A Bayesian analysis of the PPP puzzle using an unobserved components model

Richard Kleijn [*] Herman K. van Dijk [†]

Econometric Institute Report EI 2001-35

Abstract

The failure to describe the time series behaviour of most real exchange rates as temporary deviations from fixed long-term means may be due to time variation of the equilibria themselves, see Engel (2000). We implement this idea using an unobserved components model and decompose the observations on real exchange rates in long-term components, which capture the time-variation of the mean and in medium and short-term components which measure temporary deviations. A simulation-based Bayesian analysis is introduced to compute the posterior distribution of (functions) of the model parameters. A stationarity test in this setup indicates that the mean is slowly time-varying. Subsequently, we use our flexible model to derive the implied distributions of some key features of real exchange rates. Most notably, the half-life of deviations from the mean, which is a measure of persistence, is lowered. This provides a possible explanation for the PPP puzzle.

JEL classification: C11; C32; C52; F30

Keywords: purchasing power parity puzzle; real exchange rate; time-varying mean; Gibbs sampling

1 Introduction

In its simplest form Purchasing Power Parity (PPP) is based on the Law of One Price, which relates domestic and foreign price levels (P_t and P_t^*) to the nominal exchange rate S_t, i.e.

$$P_t = P_t^* S_t.$$

A more realistic version relaxes this equality, allowing for temporary deviations. Equivalently, this may be formulated as mean reversion of the real exchange rate Q_t. Using lower case

[*]Tinbergen Institute and Econometric Institute, Erasmus University Rotterdam, e-mail: rkleijn@few.eur.nl

[†]Econometric Institute, Erasmus University Rotterdam, e-mail: hkvandijk@few.eur.nl

symbols to denote logs, we have that

$$q_t = p_t^* - p_t + s_t$$

should be stationary.

There exists a vast literature on testing the hypothesis of PPP, often in the form of establishing whether the real exchange rates are stationary or nonstationary. Numerous unit root tests and stationarity tests such as the KPSS test of Kwiatkowski et al. (1992) and its variants have been used. In many cases these tests failed to reject the unit root or rejected stationarity, casting doubt on the validity of one of the fundamental building blocks of international economic theory. The relatively short span of available data on floating exchange rates in combination with the high persistence in real exchange rates could explain these findings. It is a known problem that it is very hard to distinguish between data containing a unit root and highly persistent but stationary data on the basis of a short span of data.

Moreover, real exchange rates typically show huge volatility in the short to medium term and at the same time it has a very high persistence with a half-life of deviations from its mean of 3 to 5 years. These features have given rise to the so-called Purchasing Power Parity Puzzle, see Rogoff (1996). The high volatility could potentially be explained by sticky prices and real shocks, but the mean reversion in the data is too slow to be consistent with these arguments.

On the other hand, mean reversion has been found for long spans of data of 100 years or more. This has convinced some researchers that the data provides evidence that real exchange rates really are stationary. However, Engel (2000) claims using a simulation study that these results may be spurious. For the real exchange rate data he finds that the unit root tests have serious size distortions while simultaneously the stationarity tests display a lack of power. He suggests that instead of temporary deviations from a fixed long-term mean it is more plausible to assume that the equilibrium rate itself is slowly changing over time. He explains this by distinguishing between traded goods for which price adjustment is likely and non-traded goods which lack an identifiable mechanism for mean reversion.

In this paper we analyze the PPP puzzle as follows. First, we implement the suggestion of Engel (2000) that the long term mean of the real exchange rate may exhibit some time variation, by using an unobserved components model for the real exchange rate behavior. A long term component captures the time variation in the mean while short term components describe temporary deviations. In Section 3 we motivate why we have chosen this particular model by assessing the pros and cons of existing approaches in the literature.

For inferential purposes we make use of a simulation based Bayesian analysis. The Bayesian approach has the advantage that one may give the hypotheses of stationarity and non-stationarity equal prior probabilities. Our posterior simulator for computing posterior distributions is introduced in Section 4. A third feature of our approach is exact inference on functions of parameters of interest, in particular we measure the half-life of deviations of the real exchange rate from its mean. Cheung and Lai (2000) conclude that classical point estimates of the half-life have a very large imprecision. They construct confidence intervals to quantify the uncertainty. Our Bayesian approach allows us to calculate the entire posterior density of the half-lives. In this paper we also find non-monotonic impulse responses, consistent with their findings. Earlier Bayesian work on PPP includes Schotman and van Dijk (1991) who use an autoregressive model with a fixed mean. Our simulation based Bayesian inference is an extension of the work of Koop and van Dijk (2000), but their focus is on testing for stationarity in an unobserved components model. Bayesian work related to nonstationarity, but not to PPP specifically, include Sims (1988), Sims and Uhlig (1991), Phillips (1991) and Zivot (1994), or see Bauwens et al. (1999) for a survey. West and Harrison (1997) give an extensive Bayesian treatment of unobserved components models, but they do not focus on nonstationarity issues.

In this paper, we can make a distinction between model selection for the PPP hypothesis using posterior odds and the analysis of PPP using the concept of half-life, which is carried out in the flexible model. The price of this flexibility is that one has to develop a tailored posterior simulator to evaluate the posterior density of functions of the parameters of interest. Given the methods of de Jong and Shephard (1995) or similar work of Carter and Kohn (1994) and Frühwirth-Schnatter (1994), one can develop such posterior simulators.

The outline of the paper is as follows: we start by describing our data set in Section 2, followed by a description and motivation of the modelling process in Section 3. The details of the full Bayesian analysis of the model are set out in Section 4. In Section 5 we report posterior results. The results include tests for a constant underlying mean of the real exchange rates, posterior densities for some parameters of interest, the impulse responses of deviations from the (time-varying) mean and their implied half-lives. Section 6 concludes with a discussion of the results and an indication of possible extensions for future research on this topic.

2 Data

The data we use in this paper are obtained from the OECD Statistical Compendium. The real exchange rates are constructed from nominal exchange rates and consumer price indices. The series have a monthly frequency and cover the post-Bretton Woods period, that is, they run from 1973:01 to 1998:12, when the internal Euro rates became fixed. We investigate the German Mark with the US dollar as numeraire currency (DEM/USD) and the French Franc against the German Mark (FF/DEM). Figure 1 displays the real exchange rates.

