Short-Memory and the PPP-Hypothesis

Mahmoud A. El-Gamal and Deockhyun Ryu*

January 2003

Abstract

Over the past decade, the purchasing-power parity (PPP) puzzle has taken two forms. Its early form arose from early tests of unit roots in real exchange rates, which failed to reject the null hypothesis, thus casting doubts on the long-term PPP hypothesis of real exchange rates' mean reversion. Following the development of more powerful tests that resulted in rejections of unit roots, the PPP-puzzle re-surfaced in the form of surprisingly slow rates of convergence of real exchange rates to their long-run means. Rogoff (1996) expressed this puzzle in terms of the estimated "half-life" of real exchange rate shocks being 3 to 5 years. Recent research has attempted to solve that second form of the puzzle by adopting non-linear stochastic models of real exchange rates. Despite this introduction of non-linearities, the literature has continued to focus on the notion of "half-life" as a measure of persistence.

We argue that the half-life measure is only appropriate in linear settings, failing to capture the richness of non-linear dynamics introduced in the more recent literature. We propose operational measures of persistence in such non-linear models, which we label short-memory in distribution (SMD, which is similar to the notion of ϕ-mixing), and short-memory in mean (SMM, c.f. Granger (1995)). We show that focusing on a simple notion such as "half-life" can be very misleading. In particular, we show that it is possible to match desired "half-lives" for any of the most popular non-linear models recently proposed in the literature, at the expense of matching their more general dynamic structure. We conclude that depending on the models and criteria selected for investigating the PPP-puzzle, the puzzle may be in the eye of the beholder.

Keywords: PPP-puzzle; half-life; SMD; SMM; ϕ-mixing; persistence measures; cross-validation; non-parametric time series analysis; Markov models.

* El-Gamal is a Professor of Economics and Statistics at Rice University, where Ryu is a Ph.D candidate in the Economics Department. Contact: Mahmoud A. El-Gamal, Economics MS 22, Rice University, 6100 Main Street, Houston, TX 77005. Emails: elgamal@rice.edu, dhryu@rice.edu.

1. Introduction

In the early 1990s, doubts were cast on the purchasing-power parity (PPP) hypothesis, following repeated failures to reject univariate unit root tests for real exchange rates. Then, a number of studies in the 1990s noted that those early failures to reject the unit root hypothesis were artifacts of the low power of those simple unit root tests. Using more powerful panel-data tests, those researchers managed to reject the unit root hypothesis, thus providing support for the PPP-hypothesis, if only in the very long-run. Developments in this literature up to the mid-1990s were summarized in the seminal survey of Rogoff (1996).

No sooner had the first PPP-puzzle been solved, a second PPP-puzzle arose. Despite rejection of the extreme form of persistence of real exchange rates that unit roots would imply, estimates still suggested a "half-life" (time for a transitory shock to the real exchange rate to be cut in half) of three to five years.[1] Common wisdom dictated that half-lives for real exchange rates should be two years or less, for the PPP-hypothesis to be operationally relevant. The emerging empirical stylized fact of three-to-five years' half-life was also problematic since existing models could not explain the estimated slow mean reversion together with the observed high short-term real exchange rate volatility.

It is easiest to illustrate this second form of the PPP-puzzle within the linear AR(1) model, which is the main building block for the literature on PPP-puzzle:

$$x_t = \rho x_{t-1} + \varepsilon_t,$$

where x_t is a real exchange rate and $\varepsilon_t \sim i.i.d.N(0,\sigma^2)$. In this model, the half-life (H) of the process is easily calculated based on the estimate of ρ: $\hat{H} = \ln(0.5)/\ln(\hat{\rho})$. Estimates of $\hat{\rho}$ that yield an estimated half-life below 3 years would be deemed acceptable, and higher estimates are considered puzzling. In order to estimate a $\hat{\rho}$ that yields such acceptable half-life estimates, researchers first looked for econometric explanations of why earlier estimates of half-life may have been artificially high. For instance, Murray and Papell (2000) investigated the potential upward bias of half-lives using approximately unbiased median estimators.

Later studies aimed to produce alternative half-life estimates using non-linear models of real exchange rate dynamics. For instance, Taylor (2001) used a regime switching threshold model (RSTM) of the data generating process to show that linearity-based half-life estimates would be biased upward. A number of other studies also introduced non-linear dynamics to explain the high half-life estimates in simple linear models, including Cheung and Lai (2000), Taylor, Peel and Sarno (2001). In later sections, we shall discuss some of those models in greater detail, especially the so-called

[1] The 3~5 years' half-life was first highlighted in the survey of Rogoff (1996). He noted that the immediately preceding literature had reached a surprising degree of consensus around that half-life estimate.

Threshold Autoregressive (TAR) and the Smooth Transition Autoregressive (STAR) models. The purpose of this paper is to investigate whether adding greater flexibility through non-linear dynamic models resolves the puzzle, or merely makes it more complex.

Our point of departure is an examination of the measure of persistence used by all earlier studies, including ones that propose non-linear models of real exchange rates. Despite introducing highly non-linear stochastic models, the literature has continued to occupy itself mainly with point estimates of "half-life" as a measure of persistence. In later sections, we shall propose a more general approach to studying the evolution dynamics for distributions of real exchange rates around their long-term PPP-levels. Utilizing that framework of distribution dynamics, we can investigate more fully convergence to long-term PPP levels, and speeds thereof.[2]

To study the persistence properties of non-linear stochastic models of real exchange rates, we use the notions of Short-Memory-in-Mean (SMM) and Short-Memory-in-Distribution (SMD), which are closely related to the statistical notion of ϕ-mixing. Our SMM measure was proposed by Granger (1995) and Granger and Teräsvirta (1993) as an alternative for the linear notion of I(0). In this regard, Granger (1995) showed through a number of examples that the I(0) notion is only appropriate in linear models, and argued for using SMM as a better measure of persistence in more general models. However, he stated that it was not clear how to operationalize the concept of SMM for estimation and testing. To the best of our knowledge, this is the first attempt to provide an operational algorithm for estimating SMM, and the more general notion of SMD discussed below.

The rest of this paper will proceed as follows. We discuss various recent non-linear models of real exchange rate processes in Section 2. We construct our non-persistence measures of SMM (a measure we call *m-life*, half-life being one point on the curve), and SMD (which will be an estimate of ϕ-coefficients) in Section 3. In Section 4, we introduce non-parametric estimation of our non-persistence measure, and apply it to five real exchange rates covering the post-1973 floating exchange rate system. We shall also discuss small-sample problems associated with such non-parametric estimation in Section 4. In Section 5, we evaluate the dynamic persistence properties of the various non-linear models discussed in Section 2, utilizing the non-parametric estimates of Section 4, and using various degrees of over- or under-smoothing. In Section 6, we conduct some Monte Carlo simulations to assess the effects of model mis-specification. In particular, we investigate whether non-linear models always yield shorter half-life measure relative to their linear counterparts. Some concluding remarks close the paper.

[2] To the best of our knowledge, a similar approach has only been utilized in the economic growth literature. Early studies in that literature focused on the means of cross-country income distribution (β-convergence) or the variance (σ-convergence) thereof. More recently, Quah (1993, 1997) and Durlauf (1993) highlighted the necessity of examining the dynamics of the entire distribution.

2. Non-linear models of real exchange rates

We have already noted that early tests of the PPP-hypothesis concentrated on unit root tests of real exchange rates.[3] In that early literature, OLS estimates of AR(1) coefficients were used for calculating half-lives of the estimated processes.[4] The validity of this linear-model framework was questioned in later studies. For instance, Obstfeld and Taylor (1997), and Taylor (2001), studied the upward bias in linear half-life estimates if the data generating process followed a threshold autoregressive (TAR) models. In particular, Obstfeld and Taylor (1997) showed that if the true model is a TAR with a half-life of 12 months, mis-specifying the model as a linear AR(1) may result in over-estimated half-lives of 2-3 years.

In addition to the linearity-induced biases, recent econometric studies have shown that linear-model half-life estimates may be biased due to using ordinary least squares (OLS). In particular, Murray and Papell (2000) showed that if the data generating process was indeed a linear AR(1), the OLS estimate of half-life is biased downward in small samples, with the bias growing as the true AR (1) coefficient approaches unity. Furthermore, they argued that the conventional point estimate half-life measure is not a sufficient measure of persistence, since it does not account for the serial correlation and estimation errors of such estimates.

A third set of problems with conventional linear-model half-life estimation stems from the monotonicity of that measure of persistence. Cheung and Lai (2000) investigated plausible models with non-monotonic real exchange rate adjustment, whereby impulse response functions (IRF) initially rise before the shock begins dissipating. In such models, a monotonic half-life estimate (usually computed from IRFs) will be inflated, and clearly becomes an improper measure of persistence.

After uncovering the above mentioned pitfalls of linear-model estimates of half-life, research in this area turned to proposing *ex ante* plausible non-linear models of real exchange rate dynamics, with an eye to estimating "reasonable" half lives. In an early contribution to the literature, Goldberg et al (1997) introduced a diffusion process model that allowed for more general patterns of mean reversion for real exchange rate process. More recently, Taylor, Peel, and Sarno (2001) introduced a smooth transition autoregressive (STAR) model, which allowed them to estimate shorter (and therefore less puzzling) half-lives. This model agrees with empirical patterns of mean reversion, where larger shocks are found to dissipate faster than small ones. This feature is also discussed in Section 2 of Taylor et al (2001). A similar model, called the exponential STAR (ESTAR) model, was proposed by Baum et al (2001), also allowing for more general types of mean reversion.

[3] This is somewhat of an oversimplification. Froot and Rogoff (1995) classified PPP tests in the unit roots literature into first, second, and third levels, including cointegration tests, which are not discussed here.
[4] For example, Darby (1983), Hakkio (1984), etc.

The above mentioned models of non-linear mean reversion were highly parametric. Shintani (2001) introduced a non-parametric framework with an alternative half-life estimates, which he called "average half-life". His estimate is obtained from non-parametric estimates of the derivative of a one-period Markovian real exchange rate mapping at various points. He argued that while half-life will be different for different levels of real exchange rates, allowing for many more general forms of mean reversion, his estimated "average" measure can provide some linear-model approximation. Unfortunately, this measure suffers from technical problems,[5] as well as the conceptual problem of attempting to approximate a non-linear-model notion with the average of its model-linearization equivalents at various points. Clearly, the latter estimate can be arbitrarily inappropriate depending on the actual level of the real exchange rate.

Indeed, one major drawback of the literature to-date is its continued focus on a linear measure of persistence (half-life), despite the introduction of a variety of non-linear mean-reversion dynamics in their models. We now turn to the task of defining a more appropriate measure of persistence in such non-linear models.

3. Laws of motion and Non-persistence measures

Granger and Teräsvirta (1993) and Granger (1995) showed that the notion of I(0), which is central to the unit-roots literature that dominated early PPP-hypothesis investigations, is essentially a linear notion, with no clear non-linear analog. In non-linear models, they suggested investigating short-memory in distribution (SMD) and/or short-memory in mean (SMM), to be defined below, as measures of persistence that apply to non-linear models. Before defining our operational measures of SMD and SMM, we find it useful to review a few definitions.

3.1. Useful definitions

Keeping in mind the eventual need for non-parametric estimation of our measures of SMD and SMM, we adopt the framework of Domowitz and El-Gamal (1993,1996,

[5] His average half-life of deviation from PPP is calculated as follows: $\widehat{HL} = \dfrac{\ln(1/2)}{T^{-1}\sum_{t=1}^{T} \ln\left|\widehat{Dm}(q_{t-1})\right|}$ where $\widehat{Dm}(q_{t-1})$ is a nonparametric estimator of the first derivative of $m(q_{t-1})$ in the mapping equation of $q_t = m(q_{t-1}) + \varepsilon_t$. Calling $\widehat{HL}$ an average half-life is a misnomer. In fact, it is the half-life of the average estimator. To calculate an "average half-life", he should have computed: $\widetilde{HL} = T^{-1}\sum_{t=1}^{T} \dfrac{\ln(1/2)}{\left|\widehat{Dm}(q_{t-1})\right|}$. Since $\widetilde{HL}$ is greater than $\widehat{HL}$ by Jensen's inequality ($E[f(Y)] \geq f(E[Y])$), we know that his latter mislabeled estimate of "average half-life" is biased downward.

2001) for a first-order Markovian univariate time series $\{x_t\}$.[6] We assume that for a given point $\xi \in X$, that given transition $P_T(\xi,\cdot)$ is a probability measure on $\mathfrak{B}(\mathbb{R})$, and for a given $A \in \mathfrak{B}(\mathbb{R})$, $P_T(\cdot, A)$ is a Borel measurable function. We shall refer to $P_T(\cdot,\cdot)$ as the one-step transition kernel. As usual, we define the *s*-step transition probability recursively by:

$$P_T^{(s)}(\xi, A) = \int_X P_T^{(s-1)}(\xi, d\eta) P_T(\xi, A)$$

We assume that the probability measure $P_T(\xi,\cdot)$ is absolutely continuous, and we denote the corresponding (estimated) transition density $p_T(\cdot,\cdot)$.

Starting from an initial density $g_0(x)$ on the state space $\mathbb{R}$, the probability of the process falling in any Borel set *A* at period *s* can be easily defined by

$$\Pr_{g_0}\{x_s \in A\} = \int_A g_0(\xi) P_T^{(s)}(\xi, d\eta) \equiv \int_A g_s(\eta) d\eta$$

This implicitly defines the Markov operator $\Pi_T : D(\mathbb{R}) \to D(\mathbb{R})$ (via $g_s(\cdot) = \Pi_T^s g_0(\cdot)$), where $D(\cdot)$ is the space of densities. We call Π_T the Frobenius-Perron (F-P) operator.[7]

Using this framework, we can directly apply the consistent tests of ergodicity and mixing constructed in Domowitz and El-Gamal (1993, 1996, and 2001) to various real exchange rate series. Those tests failed to reject the null hypotheses of ergodicity or mixing as shown in Table 8.[8] However, due to the smallness of our sample size (*T*=312), we know that our estimate of *P* (using the automatic bandwidth selection methods discussed below) is over-smoothed, thus producing low power for our tests.[9] A perennial problem in this literature is our inability to increase sample size by using higher-frequency data, since CPI data (used for computing the real exchange rate) are not available at higher frequencies (c.f. A. Taylor (2001) for a discussion of the same problem).

Moreover, even if we had failed to reject ergodicity and mixing with sufficiently large samples, we would not be able to conclude that the PPP-puzzle is solved. Indeed,

---

[6] Domowitz and El-Gamal (1993, 1996) discussed how to generalize this framework to a multi-variate and/or higher-order Markovian setting, providing an algorithm for dimension-reduction.

[7] Day and Pianigiani (1991) also used a Frobenius-Perron operator approach to study complex economic dynamics. Quah (1993) used a similar approach to study the time-evolution of income distribution across countries.

[8] Recall that the tests of Domowitz and El-Gamal (1993, 1996, 2001) are randomized tests, hence Table 8 reports the percentage of rejections at 5% and 10%. Since those percentages are close to 5% and 10%, respectively, we conclude that the randomized tests fail to reject the null hypotheses of ergodicity and mixing.

[9] The small sample problem appeared in Murray and Papell (2000) in the downward bias of OLS estimate of half-life.

such simple testing methodologies cannot deal with the new version of the PPP-puzzle, which focuses on measures of persistence of real exchange rates (in particular, half-life). Therefore, our focus should turn to the estimation of appropriate measures of persistence.

