Quantitative Finance

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rquf20

Forecasting exchange rates using asymmetric losses: A Bayesian approach

Georgios Tsiotas

Quantitative Finance, 2022
Vol. 22, No. 2, 273–287, https://doi.org/10.1080/14697688.2021.1942180

Forecasting exchange rates using asymmetric losses: A Bayesian approach

GEORGIOS TSIOTAS*

Department of Economics, University of Crete, Rethymnon, Greece

(*Received 20 September 2020; accepted 4 June 2021; published online 19 July 2021*)

The forecasting of exchange rate returns has long been an issue in finance literature. The use of the best forecasting model is usually sensitive to the data frequency and the sample period used. Model evaluation is usually based on either minimizing error losses or maximizing profit strategies and other likelihood-based measures. Although, much work has been devoted to model evaluation based on maximizing profits strategies little to no work has been devoted to the issue of estimating a forecast model under the same principles. Here, we propose a Bayesian framework that estimates exchange rate models by considering measures such as directional accuracy and trading rules in a form of asymmetric loss functions. Estimation is implemented using Laplace-type estimators applied in cases where the likelihood function is not of a known form. We illustrate this method using simulated and real weekly exchange rate series. The results demonstrate that the use of profit maximizing strategies within estimation can significantly improve the forecasting ability of certain exchange rate models.

Keywords: Forecast evaluation; MCMC; Laplace-type estimator; Exchange rates; Directional accuracy; Profitability

JEL Code: C11, C15, C32, C53

1. Introduction

Exchange rate return predictability has been the interest of many researchers in the past. The out-of-sample forecasting accuracy usually depends on the frequency of the series and on the models adopted. Meese and Rogoff (1983) argued that all exchange rate models, including those that incorporate economic fundamentals, do less well in out-of-sample forecasting exercises than a simple driftless random walk (RW). In the same spirit, Boothe and Glassman (1987) have confirmed earlier findings that simple time-series models such as the RW rank highest in forecast accuracy when compared with autoregressive and the real-interest differential (RID) model. Dynamic models such as: linear autoregressive models, nonlinear bilinear models (Brook 1997), threshold autoregressive models (Krager and Kugler 1993, Brook 1997), nonlinear dynamic systematic filtering models (Lisi and Medio 1997) have also been used in the past. Other approaches have proposed univariate and multivariate specifications based on macroeconomic fundamentals as a way to make accurate out-of-sample predictions of spot exchange rates (Mark 1995, Mark and Sul 2001). More recently, focus has been put on combined exchange rate forecasts using various Bayesian and non-Bayesian techniques (Wright 2008, Costantini *et al.* 2016, Cuaresma *et al.* 2018). Finally, there has been advances on the use of machine learning techniques applied in financial data. These techniques provide forecasts based on a selection of large potential predictors and model specifications (see Gu *et al.* 2020).

In the above-cited literature, model estimation and evaluation have been based on the least-squares criterion. Thus, the usual criteria for selecting the best model are based on minimizing the Mean Squared Forecast Error (MSFE), the Mean Absolute Forecast Error (MAFE), the Root Mean Squared Forecast Error (RMSFE), on maximizing likelihood-based functions. However, in exchange rate forecasting, there are issues regarding the use of proper loss functions that were not taken into account when estimating parametric models. Thus, as Granger (1999) first pointed out, model estimation and evaluation should consider apart from some optimal criteria such the least squares, others related to the correct direction of the forecast. Under this principle, rather than focus on forecasts that minimize the mean forecast error, investment managers should focus on those than can predict the direction of changes. In this spirit, Granger introduces some asymmetric forecast measures, such as: the Mean Forecast Trading Returns (MFTR) and the Mean Correction

*Corresponding author Email: tsiotas@uoc.gr

Forecasts Directions (MCFD). Clements and Hendry (1993) have shown that the MSFE measure may be inadequate and potentially misleading when it comes to selecting models with transformed data. Hong and Lee (2003) have performed an out-of-sample model evaluation and a formal model comparison test using the MFTR and MCFD measures together with the MSFE and MAFE. More recently, Menkhoff and Taylor (2007) empirically examined the potential profitability of technical trading rules based on exchange rate predictions (see applications in Costantini *et al.* 2016, Cuaresma *et al.* 2018).

However, no work has been done so far to integrate these asymmetric model evaluation measures in to model estimation. In this article, we intend to incorporate directional accuracy and trading rules measures within the model estimation of some basic exchange rate models. To do so, we will adopt a Laplace-type (LT) estimator within a Markov Chain Monte Carlo (MCMC) set-up (see Gelman *et al.* 2004, Chernozhukov and Hong 2003). The MCMC is a Bayesian simulation procedure that produces highly efficient simulation results by overcoming the problem of non-existence of the likelihood in a closed form. Under the LT estimator, introduced by Chernozhukov and Hong (2003), one can derive quasi-posterior draws from a pseudo-likelihood function based on an asymmetric loss function. In practice, we adopt a quasi-likelihood function with an asymmetric kernel function based on the MFTR and/or MCFD metric that regulates the acceptance or not of some simulated draws. Thus, the objective of this study is two-fold. First, we develop the framework for estimating dynamic exchange-rate models under the presence of asymmetric objective functions based on the directional accuracy and trading strategies principles. We will do this by using an efficient Bayesian estimation framework. Second, by using simulated and real financial data, we will investigate the effect of the new estimation procedure in forecasting exchange rate returns with simple autoregressive structures. We will test the performance of this procedure using the MSFE, MAFE, MFTR and MCFD measures in the presence of correctly and incorrectly specified autoregressive models.

The paper is organized as follows. Section 2 addresses the problem of loss functions in forecast evaluation together with the issue of adopting profit maximizing measures when evaluating models. In Section 3, we develop the framework of estimating forecast models for exchange rates in the presence of asymmetric losses. An efficient Bayesian approach is used for an LT estimator within an efficient MCMC algorithm. In Section 4, we illustrate our results for real and simulated data using autoregressive models. Here, a forecasting evaluation experiment is based on both forecast error minimization and profit maximization. Finally, we present our conclusions in Section 5.

2. Loss functions and forecast evaluation criteria

Loss functions play an important role in optimal estimation since optimal estimates will rely on the type of loss function one applies. Usually, forecasters relying on the symmetric quadratic loss function would get the conditional mean as the optimal forecast. However, as stated in Granger (1969) and Christoffersen and Diebold (1997) the actual optimal predictor is the conditional mean plus a bias term which relies on prediction error higher moments and it is also affected by a non i.i.d. nature of the prediction error.

Suppose, we have the observed time-series $\boldsymbol{y} = (y_1, \ldots, y_n)$ of exchange-rate returns and a corresponding parametric model $y_t = m(y_1, \ldots, y_{t-1}, \boldsymbol{\beta}) + \epsilon_t$. Here, we assume the parameter vector $\boldsymbol{\beta} \in \Theta \subset \Re^k$ under estimation and the error term $\epsilon_t \sim ID(0, \tau)$ with a variance component τ. Usually, researcher in economics and finance use the Mean Square Forecast Error (MSFE) and the Mean Absolute Forecast Error (MAFE) as measures of forecasting evaluation. However, there are some measures that take into account the practitioners interests which are based on maximizing profits rather than minimizing forecasts errors. To this direction Granger (1999) has emphasized the importance of certain losses: the trading return loss $L_1(y_t, \hat{y}_t) = \text{sign}(\hat{y}_t) \cdot y_t$, and the correct direction loss $L_2(y_t, \hat{y}_t) = \mathbf{1}(\text{sign}(\hat{y}_t) \cdot \text{sign}(y_t) > 0)$. Here, y_t denotes the observed series at time t and $\hat{y}_t$ the forecasted one. Also, we denote $\text{sign}(y_t) = \mathbf{1}(y_t > 0) - \mathbf{1}(y_t < 0)$ where $\mathbf{1}$ takes the value 1 if the statement in the parenthesis is true and 0 otherwise.

Suppose from the n available data series, we use the most recent $n - s$ ones and we are left with a forecasting period of s observations. Thus, we have the following forecasting measures:

$$\text{MSFE} = \frac{1}{s} \sum_{t=n-s+1}^{n} (y_t - \hat{y}_t)^2,$$

$$\text{MAFE} = \frac{1}{s} \sum_{t=n-s+1}^{n} |y_t - \hat{y}_t|,$$

$$\text{MFTR} = \frac{1}{s} \sum_{t=n-s+1}^{n} y_t \cdot \text{sign}(\hat{y}_t),$$

$$\text{MCFD} = \frac{1}{s} \sum_{t=n-s+1}^{n} \mathbf{1}\left(\text{sign}(\hat{y}_t) \cdot \text{sign}(y_t) > 0\right).$$

The last two measures are to be maximized rather than minimized, unless we place a minus sign on them. As Granger points out, managers in mutual funds are interested in market timing. Thus, in the trading return loss, if $\text{sign}(\hat{y}_t) = 1$ then the mean loss expressed by the MFTR will tend to the observed data mean value when the forecasting period tends to infinity ($\frac{1}{s} \sum_{t=n-s+1}^{n} y_t \to E(\boldsymbol{y})$ as $s \to \infty$). As regards the MCFD measure, the investment managers can maximize their profits if they can predict the direction of changes, thus earning higher than the market average. An alternative trading rule strategy has been introduced by Gencay (1998) where the trading signal is based on the spot exchange rate and its forecast. Also, there are a number of articles devoted to the profitability of trading rules (see review in Menkhoff and Taylor 2007). This can be seen as a test for the efficient market hypothesis which assesses whether by using publicly available information one can forecast changes in exchange rates.

In estimation practice, one can easily estimate a parametric model in the spirit of MSFE minimization. Thus, for a within-sample period one can either use a least squared error estimation approach $\arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{n-s}(y_t - m(y_1, \ldots, y_{t-1}, \boldsymbol{\beta}))^2$ or adopt a likelihood-based approach using the Normal assumption. Thus, under the later approach,

$$\arg\max_{\boldsymbol{\theta}} \log p(\boldsymbol{y}, \boldsymbol{\theta})$$

with parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \tau)$ and $p(\boldsymbol{y}, \boldsymbol{\theta})$ to denote the likelihood function. However, if one tries to estimate a parametric model using principles similar in spirit to that of maximizing the MFTR and MCFD measures, this will create a likelihood intractability problem. In the section that follows, we will address the problem of likelihood intractability when one tries to make parametric estimation in model cases where the objective is the maximization of profits together with that of the minimization of squared errors.