Germany and France are major trade partners, neighboring countries, and early European Union members. These factors lead us to expect beforehand that trading restrictions will be

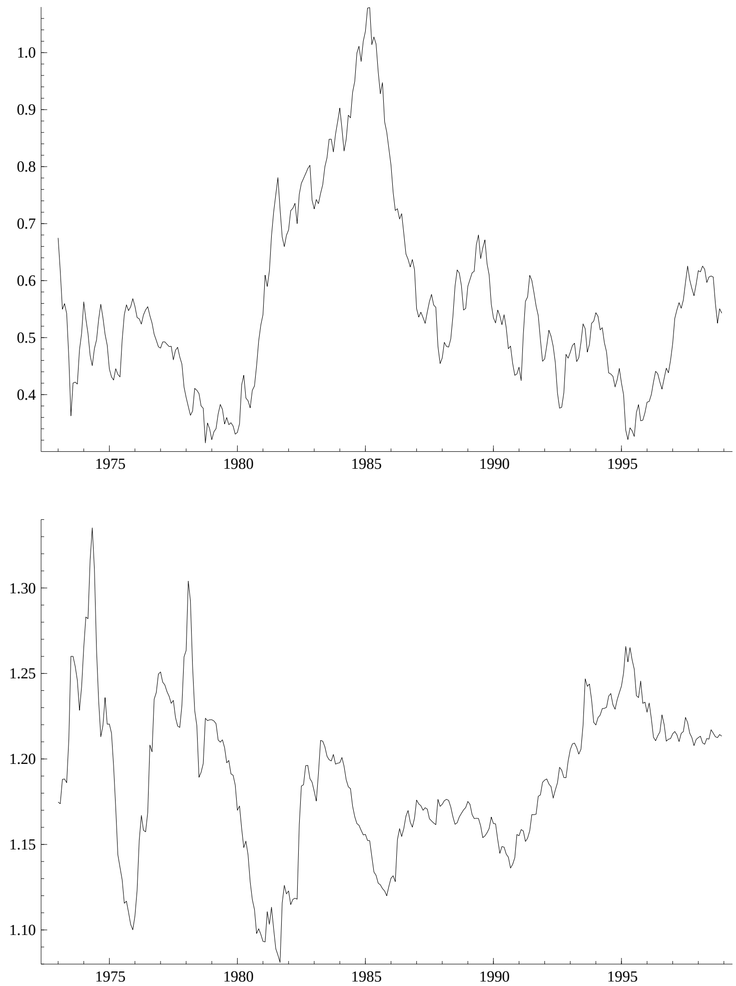

Figure 1: Real exchange rates (logs) of Germany with US as numeraire (DEM/USD) and France with Germany as numeraire (FF/DEM).

relatively minor and that the transportation costs are low. For this combination the real world circumstances are close to the assumptions underlying the theory of PPP.

3 Model Specification

In this section we motivate why we have chosen the particular model that we shall use. We do so by assessing the pros and cons of previously used approaches for testing the hypothesis of PPP and the ability of those models to describe the features of our data set adequately.

The most common approach is based on the ARIMA class of models. Unit root tests in many variants have been carried out in the context of this class. An important characteristic of the unit root tests is that they test stationarity of deviations from a fixed constant mean or from a rigid linear trend. The most common outcome for real exchange rate data is that the null hypothesis of a unit root cannot be rejected, possibly due to the high persistence of the deviations and the limited span of the data. These tests often cannot reject the and hence they are unsuitable for providing evidence of the absence of a unit root or equivalently for providing evidence in favour of PPP. In the ARIMA modelling paradigm the next logical step is to handle nonstationarity by analyzing the first differences of the series. As a result information on the level of the series is lost to some degree. An estimated ARIMA model can be decomposed in a random walk component and a stationary component by means of the Beveridge-Nelson decomposition, see Beveridge and Nelson (1981). The Beveridge-Nelson decomposition does indirectly what more structural unobserved components models can do explicitly and more transparently. An alternative for assessing whether the series is stationary or not is given by the KPSS test, see Kwiatkowski et al. (1992). They develop a test for the null hypothesis of the absence of a unit root. The setup they use is an unobserved components model, which is the sum of a random walk (possibly with drift) and a stationary component which captures the short run dynamics. The test consists of checking whether the random walk variance is close to zero such that the random walk reduces to a linear trend or a constant.

In classical testing the null and the alternative hypothesis are treated asymmetricly, implicitly favoring one outcome over the other a priori. In the Bayesian methodology equal prior probabilities can be assigned to two or more competing alternatives, in this case stationarity vs. nonstationarity. We shall treat the decomposition of the real exchange rate series into a nonstationary and a stationary component as a reasonable starting point for specifying our model. The simplest structural time series model is the so-called local level model (LLM). It consists of a random walk component with added i.i.d. normal errors. The drawback of this particular model is that under the null hypothesis the model reduces to i.i.d. normal deviations, which for our data set seems not to be very realistic. A straightforward way of improving on the behavior under H_0 is to allow for more interesting dynamics in the stationary component.

Clark (1988) suggests a decomposition into a permanent random walk and transitory components. This approach has the same problem as the KPSS test. Testing whether the variance of the random walk is zero is problematic because estimation of the variance is imprecise. This phenomenon also causes the KPSS test to have low power.

The random walk component behaves erraticly and is non-smooth. It picks up short and medium term dynamics. We consider this to be a second disadvantage of the simple LLM. We circumvent this problem by choosing a smooth I(2) trend instead of the ill-behaved random walk component. In the structural time series modelling literature a common means of imposing more smoothness is to replace the I(1) random walk with an integrated random

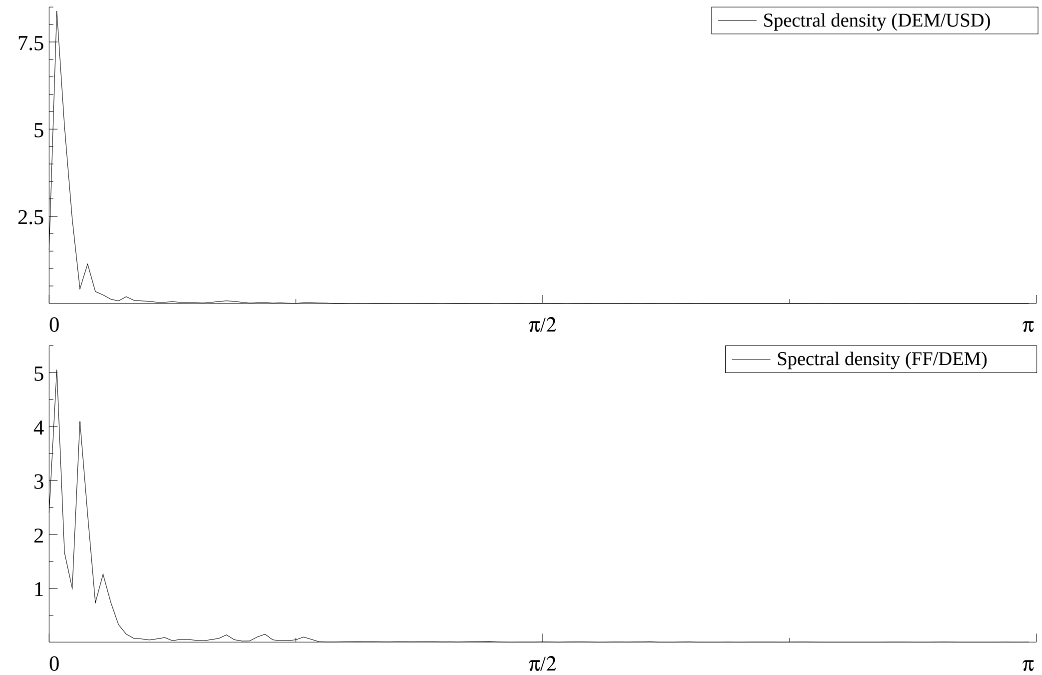

Figure 2: Sample periodogram for DEM/USD and FF/DEM.

walk of order I(2). It turns out that the variance parameter of the I(2) trend can be estimated with more precision, and hence we avoid issues associated with the low power problems of the KPSS test.

