (dollars per foreign currency; thick line, left axis) and smoothed regime probabilities (thin line, right axis)