3. Bayesian estimation

Under the Bayesian approach, one can approximate the posterior draws using the posterior density function:

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{\theta}) \cdot \pi(\boldsymbol{\theta})}{p(\boldsymbol{y})},$$

with the $p(\boldsymbol{y} \mid \boldsymbol{\theta})$ to denote the likelihood function which is known up to a normalizing constant, the $\pi(\boldsymbol{\theta})$ to denote the prior for the $\boldsymbol{\theta}$ parameter vector and the $p(\boldsymbol{y})$ to denote the marginal likelihood which is also known up to a normalizing constant. Standard approximation using MCMC methods could resolve this problem, such as $p(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto p(\boldsymbol{y}, \boldsymbol{\theta}) \cdot \pi(\boldsymbol{\theta})$. However, by introducing a likelihood function of an unknown form, the exact likelihood function $p(\boldsymbol{y}, \boldsymbol{\theta})$ will not even be known up to a normalizing constant. This is the problem of likelihood intractability.

The above problem, where parametric and other estimates are not easily computed, can be resolved via a Metropolis–Hastings algorithm within MCMC. By sampling within a Markov Chain, one targets the 'true' posterior distribution $p(\boldsymbol{\theta} \mid y)$ with the help of the ergodic property of the Monte Carlo. Under a Metropolis–Hastings algorithm, the proposed simulated value $\hat{\boldsymbol{\theta}}$ from the transition kernel $q(\cdot|\cdot)$ is accepted with probability:

$$\alpha(\boldsymbol{\theta}^{(d)}, \hat{\boldsymbol{\theta}}) = \min\left[\frac{p(y, \boldsymbol{\theta}^{(d)})\pi(\boldsymbol{\theta}^{(d)})q(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}^{(d)})}{p(y, \hat{\boldsymbol{\theta}})\pi(\hat{\boldsymbol{\theta}})q(\boldsymbol{\theta}^{(d)}|\hat{\boldsymbol{\theta}})}, 1\right].$$

Thus, even if the posterior model $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta} \mid \boldsymbol{y})$ is difficult to be generated, the posterior mean

$$\bar{\boldsymbol{\theta}} = \frac{1}{D}\sum_{d=1}^{D} \boldsymbol{\theta}_d \approx \int_{\Theta} \boldsymbol{\theta} \cdot p(\boldsymbol{\theta} \mid \boldsymbol{y})\, \mathrm{d}\boldsymbol{\theta},$$

is approximated with the above probability.

3.1. *Quasi-Bayesian estimators and asymmetric losses*

When one faces problems such as: highly non-convex objective functions, elusive loss functions or intractable likelihoods functions, a Quasi-Bayesian approach can be adopted. Thus, under (Chernozhukov and Hong 2003) the above problem can be resolved using the so-called quasi-Bayesian estimators (QBE). The definition of these estimators is similar to that of the Bayesian estimators; however, QBE use general statistical criterion functions in place of parametric likelihood functions. The class of QBEs is designed to explore the use of the Laplace approximation when the likelihood function is unknown. In practice, given a loss (or an objective) function of interest $L(\boldsymbol{y}, \boldsymbol{\theta})$, the authors take the exponential of $-L(\boldsymbol{y}, \boldsymbol{\theta})$ and combine the $\mathrm{e}^{-L(y,\boldsymbol{\theta})}$ with a prior density $\pi(\boldsymbol{\theta})$ to produce a quasi-posterior density. Thus, under a strictly positive and continuous prior, the posterior

$$p(\boldsymbol{\theta} \mid y) = \frac{\mathrm{e}^{-L(y,\boldsymbol{\theta})}\pi(\boldsymbol{\theta})}{\int_{\Theta} \mathrm{e}^{-L(y,\boldsymbol{\theta})}\pi(\boldsymbol{\theta})} \propto \mathrm{e}^{-L(y,\boldsymbol{\theta})} \times \pi(\boldsymbol{\theta})$$

is a proper one. After introducing some prior density assumptions, one can generate quasi-posterior, moments, quantiles and credible intervals using the MCMC engine.

Under a frequentist perspective, the QBE provides asymptotically normal results even if the prior density is not a proper density function. By incorporating the above likelihood function within an MCMC engine, we have the following quasi-Bayesian MCMC algorithm such as:

1. Draw a vector of starting values $\boldsymbol{\theta}^{(0)}$.
2. Generate $\hat{\boldsymbol{\theta}}$ from $q(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}^{(0)})$ from a transition kernel $q(\cdot \mid \cdot)$ kernel.
3. Update $\boldsymbol{\theta}^{(d+1)}$ from $\boldsymbol{\theta}^{(d)}$ for $d = 1, \ldots, D$, with

$$\boldsymbol{\theta}^{(d+1)} = \begin{cases} \hat{\boldsymbol{\theta}} & \text{with probability } \alpha(\boldsymbol{\theta}^{(d)}, \hat{\boldsymbol{\theta}}) \\ \boldsymbol{\theta}^{(d)} & \text{with probability } 1 - \alpha(\boldsymbol{\theta}^{(d)}, \hat{\boldsymbol{\theta}}), \end{cases}$$

where

$$\alpha(\boldsymbol{\theta}^{(d)}, \hat{\boldsymbol{\theta}}) = \min\left[\frac{\mathrm{e}^{-L(y,\boldsymbol{\theta}^{(d)})}\pi(\boldsymbol{\theta}^{(d)})q(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}^{(d)})}{\mathrm{e}^{-L(y,\hat{\boldsymbol{\theta}})}\pi(\hat{\boldsymbol{\theta}})q(\boldsymbol{\theta}^{(d)}|\hat{\boldsymbol{\theta}})}, 1\right].$$

Then, given u from $\mathcal{U}(0, 1)$, if $u \leq \alpha(\boldsymbol{\theta}^{(d)}, \hat{\boldsymbol{\theta}})$ keep $\hat{\boldsymbol{\theta}}$, otherwise reject.

The above quasi-Bayesian MCMC algorithm has the advantage, compared with the original Chernozhukov–Hong approach, of providing a direct way to construct valid posterior credible intervals as well as standard errors if the optimal weighted matrix is used (see Lise *et al.* 2016, Forneron and Ng 2018).

In practice, if one wants to generate estimates based on the minimum mean squared error criterion, the quasi-likelihood function is identical to a Gaussian likelihood with the form:

$$p_0(\boldsymbol{\theta} \mid \boldsymbol{y}) = \mathrm{e}^{-L_0(y,\boldsymbol{\theta})} \propto \exp\left(-\frac{1}{2}\frac{\psi_0}{\tau}\right),$$

where $\psi_0 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ is the squared error loss. The above method is a standard MCMC method (hereafter the MCMC algorithm).

Now, if one wants to incorporate the following auxiliary statistics such as: the trading returns loss $\psi_1 = \sum_{i=1}^{n} y_i \cdot \text{sign}(\hat{y}_i)$ and the correct directions loss $\psi_2 = \sum_{i=1}^{n} \mathbf{1}(\text{sign}(\hat{y}_t) \cdot \text{sign}(y_t) > 0)$ within an objective function, this can be a difficult task. To resolve this problem, we will integrate them within a quasi-Bayesian estimator framework.

Thus, as a second algorithm we define a quasi-likelihood function based on two auxiliary statistics: the squared error loss and the trading returns loss, such as:

$$p_1(\boldsymbol{\theta} \mid \boldsymbol{y}) = \mathrm{e}^{-L_1(y,\boldsymbol{\theta})} \propto \exp\left(-\frac{1}{2}\left[\frac{\psi_0}{\tau} - (\psi_1 - \psi_1^0)\right]\right),$$

where $\psi_1^0 = \sum_{i=1}^{n} y_i$ is the optimal trading returns measure. Thus, in the optimal case where $\psi_1 \to \sum_{i=1}^{n} y_i$, then our quasi-likelihood function is reduced to the known Gaussian likelihood function. On the other hand, under the worst scenario for the trading returns metric, the quasi-likelihood function will take the form:

$$p_1(\boldsymbol{\theta} \mid \boldsymbol{y}) = \mathrm{e}^{-L_1(y,\boldsymbol{\theta})} \propto \exp\left(-\frac{1}{2}\left[\frac{\psi_0}{\tau} - \sum_{i=1}^{n} y_i\right]\right).$$

We call the above method the Simulated Laplace-type estimator-I (hereafter the SLT-I algorithm). Secondly, we define a quasi-likelihood function based on two auxiliary statistics: the squared error loss and the correct direction measures, such as:

$$p_2(\boldsymbol{\theta} \mid \boldsymbol{y}) = \mathrm{e}^{-L_2(y,\boldsymbol{\theta})} \propto \exp\left(-\frac{1}{2}\left[\frac{\psi_0}{\tau} - (\psi_2 - \psi_2^0)\right]\right),$$

where $\psi_2^0 = n$ is the correct direction measures, respectively. Thus, in the optimal case where $\psi_2 \to n$, then our quasi-likelihood function is reduced to the known Gaussian likelihood function. On the other hand, under the worst scenario for the correct directions metric, the quasi-likelihood function will take the form:

$$p_2(\boldsymbol{\theta} \mid \boldsymbol{y}) = \mathrm{e}^{-L_2(y,\boldsymbol{\theta})} \propto \exp\left(-\frac{1}{2}\left[\frac{\psi_0}{\tau} - n\right]\right).$$

We call the above method the Simulated Laplace-type estimator-II (hereafter the SLT-II algorithm).

As a third approach, we use as loss function the Mahalanobis distance of the $\boldsymbol{\psi}(\boldsymbol{y}) = (\psi_1 - \psi_1^0, \psi_2 - \psi_2^0)'$ vector. In such a case, the quasi-likelihood function will have the form:

$$p_3(\boldsymbol{\theta} \mid \boldsymbol{y}) = \mathrm{e}^{-L_3(y,\boldsymbol{\theta})} \propto \exp\left(-\frac{1}{2}\left[\frac{\psi_0}{\tau} + g(\boldsymbol{\theta})'\boldsymbol{W}g(\boldsymbol{\theta})\right]\right),$$

where $g(\boldsymbol{\theta}) = \boldsymbol{\psi}(\boldsymbol{y}) - \frac{1}{D}\sum_{d=1}^{D} \boldsymbol{\psi}^d(\boldsymbol{y})$ and $\boldsymbol{W}$ the covariance matrix estimated by the sample covariance of the vectors $\boldsymbol{\psi}(\boldsymbol{y})$ generated after the first $D = 1000$ iterations of the chain. We call the above method the Simulated Laplace-type estimator-Mahalanobis (hereafter the SLT-Mahalanobis algorithm). The squared Mahalanobis distance has been used in the past as a distance measure in contemporary Bayesian algorithms such as the Approximate Bayesian Computation method (see Sisson 2011). Moreover, the Mahalanobis distance as a quasi-likelihood component has also been used in Forneron and Ng (2018). Another approach, such as the synthetic-likelihood introduced by Wood (2010), considers a similar idea to the above this time replacing the loss function with a Gaussian kernel as defined in the indirect likelihood approach by Jiang and Turnbull (2004).