We now proceed to define the short memory in distribution (SMD) and short memory in mean (SMM) measures used in this paper:

Definition: Consider a time series $\{x_t\}$. Let $F_s(x) = \Pr(x_{t+s} \le x \mid \mathfrak{A}_t)$ be the cumulative distribution function of x_{t+s} conditional on the past information set $\mathfrak{A}_t = \sigma(x_{t-j}; j \ge 0)$, and let $\overline{F}$ be some fixed (unconditional) distribution function. The time series is said to have short memory in distribution (SMD) if: $F_s(x) \Rightarrow \overline{F}(x)$, as $s \uparrow \infty$.

This definition implies that for an SMD process, there exists a series $\{d_s\}$ s.t. $\left|\Pr(x_{t+s} \in C_1 \mid x_{t-j} \in C_2) - \Pr(x_{t+s} \in C_1)\right| < d_s$; $d_s \xrightarrow{s\uparrow\infty} 0$, for all measurable sets C_1, C_2 with $\Pr(x_{t-j} \in C_2) \ne 0$. In this regard, the SMD property is a form of "mixing", as we shall see later. In the meantime, we consider a first-moment analog to SMD, which compares means conditional on distant information sets to unconditional means.

Definition: The time series is said to have the short memory in mean (SMM) property if $\| E[x_{t+s} \mid \mathfrak{A}_t] - E[x_{t+s}] \| < c_s$; $c_s \xrightarrow{s\uparrow\infty} 0$.

This SMM is equivalent to "mixing in mean" or "mixingales" as discussed in McLeish (1978) and Gallant and White (1988). We note that SMD implies SMM, but the opposite is not true. Even though the concepts of SMD and SMM are very useful to study convergence structure in time series context, operational methods to study d_s and c_s are not well developed. To the best of our knowledge, this paper presents the first attempt to operationalize estimation of measures of SMD and SMM in a non-parametric framework.

3.2. *m-life* and $\phi(s)$ functions

We can study SMD or SMM directly using the asymptotic independence notion of uniform, or ϕ-mixing. The ϕ-coefficient for a time series is defined by:

$$\phi(s) = \sup_{A\in\mathfrak{A}_t,\, B\in\mathfrak{A}_{t+s},\ \Pr(A)>0} \left|\Pr(B \mid A) - \Pr(B)\right|$$

The process is called uniform of ϕ-mixing if $\phi(s) \xrightarrow{s\uparrow\infty} 0$. This notion of uniform mixing was first used in econometrics by White and Domowitz (1984) to establish the asymptotic normality of various parameter estimators.

One known property of ϕ-mixing processes (c.f. Jacod and Shiryaev (1980, lemma 3.102, p.456)), is the following:

Lemma Let x be a random variable measurable with respect to $\mathfrak{A}_{t+s}$, such that $\|x\|_q \equiv [E|x|^q]^{1/q} < \infty$, for some $q > 1$, and let $1 \le p \le q$. Then,

$$\sup_{A \in \mathfrak{A}_t} \| E[x_{t+s} \mid A] - E[x_{t+s}] \|_p \le 2[\phi(s)]^{(1-\frac{1}{q})} \|x\|_q$$

Setting $p = q = 2$, we can see that a mixing process must have auto-covariances that vanish sufficiently fast. This is the link to central limit theorems exploited in earlier econometrics work. However, the previous literature has usually made mixing assumptions, without attempting to calculate or estimate the various elements of this inequality. In this paper, we shall show that given any transition kernel *P,* we can numerically calculate approximations of the LHS and RHS.

Before proceeding to our numerical approximations of the memory-in-mean and $\phi(s)$ measures, it is worthwhile putting our measures of persistence within the context of other popular measures:

- The PPP literature has to-date focused on the notion of half-life: a measure of the time until a transitory shock between the real exchange rate and its long-term PPP mean-level is cut by half.
- A finer (more general) measure of persistence may be defined as an *m-life* curve, which measures the time needed for a transitory shock between the real exchange rate and its long-term PPP-level to be cut by 1-*m,* for all $m \in (0,1)$. We shall select this notion (*m-life*) as our measure of SMM, to be defined more carefully later in the paper.[10]
- Our most general, and finest measure of persistence is SMD, as measured by $\phi(s)$. This measure looks beyond the first moment, to provide a general assessment of the dependence structure of our time series.

We can approximate our transition kernel P by finite transition matrix P_n, and approximate the fixed point f^* of the former with the latter's fixed point f_n^*,[11] hence we can approximate $\phi(s)$ with $\phi_n(s)$ as shown in the next subsection.

[10] Murray and Papell (2001) discussed the potential advantage of looking at points other than half-life. They argued that looking at estimates of three quarters-life, or 51% life, may yield different persistence implication from looking only at half-life.

[11] T. Li (1976) demonstrated that in the case $(X, \mathfrak{F}, \mu) = ([0,1], \mathfrak{F}, \mu)$, where μ is Legesgue measure, each finite approximation P_n has a non-negative fixed point f_n, and $f_n \xrightarrow{n\uparrow\infty} f^*$ weakly. Bose (1994) expanded this result to strong convergence.

3.3. Numerical approximations of *m-life* and $\phi(s)$ functions

We approximate the coefficients $\phi(s)$ by their analogs on a finite grid. As in Domowitz and El-Gamal (1993, 1996, 2001), we restrict attention to a compact subset X of the state space that covers 95% of the support of our data. We can then construct a grid X_n covering that compact set. Let P_n be the transition kernel on X_n that is induced by the transition kernel P. Since P is a bounded linear operator, it follows that $\|P_n - P\| \xrightarrow{n\uparrow\infty} 0$. This in turn implies that the finite grid analog $\phi_n(s)$ defined below converges to $\phi(s)$ as the grid size $n \uparrow \infty$ (c.f. Kress (1998), theorem 12.6, p.292):

$$\phi_n(s) = \sup_{x\in grid;\, B\in\mathfrak{A}_s^n} \left|\Pr_{P_n^s}(B\,|\,x) - \Pr_{f^*}(B)\right| \xrightarrow{n\uparrow\infty} \phi(s),$$

where $\mathfrak{A}_s^n$ is the σ-algebra induced by $\{x_t\}_{t=0}^s \in X_n^s$. Notice that the maximization over all Borel sets for initial conditions reduces in the finite grid to maximization over all single initial conditions.[12] This makes the search much more manageable.

In similar fashion, we introduce a measure of SMM, which we call the Maximum Distance in Mean (MDM) function. The finite grid $MDM_n(s)$ defined below converges to the Maximum Distance in Mean $MDM(s)$, also defined below, as $n \uparrow \infty$:

$$MDM_n(s) = \sup_{x\in grid}\left|E_{P_n^s\delta_{\{x\}}}[x_{t+s}\,|\,x_t = x] - E_{f^*}[x_{t+s}]\right|$$
$$\xrightarrow{n\uparrow\infty} MDM(s) = \sup_{f\in D(X),\, A\in\mathfrak{A}}\left|E_{P^s f}[x_{t+s}\,|\,x_t \in A] - E_{f^*}[x_{t+s}]\right|,$$

where the maximum distance in mean for each s is defined on the RHS, as the maximal distance between the conditional mean of x_{t+s} (given initial condition set and distribution), and the unconditional mean under the unique invariant distribution $f^* = Pf^*$. When the process on X is approximated by the process on the grid X_n, the analog MDM_n is obtained through maximization over all initial conditions.

The notion of half-life can now be replaced by the value of s at which $MDM_n(s) = 0.5 \times MDM_n(0)$, i.e. the number of periods needed for the worst possible transitory shock from the unconditional mean to be cut in half. This notion may then be extended beyond half-life to consider *m-life* as the number of time periods before the worst possible shock would have shrunk to (1-*m*) of its original magnitude. For a given model summarized by a transition kernel P (and approximated by the corresponding matrix P_n), our approach has the added advantage of calculating *m-life* directly numerically. In contrast, the literature on non-linear exchange rate dynamics often computes half-life by simulating impulse response functions, e.g. see Cheung and Lai (2000), and Taylor et al (2001).

[12] Proof in the Appendix.

3.4. Numerical algorithms for computing *m-life* and $\phi(s)$ functions

We now turn to the actual numerical algorithms used in this paper for computing our measures of SMD and SMM for known models and estimated transition densities. We begin with the assumption of having a known transition matrix $P_n(.,.)$ on an $n \times n$ grid (in all of our applications, we fixed the grid-size at $n = 100$).

Algorithm A ($MDM_n(s)$)

1. Fix the grid $\boldsymbol{x} = (x_1, ..., x_n)$.
2. Compute the invariant measure f_n^* by iterating on $P_n^s f$ for any initial f, and $s = 1, 2, \ldots$, until convergence (in the sup norm).
3. Compute the unconditional expectation $\mu = E_{f_n^*}[x_{t+s}] = \boldsymbol{x}' f_n^*$.
4. Define $MDM_n(0) = \max(\mu, 1-\mu)$.
5. For each s, and each point on the grid $\{x_i\}_{i=1}^n$, compute the conditional expectation $\mu_i(s) = E_{P_n^s \delta\{x_i\}}[x_{t+s} \mid x_t = x_i] = \boldsymbol{x}' P_n^s \delta_{\{x_i\}}$. Compute $MDM_n(s) = \max_i(|\mu_i(s) - \mu|)$.
6. Normalize $MDM_n(s)$ by defining $1 - m_n(s) = MDM_n(s) / MDM_n(0)$.
7. Plot *m-life* as *s* against $(1 - m_n(s))$.

Algorithm B ($\phi_n(s)$)

1. Perform steps 1-2 of Algorithm A.
2. For each s, and each point on the grid $\{x_i\}_{i=1}^n$, compute $\phi_{n,i}(s) = \| f_n^* - P_n^s \delta_{\{x_i\}} \|$.
3. Set $\phi_n(s) = \max_i(\phi_{n,i}(s))$.
4. Plot $\phi_n(s)$ against *s*.

4. Non-parametric empirical applications

In this section, we analyze the persistence properties of real exchange rates for five major currencies that are most commonly studied in investigations of the PPP-hypothesis: the Japanese Yen, the French Franc, the U.K. Pound, the Deutsche Mark, and the Swiss Franc. We studied monthly data for those five currencies over the period 1973:1 to 1998:12. The base currency for all five series is the U.S. Dollar. Following the usual practice in this literature, we constructed the series for each country's real exchange rate as the natural logarithm of its nominal exchange rate, less the difference between the natural logarithm of the CPIs of the two countries.

The early literature on the PPP-puzzle failed to reject unit root tests for those post-Bretton Woods series, thus questioning the empirical validity of the PPP-hypothesis, even in the long-run. Later studies applied higher-power tests and managed to reject the null hypothesis of unit roots, thus turning the focus of the PPP-puzzle literature towards the notion of high-persistence, or long half-lives, of real exchange rates. Following in the footsteps of the latter literature, we begin from the assumption that PPP holds in the long run. We cast this long-run PPP-hypothesis in our stochastic model as an assumption of ergodicity of real exchange rates, each process of which being assumed to have a unique invariant measure. Notice that without such an assumption, the notion of half-life (central to the recent PPP-puzzle literature) will be ill-posed, since moments need not exist, and convergence of sample moments will not be assured. This assumption of ergodicity allows us consistently to estimate the transition matrix P_n on our grid non-parametrically for each series.

4.1. Some practical problems

In the previous sections, we have shown how we compute numerical approximations of our measures of SMM and SMD, assuming that we have a given P_n on the *n*-point grid. Under the assumption of ergodicity of our time series sample of size T, we know that we can obtain consistent estimates $P_{T,n} \xrightarrow{T\uparrow\infty} P_n$ (c.f. Roussas (1969), Yakowitz (1979)). In addition, we have numerical convergence of the finite linear operator (matrix) P_n on the *n*-point grid to P as the grid size n goes to infinity. In principle, one should investigate the optimal choice of n given sample size T (=312 in our applications). However, such investigations would be mathematically cumbersome, and necessarily include constants of integration that render their usefulness for any given time series quite limited. In our applications, we simply fixed the grid-size at 101×101, and performed sensitivity analyses with 51×51 and 201×201 for known models to ensure that our P_n with n=100 provides a good approximation of P.

We decided to use the estimation approach of Roussas (1969) as adopted in Domowitz and El-Gamal (2001) to estimate $P_{T,n}$ non-parametrically using a kernel estimator. In this framework, we face the perennial problem of bandwidth selection rules. This problem was exacerbated in our applications to real exchange rate data by our small sample size of 312 monthly observations. Our initial experimentation with likelihood-based time series cross-validation (TSCV, c.f. Hart and Vieu (1990)) yielded satisfactory results, with short *m-lives* (in particular half-lives < 2 years) for all series. However, when we performed sensitivity analyses for the number of leave-out-observations in TSCV, our results proved to be excessively sensitive to this formula, rendering our initial results unreliable.[13] Reverting to the more robust Silverman-like plug-in methods (c.f. Hall, Lahiri and Truong (1995)), our sensitivity analyses showed that the results are too sensitive to the selected rule of thumb.[14]

---

[13] We thank Jeff Racine, who suggested that we replace likelihood-based TSCV with *h*-block least squares cross validation. However, the latter procedure proved equally sensitive to the number of leave-out-observations.

[14] Here the "rules of thumb" were various constant multipliers of the familiar $T^{-1/(4+d)}$ rate of convergence.

This non-robustness of our non-parametric approach to studying the PPP-puzzle does not only reflect negatively on the non-parametric approach. In fact, it highlights the importance of parametric modeling assumptions with which one may approach the real exchange rates in question. In this regard, we shall show in Section 5 how this sensitivity to bandwidth selection can provide us with a calibration-type approach to studying the persistence implications of various non-parametric models applied to the specific series in question. Before proceeding to that section, it is worthwhile discussing the effect of over- and under-smoothing on our estimates of $\phi(s)$ and *m-life*.

We begin with the estimated $\phi(s)$ and *m-life* using Silverman's rule of thumb: $h_T = \sigma_T \; T^{-1/5}$, where σ_T is the standard deviation of our series. The estimated *m-lives* with this bandwidth selection rule yielded half-lives below one year for all five series. However, the calibration/simulation exercises in Section 5 showed us that this selection rule produces an over-smoothed estimate of the transition density. The resulting estimates of $\phi(s)$ and *m-life* are consequently biased downward, and converge to zero excessively fast.

In our calibration exercise, we can ask how much under-smoothing we need to impose on our kernel estimates $P_{T,n}$ of the transition matrix P_n to avoid under-estimating our measure of persistence.[15] For parsimony, we would like to summarize the required level of under-smoothing with a single parameter *k*. An convenient way to introduce such a single-parameter "level of under-smoothing" is to place it in the denominator of Silverman's rule of thumb, yielding:

$$h_T = \left(\frac{\sigma_T}{k}\right) T^{-1/5}.$$

4.2. Empirical Applications for real exchange rates

We conclude this section with a summary of the estimated $\phi(s)$ and *m-life* functions for our five time series of real exchange rates, and for various values of the under-smoothing parameter *k*. Figures 1 through 5 show plots of the two functions for values of *k*=1,2,3,4,7 and 10. Selected curves from those figures will be used in conjunction with estimates for the same values of *k* using simulated data from the various recent non-linear models of real exchange rate dynamics reviewed in Section 2.

Before proceeding to Section 5, we need to briefly summarize the regular pattern of our results plotted in Figures 1-5:

- For *k*=1, h_T is the standard Silverman's rule of thumb, and there appears to be no PPP-puzzle. For instance, the estimated half-life for the Japanese Yen is below 10 months, and $\phi(s)$ and *m-life* show fast exponential rates of decay.