The spectral densities in Figure 2 show a large peak at a low frequency for both series. These frequencies are captured by the smooth trend component. The other peaks can be modelled adequately by two cycles. The main reason for choosing cycle components is that they are well-behaved in the sense that they do not tend to pick up the variation with a very low frequency that is associated with the trend component. We have experimented with stationary ARMA components for the short-run dynamics, but in our experience the ARMA parameters tend to move in the direction of the unit root in such a way that they start to interfere with the trend component. Although the cycles can be represented by an ARMA model, the implicit restrictions the cycle imposes on the ARMA reduced form prevent this pathological behaviour. Our preferred model thus contains a smooth trend and two cyclical components:

$$
\begin{aligned}
y_t &= \mu_t + \psi_{1,t} + \psi_{2,t} \\
\mu_t &= \mu_{t-1} + \beta_{t-1} \\
\beta_t &= \beta_{t-1} + \zeta_t \\
\begin{pmatrix} \psi_1 \\ \psi_1^* \end{pmatrix}_t &= \rho_1 \begin{pmatrix} \cos\lambda_1 & \sin\lambda_1 \\ -\sin\lambda_1 & \cos\lambda_1 \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_1^* \end{pmatrix}_{t-1} + \begin{pmatrix} \kappa_1 \\ \kappa_1^* \end{pmatrix}_t \\
\begin{pmatrix} \psi_2 \\ \psi_2^* \end{pmatrix}_t &= \rho_2 \begin{pmatrix} \cos\lambda_2 & \sin\lambda_2 \\ -\sin\lambda_2 & \cos\lambda_2 \end{pmatrix} \begin{pmatrix} \psi_2 \\ \psi_2^* \end{pmatrix}_{t-1} + \begin{pmatrix} \kappa_2 \\ \kappa_2^* \end{pmatrix}_t
\end{aligned}
\tag{3.1}
$$

with $\zeta_t \sim N(0, \sigma_\zeta^2)$, and $\begin{pmatrix} \kappa_i \\ \kappa_i^* \end{pmatrix}_t \sim N(0, (1-\rho_i^2)\sigma_{\psi_i}^2 I_2)$. The parameters λ_1 and λ_2 are the frequencies and ρ_1 and ρ_2 are the damping factors of the damped stochastic sinusoids. The local mean of the series is provided by μ_t. In the PPP analysis we shall test for $\sigma_\zeta^2 = 0$.

In a Bayesian analysis the initial conditions play a crucial role for correcting the pathological behaviour of the likelihood near the unit root, see e.g. Schotman and van Dijk (1991) on initial conditions in the context of an AR(1) model. Hence, we assume that the initial values come from the unconditional distribution as recommended in e.g. Harvey (1989). Since β_t and μ_t are non-stationary we assume that β_1 and μ_1 have a diffuse distribution. The cycle components are initialized by $\begin{pmatrix} \psi_i \\ \psi_i^* \end{pmatrix}_{t=1} \sim N(0, \sigma_{\psi_i}^2 I_2)$. Note that the model can be represented as the sum of one ARIMA(0,2,0) and two ARMA(2,2) components, which results in a complicated ARIMA model when written in single-equation form.

4 Posterior simulator

We are primarily interested in conducting posterior inference on (functions of) the model parameters. For that purpose we need the posterior density $p(\theta|Data)$ of the parameters $\theta = (\sigma^2, \rho, \lambda)$ with $\sigma^2 = (\sigma_\zeta^2, \sigma_{\psi_1}^2, \sigma_{\psi_2}^2)$, $\rho = (\rho_1, \rho_2)$ and $\lambda = (\lambda_1, \lambda_2)$.

The posterior density is proportional to likelihood multiplied by prior. The likelihood for these unobserved components models involves an integral over the unobserved states $L(y|\theta) = \int L(y, \alpha|\theta) d\alpha$, where $\alpha_t = (\mu_t, \beta_t, \psi_{1,t}, \psi_{1,t}^*, \psi_{2,t}, \psi_{2,t}^*)$ and quantities without a subscript t represent the vector containing all the elements. It can be calculated in terms of the prediction error decomposition which can be obtained from the Kalman filter.

Unfortunately, calculating posterior expectations analytically or sampling from the posterior distribution are complicated by the integral over the unobserved states that enter the posterior density through the likelihood. It is possible to circumvent these problems by recognizing that also the posterior occurs as a marginal distribution. Direct Bayesian analysis of $p(\theta|Data)$ is analytically untractable, but if we extend the parameter space, the resulting model is easier to deal with. This technique is known as data augmentation, see Tanner and Wong (1987). We then analyze the extended model and in the end the results are marginalized with respect to the additional parameters in order to translate results back to the original model, i.e.

$$p(\theta|Data) = \int p(\theta, \alpha|Data)d\alpha.$$

In order to let the data dominate the prior in the posterior, we assume the standard noninformative priors for the parameters, that is

$$\begin{aligned} p(\sigma_\zeta^2) &\propto \sigma_\zeta^{-2} \\ p(\sigma_{\psi_i}^2) &\propto \sigma_{\psi_i}^{-2} \\ p(\rho_1, \rho_2) &\propto 1 \quad \text{ for } \rho_1, \rho_2 \in [0, 1] \\ p(\lambda_1, \lambda_2) &\propto 1 \quad \text{ for } \lambda_1, \lambda_2 \in [2\pi/T, \pi] \text{ and } \lambda_1 < \lambda_2 \end{aligned}$$

When we are calculating Posterior Odds ratios in the following section we shall replace some of the noninformative priors by weakly informative natural conjugate priors because noninformative priors are known to distort the outcome of the Posterior Odds analysis.

In the remainder of this section we explain our simulation algorithm which is the basis of our posterior inference. For a recent general survey of the use of simulation methods in Bayesian analysis, see Geweke (1999). For $p(\theta, \alpha|Data)$ we use a Gibbs sampler, see e.g. Gelfand and Smith (1990) or Casella and George (1992). One iteration of the Gibbs sampler for the joint distribution of several random variables consists of going through the sequence of drawing each of the random variables conditional on the most recently obtained value of the remaining variables. The algorithm consists of drawing from the following conditional posterior densities:

1. $p(\alpha|\sigma^2, \rho, \lambda, Data)$
2. $p(\sigma^2, \rho|\alpha, \lambda, Data) = \; p(\sigma_\zeta^2|\beta)p(\rho_1|\lambda_1, \psi_1, \psi_1^*)p(\sigma_{\psi_1}^2|\rho_1, \lambda_1, \psi_1, \psi_1^*)\times$
   $p(\rho_2|\lambda_2, \psi_2, \psi_2^*)p(\sigma_{\psi_2}^2|\rho_2, \lambda_2, \psi_2, \psi_2^*)$
3. $p(\lambda_1|\alpha, \sigma^2, \rho, \lambda_2, Data) = p(\lambda_1|\sigma_{\psi_1}^2, \rho_1, \lambda_2, \psi_1, \psi_1^*)$
4. $p(\lambda_2|\alpha, \sigma^2, \rho, \lambda_1, Data) = p(\lambda_2|\sigma_{\psi_2}^2, \rho_2, \lambda_1, \psi_2, \psi_2^*)$

The algorithm consists of four blocks of parameters and it can straightforwardly be extended to more than two cyclical components. Sampling from the density $p(\alpha|\sigma^2, \rho, \lambda, Data)$ is known as the simulation smoother, see de Jong and Shephard (1995) for more details on this method. Generating a draw from $p(\sigma_\zeta^2|\beta)$ is straightforward. The situation can be described by the partial model $\Delta\beta_t \sim N(0, \sigma_\zeta^2)$, which amounts to the textbook situation of a Gaussian zero mean, unknown variance problem. Having specified a natural conjugate prior, the posterior density of $\sigma_\zeta^2|\beta$ has an inverted gamma distribution.