Alternatively, one can adopt a non-Bayesian simulation approach such as the Simulated Minimum Distance (SMD) where parametric estimates for the $\boldsymbol{\theta}$ vector are to be minimized for the distance $g(\boldsymbol{\theta})'\boldsymbol{W}g(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. Other non-simulated approaches, can be seen as minimum distance estimators such as the generalized method of moments (GMM) adopted by Hansen (1982) or the indirect likelihood method introduced by Jiang and Turnbull (2004).

Finally, as a solution to the problem of loss function uncertainty one can use a mixed quasi-likelihood function. Thus, we adopt a mixed Simulated Laplace-type estimator (hereafter the SLT-mix algorithm) under which the competing auxiliary statistics are embedded in a mixture model, defined by† :

$$p(\boldsymbol{\theta}, \boldsymbol{w} \mid \boldsymbol{y}) = \sum_{k=1}^{2} w_k p_k(\boldsymbol{\theta} \mid \boldsymbol{y}),$$

where $p_1(\boldsymbol{\theta}, \mid \boldsymbol{y}) \propto \exp(-\frac{1}{2}[\frac{\psi_0}{\tau} - (\psi_1 - \psi_1^0)])$ and $p_2(\boldsymbol{\theta}, \mid \boldsymbol{y}) \propto \exp(-\frac{1}{2}[\frac{\psi_0}{\tau} - (\psi_2 - \psi_2^0)])$ with weights $w_k \in [0, 1]$ and $\sum_{k=1}^{2} w_k = 1$. Under this set-up, observed data can assign specific weights to the two competing quasi-likelihood functions based on the maximization criterion.

Therefore, in parametric model estimation we will adopt two approaches. First, we use an adaptive MCMC algorithm to estimate the dynamic exchange rate models under the Gaussian likelihood function (hereafter MCMC algorithm). Secondly, we use a quasi-likelihood function where some asymmetric kernel function has been incorporated as in (the cases of: SLT-I, SLT-II, SLT-Mahalanobis and SLT-mix algorithms).

The parameter vector we need to sample from are: the $\boldsymbol{\theta}$ in the MCMC, SLT-I, SLT-II and SLT-Mahalanobis algorithms and the $(\boldsymbol{\theta}, \boldsymbol{w})$ in the SLT-mix algorithm case. Here, posterior draws from $\boldsymbol{\theta}, \boldsymbol{w}$ can be generated using the approximation:

$$p(\boldsymbol{\theta}, \boldsymbol{w}, \mid \boldsymbol{y}) \propto p(\boldsymbol{y}, \boldsymbol{\theta}) \cdot \pi(\boldsymbol{\theta}) \cdot \pi(\boldsymbol{w})$$

where $\pi(\boldsymbol{w})$ and $\pi(\boldsymbol{\theta})$ are the priors for the $\boldsymbol{w}$ parameter and the $\boldsymbol{\theta}$ parameter vector, respectively.

The transition proposals $q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(d)})$ or $q(\boldsymbol{\theta}', \boldsymbol{w}'|\boldsymbol{\theta}^{(d)}, \boldsymbol{w}^{(d)})$ for all parameters vectors are drawn independently and are based on a component-wise Gaussian RW. Draws are generated using the adaptive Metropolis RW with Gaussian increments (Haario *et al.* 2001, Vihola 2012).

3.2. *Choosing priors*

The choice of parametric priors is very important in Bayesian sampling. Sometimes, if one wants to generate posteriors for

† It has been shown by Rousseau and Mengersen (2011) that if all observations are distributed according to a single model, then the posterior distribution of the corresponding mixture weight w_k concentrates around 1 as n goes to infinity.

a very complex model a careless choice of priors could lead to unrealistic posteriors. Therefore, a prior knowledge of parameters' 'real' space should lead to realistic posterior draws. Given a dynamic parametric model of the form $y_t = \beta_0 + \beta_1 y_1 + \cdots + \beta_k y_{t-k} + \epsilon_t$ with the error term $\epsilon_t \sim \mathrm{NID}(0, \tau)$, we set them around seemingly realistic mean prior values that follows stationarity regions but allow values outside it in order for the algorithm to explore a wide parameter set. Therefore, we set† :

$$\beta_i \sim N(0, 1), \quad i = 0, \ldots k$$

with a small variance value such as 1. Also, we set a flat prior on τ, such that $\log(\tau) \propto 1$.

Regarding the weights parameter $\boldsymbol{w}$ in the SLT-mix algorithm, we set a Dirichlet prior, such that

$$\pi(\boldsymbol{w} \mid a_1, a_2) \propto \prod_{k=1}^{2} w_i^{a_i - 1} \sim Dir(a_1, a_2).$$

For small values of a_i the Dirichlet process is likely to be composed of a single auxiliary statistic, whereas for large values there is more probability for mixed auxiliary statistics within the likelihood function. Here, we set for the scalar a_i a value that equals 1. At the end of Section 4, we will relax the above prior assumption by increasing the flatness of the prior choice.

4. Results

In this section, we investigate the effect of the use of asymmetric loss functions in various forecasting measures for a range of simulated and real data cases. Our intention is to reveal whether the incorporation of directional accuracy and/or trading rules within posterior estimation can affect the forecasting ability of weekly exchange rate returns. To do so, we first describe some alternative simulated data experiments applied in autoregressive models. In Section 4.1, after an initial simulation experiment which will test the invariance of the MSE, MAE, MTR and MCD measures, we will investigate the forecasting ability of the competing algorithms under the presence of correctly and incorrectly specified autoregressive models. Under this set-up, we will diagnose the effect of misspecification due to data generated from a process different from the estimated autoregressive. Section 4.2 is dedicated to real financial data analysis. Here, we will examine whether the introduction of asymmetric loss functions can improve autoregressive models' forecasting ability under various data series and forecasting horizons.

4.1. Simulation data results

Here, we investigate the sampling performance of the four competing algorithms: the MCMC, the SLT-Mahalanobis, the SLT-I and the SLT-II. This is going to reveal which among these competing can estimate better autocorrelated data with similar dependence to the weekly exchange rate returns. Our intention is to reveal which among the competing algorithms can minimize the MSE and MAE and maximize the MTR and MCD measures.

Initially we simulate the data from a range of autoregressive models. These models display conditional dependence similar to the weekly exchange rate returns. So, we generate data from alternative correctly and incorrectly specified models and then we make parametric estimation using the four competing algorithms. Finally, for these alternative estimation methods we derive the MSE, MAE, MTR and MCD measures.

In particular, using the $AR(2)$ model this takes the form of:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \epsilon_t, \tag{1}$$

with $\epsilon_t \sim N(0, \tau)$ and parameters under specification $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \tau)$. In the first simulation experiment, we specify $\theta = (0.0, 0.5897, -0.1858, 0.1796)$ and assume Normality such as: $\epsilon_t \sim N(0, 0.1796)$ (hereafter known as Sim-I). In the second simulation experiment we again specify $\boldsymbol{\theta} = (0.0, 0.5897, -0.1858, 0.1796)$ but this time we assume a fat-tailed Student-t distribution such as $\epsilon_t \sim t_5(0, 0.1796)$ with variance equal to 0.1796 and five degrees of freedom (hereafter known as Sim-II). In the third simulation experiment, we will introduce an $AR(2)$ mixture of Normals model such as:

$$pN\left(\frac{\beta_{0,1}}{(1 - \beta_{1,1} - \beta_{2,1})}, \frac{\tau_1}{(1 - \beta_{1,1}^2 - \beta_{2,1}^2)}\right. \\ \left. + (1 - p)N\left(\frac{\beta_{0,2}}{1 - \beta_{1,2} - \beta_{2,2}}, \frac{\tau_2}{1 - \beta_{1,2}^2 - \beta_{2,2}^2}\right)\right)$$

where $\boldsymbol{\theta}_1 = (\beta_{0,1}, \beta_{1,1}, \beta_{2,1}, \tau_1) = (0.0, 0.5897, -0.1858, 0.1796)$ and $\boldsymbol{\theta}_2 = (\beta_{0,2}, \beta_{1,2}, \beta_{2,2}, \tau_2) = (0.0, 0.2897, -0.1858, 1)$ and $p = 0.70$ (hereafter known as Sim-III).

Here, under the Sim-I procedure, we have a correct specification whereas under the Sim-II and Sim-III procedures we have cases of misspecification. In all simulation experiments we generate a simulated data sample size of $n = 100$, 300 and 500 observations. For these simulated series, we estimate the Normal $AR(2)$ model by generating 60 000 posterior parameter draws for each competing algorithm. Here, we adopt a burn-in period of 30 000 draws. This experiment is replicated 300 times from which posterior estimates are drawn for all parameter vectors together with the posterior averaged MSE, MAE, MTR and MCD measures.

Results are reported in Table 1. Here, we report the averaged MSE, MAE, MTR and MCD measures generated from the 300 replications. Under the Sim-I experiment the SLT-I and II seem to dominate the remaining algorithms. Thus, for $n = 100$ the SLT-II algorithm is the best performer in two out of four measures, for $n = 300$ in all the measures and for $n = 500$ in just one. Under the Sim-II again the SLT-II algorithm is the best performer by getting optimal measures for $n = 100$ in two cases for $n = 300$ in two cases and for $n = 500$ in three cases. Finally, under the Sim-III experiment

† Here, $N(a, b)$ denotes a Gaussian distribution with mean a and variance b. Alternatively, one can impose constrained priors to count for stationarity. Thus, in the case of an $AR(2)$ model we can set the priors: $\beta_1 \sim U(-1, 1)$, $\beta_2 \mid \beta_1 \sim U(-(1 - \beta_1), 1 - \beta_1)$, where $U(a, b)$ denotes a Uniform distribution with an upper and lower bound being the b and a, respectively (see Johnson and Hoeting 2003).