[15] A similar approach may be applied to the non-parametric estimation in Shintani (2001), who seemed to use an arbitrary plug-in bandwidth.

Moreover, each of the two functions is fitted very well with an exponentially declining curve.

- In contrast, consider the results for k=3. In this case the half-life for the Japanese Yen is more than 60 months and the *m-life* function is much steeper. Even though $\phi(s)$ seems to be exponentially declining in this case as well, its speed of decay is much slower for this higher value of k.
- A similar pattern can be observed for various values of k.

5. Numerical investigations of recent non-linear models

As we have seen, the recent literature on the PPP-puzzle has turned to non-linear dynamic modeling of the real exchange rate processes for solutions of the puzzle of long half-lives. That literature has focused to date on single point estimates of half-lives from impulse response functions. The rich dynamics of non-linear models offer this literature an infinite number of degrees of freedom to fit this single point to any desired value. In this regard, one may wonder if the infinite flexibility of model selection that has produced different results makes confirmations of the PPP-puzzle, or lack thereof, mere artifacts of the selected models. In other words, the PPP-puzzle in this non-linear modeling world may be in the eye of the beholder.

To study this problem more systematically, we look at the persistence properties of some of the most popular non-linear models of real exchange rates recently investigated in the literature. Instead of focusing on the single point estimate of half-life, we look at the full dynamics of such models in terms of *m-life* and $\phi(s)$ functions. We find in those models that there is a tradeoff between fitting the overall dynamics, and obtaining a satisfactory estimate of half-life for any given series. This tradeoff can be summarized in terms of the degrees of "over- or under-smoothing" implicit in the model. Those measures of the degree of smoothness will be obtained by matching the measures of persistence of the model to estimates using various values of the under-smoothing parameter k.

5.1. Calibration of the under-smoothing parameter using an AR (1) benchmark

Before proceeding to the analysis of implicit over- and under-smoothing in various non-linear models, we calibrate the level of smoothing (value of k) in non-parametric estimation of *m-life* from simulated data generated by an AR(1) process. Clearly, we can solve for the *m-life* function for any AR(1) process with a known parameter. In figure 6(a), we show three *m-life* functions:

1. The first function we plot is the theoretical *m-life* function for a given value of the AR(1) parameter. In particular, we choose values of the AR(1) parameter to produce particular values of half-life. For instance, at ρ=0.9439, we obtain a theoretical half-life of exactly 1 year.[16]

[16] This is a simple calculation for monthly data: $HL = \ln(1/2)/\ln(0.9439) = 12(\text{months})$=1 year

2. The second function we plot is a numerically calculated *m-life* on a fixed 101×101 grid. Assuming an i.i.d. Gaussian error for the AR(1) process, the transition matrix P_n can easily be computed for any given AR(1) coefficient ρ and standard error σ. The goodness of fit of our numerical P_n as an approximation of the theoretical kernel P using this grid-size can be ascertained by comparing the two *m-life* curves. In Table 2, we list various combinations of ρ and standard error σ that produce half-lives of approximately one year, two years, and three years.[17]
3. The third function we plot is a best-fitting non-parametric estimates of *m-life* from simulated data using the same AR(1) model. For the parameters used in Figure 6(a), the non-parametric estimate of *m-life* using an under-smoothing factor *k*=6 matches both theoretical and numerical *m-lives*.

In Figure 7, we show a similar pattern for an AR(1) process with a 2-year half-life. In Table 3, we summarize the various combinations of smoothing factor *k* and resulting half-lives. It appears that a smoothing factor *k*=6 correctly matches the estimated half-life to its theoretical and numerical counterparts. Hence it appears rather easy within the AR(1) framework to match the half-lives of series with varying half-lives with a single choice of the smoothing parameter *k*.

However, if we wish to match more features of the model dynamics than merely the half-life point, or even the entire *m-life* function, the story is quite different. Figures 6(b) and 7(b) show the numerical and estimated $\phi(s)$ functions whose *m-life* functions are shown in figures 6(a) and 7(a), respectively.[18] Using the numerical $\phi(s)$ as our benchmark, we can see that the non-parametrically estimated $\phi(s)$ is too low at *k*=6 (implying over-smoothing of our estimated P_n). However, if we use a higher value of *k* to match the $\phi(s)$ function, we would over-estimate the *m-life.* In other words, in order to match the general dynamics of the process as measured by the $\phi(s)$ function, we would have to admit to the existence of a PPP-puzzle in the sense of an excessively long half-life, even if the actual half-life was quite short. The latter over-estimation of half-life is typically caused by under-smoothing of the estimated P_n. In other words, we have a small-sample tradeoff between over-estimation of the *m-life* function and under-estimation of the $\phi(s)$ functions. Given our sample size, it is impossible to match both functions simultaneously. Of course, the fact that this trade-off is a small sample problem, which vanishes asymptotically, offers little consolation.

As we shall shortly see, tradeoffs of the same sort appear for non-linear models as well. However, not all of the tradeoffs are in the same direction. In particular, some models will seem to have a tendency for over-estimating *m-life*, while others seem to

[17] Quah (1993)'s Markov operator *M* is approximated by discretizing the set of possible values of relative incomes into 5 intervals, i.e., his transition kernel is approximated on a 5x5 grid. Such a small grid-size would induce significant biases in our approximation.
[18] We cannot construct a theoretical $\phi(s)$ function.

have an intrinsic tendency to under-estimate it. Notice that those tendencies are also small-sample problems that we have investigated using simulated data from each maintained model. Asymptotically, the non-parametric estimator is consistent for all of the data-generating models in our investigation. However, for small samples, we can classify models as over-smoothers or under-smoothers depending on the value of k required to match the numerical *m-life* and $\phi(s)$ functions. The tradeoffs between fitting the two functions will also prove instructive regarding the dynamics implied by each of the two non-linear models that we consider below, and their relationship to the benchmark AR(1) model that initially resulted in the puzzling high estimates of half-life.

5.2. The threshold auto-regression (TAR) model

The first nonlinear model we consider is the TAR model proposed to solve the "long half-life" PPP-puzzle in Obstfeld and Taylor (1997), A.Taylor (2001), and Shintani (2001). This model assumes that trade cost frictions induce a “band of inaction” around the long-term PPP level of the exchange rate. Outside this band, the model assumes that significant arbitrage opportunities cause real exchange rates to converge quickly towards their long-term PPP levels. The time series process is assumed to follow a random walk inside the compact band, but to be covariance stationary overall due to mean-reversion outside the band:

$$x_t = \begin{cases} c + \rho(x_{t-1} - c) + \varepsilon_t & \text{if } x_{t-1} \geq c, \\ x_{t-1} + \varepsilon_t & \text{if } -c \leq x_{t-1} \leq c, \\ -c + \rho(x_{t-1} + c) + \varepsilon_t & \text{if } x_{t-1} \leq -c, \end{cases}$$

where $\varepsilon_t \sim N(0, \sigma^2)$. This model is thus parameterized by the AR (1) coefficient outside the band edge: ρ, and the radius of the inaction band: c.

While theoretical *m-life* cannot be easily computed for this model, we can calculate *m-life* and $\phi(s)$ numerically as before, for any given values of ρ and c. We can also use simulated data for various values of those parameters to obtain nonparametric estimates of *m-life* and $\phi(s)$. In Tables 4 and 5, we summarize the numerical half-life and non-parametrically-estimated half-lives for various parameter values and smoothing factor k.

The 2-year half-life case is shown in Figure 8. For the parameters shown in that figure, we need to use an under-smoothing factor of k=170 to match the numerical half-life of the process. However, this value of k leads to excessive under-smoothing, as shown in the over-estimation of $\phi(s)$, plotted in Figure 8(b). Thus, despite the added complexity in this model, there is still a small sample tradeoff between matching the half-life of the process, and matching its more general dependence structure as measured by the $\phi(s)$ function. The same pattern is also seen in the 3-year half-life parameterization, shown in Figure 9 for k=185.

More importantly, the small sample tradeoff in this model is the opposite of that for the AR(1) model. For the TAR(1) model, the tradeoff is between over-estimating $\phi(s)$ or under-estimating *m-life*. In other words, the TAR model tends to under-estimate *m-life* relative to the AR model! It is thus not surprising that the studies utilizing this model estimated shorter half-lives, and the Monte Carlo analyses in that literature suggested that using an AR(1) model to estimate half-life when the DGP was a TAR(1) will tend to produce over-estimates. For instance, Obstfeld and Taylor (1997) have shown this tendency in their investigation of the law of one price across various cities in the U.S.A.

5.3. The smooth transition autoregression (STAR) model

We now consider the STAR model, introduced in Granger and Teräsvirta (1993) as a promising non-linear model, and recently employed by Michael et al (1997), Taylor and Sarno (1998), Taylor et al (2001), and Baum et al (2001) to address the PPP-puzzle. The STAR model takes the form:

$$q_t - \mu = \sum_{j=1}^{p} \beta_j [q_{t-j} - \mu] + \sum_{j=1}^{p} \beta_j^* [q_{t-j} - \mu]\, \Phi[\theta:\ q_{t-d} - \mu] + \varepsilon_t,$$

where $\{q_t\}$ is a stationary and ergodic process with mean μ, $\varepsilon_t \sim iid(0,\sigma^2)$, and $(\theta,\mu) \in \mathbb{R}^+ \times \mathbb{R}$. The transition function $\Phi[\theta:\ q_{t-d} - \mu]$ determines the degree of mean-reversion,[19] and is controlled by the parameter θ. We shall consider a similar model with an exponential transition function, which is called the Exponential STAR (ESTAR) model. Taylor et al (2001) proposed this model with $p = d = 1$:

$$q_t - \mu = \beta_1[q_{t-1} - \mu] + \beta_1^*[q_{t-1} - \mu](1 - \exp\{-\theta^2[q_{t-1} - \mu]^2\}) + \varepsilon_t$$

To further simplify the model, we consider the case where $\beta_1 = 1,\ \beta_1^* = -1,\ \mu = 0$:

$$q_t = \mu + (q_{t-1} - \mu)(1 - \exp\{-\theta^2[q_{t-1} - \mu]^2\}) + \varepsilon_t$$

This leaves us with the single parameter θ for the numerical calculation and estimation of *m-life* and $\phi(s)$ functions at various parameter values.[20] The various parameter values yielding different half-lives are summarized in Table 6. Consider the case of θ=0.16, with a 3-year half-life. In this case, we can see in Figure 10 that the

[19] The transition parameter θ determines the speed of transition between two extreme regimes. When $\Phi = 0$ (no transition), the model becomes a linear AR (p). At the other extreme of $\Phi = 1$, the model becomes a different AR(p) with a different speed of mean reversion if $\beta_j^* \neq 0$ for some *j*.

[20] We consider $\theta \in [0,1]$. As $\theta \downarrow 0$, the process approaches a random walk. As $\theta \uparrow 1$, the process becomes purely stationary. Thus we can expect higher rates of mean reversion for higher values of θ.

nonparametric estimates with k=3.5 can match the numerical *m-life* function (see also Table 7). However, at this value of k, $\phi(s)$ is under-estimated.

In other words, the ESTAR model produces the same trade-off as the AR(1) model: we can either over-estimate *m-life* or under-estimate $\phi(s)$. In fact, we can rank-order the inherent levels of implicit smoothing in the three models, and conclude that the ESTAR model tends to over-estimate *m-life* relative to the TAR model, and under-estimate it relative to the AR model. For instance, Taylor et al (2001) have shown that the estimated half-life in the ESTAR model with a 20% shock was in the range of 18~24 months, which is considerably shorter than the stylized fact of 3~5 years half-life.

6. Monte Carlo and empirical analysis of AR, TAR and ESTAR estimators

In section 5, we have shown that there are various tradeoffs between fitting the overall dynamics and obtaining a satisfactory estimate of half-life. In recent years, non-linear models of real exchange rate processes were proposed, and the literature has focused on showing that linear AR estimates of half-life are biased when the true data generating process is non-linear. For instance, Taylor (2001) and Shintani (2002) hypothesized a TAR(1) model of the data generating process, and performed Monte Carlo analyses for AR(1) estimates under that maintained hypothesis. Those Monte-Carlo analyses suggested that the AR(1) model over-estimates the process half-life when the true DGP is a TAR(1). Similar results were obtained by Taylor and Sarno (1998) and Taylor et al (2001) using a STAR specification of the DGP.

To assess the effects of model mis-specification more fully, we perform three sets of Monte Carlo analyses, with the DGP being specified once as an AR(1), once as a TAR(1), and once as an ESTAR (1). Under each of the three DGP specifications, we estimate all three models, to see if TAR(1) estimates of half-life are always lower than their AR(1) counterparts, etc.

For easy comparison with earlier Monte Carlo results in the literature, first consider the results under the maintained hypothesis of a TAR(1) DGP, reported in Table 9(b). We generate simulated data samples of length 312, and for each simulated sample we estimate the parameters of each of our three models: AR(1), TAR(1) and ESTAR(1).[21] For each given parameterization of the TAR(1) DGP, allowing for different values of ρ and c (see section 5) and a fixed value of σ, we obtain four numerical calculations of *m-life* (one for the true parameters and DGP, and one for each of the three estimated models). We can see in Table 9(b) that the AR(1) model overestimates half-life quite significantly, in accordance with the earlier Monte Carlo results found in the literature. Since the TAR(1) model is correctly specified, estimated half-lives under that model do not differ significantly from their numerical counterparts when the parameters are known. Furthermore, since the ESTAR(1) model is a more flexible version of the

[21] For estimation of the TAR(1) model, we employ the best grid search and maximum likelihood estimation method used in Obsfeldt and Taylor (1997).

TAR(1) model, half-life estimates under the former are also very close to their numerical counterparts.

The results are different, however, if we postulate the ESTAR(1) model as the DGP. Table 9(c) shows in this case that the TAR(1) model will have a higher tendency of over-estimating half-life, even compared to the AR(1) model. Similarly, if the true DGP is AR(1), we find in Table 9(a) that TAR(1) model also overestimates the most. In other words, within our class of models, the TAR(1) model only obtains a relatively short half-life if it is the correctly specified model, otherwise it would tend to over-estimate half-life more severely than other linear and non-linear models under consideration. This Monte Carlo result is of particular interest given the empirical findings that follow.

Figure 12 shows maximum likelihood estimates of the AR(1), TAR(1), and ESTAR(1) models for each of the five real exchange rates analyzed earlier in the paper. It is interesting to note that the TAR(1) model produces the longest *m-life* for all five series. In some cases (e.g. for the Japanese Yen), the level of over-estimation of *m-life* by the TAR(1) model relative to the ESTAR model is quite significant. The cases of the French Franc and the Deutsche Mark are particularly interesting, since both cases produce lower estimated *m-lives* under the AR(1) model relative to both the TAR and ESTAR models. Our Monte Carlo studies suggested that the TAR model would produce shorter half-life estimates than the AR model if the DGP was indeed a TAR process. Since the TAR model produced longer half-life estimates than the AR model for all of our series, this suggests that the TAR model is incorrect. More importantly, the estimated results from all three models tend to be qualitatively similar, in the sense that the difference between the shortest and the longest estimated half-lives is rarely large enough to make a qualitative difference for the validity of the medium-term PPP-hypothesis. Qualitatively, the estimated half-lives under all three models are less than two years for all series with the exception of the Japanese Yen. For the latter, the half-life is roughly two years under the ESTAR model, and three or more years under the AR and TAR models. In other words, if the goal was to obtain half-lives at or below two years for all series, one would only need to use the ESTAR model to resolve the PPP-puzzle.