If it is hard to sample directly from the conditional density, but sampling from an approximating density is possible, a Metropolis-Hastings (M-H) step can be embedded in the Gibbs sampler. See e.g. Chib and Greenberg (1995) for an intuitive treatment of the M-H algorithm. Sampling $\rho_i, \sigma_{\psi_i}^2|\lambda_i, \psi_i, \psi_i^*$ is done using a M-H step with a candidate for $\rho_i|\lambda_i, \psi_i, \psi_i^*$ that resembles the actual density closely and an exact draw from $\sigma_{\psi_i}^2|\rho_i, \lambda_i, \psi_i, \psi_i^*$. For $\lambda_1|\rho_1, \sigma_{\psi_1}^2, \lambda_2, \psi_1, \psi_1^*$ and $\lambda_2|\rho_2, \sigma_{\psi_2}^2, \lambda_1, \psi_2, \psi_2^*$ we use a M-H step based on a carefully constructed Student-t candidate whose Taylor expansion around its mode matches the target density up to order 6. In the Appendix we provide more details on the steps of the algorithm which involve sampling the parameters in the cyclical components.

We have set up the simulation as follows. We ran the Gibbs sampler 255,000 times discarding the first 5,000 as a burn-in period. Of the remaining 250,000 we only used every 25th draw in order to eliminate most of the correlation between successive values. All posterior results have been calculated on the basis of the resulting 10,000 values of the parameters θ and the states α. In each 25th iteration also the full impulse response function for that parameter realization is calculated. On average the algorithm runs at more than 100 iterations/second on an AMD 1300MHz PC.

5 Posterior Results

5.1 Posterior distribution of the model parameters

The posterior densities of the parameters are graphed in Figure 3. Table 1 lists some summary statistics of the posterior distributions of the model parameters, namely means, standard deviations and medians. The most striking difference between the DEM/USD data and the FF/DEM data is the variance of the trend component. We shall come back to the implications of this in Subsection 5.3.

The medians of the periods of the cycles are 12.3 and 71.6 months, respectively, for the DEM/USD real exchange rate and we find 9.6 and 57.1 months for the FF/DEM data. The period of the shorter cycle may be related to a seasonal or calender effect. The longer cycle may have a connection to the business cycle. However, including the cycle components serves merely as a means of providing the model with the required flexibility to capture any relevant short-run dynamics. Attaching a structural or economic interpretation to the individual cycles is not our primary focus.

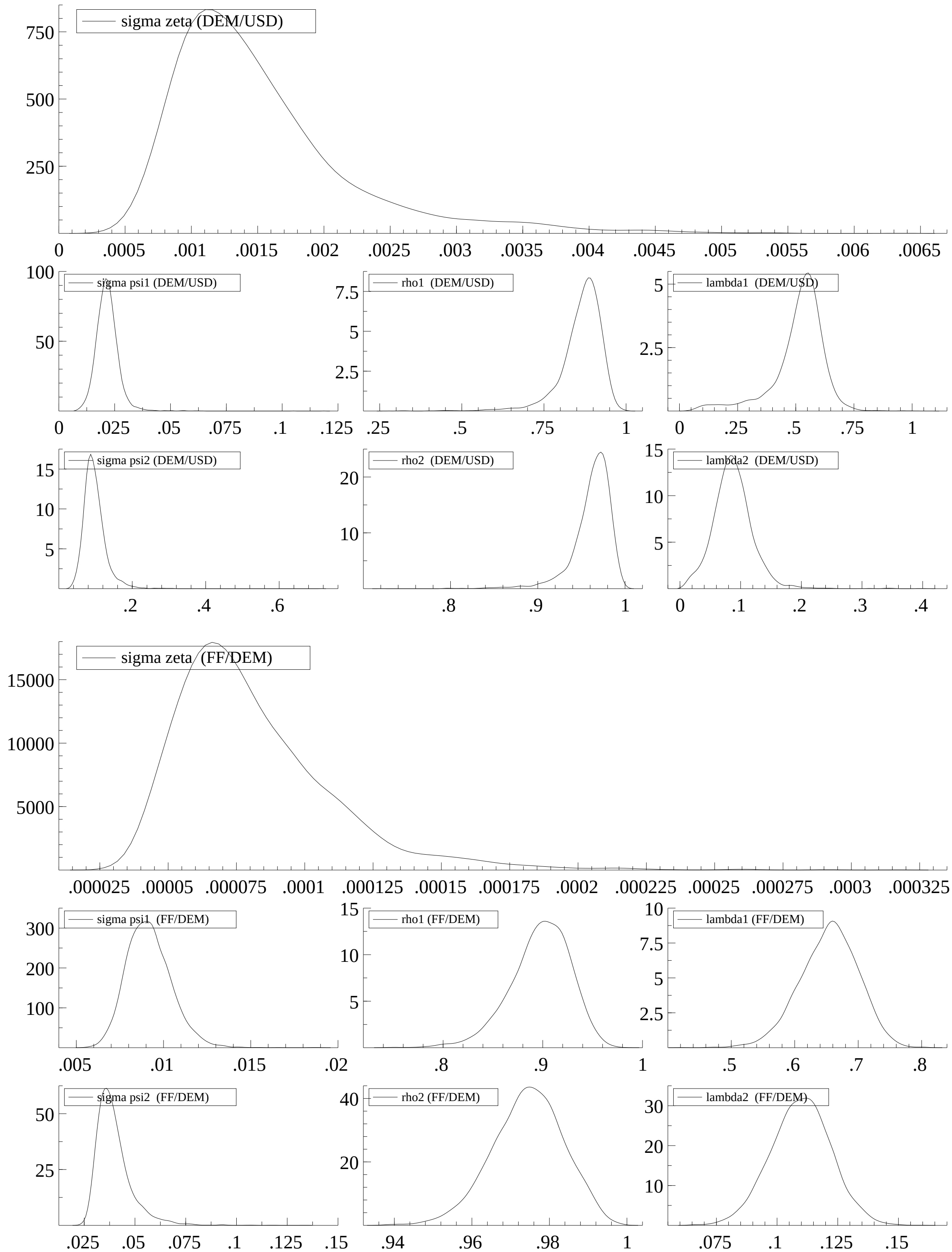

Figure 3: Posterior densities of model parameters for DEM/USD and FF/USD.

Table 1: Posterior results.

