Forecasting the intraday market price of money[*]

Andrea Monticini[†]

Università Cattolica - Milano

Francesco Ravazzolo[‡]

Norges Bank and BI Norwegian Business School

June 10, 2013

Abstract

Central banks' operations and efficiency arguments would suggest that the intraday interest rate should be set to zero. However, a liquidity crisis introduces frictions related to news, which can cause an upward jump of the intraday rate. This paper documents that these dynamics can be partially predicted during turbulent times. Long memory approaches or a combination of them to account for model uncertainty outperform random walk, autoregressive and moving average benchmarks in terms of point and density forecasting. The relative accuracy is higher when the full distribution is predicted. We also document that such statistical accuracy can provide economic gains in investment strategies based on lending in the intraday market.

Keywords: interbank market, intraday interest rate, forecasting, density forecasting, policy tools.

JEL codes: C22, C53, E4, E5.

1 Introduction

An explicit market for intraday loans does not exist. However, we can observe an intraday interest rate by the spread between the interest rates on two overnight loans delivered at different times within the same day (provided they are repaid at the same time next day). Furfine (2001), Baglioni and Monticini (2008), Baglioni and Monticini (2010) and Jurgilas and Zikes (2011) find empirical evidence for the existence of such a market in the US, in EU and in the UK. That market is partially unexplored and rich in aspects worth to analyzing: efficiency, the microstructure, arbitrage opportunities and so on. A zero level for the intraday interest spread, and therefore a flat intraday pattern for the rate, should be set for at least two reasons, as discussed in Baglioni and Monticini (2010). The first one is related to the role of the policymakers. A positive intraday spread might induce individual banks to delay payments, imposing a negative externality on the banking system, see Angelini (1998), Bech and Garratt (2003), Mills and Nesmith (2008), Martin and McAndrews (2008), and consequently increasing the operational risk in the payment systems, see e.g. FED (2006) and FED (2007). The second one refers to the role of money as a medium of exchange. The intraday rate is just a transaction cost to settle debt which should be minimized, see Zhou (2000). Moreover, a zero level for the intraday spread provides an insurance for consumers against liquidity shocks (see e.g. Martin (2004) and Bhattacharya et al. (2007)). Central banks' daily market operations seem to support these arguments and indeed central banks often provide free daylight credit to the banking system. For example, the Eurosystem does not charge any fee on daylight overdrafts, and cash settlements must be cleared late in the afternoon and not early.

Baglioni and Monticini (2008) show that thanks to central bank interventions the intraday markets function fairly well in normal times with interest rates close to zero. However, liquidity crises change the functioning of the markets enormously. Baglioni and Monticini (2010) find that the ability of central banks to reduce the market price of intraday liquidity partially vanishes during crises. Baglioni and Monticini (2012) build up a simple model to explain why in normal times the only friction in action is related to settlement procedures and to the cost of central bank intraday credit (see the above references and VanHoose (1991)), while a liquidity crisis introduces a second component related to the chance of an upward jump of the intraday rate within the day due to some news (e.g., liquidity problems for some players in the market). Furthermore, Brunetti et al. (2011) find that central bank interventions during the recent crisis introduced uncertainty and pushed up the intraday money market rate further than (negative) economic news. Durré and Nardelli (2008) show that money market rates have been more sensitive to fine-tuning operations in recent

years and Brunetti et al. (2011) claim that central banks either did not fully grasp the crowding effect, meaning commercial banks replace money market liquidity with central bank liquidity so that market conditions did not improve, see Heider et al. (2009), or consistently underestimated funding liquidity demand.

Using a database from the e-Mid market similar to Baglioni and Monticini (2012), we document that positive intraday spreads are often observed in the euro area market from January 2007 to April 2009, when our database stops. Moreover, we show that the dynamics of the series over our sample period are not random, but both in-sample and out-of-sample predictability seems to exist, suggesting positive rates are not just due to measurement errors. In particular, our results find that long memory approaches, represented by ARFIMA(p,d,q) models where d is the order of integration, provide superior fit-measures and statistically outperforms, in terms of point and density forecasting, random walk, autoregressive and moving average models during periods of high volatilities. Brunetti et al. (2011) do not find mean reversion, but their linear specifications might not capture high persistence and nonstationarity modelled by our ARFIMA model. Moreover, our more recent sample where the intraday interest rates reduce in the final part of the sample and lower frequency data could explain the difference. Our results also indicate that intraday interest rates behave somewhat differently than longer maturity interest rates for which predictability is often not found and a random walk model is very difficult to beat, see Ang and Piazzesi (2003), Diebold and Li (2006) and de Pooter et al. (2010). Hamilton (2009) finds similar evidence of predictability for the daily changes in the Fed Funds. Finally, adding exogenous variables which could proxy funding liquidity and counterparty risks in financial markets as the spread between the three-month Euribor and the three-month Eonia swap rates does not improve forecast accuracy, suggesting that predictability might derive from the econometric properties of the series more than from economic news available in real-time to market participants, see, for example, Robertson and Wright (2012).