7. Concluding Remarks

A number of non-linear models have been proposed in the late 1990s to solve the second-incarnation of the PPP-puzzle: half-lives exceeding two years. In the immediately preceding analysis, we saw that the ESTAR model produces half-lives at or below two years for all five real exchange rate series considered in this paper. However, selecting a model to estimate *ex post*, based on meeting a half-life target that solves a supposed puzzle, is dubious at best. The space of non-linear models is quite large, and there is no doubt that for any fixed data set, one can find a model that produces any desired pattern of half-lives. Indeed, as we have seen in our Monte Carlo studies and in some of the empirical results, there are instances in which the ESTAR model would produce estimated half-lives that are longer than those obtained with an AR model. If we happened to observe such a series, and if the AR model's half-life estimate was at the boundary of our acceptable set, we might be tempted to revert back to linear models of real exchange rate processes! In fact, if one fixes a single criterion (e.g. a half-life less than a particular value), and allows for an infinite class of models, the puzzle can only be in the eye of the beholder.

Instead, we propose that if non-linear models are entertained, their full dynamical structure should be analyzed, instead of focusing on a single linear measure of dependence, such as the half-life concept. Towards that end, we have devised numerical methods for computing measures of short memory in mean and short memory in distribution (*a la* Granger (1995)) to analyze the full dependence structure of any estimated non-linear model. In this regard, we would argue that the choice of a non-linear model to study any particular phenomenon (e.g. the behavior of real exchange rates) should be driven by theoretical and institutional considerations, rather than the model's ability to yield desirable parameter values. While the latter can be easily attained when using the infinite number of degrees of freedom afforded by selection from a very large class of models, it is also uninformative for the same reason.

Appendix

We wish to show that maximization over all Borel sets A of initial conditions may be reduced to maximization over all singleton initial conditions. Since we have restricted attention to a finite grid approximation, we may construct a proof by induction. Thus, we first consider conditioning on two initial conditions relative to conditioning on one, in order to show that:

$$\{\Pr(x_s \in B \mid x_0 = i \text{ or } x_0 = j) - \Pr_{f*}(x_s \in B)\} \;\le\; \{\Pr(x_s \in B \mid x_0 = i) - \Pr_{f*}(x_s \in B)\}$$

We can re-write both the LHS and the RHS as follows:

$$\text{LHS } = \frac{\Pr(x_s \in B \text{ and } (x_0 = i \text{ or } x_0 = j))}{\Pr(x_0 = i \text{ or } x_0 = j)} - \Pr_{f*}(x_s \in B)$$

$$= \frac{\Pr(x_s \in B \text{ and } x_0 = i) + \Pr(x_s \in B \text{ and } x_0 = j) - \Pr_{f*}(x_s \in B)\{\Pr(x_0 = i) + \Pr(x_0 = j)\}}{\Pr(x_0 = i) + \Pr(x_0 = j)}$$

$$= \frac{\{\Pr(x_s \in B \text{ and } x_0 = i) - \Pr(x_0 = i)\Pr_{f*}(x_s \in B)\} + \{\Pr(x_s \in B \text{ and } x_0 = j) - \Pr(x_0 = j)\Pr_{f*}(x_s \in B)\}}{\Pr(x_0 = i) + \Pr(x_0 = j)}$$

$$\text{RHS } = \frac{\Pr(x_s \in B \text{ and } x_0 = i) - \Pr(x_0 = i)\Pr_{f*}(x_s \in B)}{\Pr(x_0 = i)}$$

We first verify the desired result for the case of positive LHS, the other case follows by symmetry: $\Pr(x_s \in B \mid x_0 = i \text{ or } x_0 = j) > \Pr_{f*}(x_s \in B)$.

We further assume without loss of generality that:

$$\Pr(x_s \in B \mid x_0 = i\ ) > \Pr(x_s \in B \mid x_0 = j\ ) \Leftrightarrow \frac{\Pr(x_s \in B \text{ and } x_0 = i)}{\Pr(x_0 = i)} > \frac{\Pr(x_s \in B \text{ and } x_0 = j)}{\Pr(x_0 = j)}$$

Note that: $\frac{A}{a} > \frac{B}{b} \;\Rightarrow\; \frac{B}{b} < \frac{A+B}{a+b} < \frac{A}{a}$ for $A, B, a, b > 0$.

Setting $\frac{A}{a} = \frac{\Pr(x_s \in B \text{ and } x_0 = i)}{\Pr(x_0 = i)}$ and $\frac{B}{b} = \frac{\Pr(x_s \in B \text{ and } x_0 = j)}{\Pr(x_0 = j)}$, the desired inequality holds.

Proceeding by induction, we can see that maximization over all Borel sets of initial conditions reduces on the finite grid to maximization over all single initial conditions.

References

Baum, C., Barkoulas, J., and Caglayan, M., 2001. Nonlinear adjustment to purchasing power parity in the post-Bretton Woods era. *Journal of International Money and Finance* **20**, pp.379-399.

Bose, C., 1994. On the existence and approximation of invariant densities for nonsingular transformations on $\mathbb{R}^d$. *Journal of Approximation Theory* **79**, pp. 260-270.

Cheung, Y. and Lai, K., 2001. On the purchasing power parity puzzle. *Journal of International Economics* **52**, pp.321-330.

Darby, M., 1983. Movements in purchasing power parity: The short and long runs, in M. Darby and J. Lothian, eds. *The international transmission of inflation* , University of Chicago Press, Chicago.

Day, R. and Pianigiani, G., 1991. Statistical dynamics and economics. *Journal of Dynamics and Control* **16**, pp.37-83.

Domowitz, I. and El-Gamal, M.1993. A consistent test of stationary ergodicity. *Econometric Theory* **9**, pp.589-601.

Domowitz, I. and El-Gamal, M.1996. Is your model chaotic ? A consistent test simulation-based test, in Lakshmikantham, V. eds. *Proceedings of the First World Congress of Nonlinear Analysts*. Walter de Gruyter, Berlin.

Domowitz, I. and El-Gamal, M.2001. A consistent nonparametric test of ergodicity for times series with applications. Journal of *Econometrics* 102, pp.365-398.

Durlauf, S., 1993. Nonergodic economic growth. *Review of Economic Studies* **60**, pp.349-366.

Frankel, J.A. and Rose, A., 1996. Mean reversion within and between countries: a panel project on purchasing power parity. *Journal of International Economics* **40**, pp. 209–224.

Froot, K.A. and Rogoff, K., 1995. Perspectives on PPP and long-run real exchange rates. In: Grossman, G. and Rogoff, K., Editors, 1995. *Handbook of International Economics*, North-Holland, Amsterdam (chap. 32).

Gallant, A. and White, H., 1988. *A unified theory of estimation and inference for nonlinear dynamic models*. Bail Blackwell, New York

Goldberg, L.G., Gosnell, T.F. and Okunev, J., 1997. Purchasing power parity: modeling and testing mean reversion. *Journal of Banking and Finance* **21**, pp. 949–966.

Granger, C.W.J., 1995. Modelling Nonlinear Relationship between Extended-Memory Variables. *Econometrica* **63** 2, ppp.265-279

Granger, C.W.J. and Teräsvirta, T., 1993. *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford, UK.

Hakkio, C., 1984. A reexamination of purchasing power parity. *Journal of International Economics* **17**, pp.265-277.

Hart, J. and Vieu, P., 1990. Data-driven bandwidth choice for density estimation based on dependent data. *Annals of Statistics* **18**, pp.873-890.

Hall. P., Lahiri,S., Troung, Y., 1995. On bandwidth choice for density estimation with dependent data. *Annals of Statistics* **23**, pp. 2241-2263.

Jacod, J. and Shiryaev, A., 1987. *Limit Theorems for Stochastic Processes*, Springer-Verlag, Berlin.

Kress, R., 1998. *Numerical Analysis*, Springer.

Lasota, A. and Mackey, M., 1994. *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamic,* Springer-Verlag, New York.

Li., T., 1976. Finite approximation for the Frobenius-Perron operation: a solution to Ulam's conjecture. *Journal of Approximation Theory* **17**, pp. 177-186.

Lothian, J.R. and Taylor, M.P., 1996. Real exchange rate behavior: the recent float from the perspective of the past two centuries. *Journal of Political Economy* **104** 3, pp. 488–510.

Mankiw, G., Romer, D., and Weil, D., 1992. A contribution to the empirics of economic growth. *Quarterly Journal of Economics* **107**, pp.407-437.

McLeish, D., 1978. A maximal inequality and dependent strong laws. *Annals of Probability* **3**, pp.264-290.

Michael, P., Nobay, A.R. and Peel, D.A., 1997. Transactions costs and nonlinear adjustment in real exchange rates: an empirical investigation. *Journal of Political Economy* **105** 4, pp. 862–879.

Murray, C.J., Papell, D.H., 2001. The Purchasing Power Parity is worse than you think: A note on long-run real exchange rate behavior. mimeo

Murray, C.J., Papell, D.H., 2002. The Purchasing Power Parity Persistence Paradigm. *Journal of International Economics* **56** 1, pp. 1-19.

Obstfeld, M. and Taylor, A., 1997. Nonlinear aspects of goods-market arbitrage and adjustment: Heckscher's commodity points revisited. *Journal of the Japanese & International Economies* **11** 4, pp. 441–479.

Papell, D., 1997. Searching for stationarity: purchasing power parity under the current float. *Journal of International Economics* **43**, pp. 313–332.

Papell, D. and Theodoridis, H., 1998. Increasing evidence of purchasing power parity over the current float. *Journal of International Money and Finance* **17**, pp. 41–50.

Quah, D., 1993. Empirical cross-section dynamics in economic growth. *European Economic Review* **37**, pp.426-434.

Quah, D., 1997. Empirics for growth and distribution: polarization, stratification, and convergence clubs. *Journal of Economic Growth* **2**, pp.27-59

Rogoff, K., 1996. The purchasing power parity puzzle. *Journal of Economic Literature* **34**, pp. 647–668.

Roussas, G., 1969. Nonparametric estimation in Markov processes. *Annals of the Institute of Statistical Mathematics* **21**, pp.73-87.

Sala-i-Martin, X., 1996. The classical approach to convergence analysis. *Economic Journal* **106**, pp.1019-1036.

Shintani, M., 2001. A Nonparametric measure of convergence toward purchasing power parity. Mimeo.

Tayolr, A., 2001. Potential pitfalls for the purchasing power parity puzzle? Sampling and specification biases in mean-reversion tests of the law of one price. *Econometrica* **69**, pp.473-498.

Taylor, M. and Sarno, L., 1998. The behavior of real exchange rates during the post-Bretton Woods period. *Journal of International Economics* **46**, pp. 281–312.

Taylor, M., Peel,D., and Sarno, L., 2001. Nonlinear mean-reversion in real exchange rates: towards a solution to the purchasing power parity puzzle. *International Economic Review* **42**, pp.1015-1042.

Tong, H., 1993. *Non-linear Time Series: A Dynamical System Approach*, Oxford Science Publications, Oxford, UK.

White, H. and Domowitz, I., 1984. Nonlinear regression with dependent observations. *Econometrica* **52**, pp.143-162.

Yakowitz, S., 1979. Nonparametric Estimation of Markov Transition Functions. *Annals of Statistics* **7**, pp.671-679.

< Table 1. *k* rule of thumbs' bandwidth and half-lives >

K	Japan	France	U.K.	Germany	Swiss
1	**5**	**5**	**8**	**5**	**7**
2	33	**15**	**22**	**20**	**23**
3	> 60	37	29	40	37
4	> 60	55	32	56	45
5	> 60	> 60	34	> 60	45
6	> 60	> 60	35	> 60	53
7	> 60	> 60	36	> 60	55
8	> 60	> 60	37	> 60	56
9	> 60	> 60	37	> 60	56
10	> 60	> 60	37	> 60	57

< Table 2. Numerical calculation of half-lives and AR(1) model >

(unit, months)

ρ=0.9439	σ	0.010	0.015	0.020	0.025
1-year	Half-life	**12**	12	12	11
ρ=0.9715	σ	0.008	0.011	0.014	0.017
2-year	Half-life	**23**	23	23	22
ρ=0.9809	σ	0.008	0.010	0.012	0.014
3-year	Half-life	**35**	34	34	33

< Table 3. Smoothing Factor *k* and and half-lives in AR(1) model >

(unit, months)

ρ=0.9439	*K*	1	2	3	4	5	**6**
σ=0.010	Half-life	7	11	11	11	11	**11**
ρ=0.9715	*K*	1	2	3	4	5	**6**
σ=0.008	Half-life	8	17	19	20	21	**21**
ρ=0.9809	*K*	1	2	3	4	5	**6**
σ=0.008	Half-life	8	20	25	27	28	**30**

< Table 4. Numerical calculation of half-lives and TAR(1) model >

(unit, months)

ρ=0.92	C	**0.75**	0.73	**0.72**	0.65	**0.60**
	I(1) interval	**[0.25, 0.75]**	[0.27, 073]	**[0.28, 0.78]**	[0.35, 0.65]	**[0.40, 0.60]**
	Half-life	**36**	28	**25**	14	**12**
ρ=0.94	C	0.75	**0.73**	0.72	0.69	**0.68**
	I(1) interval	[0.25, 0.75]	**[0.27, 0.73]**	[0.28, 0.72]	[0.31, 0.69]	**[0.32, 0.68]**
	Half-life	47	**37**	34	26	**24**
ρ=0.96	C	0.75	0.69	**0.68**	0.65	**0.60**
	I(1) interval	[0.25, 0.75]	[0.31, 0.69]	**[0.32, 0.68]**	[0.35, 0.65]	**[0.40, 0.60]**
	Half-life	> 60	39	**36**	30	**24**

< Table 5. Smoothing Factor *k* and half-lives in TAR model >

(unit, months)

ρ=0.92	c=0.75	*k*	150	170	**185**	190	200
		Half-life	14	23	**35**	40	54
	c=0.72	*k*	150	170	185	190	200
		Half-life	14	**22**	34	40	58
	c= 0.60	*k*	150	170	185	190	200
		Half-life	16	27	42	49	> 60
ρ=0.94	c=0.73	*k*	100	120	**123**	**126**	129
		Half-life	20	32	**35**	**35**	38
	c=0.68	*k*	100	120	123	126	129
		Half-life	20	29	32	35	38
	c=0.55	*k*	100	120	123	126	129
		Half-life	20	29	32	**35**	37
ρ=0.96	c=0.68	*k*	25	30	32	**33**	35
		Half-life	24	31	35	**36**	41
	c=0.60	*k*	**25**	30	32	33	35
		Half-life	**24**	31	35	37	42

< Table 6. Numerical calculation of half-lives and ESTAR(1) model >

(unit, months)

θ	half-life
0.12	57
0.14	44
0.16	**36**
0.18	29
0.20	**24**
0.22	20
0.24	17
0.26	14
0.28	**12**

< Table 7. Smoothing Factor *k* and half-lives in ESTAR model>

(unit, months)

k	θ=0.16	θ=0.20	θ=0.28
1	5	5	5
1.5	11	11	**10**
2	18	18	17
2.5	25	**25**	24
3	32	32	31
3.5	**37**	38	37
4	43	43	43
5	51	51	51
6	57	57	57
10	> 60	> 60	> 60

< Table 8(a). Percentiles for ergodicity test >

Series	% p- value < 0.05	% p- value < 0.1
Japanese Yen	16	25
French Franc	3	8
British Pound	8	16
Deutsche Mark	5	12
Swiss Franc	5	10

< Table 8(b). Percentiles for mixing test >

Series	% p- value < 0.05	% p- value < 0.1
Japanese Yen	3	8
French Franc	3	7
British Pound	3	8
Deutsche Mark	3	8
Swiss Franc	3	8

< Table 9(a) Simulation results: True DGP is AR(1) >

(unit, months)

ρ	0.95	0.96	0.97	0.98	0.99
AR(1) known parameters	8 (7,9)	9 (8,11)	12 (10,14)	16 (12,19)	23.5 (16,31)
AR(1) Estimated parameters	8 (7,9)	9.5 (7,11)	12 (9,15)	16 (11,20)	22.5 (14,30)
TAR(1) Estimated parameters	9 (7,11)	11 (8,14)	14 (10,19)	18 (13,26)	23 (15,33)
ESTAR(1) Estimated parameters	8 (7,10)	9 (7,12.5)	12 (8,16)	14 (10,20)	18.5 (12,27)

Note : Fix σ=0.01. Median of calculated half-lives with 25th and 75th percentiles in parenthesis. Total 100 replications.