	DEM/USD		FF/DEM	
	mean	median	mean	median
$\sigma_\zeta(\times 10^{-3})$	1.52	1.355	0.0816	0.0751
	(0.694)		(0.0296)	
$\sigma_{\psi_1}(\times 10^{-3})$	21.7	21.4	9.18	9.09
	(5.31)		(1.28)	
ρ_1	0.863	0.876	0.897	0.900
	(0.0666)		(0.0302)	
λ_1	0.512	0.535	0.654	0.656
	(0.114)		(0.0458)	
$\sigma_{\psi_2}(\times 10^{-3})$	97.5	92.5	39.8	37.9
	(31.8)		(9.07)	
ρ_2	0.962	0.966	0.974	0.974
	(0.0234		(0.00919)	
λ_2	0.0878	0.0859	0.110	0.110
	(0.0342)		(0.0123)	

Note: The reported values are posterior means and medians obtained from the posterior simulator. Posterior standard deviations are given in parentheses.

5.2 Testing the sharp null hypothesis of stationarity

Posterior Odds ratios are the preferred Bayesian approach for testing a sharp null against a composite alternative. In this section we report the closely related Bayes Factors for the hypothesis $\sigma_\eta^2 = 0$ which is a necessary condition for PPP. There are complications for interpreting Bayes Factors for a restriction if an improper prior is specified on the parameters involved in the restriction of interest. Hence, we use weakly informative but proper inverted gamma priors for the variance parameters.

Chib's method for calculating Bayes factors, see Chib (1995), uses the output of a Gibbs sampler efficiently. Unfortunately, is not easily applicable because the constants of integration are not known for all conditional densities. Therefore, the marginal posteriors are computed using the Laplace approximation. The posterior distribution is approximated by a multivariate normal which has the posterior mode as its mean and its covariance matrix is minus the inverse of the Hessian at the posterior mode, see Kass et al. (1990). We transformed each parameter such that its support becomes the entire real line. As a side effect of this, the skewness of parameters is reduced, especially of the variance parameters. This helps making the approximation by a normal distribution more precise.

The Bayes factors in Table 5.3 provide very strong evidence against $\sigma_\zeta^2 = 0$. The ap-

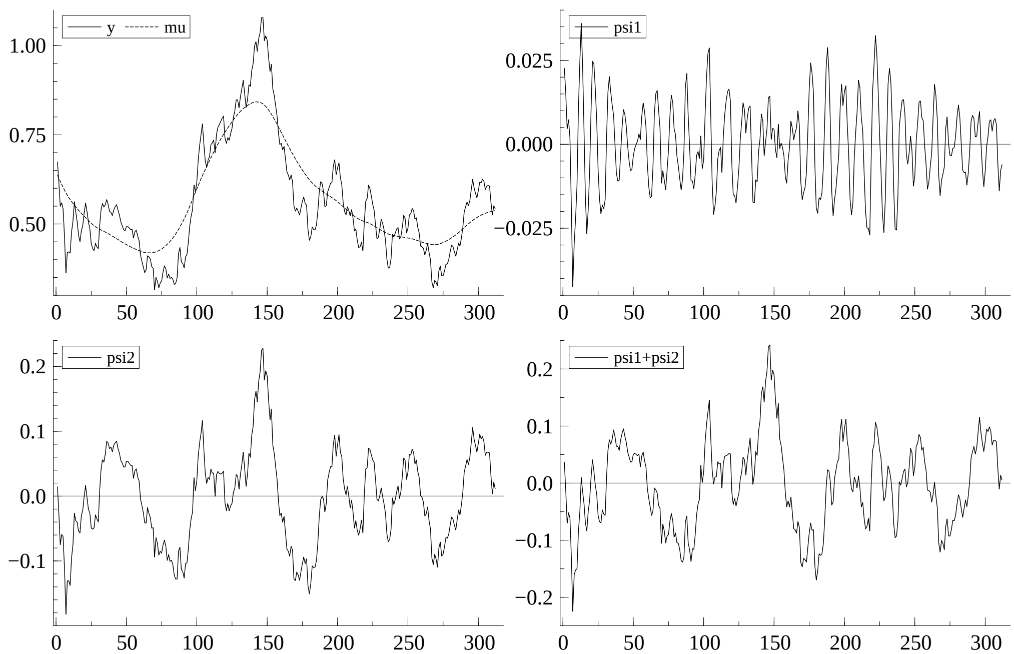

Figure 4: Posterior means of μ_t, $\psi_{1,t}$ and $\psi_{2,t}$ for DEM/USD.

proximation error introduced by the Laplace expansion is unlikely to change this conclusion substantially.

5.3 Implied features of the model

An important byproduct of the Gibbs sampler is a sample of the smoothed components, for which the sample average over all draws at each point in time gives the Bayesian equivalent of the state smoother. The most important difference however is that parameter uncertainty is accounted for. The graphs of the expected value of the unobserved components are shown in Figure 4. Note that although the variation in the DEM/USD real exchange rates is much larger than for the FF/DEM data, the stationary component of both series have comparable variance. For the DEM/USD data more of the variation in the time series has been absorbed by the trend.

In the PPP literature there is a growing interest in the half-life of deviations from PPP. The half-life is defined as the number of years before the effect of a shock becomes permanently less than half the size of the original shock and therefore has a natural connection with impulse response functions. It is an important summary statistic of the full impulse response function. The half-life is often found to be between three and five years, see e.g. Cheung and Lai (2000) and the references cited therein. Note that this finding is based on models that impose PPP

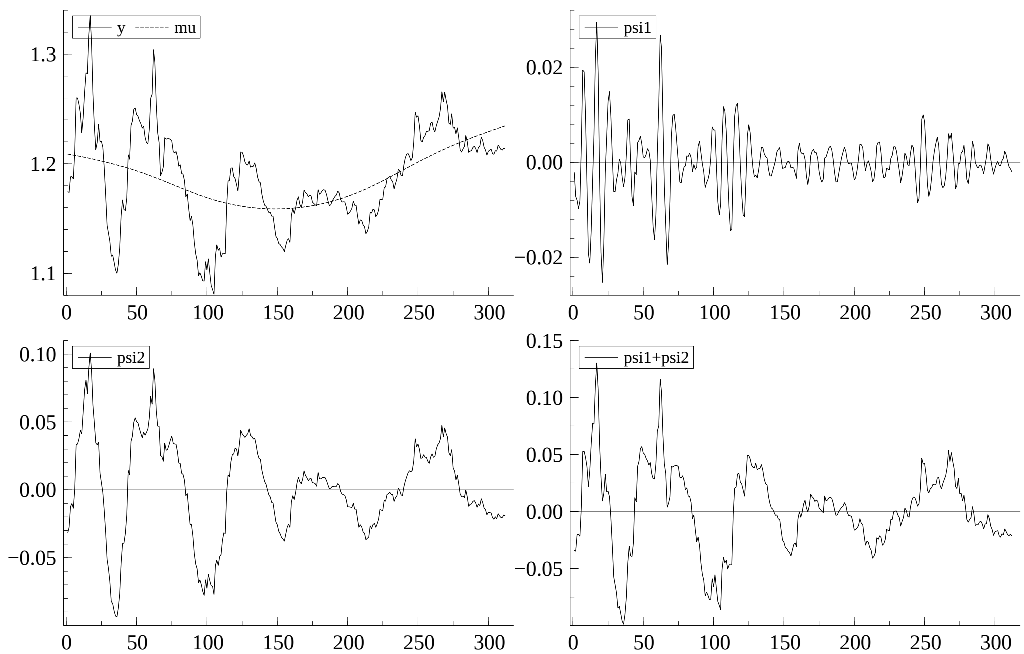

Figure 5: Posterior means of μ_t, $\psi_{1,t}$ and $\psi_{2,t}$ for FF/DEM.