We believe that our findings are very important for at least two players in the intraday market. Firstly, central banks could plan supplementary interventions to keep intraday spreads close to zero when forecasts indicate severe deviations from the zero level. The policy implication for interventions may be found in the following arguments. Intuitively, a bank short of liquidity say at 9 a.m. has two alternatives: (i) borrow in the interbank ON market, (ii) obtain intraday credit from the central bank and borrow later (say at 3 p.m.) in the ON market. If these two alternatives were substitutes, such bank would not be willing to pay an implicit intraday interest charge larger

than the cost of a six hour loan from the central bank. This is the reason why the cost of daylight liquidity provided by the central bank may be seen as an upper bound for the implicit intraday interest rate. The ECB does not charge any fee on intraday credit. The implicit cost comes from the collateral requirement: if a bank has to borrow eligible securities, she incurs in an explicit cost; to the contrary, if a bank holds eligible securities in her portfolio, she bears only an opportunity cost, as she is not free to trade such securities. A positive spread may be an indication that the central bank should provide (unlimited) free daylight overdrafts on a uncollateralized basis. Secondly, an investor who has collateral or liquidity to invest can implement active strategies based on lending in the intraday market and using information given by predictive densities. We document that the investor can receive economic gains by lending at the intraday rate when the forecast predicts a positive rate. She can also use the full distribution and derive some rules to exit the strategy when predicted quantiles are too high because, for example, they are associated to high counterparty risk or market dysfunctional.

The paper is organized as follow. Section 2 describes the data set, while section 3 introduces long memory models for the intraday rate and compares them with the short memory ARMA processes, our benchmarks. A forecast combination based on linear pooling is also presented. Section 3.3 describes forecasting results and implement active investment strategies based on the forecasts. Section 5 concludes.

2   Data

Our data set includes all overnight (ON) trades taking place on the e-MID interbank market over the period January 2nd 2007 to April 30th 2009 for a total of 593 working days (see Figure 1).

The e-MID represents the only readily available source of micro data on interbank transactions in the euro area, as most of the other transactions in this category are conducted in the over-the-counter (OTC) market. The e-MID is located in Italy, but we believe it can be considered as a proxy for the euro area money market (more than 200 counterparties from all over Europe have access to the system). As banks can arbitrage between the e-MID and the OTC market, the ON interest rate in the e-MID market is closely related to the Eonia rate, which is the euro ON index computed daily by the European Central Bank and published by the European Banking Federation. Furthermore, e-MID interest rates reflect actual transactions, and therefore they do not suffer from the potential distortions affecting offered rates, such as Euribor rates.