< Table 9(b) Simulation results: True DGP is TAR(1) >

(unit, months)

c	0.90	0.80	0.70	0.60	0.50
TAR(1) Known parameters	17 (10,28)	17 (10,29)	16 (11,28)	17 (10,30)	19.5 (12,32)
AR(1) Estimated parameters	22 (10,28)	22 (10,49)	22.5 (13,49)	22 (11,53)	24 (15,46)
TAR(1) Estimated parameters	18.5 (10,35)	18.5 (10,34)	19.5 (13,31)	18.5 (10,33)	23 (14,33)
ESTAR(1) Estimated parameters	15.5 (9,30)	16.5 (9,30)	16 (11,29)	16 (8,33)	18.5 (12,31)

Note : Fix ρ=0.97 σ=0.01.

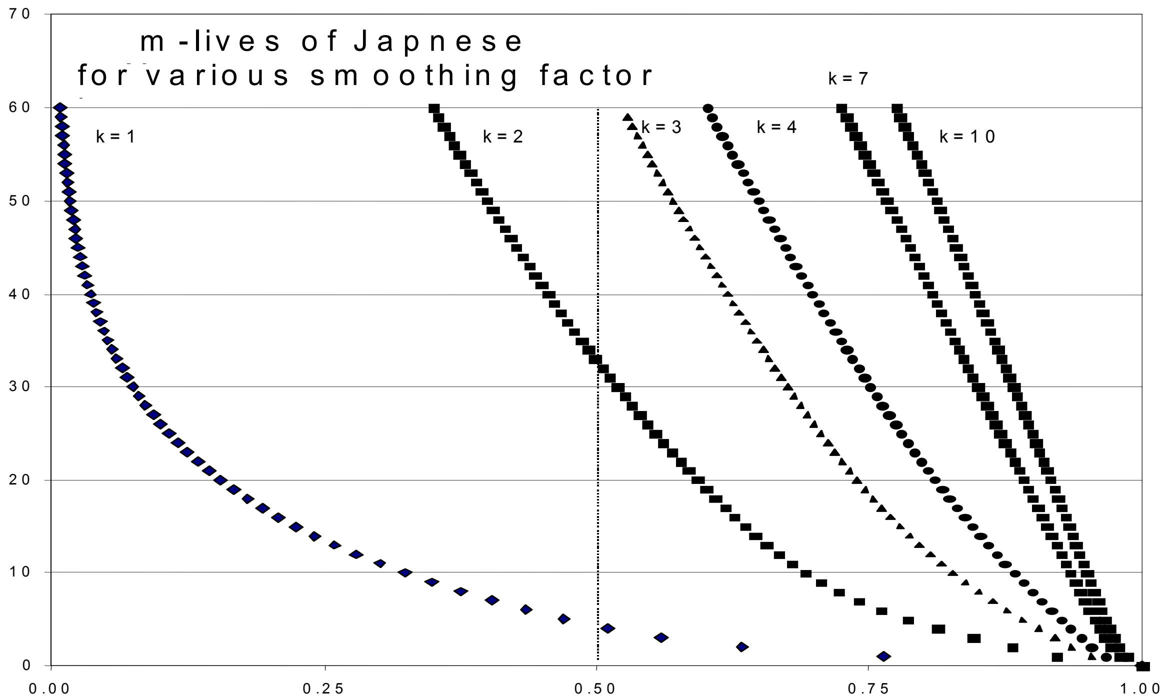
< Table 9(c) Simulation results: True DGP is ESTAR(1) >

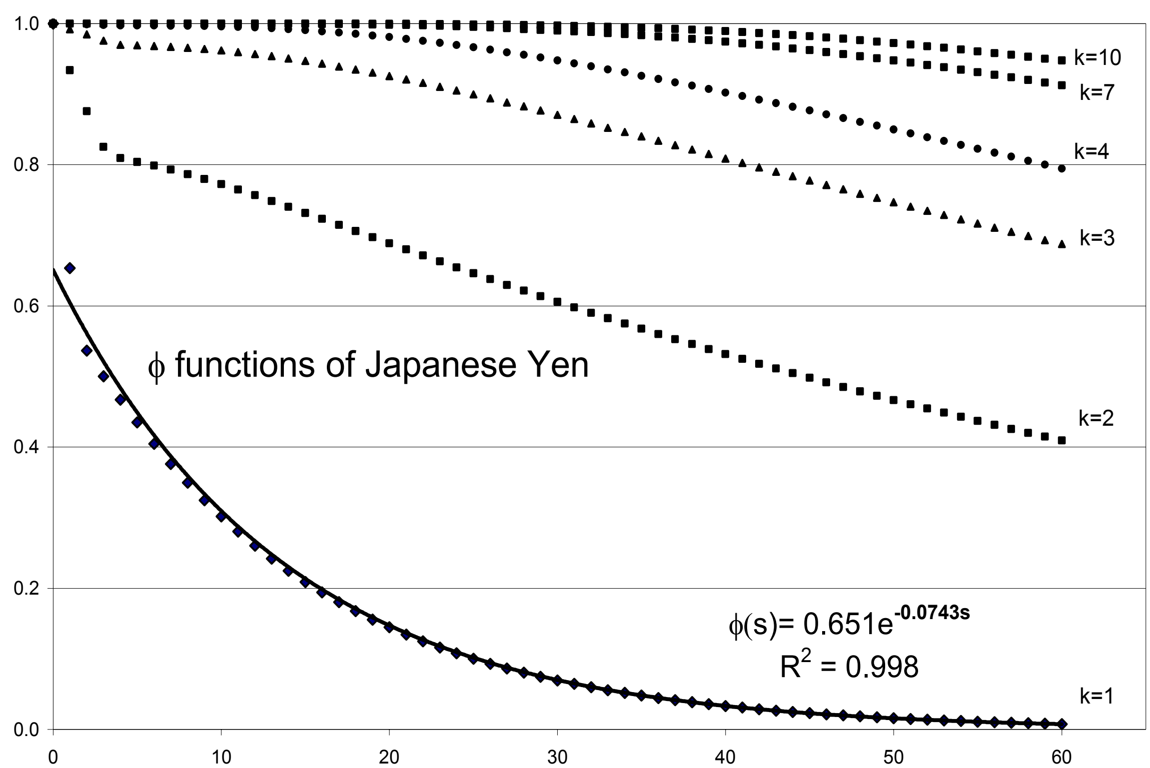
(unit, months)

θ	0.50	0.40	0.30	0.20	0.10
ESTAR (1) Known parameters	7 (6,8)	9 (8,9)	11 (9,12)	15 (11,17)	20 (15,28)
AR(1) Estimated parameters	10 (8,11)	11 (10,13)	14 (11,16)	17 (12,22)	21.5 (15,32)
TAR(1) Estimated parameters	10 (9,11)	11 (10,14)	14 (11,17)	17.5 (12,22)	21.5 (16,33)
ESTAR(1) Estimated parameters	7 (6,8)	8 (7,10)	10 (8,13)	13 (10,17)	17 (12,27)

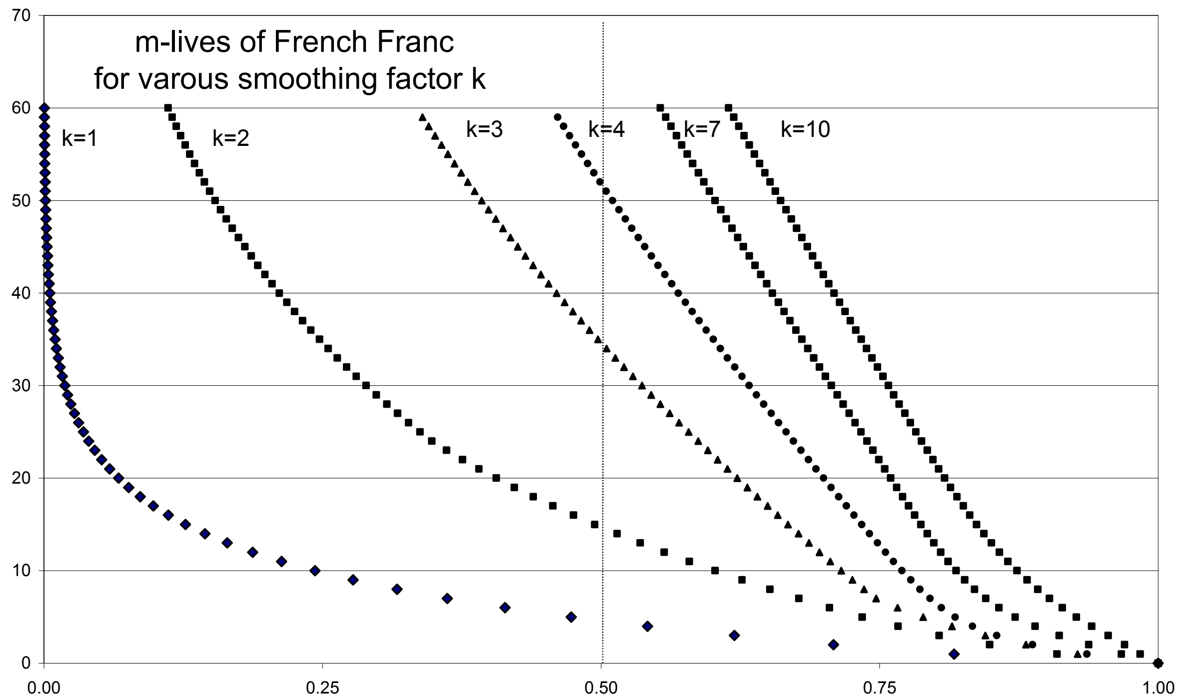
Note : Fix μ=0.5, σ=0.1.

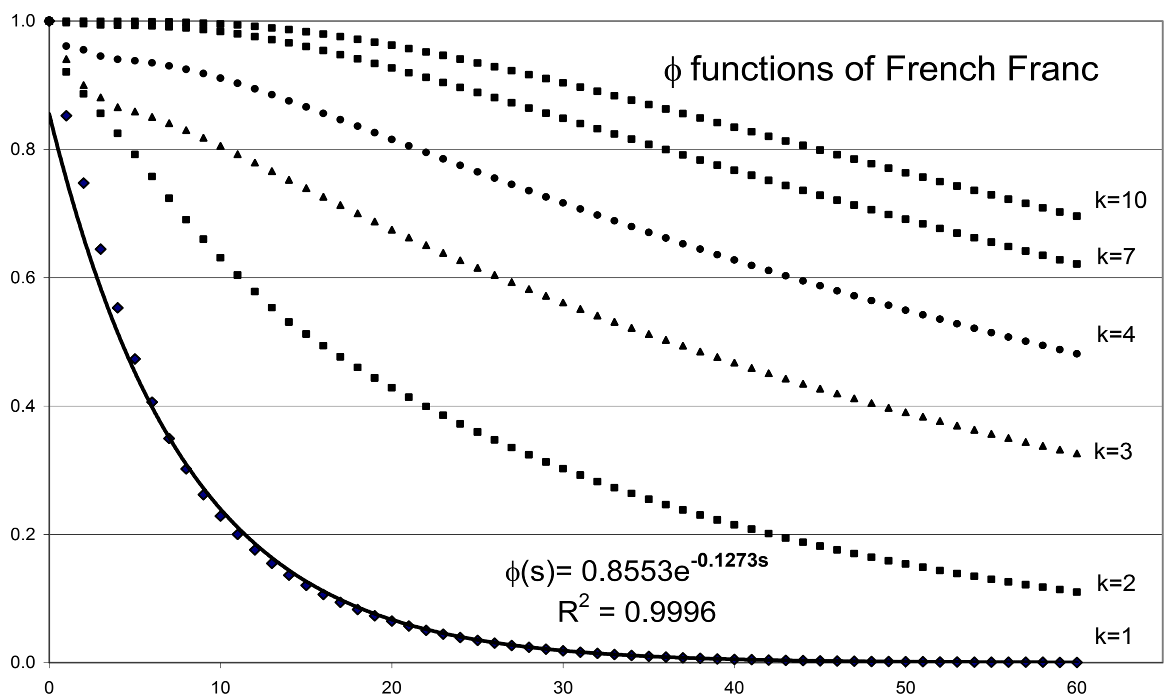
< Figure 1. *m-life* and $\phi(s)$ of Japanese Yen >