in a rigid way. In our model a fixed mean is not imposed. In our model, we first calculate the impulse response function of the deviations from the (time-varying) mean by means of the infinite moving average or Wold representation, see e.g. Hamilton (1994),

$$\Psi_i = ZT^{i-1}K, \tag{5.1}$$

where Z is the loading vector $\begin{pmatrix} 1 & 0 & 1 & 0 \end{pmatrix}$ in the measurement equation, T is the block diagonal state transition matrix

$$\rho_1 \begin{pmatrix} \cos\lambda_1 & \sin\lambda_1 \\ -\sin\lambda_1 & \cos\lambda_1 \end{pmatrix} \oplus \rho_2 \begin{pmatrix} \cos\lambda_2 & \sin\lambda_2 \\ -\sin\lambda_2 & \cos\lambda_2 \end{pmatrix}, \tag{5.2}$$

and K is the 4×1 steady-state Kalman gain which is straightforwardly obtained from the Kalman filter.

A unit shock is then assumed for $t = 0$. It is important to note that it matters which error term in our model is subjected to a shock. The value of the frequency parameter and the damping factor of that cycle influence how the effect of the shock will evolve over time. An unambiguous way of attributing the initial shock to the different components is to do so according to the Kalman gain. This provides an intuitive explanation of (5.1). Given the impulse response function, the half-life is trivially the time it takes for the impulse response to fall permanently below a half.

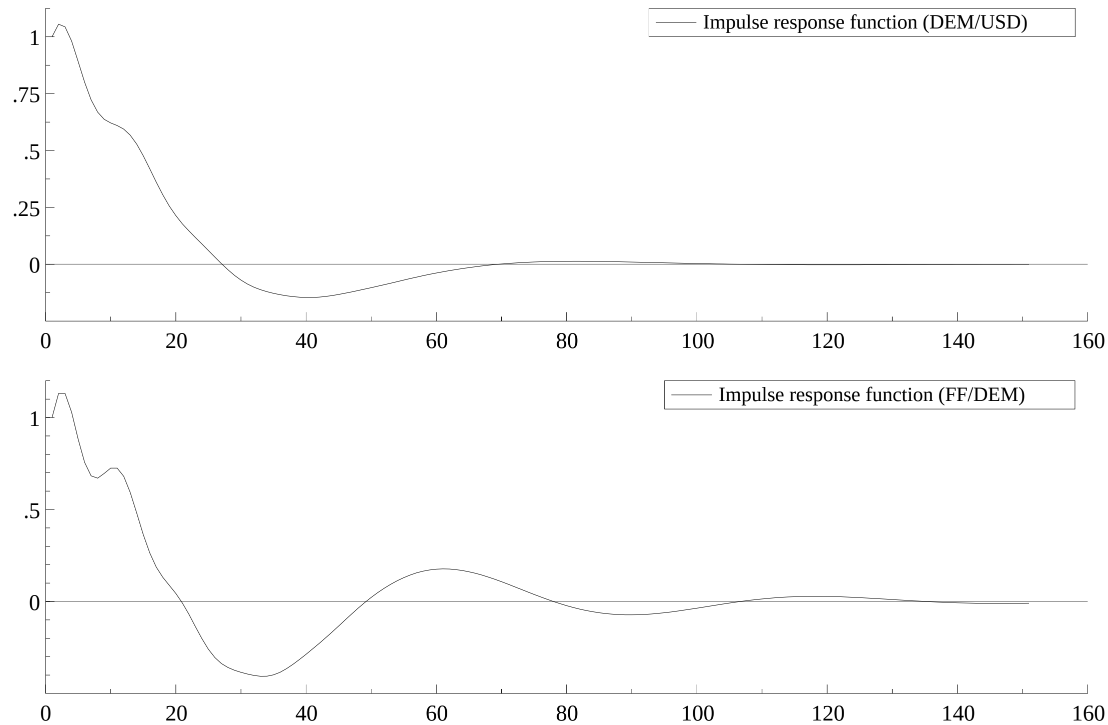

Figure 6: Posterior mean of the impulse response function of $\psi_1 + \psi_2$ for DEM/USD and FF/DEM.

Different possible parameter configurations result in a variety of impulse responses. The Bayesian methodology enables us to integrate out the parameter uncertainty, weighting each possible impulse response and its implied half-life according to the posterior probability of the particular outcome. Note that the posterior expectation of the half-life is not the same as the half-life of the expected impulse response, because the half-life is a highly nonlinear function of the impulse response.

In Figure 7 we see that the posterior density of the half-life even almost entirely excludes half-lives of 3 to 5 years. The half-lives we find are more likely to be attributable to sticky prices and other rigidities than the earlier findings. Including a flexible mean is the main reason that we find half-lives around one year for both real exchange rates. This is much shorter than the half-lives reported in previous studies which are in the range of 3 to 5 years. Surprisingly, we do not find substantial differences between the half-lives for the DEM/USD data and the FF/DEM data in the mean or median. From the densities of the half-lives, we find that for DEM/USD the distribution is much more dispersed than for FF/DEM, indicating less uncertainty for the intra-European combination of countries.

Note that we cannot use the full model to calculate half-lives. In that case the impulse responses are related to shocks to the entire system instead of shocks to the deviations from the mean. Part of the shock will be attributed to the trend component with the effect that