Table 1. Mean MSE, MAE, MTR and MCD measures estimated using simulated time-series data under the MCMC, SLT-Mahalanobis, SLT-I and SLT-II algorithms.

		S = 100				S = 300				S = 500			
	measures	MCMC	SLT-Mahalanobis	SLT-I	SLT-II	MCMC	SLT-Mahalanobis	SLT-I	SLT-II	MCMC	SLT-Mahalanobis	SLT-I	SLT-II
Sim-I	MSE	0.1770	0.1980	0.1806	$\boxed{0.1768}$	0.1835	0.1827	0.1884	$\boxed{0.1768}$	0.1784	0.1789	$\boxed{0.1759}$	0.1795
	MAE	0.3378	0.3535	0.3365	$\boxed{0.3325}$	0.3442	0.3428	0.3457	$\boxed{0.3364}$	0.3370	0.3407	$\boxed{0.3347}$	0.3386
	MTR	0.2164	0.2287	$\boxed{0.2512}$	0.1888	0.2071	0.2111	0.2265	$\boxed{0.2285}$	0.2069	$\boxed{0.2179}$	0.2087	0.2142
	MCD	0.6851	0.7038	$\boxed{0.7205}$	0.6737	0.6688	0.6739	0.6847	$\boxed{0.6976}$	0.6759	0.6821	0.6821	$\boxed{0.6962}$
Sim-II	MSE	0.2910	$\boxed{0.2837}$	0.2941	0.3190	0.3141	$\boxed{0.3095}$	0.3109	0.3299	0.2976	0.3156	0.3000	$\boxed{0.2850}$
	MAE	$\boxed{0.4010}$	0.4056	0.4060	0.4230	$\boxed{0.4024}$	0.4045	0.4060	0.4092	0.4081	0.4121	0.4064	$\boxed{0.3986}$
	MTR	0.2020	0.2543	0.2801	$\boxed{0.3109}$	0.2414	0.2707	0.2689	$\boxed{0.2717}$	0.2408	$\boxed{0.2670}$	0.2560	0.2606
	MCD	0.6555	0.6701	0.6995	$\boxed{0.7415}$	0.6771	0.6939	0.6952	$\boxed{0.7080}$	0.6749	0.6898	0.6872	$\boxed{0.7055}$
Sim-III	MSE	$\boxed{0.1964}$	0.2028	0.1989	0.1972	0.2092	$\boxed{0.1944}$	0.2053	0.2032	0.2031	$\boxed{0.1995}$	0.2065	0.2041
	MAE	0.3587	0.3553	0.3535	$\boxed{0.3526}$	0.3676	$\boxed{0.3511}$	0.3589	0.3565	0.3589	$\boxed{0.3541}$	0.3621	0.3603
	MTR	0.0826	0.0781	0.1174	$\boxed{0.1192}$	0.0719	0.0837	0.0818	$\boxed{0.0902}$	0.0627	0.0725	0.0833	$\boxed{0.0905}$
	MCD	0.5693	0.5731	0.5994	$\boxed{0.6314}$	0.5583	0.5725	0.5798	$\boxed{0.5950}$	0.5568	0.5658	0.5757	$\boxed{0.5781}$

Note: Entries in each table represent the simulated mean estimates. Boxed numbers indicate the favoured estimation case function for each simulated data series.

again the SLT-II algorithm is the best performer for $n = 100$ in three cases for $n = 300$ in three cases and for $n = 500$ in two cases. The MCMC algorithm can only beat the remaining of the algorithms two times and only under the MSE or the MAE measures.

Simulated data analysis illustrates a number of general results:

1. The MCMC algorithm can only beat the remaining of the algorithms two times and only under the MSE or the MAE measures.
2. The use of asymmetric loss functions such as the directional accuracy and the trading rule measures within estimation does not only improve the values of the MTR and MCD measures but in most cases, improves the values of the MSE and the MAE measures.
3. Under the MTR and MCD measures the SLT-II algorithm is the best performer among the algorithms that use directional accuracy and the trading rule measures.

4.2. *Empirical data results*

Here, we will investigate whether certain algorithms with asymmetric loss functions, can improve the forecasting ability of certain autoregressive models when using real exchange rate data. In practice, we will show whether autoregressive models with mean trading returns and/or the mean correct directions as components in loss functions can outperform those without such loss functions when forecasting exchange rate returns.

Our empirical application uses weekly exchange rate series taken from the Federal Reserve Bank of St. Louis. We consider the weekly returns from $n = 521$ observations of the TWEXO, TWEXM and TWEXB indices† , starting from the 8th of June 2008 and ending at the 3rd of June 2019 covering a period of 11 years including that of the Global Financial Crises‡. We examine a forecasting period scenario of $s = 52$ observations starting from the 11th of June 2018 and ending on the 3rd of June 2019 (see Figure 1 for all the return series). In Table 2, we describe several statistical properties of the return series analyzed. These include the mean, standard deviation, minimum, maximum, skewness, excess kurtosis coefficients and the Jarque–Bera normality test result for each series. We observe that the TWEXM and TWEXB return series exhibit Normality violation expressed by leptokurtosis and positive skewness. In Figure 2, we present the autocorrelation function and the partial autocorrelation function of the TWEXO, TWEXM and TWEXB return series. These signify

† The TWEXO, TWEXM and TWEXB indices are all Trade Weighted U.S. Dollar Weekly Indices. The TWEXO index includes other important trading partners and goods with January 1997 = 100, the TWEXM index includes major currencies and goods with March 1973 = 100 and the TWEXB index includes broad and goods, with January 1997 = 100. All these indices are weekly and not seasonally adjusted.

‡ Here, we need to point out that the use of weekly data instead of daily or intraday ones stems from the fact that the latter series do not usually have a first-order serial dependence. In other words, the observed series are white noise processes and conditional correlation is observed in the second order (in variance).

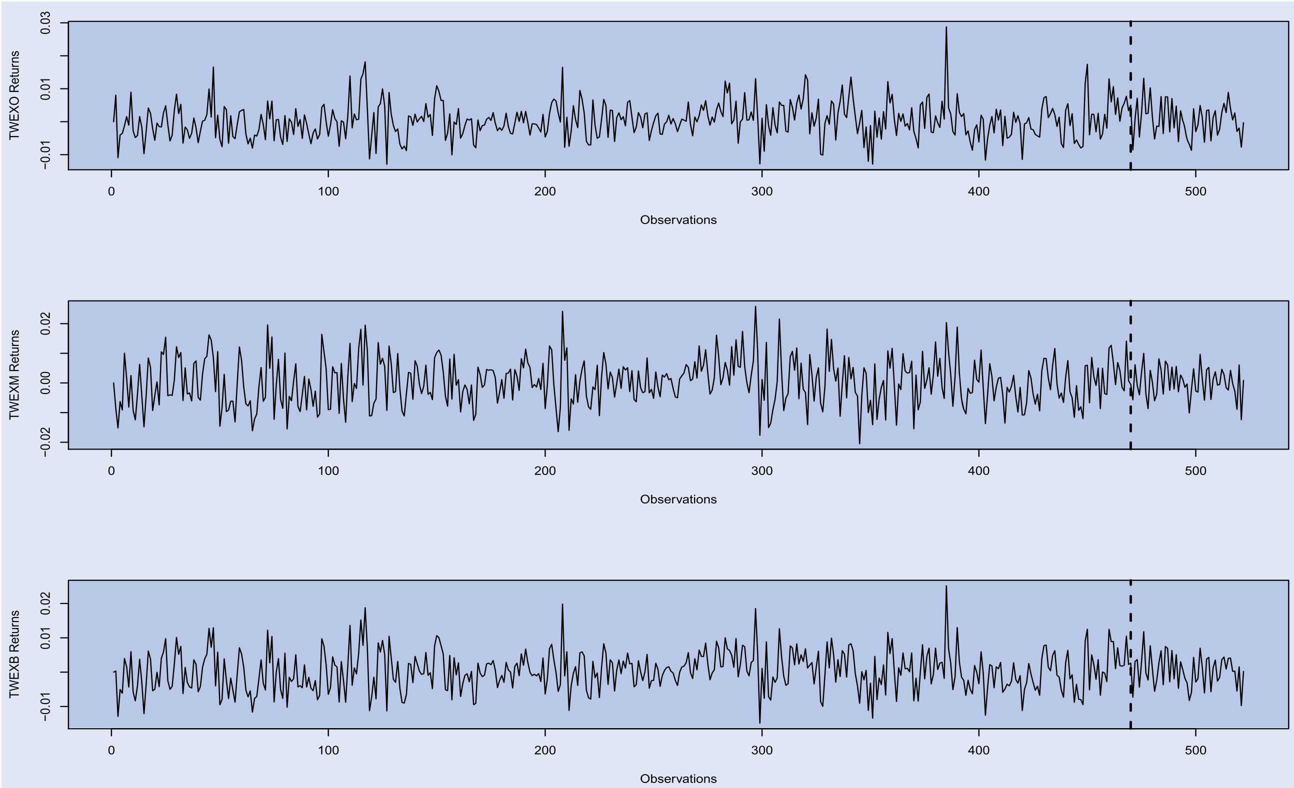

Figure 1. Plots of the whole TWEXO, TWEXM and TWEXB return data series and the forecasting period of 52 observations.

Table 2. Summary statistics of Trade Weighted U.S. Dollar Index returns.

	Statistics for data set						
	mean	sd	skewness	excess kurtosis	min.	max.	Jarque–Bera test
TWEXO	0.0300	0.7583	0.1911	−0.0324	−2.05	2.584	3.2197(0.1999)
TWEXM	0.0411	0.5336	0.5641	1.6758	−1.291	2.877	86.7819(< 2.2e–16)
TWEXB	0.0362	0.5673	0.3007	0.5725	−1.487	2.515	14.5850(0.00068)

a first-order correlation of order one. This preliminary results can be seen as a proxy for the type of autoregressive model we are going to use for the return data series.

Using the learning period of $n-s$ observations for each series, we derive posterior simulation results using a Gaussian $AR(1)$ model under the MCMC, the SLT-Mahalanobis, the SLT-I, the SLT-II and the SLT-mix algorithms. From the 60 000 draws we generate, the first 30 000 are discarded (burn-in) and we are left with around 30 000 draws which are then used for summary statistics.