4

Figure 1: Data
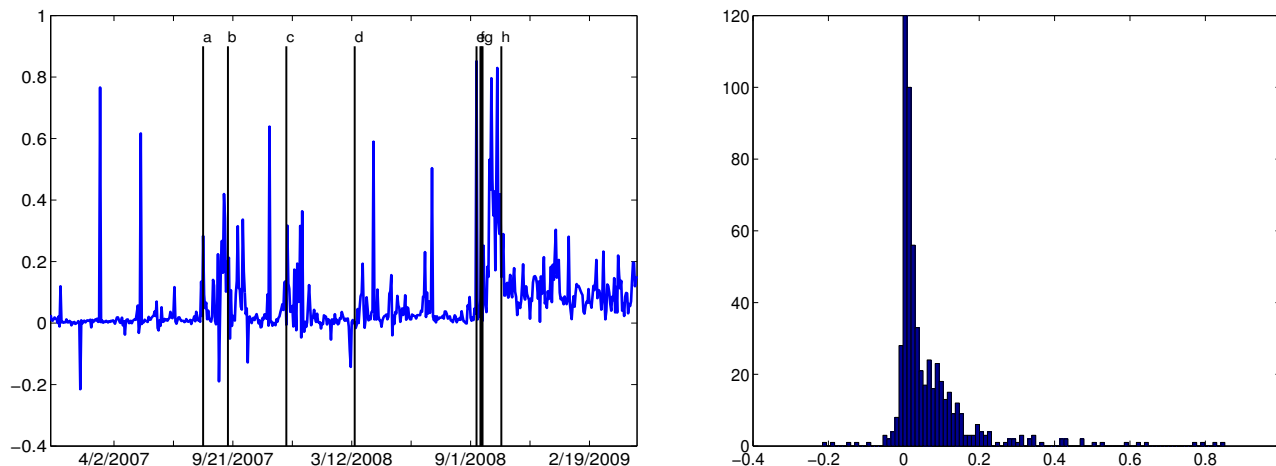
Note: The figure shows in the left panel the intraday interest rate in % and in the right its histogram. Timeline legend for left panel: a - 8/9/2007, BNP Paribas redemptions on three investment funds; b - 9/14/2007, liquidity support for Northern Rock; c - 12/6/2007, bank writedowns (UBS, Lehman); d - 3/17/2008, collapse of Bear Stearns; e - 9/15/2008, Lehman banckruptcy; f - 9/16/2008, loan to AIG; g - 9/17/2008, money market funds trouble; h - 10/14/2008, US TARP announced; 10/15/2008, ECB extraordinary measures.

We use our data to build up a daily time series of the intraday interest rate as in Baglioni and Monticini (2012). For each business day in our sample, we compute the average interest rate for the ON trades taking place between 9 a.m. and 1 p.m.: this is the morning rate (R1). In the same fashion, we compute the average rate of the ON trades taking place between 2 p.m. and 6 p.m.: this is the afternoon rate (R2). The difference between R1 and R2 is the intraday interest spread. The average value of the series is around 6.4 basis point, with a standard deviation of 0.112, it is positively skewed (the skewness is equal to 3.43) and not surprisingly, due to spikes most of them related to reserve requirements, see discussion in the next section, and to the money market financial turmoil, leptokurtic, see descriptive statistics in Table 1.[1] Figure 1 shows its erratic pattern distinguished by fat tails, and an increase in the mean and volatility after Lehman banckruptcy in September 2008 when money markets all around the World experienced liquidity issues.

As theoretically explained in Baglioni and Monticini (2012), the difference between R1 and R2 is due to both funding liquidity (hereafter we refer to funding liquidity as 'liquidity') and counterparty

---

[1]We have also computed the ON series using different hours, such as the difference between the average rate of the opening value computed over the first hour of trading and the average closing value over the last hour, and differences are marginal.

Table 1: Summary Statistics for the variable intraday interest rate

Mean	Median	Minimum	Maximum
0.063	0.022	-0.215	0.851
Std. Dev.	C.V.	Skewness	Ex. kurtosis
0.112	1.762	3.432	16.3199

risk. Forecasting the intraday spread requires the use of a proxy to take into account these risks. An obvious candidate is the spread between the three-month Euribor and the three-month Eonia swap rate (thereafter referred to Eonia-Euribor spread). This is a well known indicator, often used in the analysis of the liquidity crisis. It reflects both the liquidity and the counterparty risks perceived by the participants in the money market; at the same time, it is not affected by changes in interest rate expectations. Both the Euribor and the Eonia swap rate are calculated at 11 a.m., using the information provided by a panel of primary European banks[2]. To explain why this spread is a good risk indicator, suppose this morning the spread is larger than yesterday: this may be taken as an indication that the liquidity risk and the counterparty risk perceived by market participants have gone up, presumably reflecting the release of some negative news. For this reason, we take the daily change of the spread ('ω_1') as an indicator of movements of the ON rate within the day.[3]

3 Model

As a preliminary check, we test the intraday interbank market spread (y_t) time-series for a unit root. The unit root hypothesis is rejected (GLS-ADF (Elliott et al. (1996))) in favour of stationary ARMA models, although with autoregressive coefficients close to unity. Then we test for long memory, and the Lo's RS test (Lo (1991)) (3.062 with p-value < 0.005) provides evidence in favour of long memory. Therefore we focus on ARMA and ARFIMA models.

---

[2]They are provided by the European Banking Federation. See http://www.euribor.org/ for detailed information and for daily data.