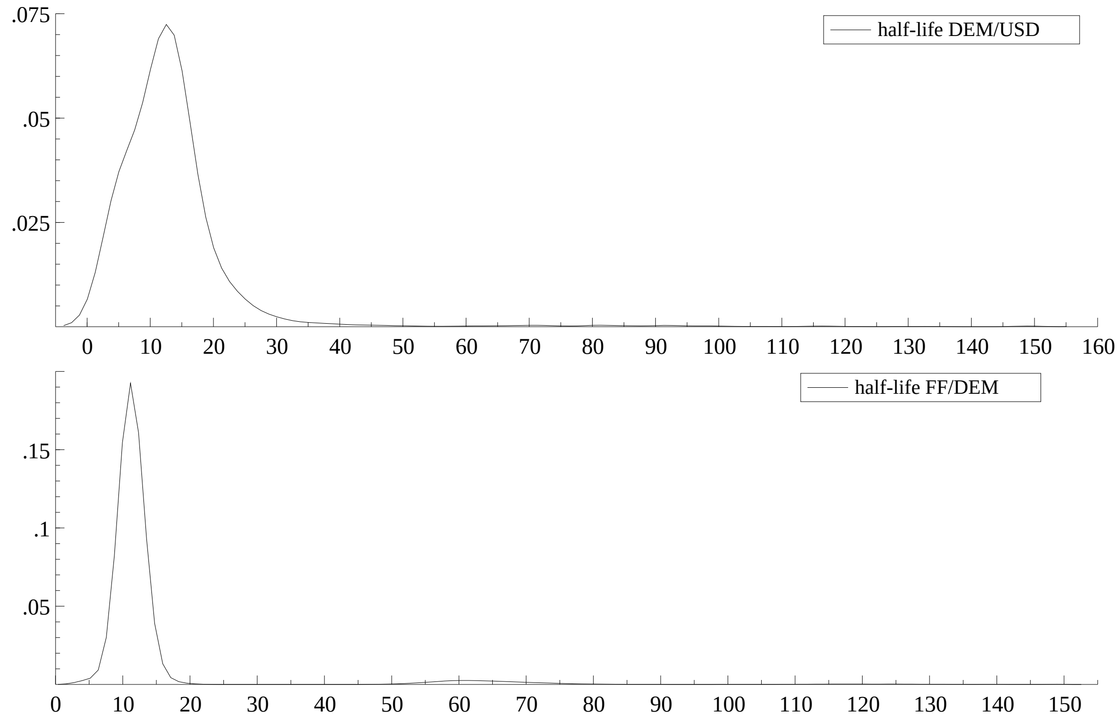

Figure 7: Posterior density of the half-life for DEM/USD and FF/DEM.

the slope is changed. Thus, the initial shock will have a permanent and even increasing effect over time. The impulse response function will therefore diverge and no finite half-life exists in that case.

Table 2: More posterior results.

	DEM/USD	FF/DEM
BF	0.000357	0.000553
Half-life (mean)	13.6	14.8
Half-life (median)	12.0	11.0

BF: the Bayes Factor for testing $\sigma_\zeta^2 = 0$. Half-life: mean and median of the posterior distribution of the half-life of the stationary components.

6 Final remarks

While the debate on whether long-run PPP holds or not will continue, we have provided some evidence for short run PPP for the recent period of floating exchange rates. After correcting for the long-term variation, the half-lives drop to reasonable levels of about one year. It seems that replacing the fixed-mean assumption of PPP by a more flexible slowly varying mean is definitely helpful in this respect. Our Bayesian framework can fully quantify the uncertainty of the half-life estimates that Cheung and Lai (2000) report by providing the entire posterior density. We find the same kind of non-monotonicity in the impulse response function as they do. A shock is initially amplified before it starts to die out.

Of course, the approach taken in this paper is pragmatic in nature but at least it suggests a direction in which an explanation for the PPP puzzle may be found. It still remains a challenging task to explain and model the mechanisms that determine the time-varying behavior of the real exchange rates. Non-traded goods might not be the only factors responsible for the time-variation of the equilibrium rate. A viable other possibility would be to use a monetary exchange rate model that includes interest rates and money demand and supply. This is an interesting topic of further research. Furthermore, in the PPP literature panel models have aided a lot to compensate for the limited time span of the recent period of floating exchange rates. Hence, an extension of our model to panels of countries will be the topic of our future research.

* * *

All computations in this paper were carried out using Ox, see Doornik (1999) and the SsfPack package (Koopman et al. (1998)).

* * *

R. Kleijn gratefully acknowledges financial support from N.W.O.

A Sampling the cycle parameters

The cycle has three parameters, the variance σ^2, the damping factor ρ and the frequency parameter λ. We describe how to sample from their posterior distributions. The procedure is a Metropolis-Hastings (M-H) within Gibbs algorithm with carefully chosen candidate distributions. This algorithm for sampling of the cycle parameters can be used as a block of an encompassing Gibbs sampler when the cycle is a component of a larger model. For simplicity of the exposition we assume that a Jeffreys' prior is used for σ^2 and for ρ and λ we specify uniform priors on $[0, 1]$ and $[0, \pi]$, respectively. An inverted gamma prior on the variance is natural conjugate and can be handled analytically. Other priors can easily be implemented by taking account of them in the acceptance probabilities of the relevant Metropolis-Hastings (M-H) steps. Regarding the initial conditions we assume that ψ_1 and ψ_1^* are initialized by a draw from the unconditional distribution of $\psi = (\psi_1, \ldots, \psi_T)$ and $\psi^* = (\psi_1^*, \ldots, \psi_T^*)$. We introduce some other notation: $Z' = (\psi', \psi^{*\prime})$, $Z_{-1} = LZ$, with L the lag operator, and $C_\lambda = \begin{pmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{pmatrix}$.

The joint posterior density is obtained from combining the likelihood of a 2-dimensional restricted VAR(1) with the prior information and the initial conditions,

$$\begin{aligned}
p(\rho, \lambda, \sigma_\kappa^2|\psi, \psi^*) &\propto p(\rho, \lambda, \sigma_\kappa^2|\psi, \psi^*) \\
&\propto L(\psi, \psi^*|\psi_0, \psi_0^*, \rho, \lambda, \sigma_\kappa^2) p(\psi_0, \psi_0^*|\rho, \sigma_\kappa^2) p(\rho, \lambda, \sigma_\kappa^2) \\
&\propto (\sigma_\kappa^2)^{-T} \exp(-\frac{1}{2\sigma_\kappa^2} \operatorname{tr}((Z - \rho C Z_{-1})'(Z - \rho C Z_{-1}))) \times \\
&(1-\rho^2)\sigma_\kappa^{-2} \exp(-(1-\rho^2)\frac{\psi_0^2 + \psi_0^{*2}}{2\sigma_\kappa^2}) \times \sigma_\kappa^{-2} \\
&= (1-\rho^2)(\sigma_\kappa^2)^{-(T+2)} \exp(-\frac{1}{2\sigma_\kappa^2}\left[\operatorname{tr}((Z - \rho C Z_{-1})'(Z - \rho C Z_{-1})) + \right. \\
&\qquad \left. (1-\rho^2)(\psi_0^2 + \psi_0^{*2})\right])
\end{aligned}$$

One can recognize the kernel of an inverted gamma distribution for σ^2, so the conditional posterior of σ^2 is described by

$$\left. \frac{\operatorname{tr}((Z - \rho C Z_{-1})'(Z - \rho C Z_{-1})) + (1-\rho^2)(\psi_0^2 + \psi_0^{*2})}{\sigma_\kappa^2} \right| \rho, \lambda, \psi, \psi^* \sim \chi^2_{2(T+1)}.$$

Using an inverted gamma integration step, the joint marginal posterior density of ρ and λ is found to be

$$\begin{aligned}
p(\rho, \lambda|\psi, \psi^*) &= \int p(\rho, \lambda, \sigma_\kappa^2|\psi, \psi^*) d\sigma_\kappa^2 \\
&\propto (1-\rho^2)\left[\operatorname{tr}((Z - \rho C Z_{-1})'(Z - \rho C Z_{-1})) + (1-\rho^2)(\psi_0^2 + \psi_0^{*2})\right]^{-(T+1)}
\end{aligned}$$