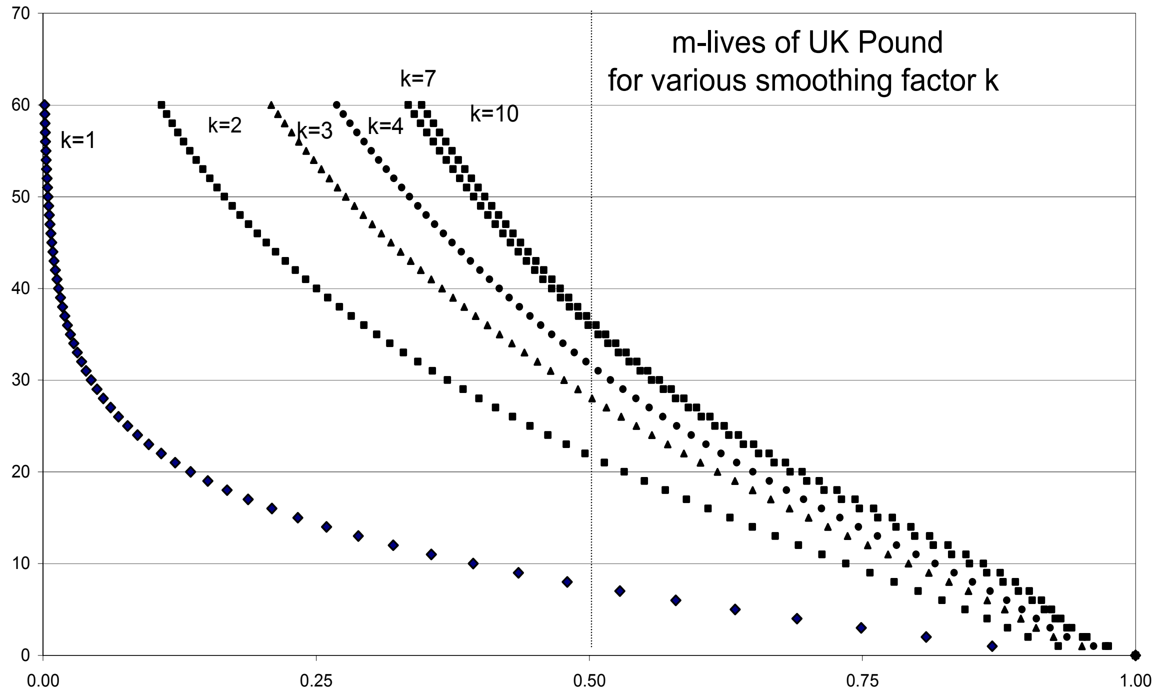


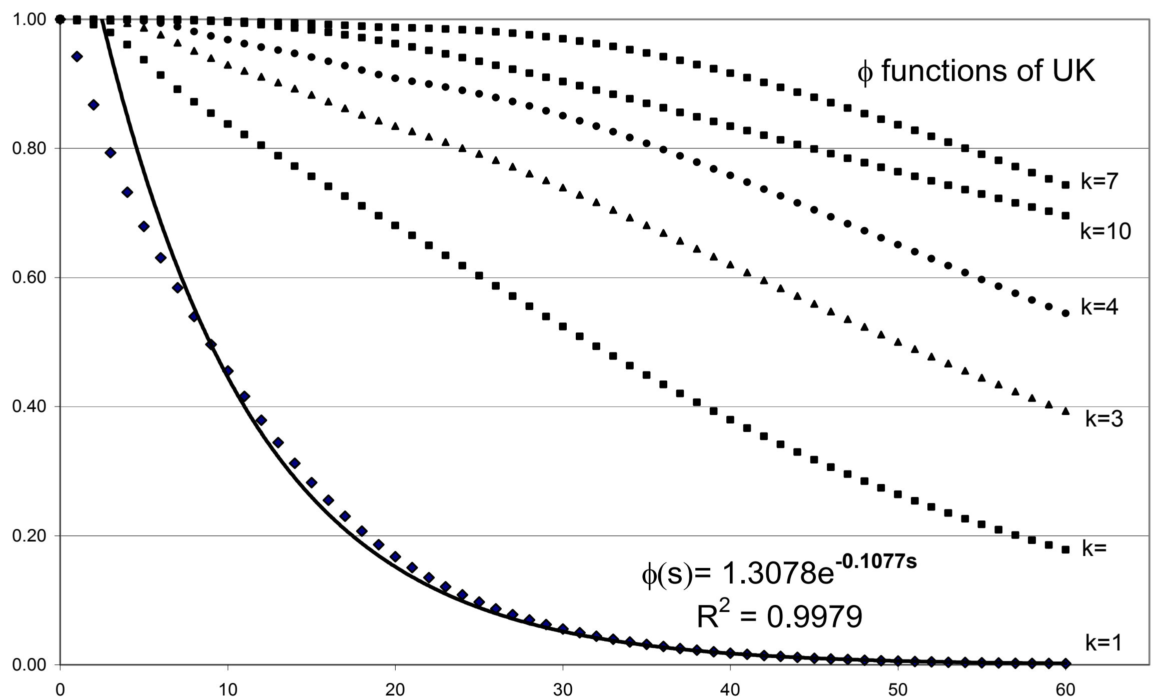
< Figure 2. *m-life* and $\phi(s)$ of French Franc >



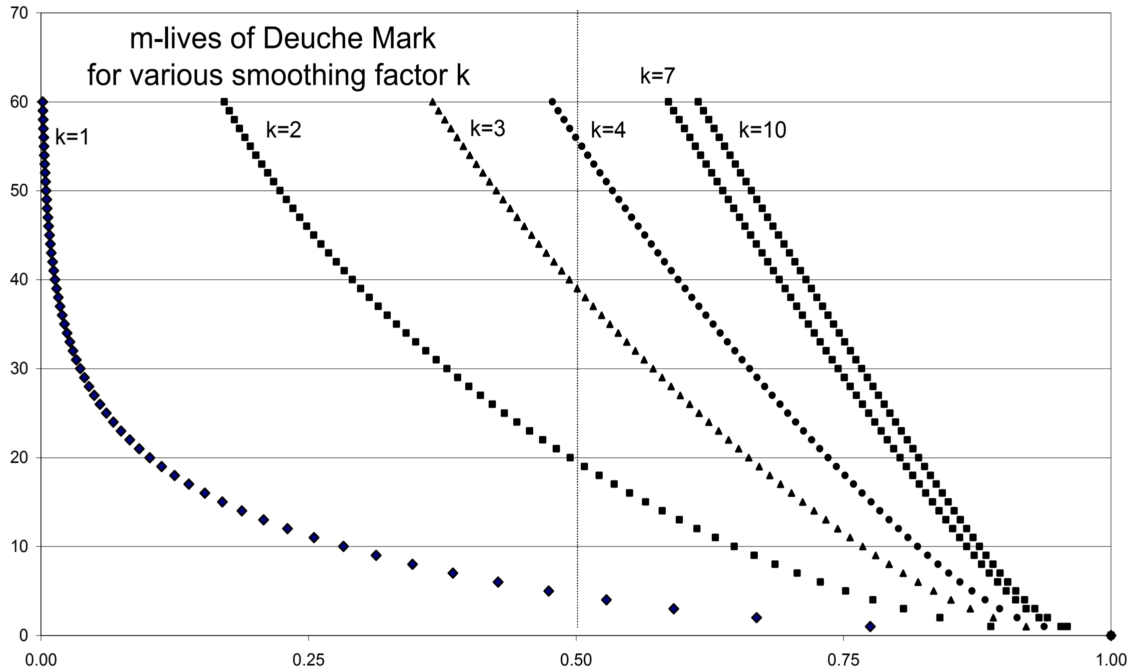


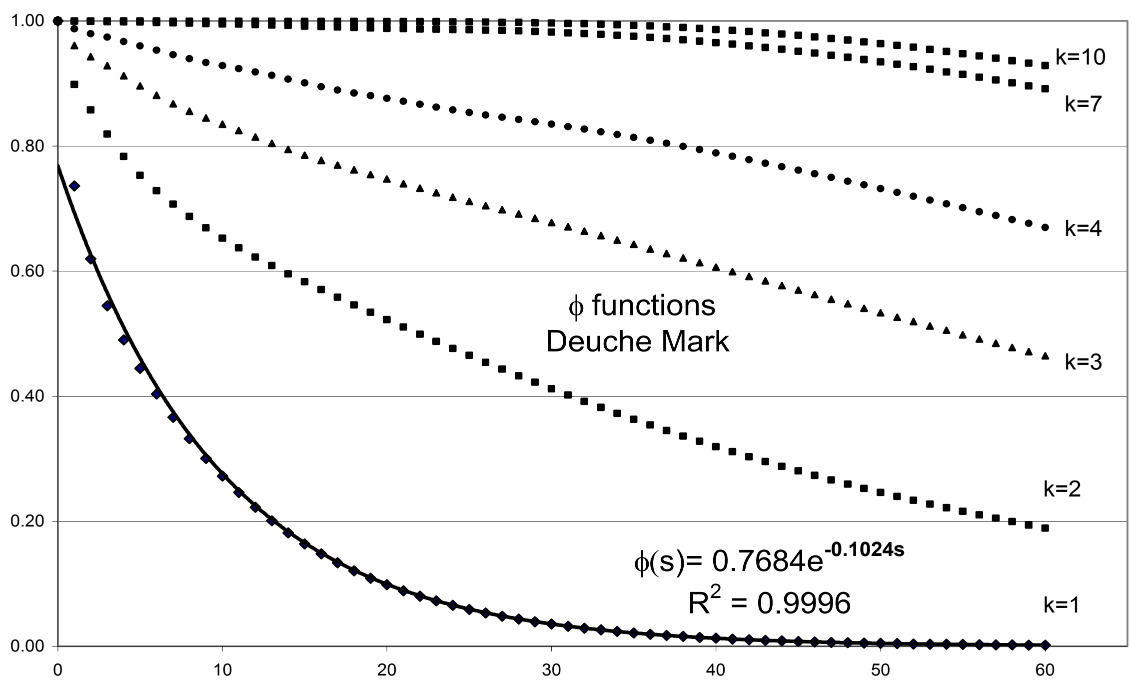
< Figure 3. *m-life* and $\phi(s)$ of U.K. Pound >



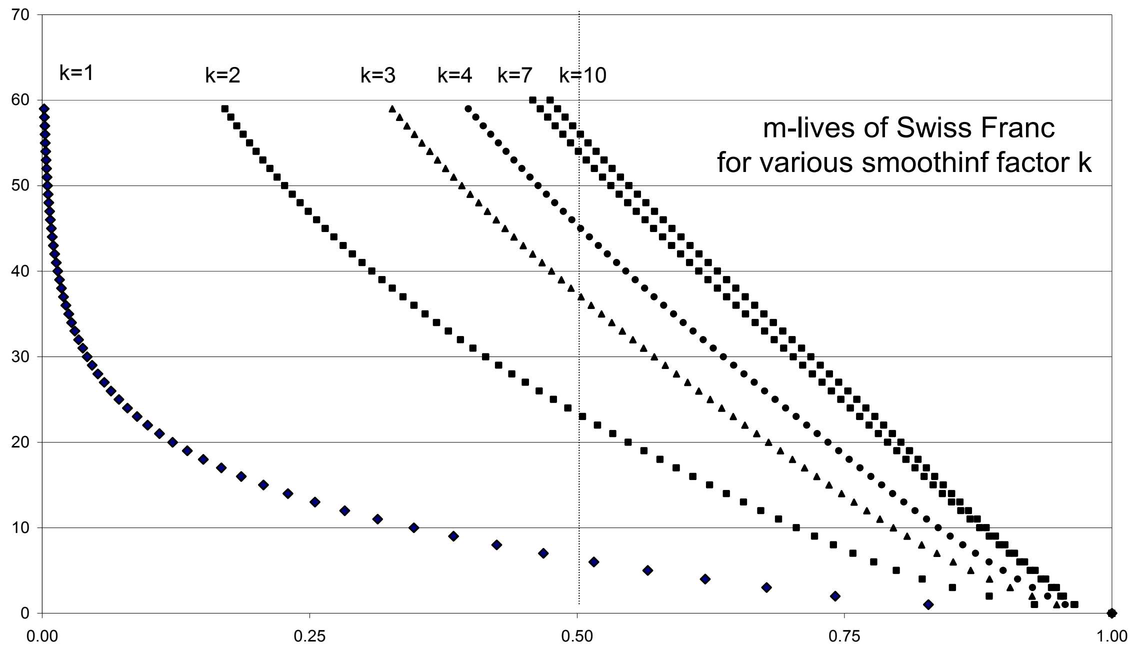


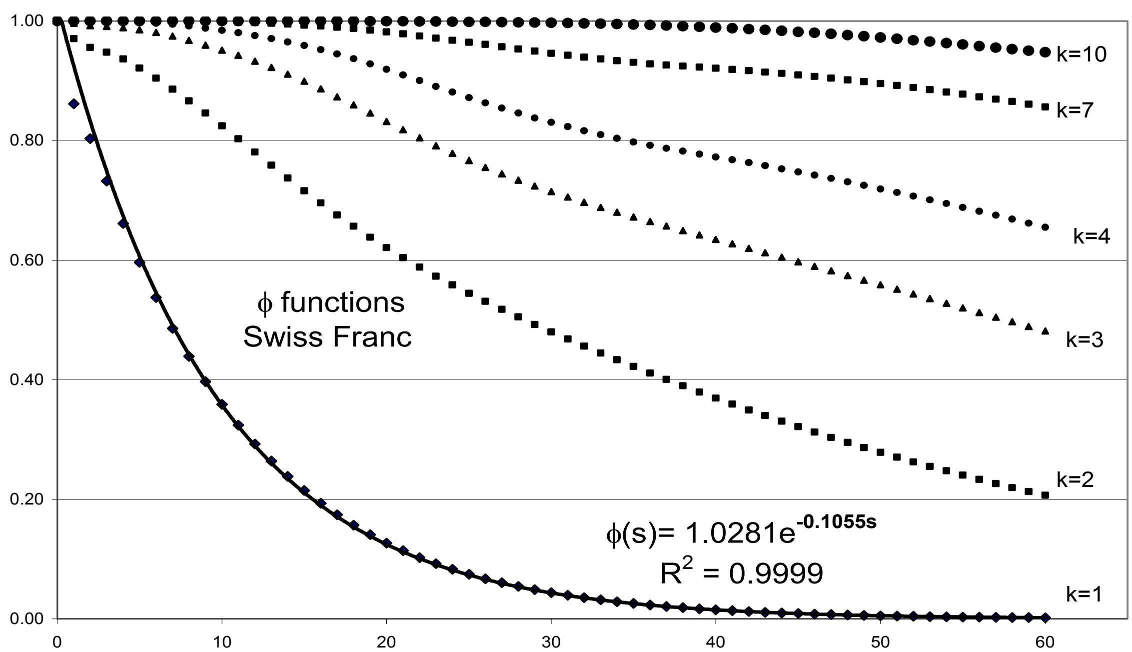
< Figure 4. *m-life* and $\phi(s)$ of Deutsche Mark >





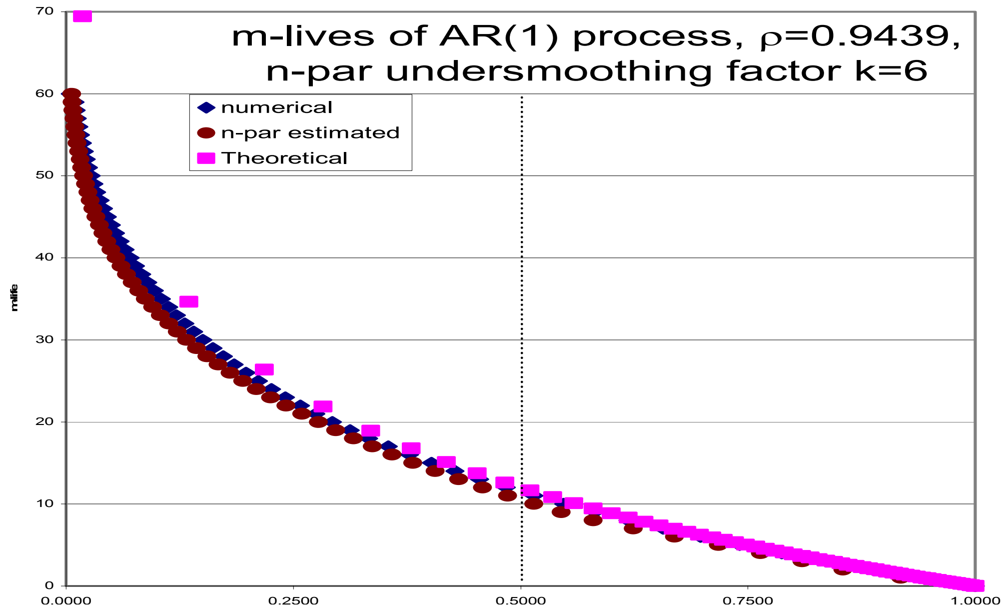
< Figure 5. *m-life* and $\phi(s)$ of Swiss Franc >