[3]We also considered a generic Credit Default Swap Index, but predictive results are inferior to the results using the Eonia-Euribor spread. Results are available upon request.

3.1 Individual models

ARMA The first model is a traditional time series approach to modelling interest rates, the autoregressive moving average (ARMA) model (see, for example, Hamilton (1994)). The $ARMA(p,q)$ model implies that the current value of the investigated process (say, the intraday interbank market spread y_t) is expressed linearly in terms of its past p values (autoregressive part) and in terms of the q previous values of the process ϵ_t (moving average part) and a possible set of k deterministic terms and l stochastic variables as predictors. We specify the $ARMA(p,q)$ as:

$$\phi(L)y_t = \theta_0 + \sum_{i=1}^{k} \psi_i x_{i,t} + \sum_{j=1}^{l} \varphi_j w_{j,t} + \theta(L)\epsilon_t, \tag{1}$$

where θ_0 is a constant term, $\phi(L)$ and $\theta(L)$ are the autoregressive and moving average polynomials in the lag operator L respectively, defined as:

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p, \tag{2}$$

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - ... - \theta_q L^q, \tag{3}$$

and ϵ_t is an independent and identically distributed (iid) noise process with zero mean and finite variance σ^2;[4] $x_t = (x_{1,t}, x_{2,t}, ..., x_{k,t})'$ is the $(k \times 1)$ vector of deterministic terms (dummies) at time t and $\psi = (\psi_1, \psi_2, ..., \psi_k)'$ is a $(k \times 1)$ vector of coefficients; $w_t = (w_{1,t}, w_{2,t}, ..., w_{l,t})'$ is the $(l \times 1)$ vector of stochastic predictors for time t, and $\varphi = (\varphi_1, \varphi_2, ..., \varphi_l)'$ is a $(l \times 1)$ vector of coefficients.[5]

Depository institutions in the Eurosystem ought to meet some reserve requirements. The reserve requirements are the amount of funds that a depository institution ought to hold in reserve against specified deposit liabilities. The requirement has to be satisfied on average over a maintenance period. See Durré and Nardelli (2008) for a description of the current Eurosystem monetary policy operational framework. There is empirical evidence, as in Angelini (2000) for the previous Eurosystem monetary policy operational framework and Durré and Nardelli (2008) for the current one, that on last day of the reserve requirement maintenance period, the intraday volatility of the interbank rate is substantially higher than during the rest of the month. For this reason, we

---

[4]We also investigate time-varying volatility residuals, modelled with a Garch(1,1) specification, for the ARMA and other models. Forecast accuracy, on average, decreases and we do not report results. A (continuous) time-varying volatility model would imply ECB does not pursue any daily market operation to smooth intraday rates.

[5]An ARMA model collapses to a random walk (RW), model for absence of predictability, by assuming $\phi_1 = 1$, $\phi_2 = ... = \phi_p = 0$ and $\theta_1 = \theta_2 = ... = \theta_q = 0$. We use it as benchmark in our analysis.

introduce a 0, 1 dummy variable. The dummy variable takes 1 on settlement days and 0 otherwise. Moreover, we include a 0, 1 dummy variable for the end of month and end of quarter effects like in (Ángel León et al., 2006). We apply two ARMA(p, q) specifications: in the first model $k = 2$ $(l = 0)$, which corresponds to the two dummy variables. In the second specification $k = 2$ and $l = 1$, referring to the use of the liquidity and counterparty risk variable defined by the Eonia-Euribor spread as stochastic predictors. The value of the spread at time t is used to predict the intraday rate at time $t + 1$. The second class of models is denoted with suffix X in the remainder of the text.[6]

ARFIMA   Long memory behaviors found in section 3 can be modelled in a more rigorous way than by using p lags in ARMA models. We propose the following ARFIMA(p, d, q) model:

$$\phi(L)(1 - L)^d y_t = \theta_0 + \sum_{i=1}^{k} \psi_i x_{i,t} + \sum_{j=1}^{l} \varphi_j w_{j,t} + \theta(L)\epsilon_t, \tag{4}$$

where we recall $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$, $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - ... - \theta_q L^q$ and ϵ_t is an independent and identically distributed (iid) noise process with zero mean and finite variance σ^2. The parameter d specifies the order of integration. We require $0 < d < 1$, $\sum_{i=1}^{p} |\phi_i| < 1$, $\sum_{i=1}^{q} |\theta_i| < 1$. We implemented two different estimation methods for d over the full sample, a variant of the log-periodogram regression proposed by Geweke and Porter-Hudak (1983) ($\hat{d} = 0.326$ (0.044)), and the maximum likelihood (ML) estimator implemented by Sowell (1992) ($\hat{d} = 0.279$ (0.028)). Similar results were obtained in each case. We apply two ARFIMA(p, d, q)s depending on whether the explanatory stochastic variable is used (referred as ARFIMAX(p, d, q)) or not (referred as ARFIMA(p, d, q)).