We can rewrite the trace expression as a quadratic expression in ρ

$$\begin{aligned}\operatorname{tr}((Z-\rho CZ_{-1})'(Z-\rho CZ_{-1})) &= \operatorname{tr}(Z'Z) - 2\rho\operatorname{tr}(Z'CZ_{-1}) + \rho^2\operatorname{tr}(Z_{-1}'Z_{-1}) \\ &= \operatorname{tr}(Z_{-1}'Z_{-1})\left(\rho - \frac{\operatorname{tr}(Z'CZ_{-1})}{\operatorname{tr}(Z_{-1}'Z_{-1})}\right)^2 + \\ &\quad \left(\operatorname{tr}(Z'Z) - \frac{\operatorname{tr}(Z'CZ_{-1})^2}{\operatorname{tr}(Z_{-1}'Z_{-1})}\right)\end{aligned}$$

The factor $(1-\rho^2)$ and the term $(1-\rho^2)(\psi_0^2+\psi_0^{*2})$ represent the initial conditions of cycle. If enough observations are available, they tend to be dominated by the information in the data in the posterior density. Without these subexpressions one can recognize a Student-t kernel. Hence, it is sensible to use a (truncated) Student-t candidate in a M-H step for $\rho|\lambda,\psi_1,\psi_2$.

So, $\rho^{cand}|\lambda,\psi,\psi^* \sim t(\mu,s^2,\nu)$, i.e. $p(\rho^{cand}) \propto \left(1+\frac{(\rho^{cand}-\mu)^2}{\nu s^2}\right)^{-\frac{1}{2}(\nu+1)}$, with

$$\begin{aligned}\mu &= \frac{\operatorname{tr}(Z'CZ_{-1})}{\operatorname{tr}(Z_{-1}'Z_{-1})} \\ s^2 &= \frac{1}{2T+1}\left[\frac{\operatorname{tr}(Z'Z)}{\operatorname{tr}(Z_{-1}'Z_{-1})} - \frac{\operatorname{tr}(Z'CZ_{-1})^2}{\operatorname{tr}(Z_{-1}'Z_{-1})^2}\right] \\ \nu &= 2T+1\end{aligned} \tag{A.1}$$

In a single M-H step we accept or reject the proposed values for ρ and $\sigma^2|\rho$.

Finally, the full conditional density of λ is given by

$$\begin{aligned}p(\lambda|\rho,\sigma_\kappa^2,\psi,\psi^*) &\propto p(\rho,\lambda,\sigma_\kappa^2|\psi,\psi^*) \\ &\propto \exp\left(\frac{\rho}{\sigma_\kappa^2}\operatorname{tr}(Z'CZ_{-1})\right) \\ &= \exp\left(\frac{\rho}{\sigma_\kappa^2}\left[(\psi\psi_{-1}' + \psi^*\psi_{-1}^{*}{}')\cos\lambda + (\psi\psi_{-1}^{*}{}' - \psi^*\psi_{-1}')\sin\lambda\right]\right)\end{aligned}$$

We may combine the sine and cosine into one cosine with phase shift, and do a Taylor expansion as follows,

$$\begin{aligned}p(\lambda|\rho,\sigma_\kappa^2,\psi,\psi^*) &= \exp(A\cos\lambda + B\sin\lambda) \\ &= \exp(R\cos(\lambda-\varphi)) \\ &\propto 1 - \frac{1}{2}R(\lambda-\varphi)^2 + \frac{1}{24}R(3R+1)(\lambda-\varphi)^4 + O((\lambda-\varphi)^6) \\ &\approx \left(1+\frac{(\lambda-\varphi)^2}{6}\right)^{-3R}\end{aligned}$$

where $R^2 = A^2+B^2$ and $\varphi = \arctan(B/A)$. The last expression is the kernel of a Student-t density with location φ, scale parameter $\frac{6}{6R-1}$ and $6R-1$ degrees of freedom. We use this Student-t distribution as a M-H candidate. The bounded support of λ introduces truncation, which can be accounted for either by using a truncated Student-t candidate or by including it in the acceptance probability of the independence Metropolis-Hastings step.

References

L. Bauwens, M. Lubrano, and J.-F. Richard. *Bayesian inference in dynamic econometric models*. Advanced texts in Econometrics. Oxford University Press, Oxford, 1999.

S. Beveridge and C.R. Nelson. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "business cycle". *Journal of Monetary Economics*, **7**:151–174, 1981.

C.K. Carter and R. Kohn. On Gibbs sampling for state space models. *Biometrika*, **81**: 541–553, 1994.

G. Casella and E.I. George. Explaining the Gibbs sampler. *The American Statistician*, **46**: 167–174, 1992.

Y-W. Cheung and S.L. Lai. On the purchasing power parity puzzle. *Journal of International Economics*, **52**:321–330, 2000.

S. Chib. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, **90**(432):1313–1321, 1995.

S. Chib and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49:327–335, 1995.

P.K. Clark. Nearly redundant parameters and measures of persistence in economic time series. *Journal of Economic Dynamics and Control*, **12**:447–461, 1988.

P. de Jong and N. Shephard. The simulation smoother for time series models. *Biometrika*, **82**:339–350, 1995.

J.A. Doornik. *Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants Press, London, 3rd edition, 1999. Oxford: www.nuff.ox.ac.uk/Users/Doornik.

C. Engel. Long-run PPP may not hold after all. *Journal of International Economics*, **57**: 243–273, 2000.

S. Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15**:183–202, 1994.

A.E. Gelfand and A.F.M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:389–409, 1990.

J. Geweke. Using simulation methods for Bayesian econometric models: inference, development and communication. *Econometric Reviews*, **18**(1):1–73, 1999.

J.D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, NJ, 1994.

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

R.E. Kass, L. Tierney, and A.E. Raftery. The validity of posterior expansions based on Laplace's method. In S. Geisser, editor, *Bayesian and Likelihood methods in statistics and econometrics: Essays in honor of George A. Barnard*, pages 473–488. North Holland, Amsterdam, 1990.

G. Koop and H.K. van Dijk. Testing for integration using evolving trend and seasonals models: A Bayesian Approach. *Journal of Econometrics*, **97**:261–291, 2000.

S.J. Koopman, N. Shephard, and J.A. Doornik. Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal*, **1**:1–55, 1998.

D. Kwiatkowski, P.C.B. Phillips, P. Schmidt, and Y. Shin. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, **54**:159–178, 1992.

P.C.B. Phillips. To criticize the critics: An objective Bayesian analysis of stochastic trends. *Journal of Applied Econometrics*, 6:333–364, 1991.

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

P.C. Schotman and H.K. van Dijk. A Bayesian analysis of the unit root in real exchange rates. *Journal of Econometrics*, **49**:195–238, 1991.

C.A. Sims. Bayesian skepticism on unit root econometrics. *Journal of Economic Dynamics and Control*, **12**:463–474, 1988.

C.A. Sims and H. Uhlig. Understanding unit rooters: a helicopter tour. **59**:1591–1599, 1991.

M.A. Tanner and W.H. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**:528–550, 1987.

M. West and J. Harrison. *Bayesian forecasting and dynamic models, 2 edn.* Springer-Verlag, New York, 1997.

E. Zivot. A Bayesian analysis of the unit root hypothesis within a unobserved components model. *Econometric Theory*, **10**:552–578, 1994.