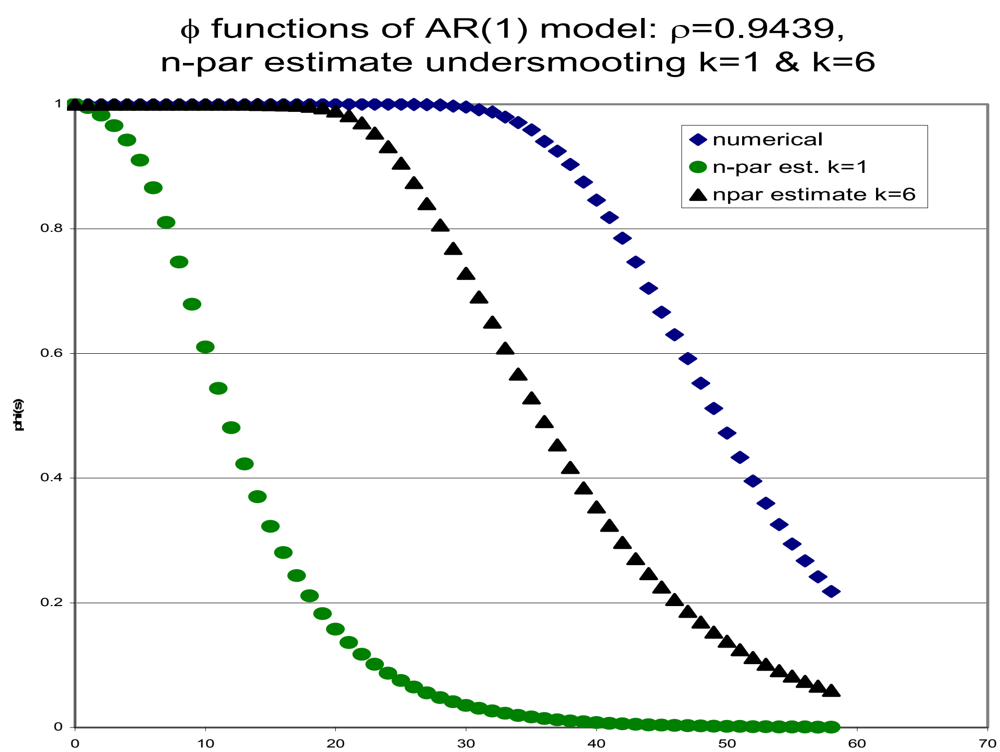


< Figure 6. PPP and Non-persistence measures: AR(1) –1 year HL>

(a) *m-life*

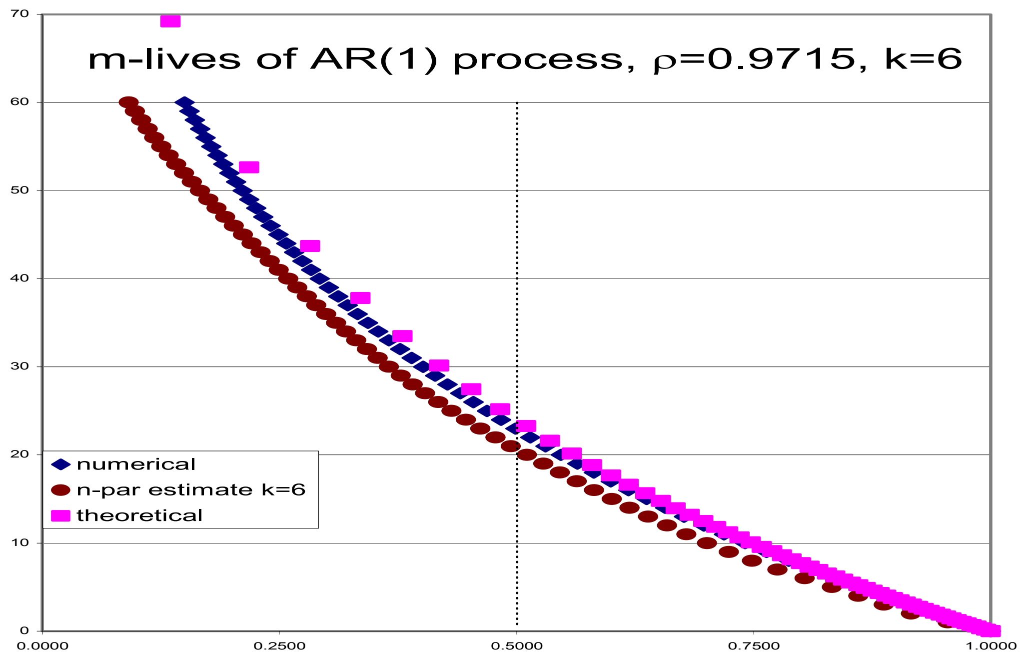


(b) $\phi(s)$

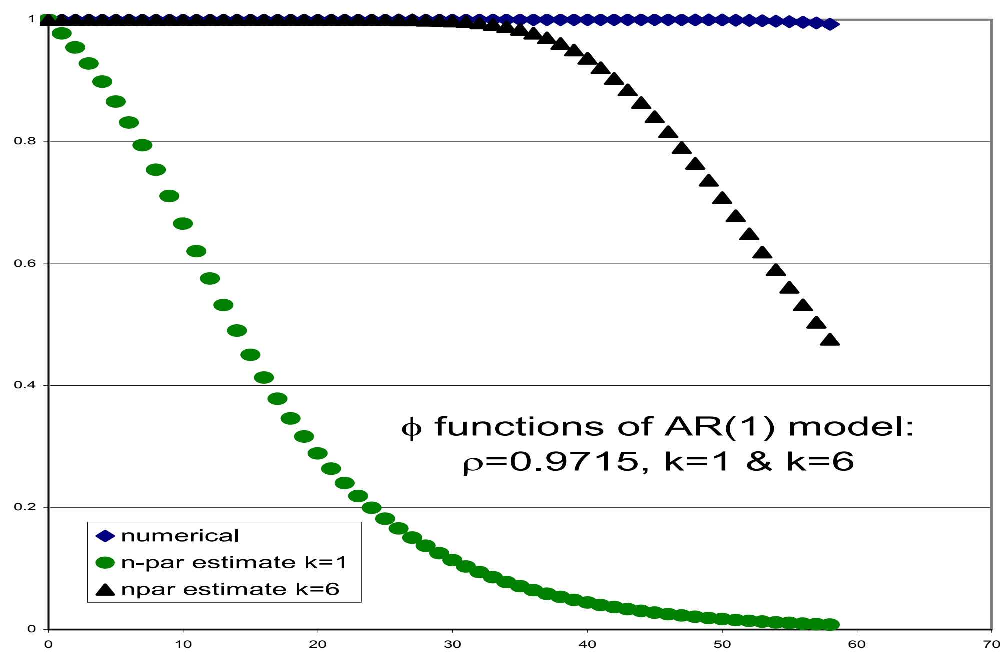


< Figure 7. PPP and Non-persistence measures: AR(1) –2 year HL>

(a) *m-life*

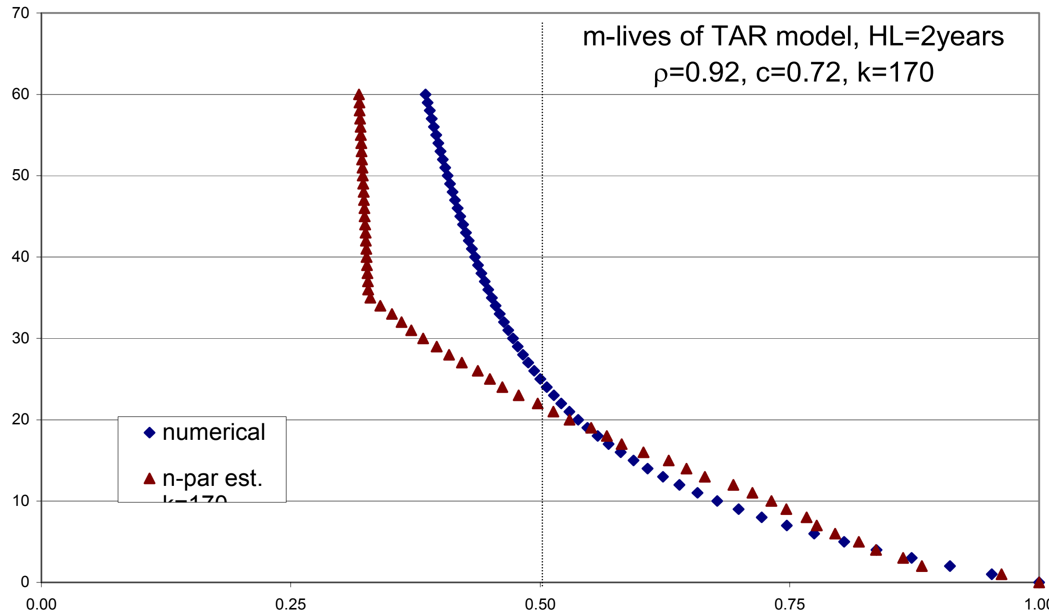


(b) $\phi(s)$

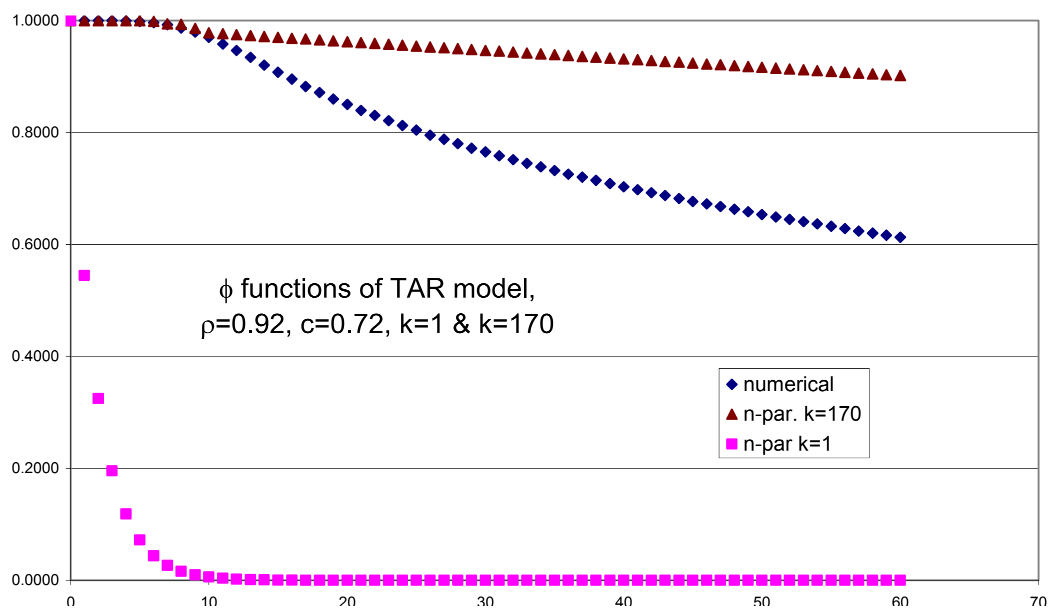


< Figure 8. PPP and Non-persistence measures: TAR (1) –2 year HL>

(a) *m-life*

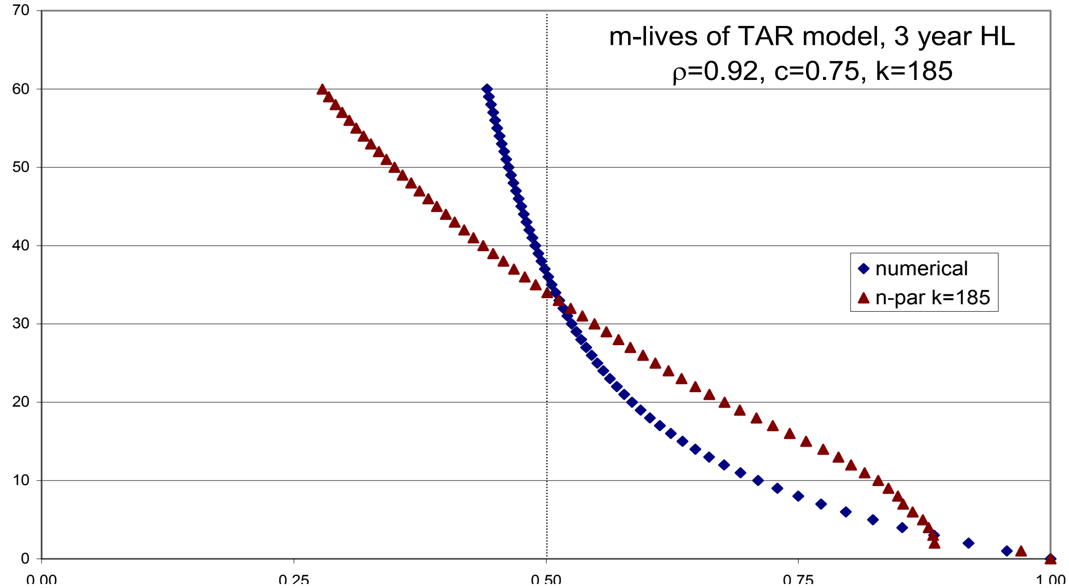


(b) $\phi(s)$

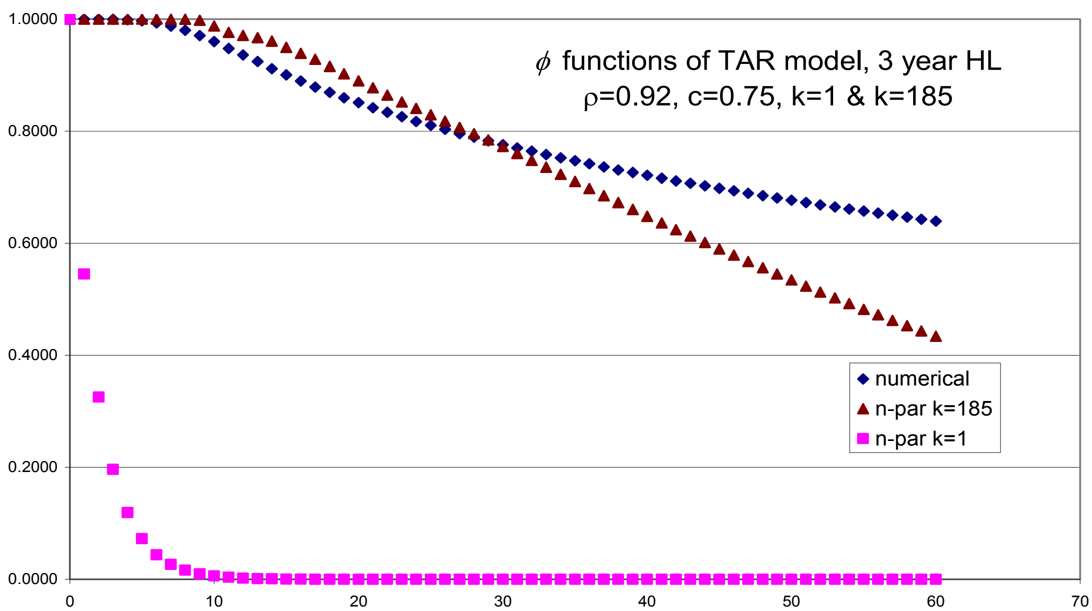


< Figure 9. PPP and Non-persistence measures: TAR (1) –3 year HL>

(a) *m-life*