Bhardwaj and Swanson (2006) discuss how ARFIMA(p, d, q) processes perform better in (point) forecasting exercises when the data sample is small. Moreover, spurious long memory behaviors arise in many contexts, such as when there are (stochastic) structural breaks in linear and non-linear models, regime switching models, and when forming models using variables that are non-linear transformations of underlying "short memory" variables, see for example Byers et al. (1997), Diebold and Inoue (2001), Engle and Smith (1999) and Bhardwaj and Swanson (2006).

---

[6]We have also tested ARIMA models, but results are substantially inferior to those for the ARMA models and not reported.

Estimation   We estimate models using an expanding window approach, implying that every day a new observation is available we expand the in-sample period with one observation and produce a new forecasts.[7] We estimate our models using maximum likelihood estimators in Ox, see Doornik and Ooms (2006). For computing 1-step ahead density forecasts, we use a conditional normal approximation as in Huurman et al. (2010). Assuming that the past errors and coefficients are known, the conditional expectation corresponds to the point forecast of each individual model. The forecast variance is computed by approximating the forecast error variance with the in-sample estimate of the error variance σ^2. The predictive density given by any of the models in the suite is then

$$f_{t+1,i}(y_{t+1}) \sim N(\mu_{t+1,i}, \sigma^2_{t+1,i}), \tag{5}$$

where $\mu_{t+1,i}$ is the point forecast and $\sigma^2_{t+1,i}$ is the variance forecast for model i made at time t for $t + 1$.

3.2   Linear opinion pooling

Our model set contains two classes of models, ARMA(p, q) and ARFIMA(p, d, q); in each of them the exogenous Eonia-Euribor spread (labeled with a 'X' suffice) can be added; and number of p AR and q MA lags have also to be estimated. When restricting p and q to assume as maximum value 1, we result in 16 individual models: RW, AR, MA, ARMA, RWX, ARX, MAX, ARMAX, ARFIMA(0,d,0), ARFIMA(1,d,0), ARFIMA(0,d,1), ARFIMA(1,d,1), ARFIMAX(0,d,0), ARFIMAX(1,d,0), ARFIMAX(0,d,1), ARFIMAX(1,d,1). Ex ante selection of the best model can be difficult, see, for example, discussion in Robertson and Wright (2012) on how in the ARMA class, reduced models that exclude exogenous variables can encompass larger models with exogenous variables even if the exogenous variables contain useful information and might represent the true model. Forecast combination is a possible alternative which account for model uncertainty. We follow, among others, Mitchell and Hall (2005), Jore et al. (2010), Geweke and Amisano (2010), Kascha and Ravazzolo (2010) and Ravazzolo and Vahey (2013), and produce predictive densities

---

[7]We have also investigated several moving window approaches, and in particular a 65-days window to account for the possibility of frequent breaks, see, for example, Pesaran and Timmermann (2007) for a discussion on the selection of the estimation window in the presence of structural instability. Forecast accuracy decreases with such rolling windows.

by combining the individual predictions using a linear opinion pooling (LOP):

$$f_{t+1,LOP}(y_{t+1}) = \sum_{i=1}^{N} w_{i,t+1} f_{t+1,i}(y_{t+1}), \ , \ w_{i,t+1} > 0, \sum_{i=1}^{N} w_{i,t+1} = 1 \tag{6}$$

where $f_{t+1,i}(y_{t+1})$ is the predictive density for model i described in the above sections; and $w_{i,t+1}$ is the model i specific weight computed using information up to time t. LOP has several advantages, including that time-varying weights allow different relative contributions from individual models over time; weights can be based on past performance; and if individual predictive densities are normally distributed, LOP results in a mixture of normals, therefore fitting more general distributions; see discussion in Kascha and Ravazzolo (2010).