Representative posterior simulation results using a normal $AR(1)$ model estimated under the MCMC, the SLT-Mahalanobis, the SLT-I, the SLT-II and the SLT-mix algorithms using the TWEXO return series are reported in Table 3. These depict the mean, standard deviation, median, 95% credible interval, Monte Carlo standard error (MC S.E.) the Heidelberg–Welch convergence diagnostics (Heidelberger and Welch 1983), the inefficiency factor and the correlation function of all estimates†. Posterior results are in line with the literature on parametric estimates in weekly exchange rate data (Hong and Lee 2003) and with the preliminary analysis using autocorrelation functions. Also, the reported mean estimates are all of high level of significance with the exception of constant term in all the competing algorithms. The reported medians seem to coincide with the mean estimates in most of the cases. Also, the 95% credible interval seems within realistic values for all estimated parameters. The Heidelberg–Welch convergence diagnostics tool, using the Cramer–von-Mises statistic, tests the null hypothesis that the simulated values come from a stationary distribution. The p-value results show considerable stationarity in all posterior simulation cases. Concerning the inefficiency factor (IF) reported, the $AR(1)$ model estimated under the MCMC show excellent efficiency in the range of $4.639 \sim 4.705$. It implies that we need to sample from the MCMC algorithm about 4.705 as many times as the hypothetical uncorrelated sampler to obtain the same variance of the posterior sample mean. As regards the SLT-Mahalanobis, the SLT-I, the SLT-II and the SLT-mix algorithms we have small to moderate IF values

† Here, we need to point out that the use a Student-t $AR(1)$ model has given us close to normal mean posterior results for the degrees of freedom parameter. Also, the MSFE, MAFE, MTFR and MCFD forecasting measures are not better than the corresponding ones derived from the normal $AR(1)$ model.

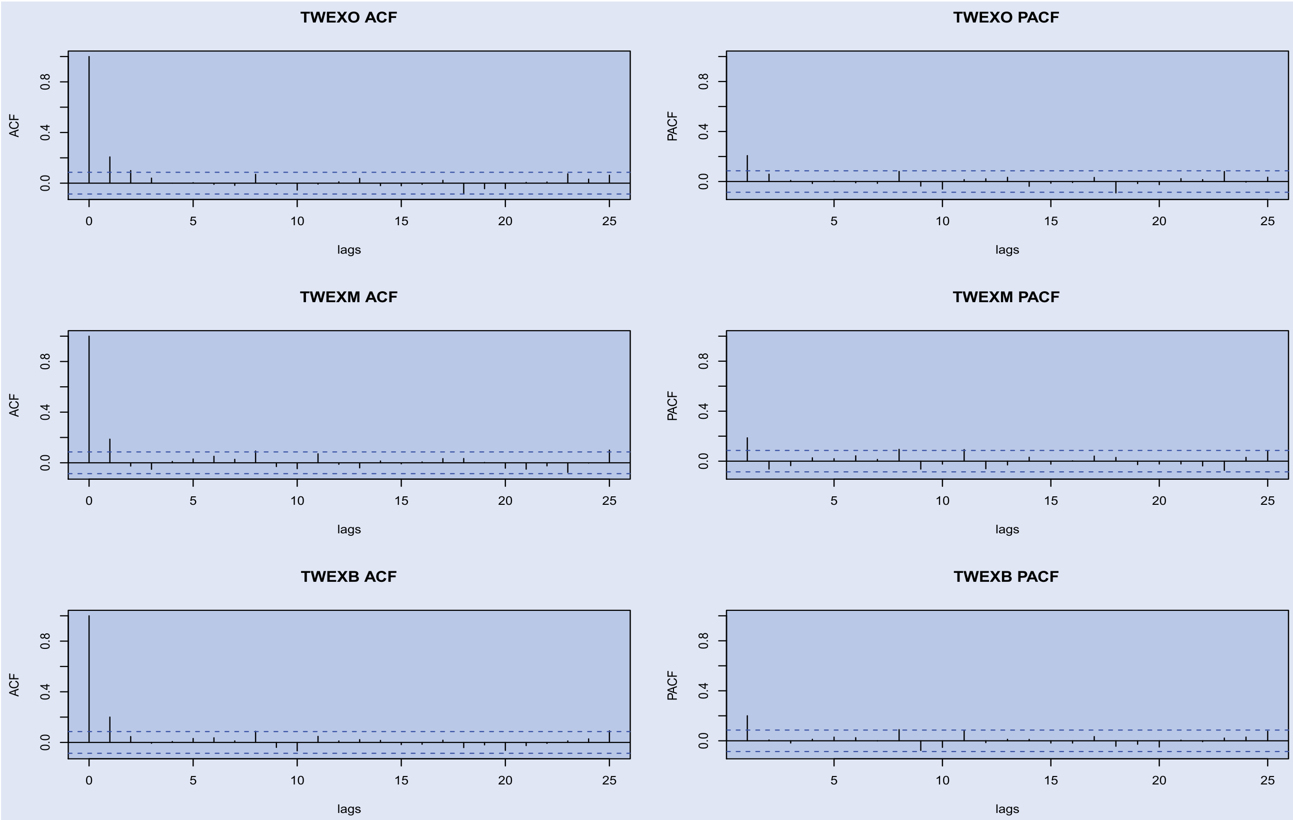

Figure 2. Autocorrelation (ACF) and Partial Autocorrelation functions (PACF) of the TWEXO, TWEXM and TWEXB return data series.

with ranges 8.212 ∼ 11.479, 4.630 ∼ 7.746, 7.133 ∼ 10.606 and 8.074 ∼ 21.092 , respectively†. Here, although we have higher IF compared to the initial MCMC algorithm, the results show moderate inefficiency for the general MCMC practice. As Kim *et al.* (1998) point out, the IF is a useful, but by no means, the only diagnostic for measuring how well the chain mixes. An additional diagnostic is the correlation function also reported in Table 3. Our results show moderate correlation values among the estimated parameters.

To graphically demonstrate the above results, Figures 3 and 4 demonstrate the plots, histograms and autocorrelation functions of the posterior parametric estimates in the MCMC and SLT-mix algorithms respectively applied for the TWEXO return series. In these results we see considerable posterior convergence. The autocorrelation function values appear to have statistical significance, but they soon fall. Also, Figures 5 and 6 present the trace plot of all the 60 000 posterior draws for the MCMC and SLT-mix algorithms. The results show a nearly instant convergence to a specific posterior distribution for the MCMC algorithm and a slower convergence in the SLT-mix algorithm. We need to state that, we have generated similar results to the above posterior simulation when using the TWEXM and TWEXB return series under the MCMC, SLT-Mahalanobis, SLT-I, the SLT-II and the SLT-mix algorithms.

Based on the aforementioned representative posterior simulation results, Tables 4–6 report the one-, two-, three- and four-step-ahead MSFE, MAFE, MTFR and MCFD measures under the four competing algorithms plus an RW model which is set as benchmark. These results are taken using the TWEXO, TWEXM and TWEXB return series. Initially, we observe that the MAFE results are all equal across algorithms for each return series case. Also, the MTFR and MCFD results for both SLT-II and SLT-mix algorithms are identical due the fact that the latter algorithm put nearly all posterior weights on the ψ_2 auxiliary statistic. Now, results for the TWEXO return series reported in Table 4 show that the SLT-II algorithm is ranked first in ten cases with the SLT-mix algorithm following in six cases. The SLT-Mahalanobis and the SLT-I algorithms top scores two times in MAFE and MTFR at the two-step-ahead forecasts. The MCMC algorithm and the RW model do not have any top scoring case. In Table 5 we present the results under the TWEXM return series. These show that the SLT-II algorithm is ranked first in seven cases with the SLT-Mahalanobis algorithm following in four cases while the SLT-mix algorithm is ranked third with three top scoring cases. The remaining MCMC and SLT-I algorithms together with the RW model do not have any top scoring case. In Table 6, we present the results under the TWEXB return series. Here, the SLT-II algorithm is ranked first in nine cases with the SLT-mix to follow in seven cases and the SLT-I and SLT-Mahalanobis algorithms are ranked third with two top scoring cases. Here, the RW model gains one top scoring case

† The inefficiency factor (IF) evaluated for each simulated parameter is defined as $1 + \sum_{s=1}^{\infty} p_s$ where p_s is the sample autocorrelation function at lag s calculated from the sample value (see Chib 2001). It represents the ratio of the simulation variance to the variance of the sampling mean from the hypothetical sampler which draws independent random variables from the posterior.

Table 3. Posterior simulation results for TWEXO data using the $AR(1)$ model under the MCMC, SLT-Mahalanobis, SLT-I, SLT-II, SLT-mix algorithms.

	mean	sd.	median	95% CI	MC S.E.	IF	Heidel.diag	Corr			
					Method: MCMC						
β_0	0.0003	0.0002	0.0003	$(-0.0001, 0.0007)$	1.742e–6	4.705	0.575			−0.076301178	0.008242579
β_1	0.2280	0.0447	0.2284	$(0.1402, 0.3174)$	3.166e–4	4.955	0.729				0.001686165
$\log\tau$	−5.2455	0.0329	−5.2460	$(-5.3108, -5.1825)$	2.328e–4	4.639	0.200				
					Method: SLT-Mahalanobis						
β_0	0.00030	0.0002	0.0003	$(-0.0001, 0.0007)$	2.358e–6	8.505	0.974			0.25684522	−0.09318146
β_1	0.2286	0.0440	0.2291	$(0.1442, 0.3154)$	4.408e–4	11.479	0.322				0.015755202
$\log\tau$	−5.2462	0.0333	−5.2477	$(-5.3100, -5.1822)$	3.336e–4	8.212	0.193				
					Method: SLT-I						
β_0	0.0003	0.0002	0.0002	$(-0.0001, 0.0007)$	2.348e–6	7.746	0.6329			0.30914518	0.05751349
β_1	0.2278	0.0449	0.2289	$(0.1399, 0.3159)$	4.495e–4	5.589	0.2330				0.03019451
$\log\tau$	−5.2460	0.0332	−5.2462	$(-5.3091, -5.1797)$	3.328e–4	4.630	0.0509				
					Method: SLT-II						
β_0	0.0002	0.0001	0.000213	$(0.00002, 0.0004)$	1.429e–6	9.584	0.1020			0.12792320	−0.002028129
β_1	0.2257	0.0452	0.225149	$(0.1352, 0.3199)$	4.526e–4	7.133	0.0653				−0.022402335
$\log\tau$	−5.2468	0.0332	− 5.24694	$(-5.30855, -5.1805)$	3.320e–4	10.606	0.1679				
					Method: SLT-mix						
β_0	−0.0001	3.519e–5	−0.0001	$(-1.847e--4, -5.266e--5)$	2.032e–7	8.074	0.197	−0.414402442	−0.009013941	0.0008035395	−0.0008034718
β_1	0.2289	4.458e–2	0.2278	$(0.1426, 0.3170)$	2.574e–4	9.194	0.314		−0.015348256	−0.0031882660	0.0031882406
$\log\tau$	−5.1684	3.236e–2	−5.1691	$(-5.231, -5.102)$	1.868e–4	10.768	0.867			−0.0320626534	0.0320626988
π_1	0.0035	3.760e–3	0.0023	(7.785e–5, 0.0139)	2.171e–5	21.092	0.103				
π_2	0.9964	3.760e–3	0.9976	$(0.9860, 0.9999)$	2.171e–5	21.092	0.103				

Note: Entries in this table represent the sampled mean, the standard deviation (s.d.), the median, the 95% credible interval (95% CI), the Monte Carlo S.E. (MC S.E.), the p-value of the Heidelberger and Welch convergence diagnostic test (Heidel.diag), the inefficiency factor (IF) and the correlation function (Corr) of stationarity of posterior parameter estimates.