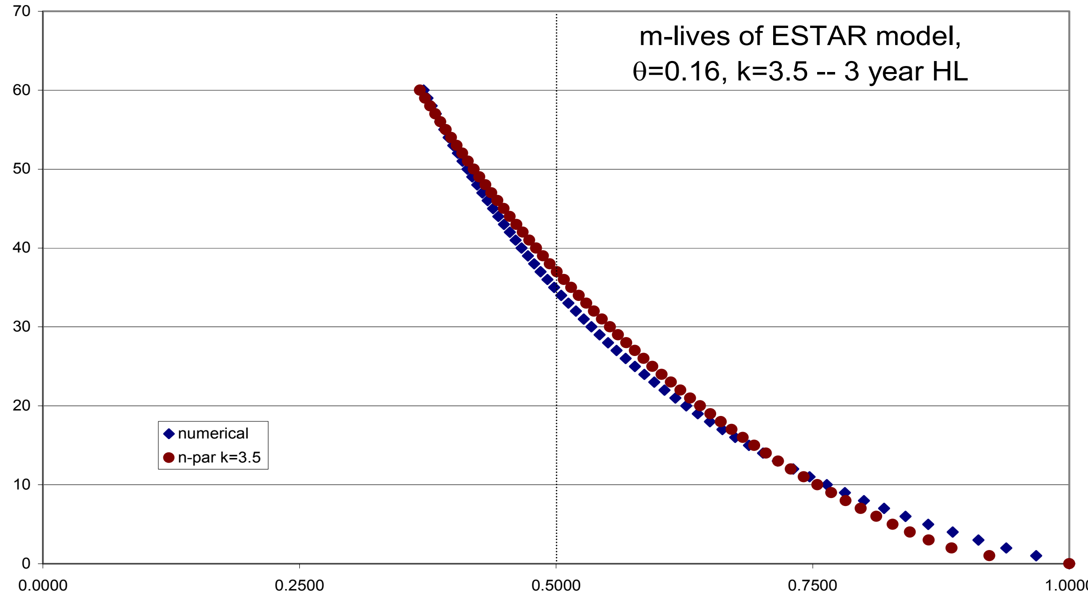


(b) $\phi(s)$

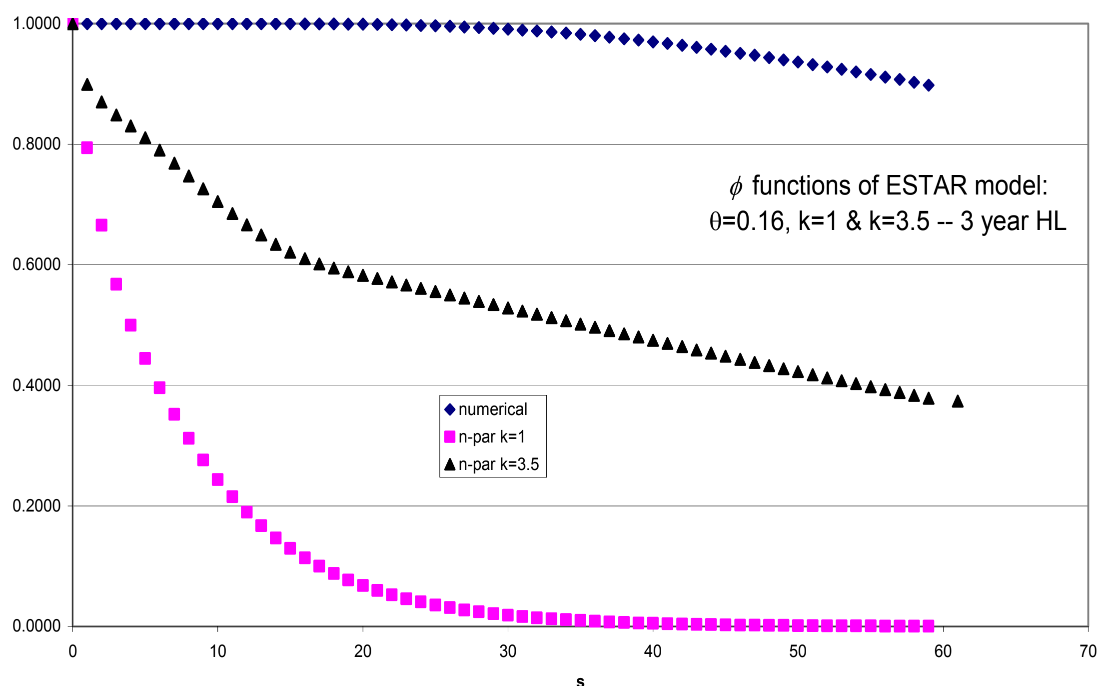


< Figure 10. PPP and Non-persistence measures: ESTAR (1) –3 year HL>

(a) *m-life*

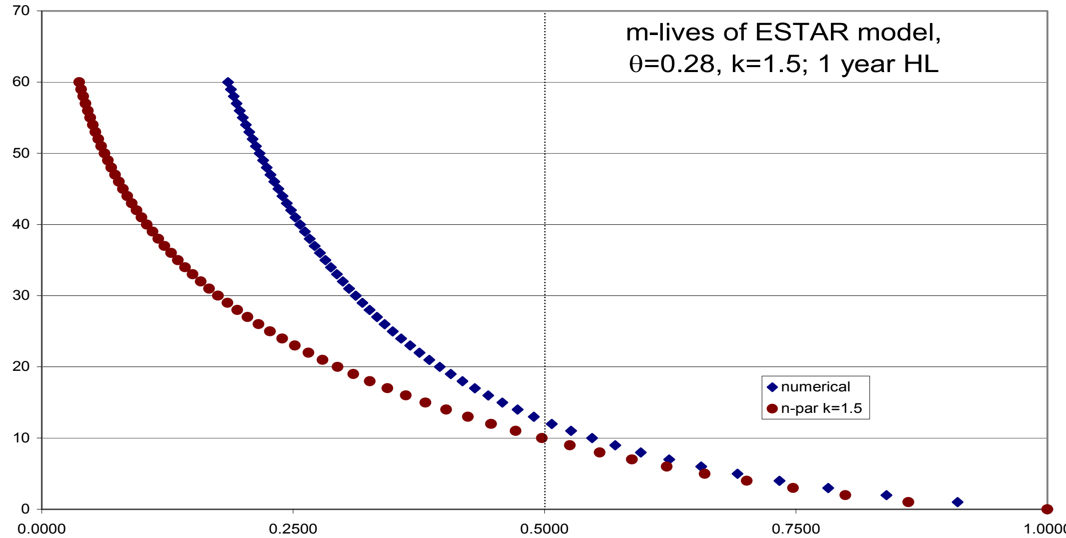


(b) $\phi(s)$

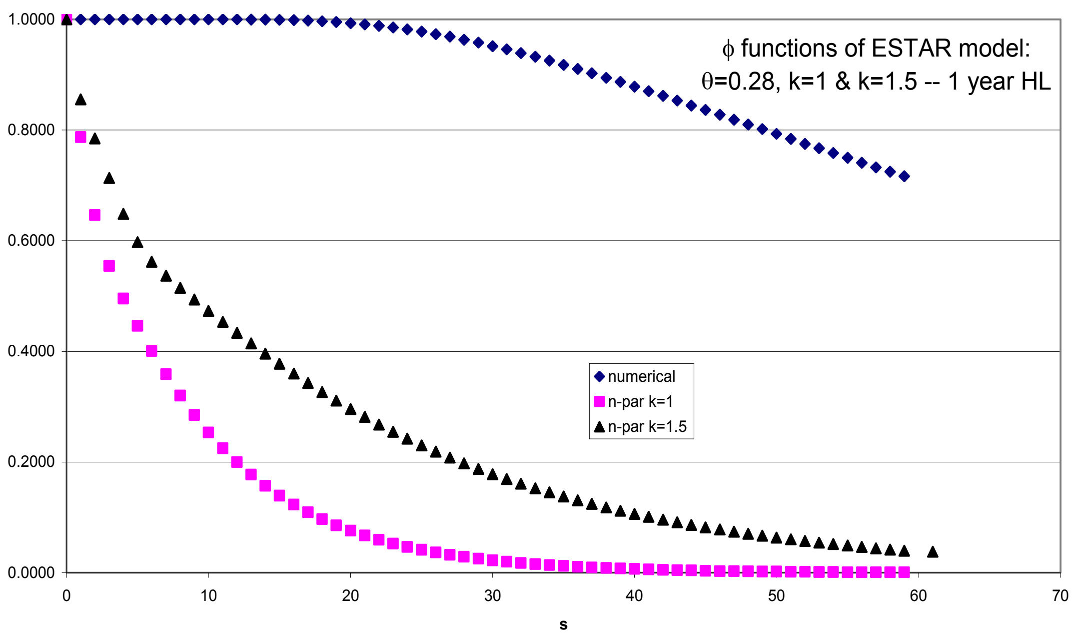


< Figure 11. PPP and Non-persistence measures: ESTAR (1) –1 year HL>

(a) *m-life*

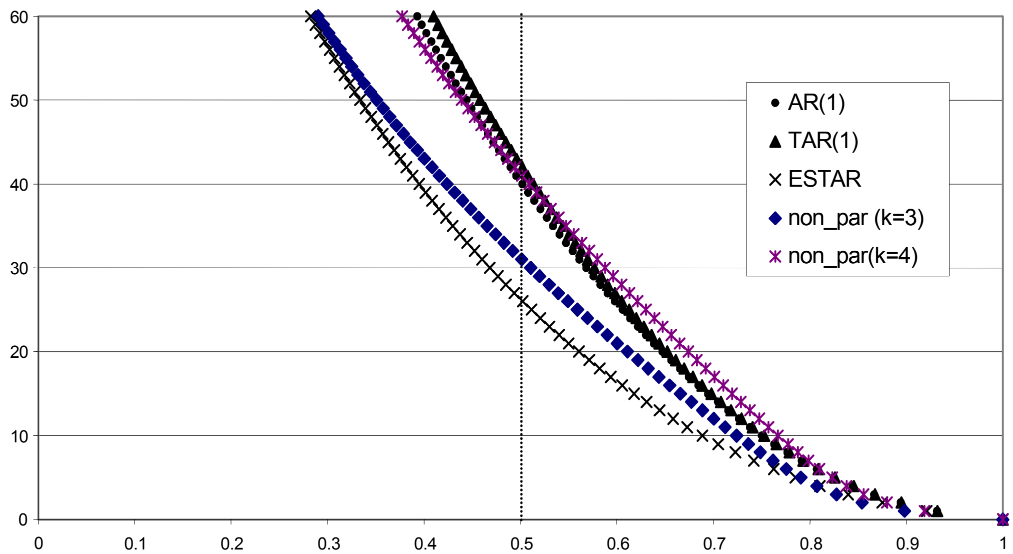


(b) $\phi(s)$

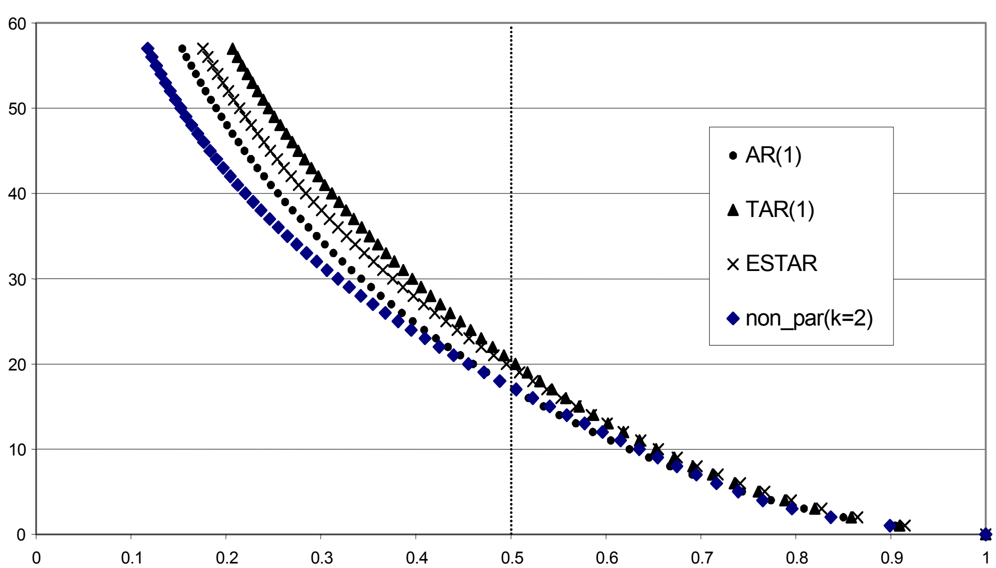


< Figure 12. *m-lives* and non-linear models >

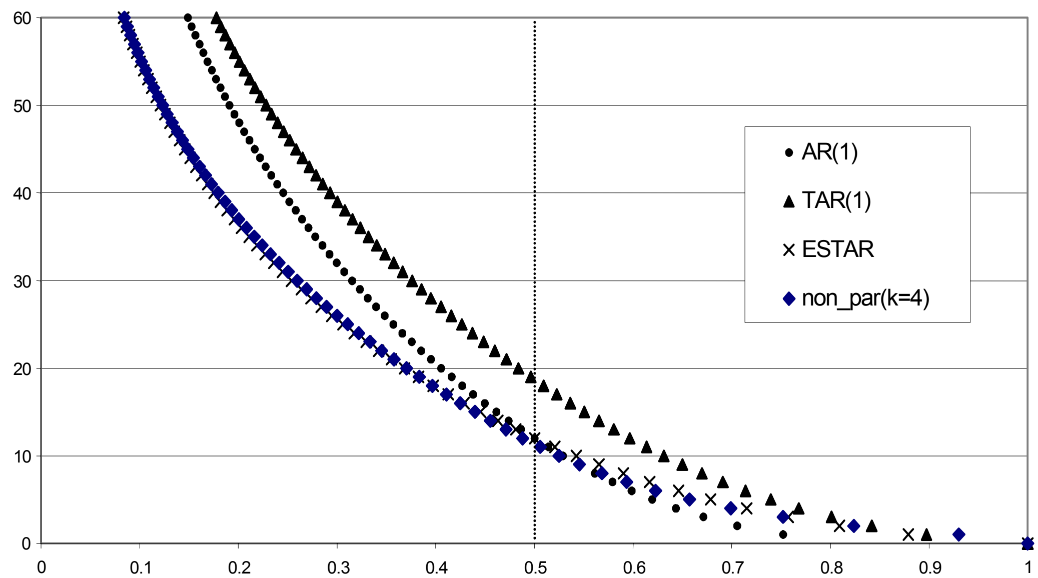
(a) Japanese Yen



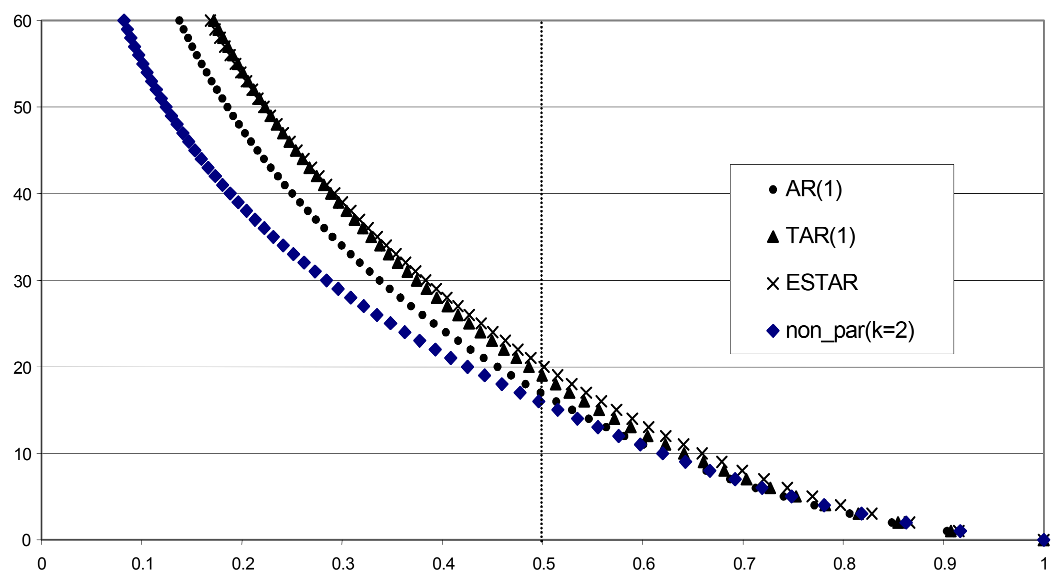
(B) French Franc



(c) British Pound



(d) German Mark



(e) Swiss Franc