The choice on how to derive weights is an important element for LOP. The simple and easier solution is applying equal weights, $w_{i,t+1} = 1/N$. Equal weights have been shown to perform well in several financial and macroeconomic applications, see Timmermann (2006) for a review on forecast combinations. However, Jore et al. (2010) and Kascha and Ravazzolo (2010) find that equal weights are not necessarily the best strategy when the focus of the exercise is on density forecasting and, in general, on higher moments. They propose to use weights based on the evaluation criterion and the associated loss function applied in the (density) evaluation. We devote next section to the description of our evaluation methodology, but let assume that $l(I_{t,i})$ is the function to maximize, such as the inverse mean square error or Kullback Leiber Information Criterion, using information set I_t available at time t and model i, LOP weights can be computed as:

$$w_{i,t+1} = \exp(l(I_{t,i})) / \sum_{i=1}^{N} \exp(l(I_{t,i})) \tag{7}$$

Following, again, Jore et al. (2010) and Kascha and Ravazzolo (2010), we apply the log score described in equation 11 in the next section as value for $l(I_t)$ to derive combination weights.

3.3 In-sample and out-of-sample Evaluation

In-sample evidence can be interpreted as an ex-post analysis of the relevance of long memory properties and predictability in the data. Inoue and Kilian (2004) suggest that in-sample tests are likely to have greater power than out-of-sample tests. They examine the question of in-sample versus out-of-sample testing of predictability, motivated by the finding that positive in-sample evidence of predictability is often not associated with out-of-sample predictability. They argue that the claim

made by Ashley et al. (1980), i.e. that in-sample inference without out-of-sample verification is likely to be spurious, with an out-of-sample approach inherently involving less overfitting, is not compelling since there is ample opportunity for the researcher to data mine in a simulated out-of-sample study, and because data snooping adjustments can be made to both tests. However, the evidence presented by Inoue and Kilian (2004) refers to point forecasting, and as the models under consideration are chosen by data-drive model selection procedures, the "impossibility" theorem in Leeb and Potscher (2005) might imply that the true distribution of the test statistics might be unknowable even by standard simulation methods. Therefore, we view the results we obtain as a natural complement to the set of mixed and conflicting results reported by leading scholars in the literature and refer to the argument of Welch and Goyal (2008) that out-of-sample tests provide "useful diagnostic" information about the underlying dynamic relationship.

Figure 2 presents in-sample evidence via Akaike Information Criterion (AIC) on the predictability of alternative models with respect to the RW benchmark. We have also computed the Bayesian Information Criterion and results are similar. The models provide similar AIC in the first part of the sample; but alternative models provide lower values of the criteria around unstable periods. The increases is marginally in August 2007, the beginning of turbulent period; and for Bearn Stearns crisis in March 2008; but much larger around the collapse of Lehman Brothers in September 2008 and afterwards. ARMA, ARMAX and most of ARFIMA models seem to provide the best fitting.

To shed light on the predictive power of individual models, we use a number of evaluation statistics for point and density forecasts previously proposed in literature. Our ARFIMA models, of order (0,d,0), are not nested in the ARMA benchmark, of order (1,0). We compare point forecasts in terms of mean absolute prediction error (MAPE) and root mean square prediction errors (RMSPE). Following evidence and discussion in Clark and McCracken (2012) [section 3.2], we test whether mean square prediction errors are statistically different using an encompassing t-type of test (hereafter ECN-t) and compare all the models to the benchmark ARMA model. Moreover, following Welch and Goyal (2008), we graphically analyze what we call the Cumulative Squared Prediction Error Difference (CSPED):

$$CSPED_{t+1,i} = \sum_{s=\underline{t}}^{t} \widehat{d}_{s+1,i}, \quad t = \underline{t}, ..., T-1, \tag{8}$$

where $\widehat{d}_{t+1,i} = e_{t+1,RW} - e_{t+1,i}$, $i =$ AR, MA, ARMA, RWX, ARX, MAX, ARMAX, ARFIMA(0,d,0), ARFIMA(1,d,0), ARFIMA(0,d,1), ARFIMA(1,d,1), ARFIMAX(0,d,0), ARFIMAX(1,d,0), ARFI-

MAX(0,d,1), ARFIMAX(1,d,1), LOP; and $e_{t+1,i}$ is the MSPE for model i. Increases in $CSPED_{t+1,i}$ indicate that the alternative to the benchmark (RW model) predicts better the out-of-sample observation t. Such graph allows to split the analysis on forecast performance over different subsamples.