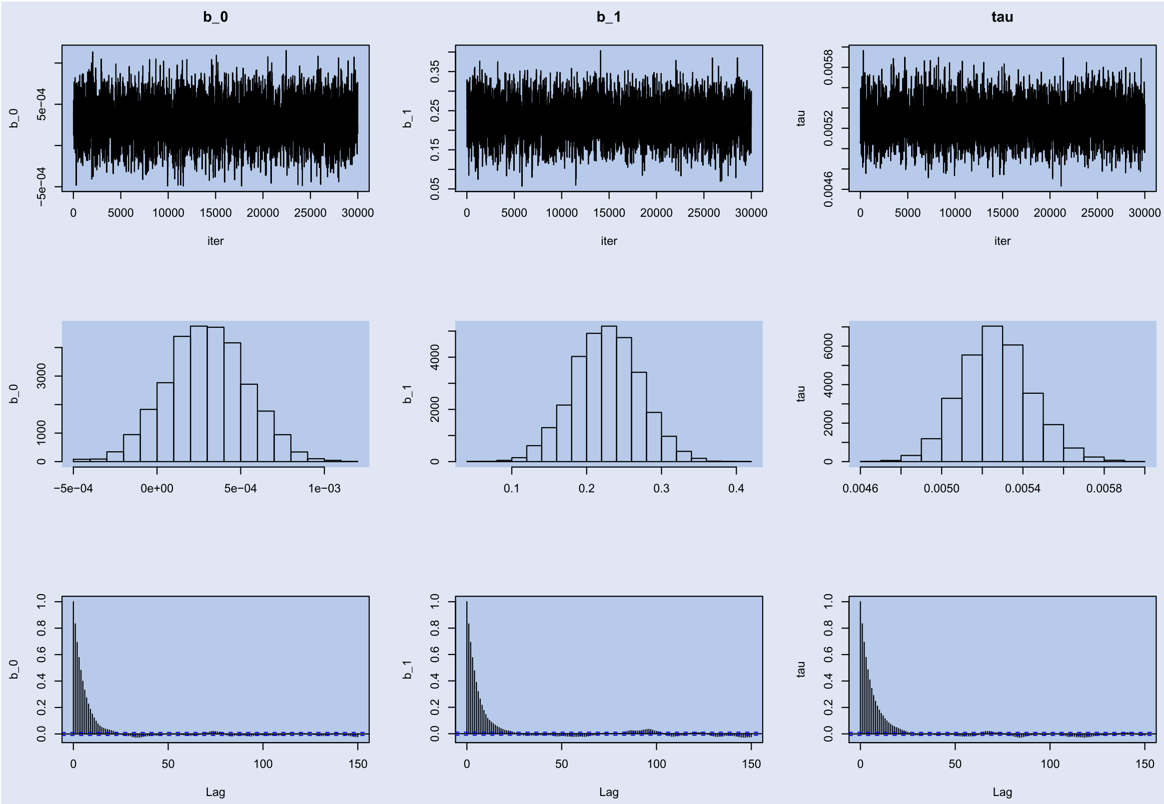

Figure 3. Plots, histograms and autocorrelation functions of the parameters' posterior simulations using the MCMC algorithm for the $AR(1)$ model for TWEXO data series. These represent, from the top to the bottom, the parameters β_0, β_1 and τ.

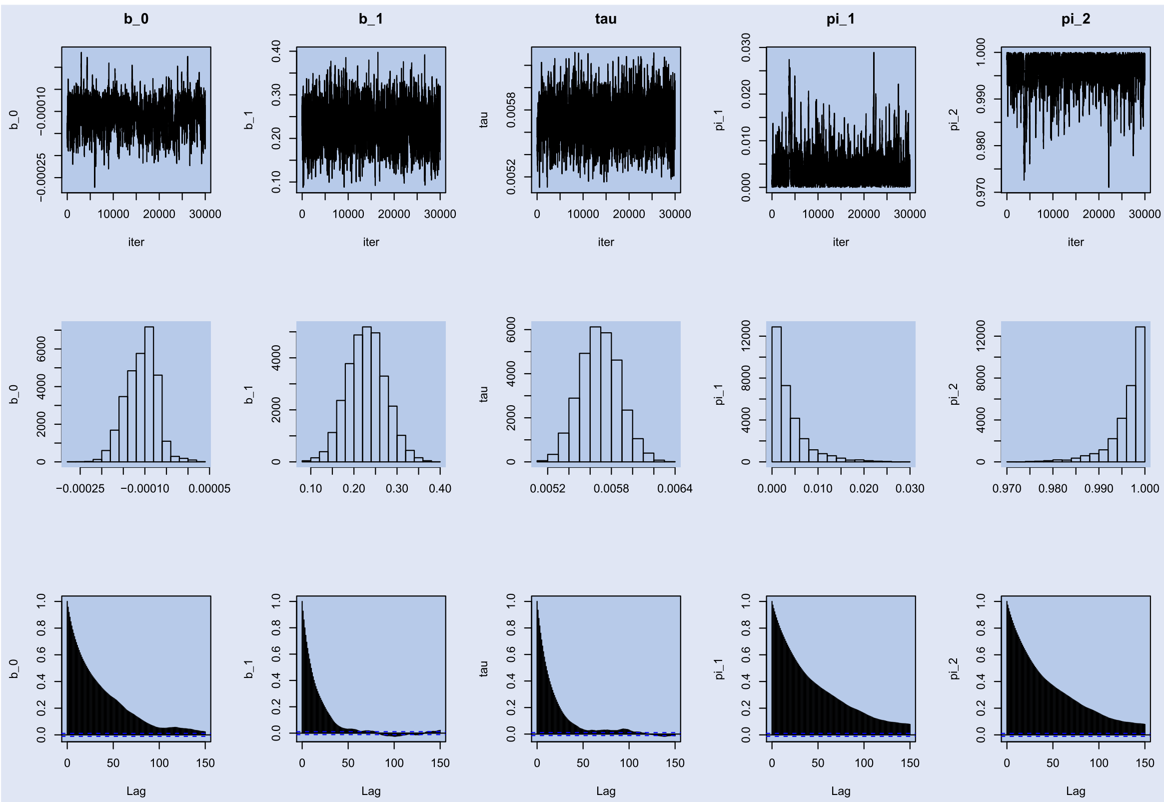

Figure 4. Plots, histograms and autocorrelation functions of the parameters' posterior simulations using the SLT-mix algorithm for the $AR(1)$ model for the TWEXO data series. These represent, from top to bottom, the parameters β_0, β_1, τ, π_1 and π_2.

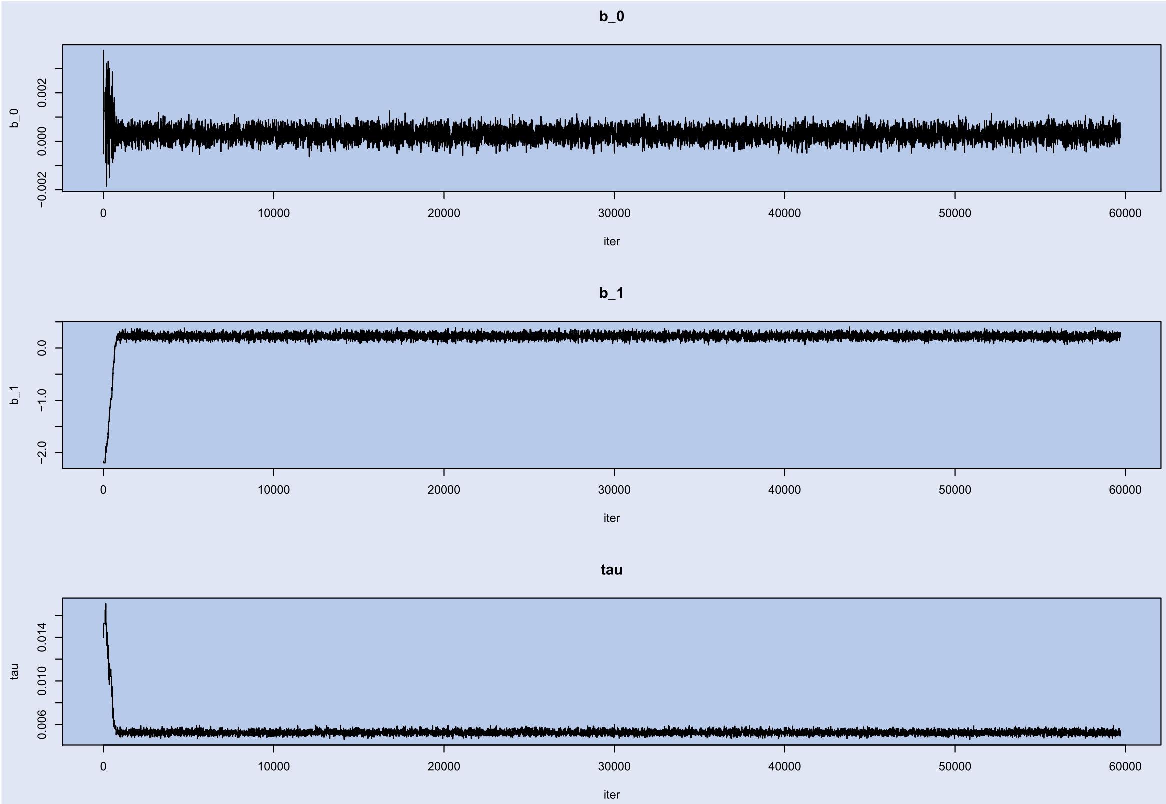

Figure 5. Plots of the parameters' posterior simulations using the MCMC algorithm for the $AR(1)$ model for the TWEXO data series. These represent, from top to bottom, the parameters β_0, β_1 and τ.

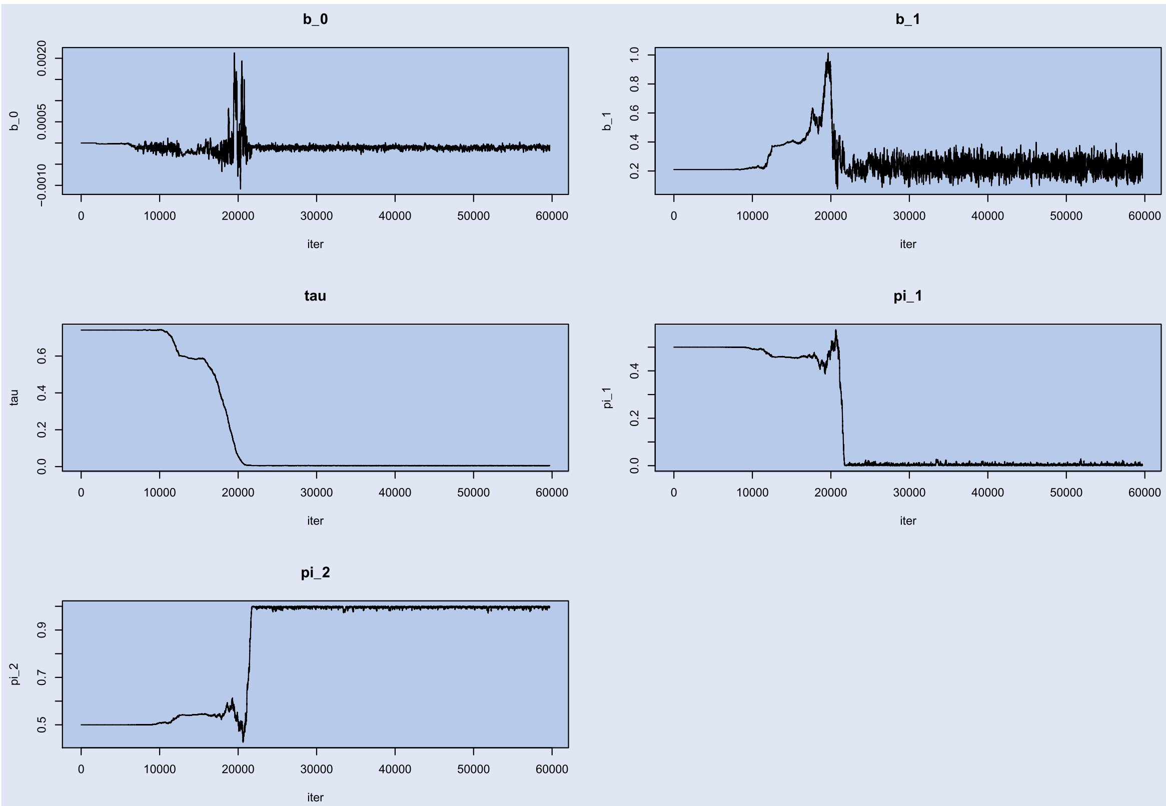

Figure 6. Plots of the parameters' posterior simulations using the SLT-mix algorithm for the $AR(1)$ model for the TWEXO data series. These represent, from top to bottom, the parameters β_0, β_1, τ, π_1 and π_2.

Table 4. Mean MSFE, MAFE, MTFR and MCFD measures estimated using TWEXO time-series data estimated under the MCMC, SLT-Mahalanobis, SLT-I, SLT-II, SLT-mix algorithms and the RW model.

measures	RW	MCMC	SLT-Mahalanobis	SLT-I	SLT-II	SLT-mix	RW	MCMC	SLT-Mahalanobis	SLT-I	SLT-II	SLT-mix
	1-step-ahead						**2-step-ahead**					
MSFE	5.165e–5	2.712e–5	2.711e–5	2.714e–5	$\boxed{2.702e{-}5}$	2.709e–5	5.086e–5	2.571e–5	2.570e–5	2.573e–5	$\boxed{2.557e{-}5}$	2.612e–5
MAFE	0.0061	0.0043	0.0043	0.0043	0.0043	0.0043	0.0057	0.0043	$\boxed{0.0042}$	$\boxed{0.0042}$	0.0043	0.0043
MTFR	0.0001	0.0006	0.0006	0.0006	0.0006	0.0006	0.0001	0.0007	$\boxed{0.0008}$	$\boxed{0.0008}$	$\boxed{0.0008}$	$\boxed{0.0008}$
MCFD	0.5192	0.5576	0.5576	0.5576	0.5576	0.5576	0.5	0.5659	0.5668	0.5664	$\boxed{0.5686}$	$\boxed{0.5686}$
	3-step-ahead						**4-step-ahead**					
MSFE	3.855e–5	2.610e–5	2.609e–5	2.612e–5	$\boxed{2.597e{-}5}$	2.614e–5	5.453e–5	2.573e–5	2.572e–5	2.575e–5	$\boxed{2.563e{-}5}$	2.579e–5
MAFE	0.0048	0.0043	0.0043	0.0043	0.0043	0.0043	0.0063	0.0042	0.0042	0.0042	0.0042	0.0042
MTFR	0.0007	0.0006	0.0007	0.0007	$\boxed{0.0008}$	$\boxed{0.0008}$	0.0003	0.0005	0.0005	0.0005	$\boxed{0.0007}$	$\boxed{0.0007}$
MCFD	0.4961	0.5510	0.5529	0.5517	$\boxed{0.5600}$	$\boxed{0.5600}$	0.4615	0.5409	0.5427	0.5413	$\boxed{0.5510}$	$\boxed{0.5510}$

Note: Entries in each table represent the simulated mean estimates. Boxed numbers indicate the favoured estimation case function for each data series.

Table 5. Mean MSFE, MAFE, MTFR and MCFD measures estimated using TWEXM time-series data estimated under the MCMC, SLT-Mahalanobis, SLT-I, SLT-II, SLT-mix algorithms and the RW model.

measures	RW	MCMC	SLT-Mahalanobis	SLT-I	SLT-II	SLT-mix	RW	MCMC	SLT-Mahalanobis	SLT-I	SLT-II	SLT-mix
	1-step-ahead						**2-step-ahead**					
MSFE	6.191e–5	2.919e–5	2.910e–5	2.922e–5	$\boxed{2.902e{-}5}$	2.935e–5	6.306e–5	2.929e–5	2.899e–5	2.912e–5	$\boxed{2.894e{-}5}$	2.926e–5
MAFE	0.0066	0.0045	0.0045	0.0045	0.0045	0.0045	0.0066	0.0045	0.0045	0.0045	0.0045	0.0045
MTFR	−0.0008	6.710e–5	5.716e–5	7.168e–5	$\boxed{0.0001}$	$\boxed{0.0001}$	−0.0015	0.0001	0.0001	0.0001	0.0001	0.0001
MCFD	0.4038	0.4931	$\boxed{0.4941}$	0.4926	0.4807	0.4807	0.3076	0.4949	$\boxed{0.4954}$	0.4946	0.4901	0.4901
	3-step-ahead						**4-step-ahead**					
MSFE	3.968e–5	2.868e–5	2.859e–5	2.871e–5	2.850e–5	2.883e–5	3.293e–5	2.915e–5	2.914e–5	2.917e–5	2.898e–5	2.930e–5
MAFE	0.0047	0.0044	0.0044	0.0044	0.0044	0.0044	0.0046	0.0045	0.0045	0.0045	0.0045	0.0045
MTFR	−0.0009	9.986e–5	9.197e–5	8.510e–5	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002
MCFD	0.4153	0.4892	$\boxed{0.4900}$	0.4888	0.48	0.48	0.4538	0.4944	$\boxed{0.4949}$	0.4942	0.4897	0.4897

Note: Entries in each table represent the simulated mean estimates. Boxed numbers indicate the favoured estimation case function for each data series.

in the MCFD measure. Again, the MCMC algorithm does not have any top scoring case. As in the previous data series for all step-ahead cases the MAFE measure remains constant for the four competing algorithms.

Finally, in order to demonstrate the stability of the $\boldsymbol{w}$ posterior results, we adopt four alternative Exponential prior scenarios, as follows:

$$\text{Prior I} : \boldsymbol{w} \sim \text{Dir}(1, 1),$$

$$\text{Prior II} : \boldsymbol{w} \sim \text{Dir}(.75, .75),$$

$$\text{Prior III} : \boldsymbol{w} \sim \text{Dir}(.5, .5),$$

$$\text{Prior IV} : \boldsymbol{w} \sim \text{Dir}(.25, .25),$$

adopted for the TWEXO return series under the SLT-mix algorithm. Posterior simulation results presenting the histograms of the $\boldsymbol{w}$ parameter appear in Figure 7. These demonstrate that alternative prior scenario concerning the prior of the $\boldsymbol{w}$ parameter have a marginal deviation in the posterior mean that they generate.

Real data analysis demonstrate the following results:

1. The SLT-II algorithm outperforms all other algorithms when it comes to access the forecasting ability of the Gaussian $AR(1)$ autoregressive model applied to weekly exchange rate returns.
2. The choice of algorithm is mostly stable when it comes to comparing one-, two-, three- and four-step ahead forecasting measures under alternative data series.
3. The introduction of algorithms with asymmetric loss functions such as directional accuracy and the trading rules improve not only the forecasting measures such as MCFD and MTFR but also standard forecasting measures such as the MSFE.

5. Concluding remarks

In this paper, we have proposed a Bayesian framework to estimate exchange rate forecast models when one considers maximizing profits criteria. The question of maximizing profits has long been an issue when one assesses the forecasting performance of an exchange rate model. However, the incorporation of profit maximization strategies within estimation creates a likelihood function intractability problem. To resolve this inference problem, we have introduced Laplace-type estimators using asymmetric components.

Our inference strategy has been based on LT estimators within a quasi-Bayesian MCMC algorithm. We have assessed this approach for its effect in forecasting exchange rate returns in comparison with standard Bayesian approaches based on a Normal likelihood function maximization which renders error minimization. This experiment is applied for various simulated data scenarios and real financial data forecasting periods.

Results are encouraging for forecasting models with asymmetric metric measures. These demonstrate that when using the SLT-II algorithm in autoregressive models, one can optimally forecast exchange rate returns when the data generating process is pre-specified. In most of the simulation results, the

Table 6. Mean MSFE, MAFE, MTFR and MCFD measures estimated using TWEXB time-series data estimated under the MCMC, SLT-Mahalanobis, SLT-I, SLT-II, SLT-mix algorithms and the RW model.

measures	RW	MCMC	SLT-Mahalanobis	SLT-I	SLT-II	SLT-mix	RW	MCMC	SLT-Mahalanobis	SLT-I	SLT-II	SLT-mix
			1-step-ahead						**2-step-ahead**			
MSFE	4.802e–5	2.387e–5	2.383e–5	2.391e–5	$\boxed{2.369e{-5}}$	2.881e–5	4.865e–5	2.304e–5	2.299e–5	2.307e–5	$\boxed{2.286e{-5}}$	2.946e–5
MAFE	0.0059	0.0041	0.0041	0.0042	0.0042	0.0042	0.0057	0.0041	0.0041	0.0041	0.0041	0.0041
MTFR	−0.0008	0.0004	0.0004	0.0004	0.0004	0.0004	−0.0009	0.0005	0.0005	0.0005	$\boxed{0.0006}$	$\boxed{0.0006}$
MCFD	0.4230	0.5767	0.5765	0.5768	$\boxed{0.5769}$	$\boxed{0.5769}$	0.4038	0.5761	0.5735	0.5759	$\boxed{0.5882}$	$\boxed{0.5882}$
			3-step-ahead						**4-step-ahead**			
MSFE	3.220e–5	2.310e–5	$\boxed{2.305e{-5}}$	2.313e–5	2.984e–5	2.416e–5	3.771e–5	2.342e–5	$\boxed{2.337e{-5}}$	2.345e–5	2.970e–5	2.371e–5
MAFE	0.0043	0.0041	0.0041	0.0041	.0041	0.0041	0.0050	0.0041	0.0041	0.0041	0.0041	0.0041
MTFR	0.0001	0.0004	0.0003	$\boxed{0.0005}$	$\boxed{0.0005}$	$\boxed{0.0005}$	0.0002	0.0003	0.0002	$\boxed{0.0004}$	$\boxed{0.0004}$	$\boxed{0.0004}$
MCFD	0.5346	0.5606	0.5597	0.5617	$\boxed{0.5700}$	$\boxed{0.5700}$	0.5714	0.5517	0.5512	0.5525	$\boxed{0.5714}$	$\boxed{0.5714}$

Note: Entries in each table represent the simulated mean estimates. Boxed numbers indicate the favoured estimation case function for each data series.

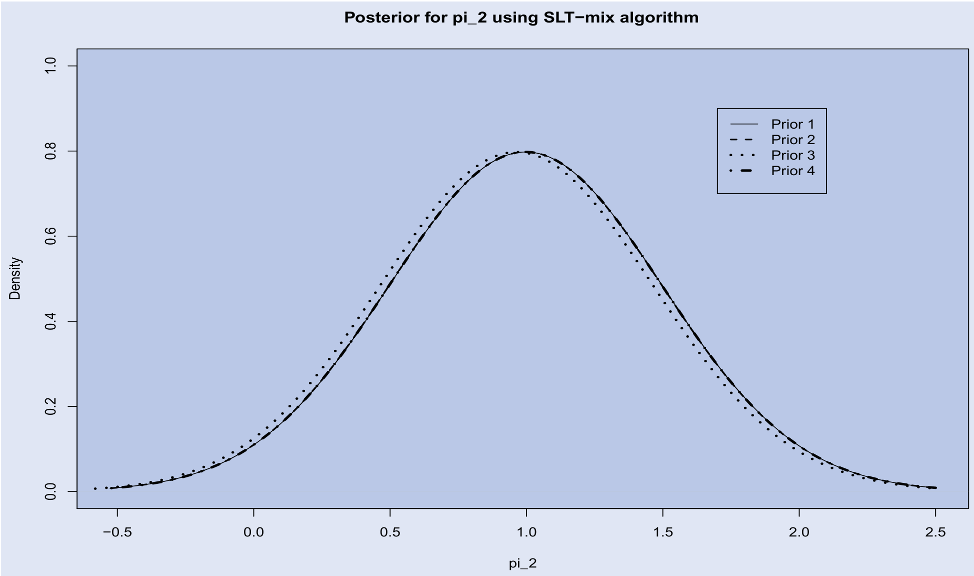

Figure 7. Density plot of the π_2 posterior results under alternative prior densities for the SLT-mix algorithm using the TWEXO data series.

introduction of profit maximization measures outperforms the existing inference strategy. In real data analysis, the dominance of the new approach is overwhelmingly in support of the SLT estimation approach when forecasting weekly exchange rate returns.

Finally, we need to state that the present approach can be easily applied to utility-based loss functions when modeling exchange rate dynamics (see West *et al.* 1993). Also, one can compare the present Laplace-type estimation approach with contemporary Bayesian approaches such as the Approximate Bayesian Computation approach or others based on a synthetic likelihood function (see Wood 2010, Sisson 2011).

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Boothe, P. and Glassman, D., Exchange rate forecasting models: Accuracy versus profitability. *Int. J. Forecast.*, 1987, **3**, 65–79.

Brook, C., Linear and nonlinear (non-)forecastability of high-frequency exchange rates. *J. Forecast.*, 1997, **16**, 125–145.

Chernozhukov, V. and Hong, H., An MCMC approach to classical estimation. *J. Econom.*, 2003, **115**, 293–346.

Chib, S., Markov chain Monte Carlo methods: Computation and inference. In *Handbook of Econometrics*, edited by J.J. Heckman and E. Leamer, Volume 5, 57, pp. 3569–3649, 2001 (Elsevier: North Holland).

Christoffersen, P. and Diebold, F.X., Optimal prediction under asymmetric loss. *Econom. Theory*, 1997, **13**, 808–817.

Costantini, M., Crespo Cuaresma, J. and Hlouskova, J., Forecasting errors, directional accuracy and profitability of currency trading: The case of EUR/USD exchange rate. *J. Forecast.*, 2016, **35**, 652–668.

Cuaresma, J.C., Fortin, I, and Hlouskova, J., Exchange rate forecasting and the performance of currency portfolios. *J. Forecast.*, 2018, **37**, 519–540.

Forneron, J-J. and Ng, S., The ABC of simulation estimation with auxiliary statistics. *J. Econom.*, 2018, **205**, 112–139.

Gencay, R., The predictability of security returns with simple technical trading rules. *J. Empir. Finance*, 1998, **5**, 347–359.

Clements, M.P. and Hendry, D.F., On the limitations of comparing mean square forecast errors. *J. Forecast.*, 1993, **12**, 617–637.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B., *Bayesian Data Analysis*, 2004 (Chapman and Hall: London).

Granger, C.W.J., Prediction with the generalized cost of error function. *Operat. Res. Q.*, 1969, **20**, 199–207.

Granger, C.W.J., Outline of forecast theory using generalized cost functions. *Spanish Econom. Rev.*, 1999, **1**, 161–173.

Gu, S., Kelly, B. and Xiu, D., Empirical asset pricing via machine learning. *Rev. Financ. Stud.*, 2020, **33**, 2223–2273.

Hansen, L.P., Large sample properties of generalized method of moments estimators. *Econometrica*, 1982, **50**, 1029–1054.

Hong, Y. and Lee, T.H., Inference on predictability of foreign exchange rates via generalized spectrum and nonlinear time series models. *Rev. Econom. Stat.*, 2003, **85**, 1048–1062.

Haario, H., Saksman, E. and Tamminen, J., An adaptive metropolis algorithm. *Bernoulli*, 2001, **7**, 223–242.

Heidelberger, P. and Welch, P.D., Simulation run length control in the presence of an initial transient. *Oper. Res.*, 1983, **31**, 1109–1144.

Jiang, W. and Turnbull, B., The indirect method: Inference based on intermediate statistics – A synthesis and examples. *Statist. Sci.*, 2004, **19**(2), 239–263.

Johnson, D. and Hoeting, J., Autoregressive models for capture-recapture data: A Bayesian approach. *Biometrics*, 2003, **59**, 341–350.

Kim, S., Shephard, N. and Chib, S., Stochastic volatility: Likelihood inference and comparison with ARCH models. *Rev. Econom. Stud.*, 1998, **65**, 361–393.

Krager, H. and Kugler, P., Nonlinearity in foreign exchange markets: A different perspective. *J. Int. Money. Finance*, 1993, **12**, 195–208.

Lisi, F. and Medio, A., Is a random walk the best exchange rate predictor?. *Int. J. Forecast.*, 1997, **13**, 255–267.

Lise, J., *et al.*, Matching, sorting and wages. *Rev. Econom. Dynam.*, 2016, **19**, 63–87.

Meese, R. and Rogoff, K., Empirical exchange rate models of the seventies: Do they fit out of sample?. *J. Int. Econ.*, 1983, **14**, 3–24.

Mark, N.C., Exchange rates and fundamentals: Evidence on long-horizon prediction. *Am. Econom. Rev.*, 1995, **85**, 201–218.

Mark, N.C. and Sul, D., Nominal exchange rates and monetary fundamentals: Evidence from a small post-Bretton woods panel. *J. Int. Econ.*, 2001, **53**, 29–52.

Menkhoff, L. and Taylor, M.P., The obstinate passion of foreign exchange professionals: Technical analysis. *J. Econ. Lit.*, 2007, **45**, 936–972.

Rousseau, J. and Mengersen, K., Asymptotic behaviour of the posterior distribution in overfitted mixture models. *J. R. Statist. Soc.: Seri. B*, 2011, **73**, 689–710.

Sisson, S.A., Likelihood-free MCMC. In *Handbook of Markov Chain Monte Carlo*, edited by S. Brooks, A. Gelman, G. Jones and X.-L. Meng, 2011 (CRC Press).

Vihola, M., Robust adaptive metropolis algorithm with coerced acceptance rate. *Statist. Comput.*, 2012, **22**, 997–1008.

West, K.D., Edison, H.J. and Cho, D., A utility based comparison of some models of exchange rate volatility. *J. Int. Econ.*, 1993, **35**, 23–45.

Wood, S.N., Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 2010, **466**, 1102–1104.

Wright, J.H., Bayesian model averaging and exchange rate forecasts. *J. Econom.*, 2008, **146**, 329–341.