We evaluate the predictive densities using tests of absolute forecast accuracy. Like Diebold et al. (1998), we utilize the probability integral transforms, *PITS*, of the realization of the variable with respect to the forecast densities. A forecast density is preferred if the density is correctly calibrated, regardless of the forecasters' loss function.

The *PITS* are:

$$z_{y_{t+1}} = \int_{-\infty}^{y_{t+1}} f_{t+1}(u)du.$$

The *PITS* should be both uniformly distributed, independently and identically distributed if the forecast densities are correctly calibrated. Hence, calibration evaluation requires the application of tests for goodness-of-fit and independence. We apply the Likelihood Ratio test proposed by Berkowitz (2001): the LR3 Berkowitz test is a three degrees of freedom variant, with a test for independence and uniformity, where under the alternative $z_{y_{t+1}}$, $t = \underline{t}, ..., T - 1$, follows an AR(1) process. A well calibrated density should give high probability values for the test—implying the null hypothesis of no calibration failure cannot be rejected.

Turning to our analysis of relative predictive accuracy, we consider a Kullback Leibler Information Criterion (KLIC)-based test, utilizing the expected difference in the Logarithmic Scores of the candidate forecast densities; see for example Kitamura (2002), Mitchell and Hall (2005), Amisano and Giacomini (2007), Huurman et al. (2010) and Caporin and Pres (2010). Geweke and Amisano (2010) and Mitchell and Wallis (2010) discuss the value of information-based methods for evaluating forecast densities that are well calibrated on the basis of *PITS* tests. The KLIC chooses the model which on average gives higher probability to events that have actually occurred. Specifically, the KLIC distance between the true density f_{t+1} of a random variable y_{t+1} and some candidate density $f_{t+1,i}$ obtained from model i (or LOP) is defined as

$$\begin{aligned} \text{KLIC}_{t+1} &= \int f_{t+1}(y_{t+1}) \ln \frac{f_{t+1}(y_{t+1})}{f_{t+1,i}(y_{t+1})} dy_{t+1}, \\ &= E[\ln f_{t+1}(y_{t+1}) - \ln f_{t+1,i}(y_{t+1})]. \end{aligned} \tag{9}$$

Under some regularity conditions, a consistent estimate can obtained from the average of the sample

information, $y_{\underline{t}+1}, \ldots, y_T$, on f_{t+1} and $f_{t+1,i}$:

$$\overline{KLIC} = \frac{1}{n} \sum_{t=\underline{t}}^{T-1} [\ln f_{t+1}(y_{t+1}) - \ln f_{t+1,i}(y_{t+1})] \tag{10}$$

where $n = T - \underline{t}$. Even though we do not know the true density, we can still compare multiple densities, $f_{t+1,i}$. For the comparison of two competing models, it is sufficient to consider only the latter term in the above sum,

$$LS_{t+1,i} = -\frac{1}{n} \sum_{t=\underline{t}}^{T-1} \ln f_{t+1,i}(y_{t+1}), \tag{11}$$

for all i and to choose the model for which the expression in (11) is minimal, or as we report in our tables, the opposite of the expression in (11) is maximal. Differences in KLIC can be statistically tested. We apply a test of finite-sample predictive ability of two density forecasts as defined in Clark and McCracken (2012) similar to Mitchell and Hall (2005) and Amisano and Giacomini (2007), based on Giacomini and White (2006) formulation. Suppose there are two 1-step ahead density forecasts, $f_{t+1,1}(y_{t+1})$ and $f_{t+1,2}(y_{t+1})$, and consider the loss differential

$$d_{t+1} = \ln f_{t+1,1}(y_{t+1}) - \ln f_{t+1,2}(y_{t+1}).$$

We apply the following WALD test:

$$GW = n(n^{-1} \sum_{t=\underline{t}}^{T-1} h_t d_{t+1})' \widehat{\Sigma}_{t+1} (n^{-1} \sum_{t=\underline{t}}^{T-1} h_t d_{t+1}), \tag{12}$$

where $h_t = (1, d_t)'$, and $\widehat{\Sigma}_{t+1}$ is the HAC estimator for the variance of $(h_t d_{t+1})$. Under the null of equal predictability, $GW \sim \chi_2^2$.

Analogous to our use of the CSPED for graphically examining relative MSPEs over time, and following Kascha and Ravazzolo (2010), we define the Cumulative Log Score Difference ($CLSD$):

$$CLSD_{t+1,i} = -\sum_{s=\underline{t}}^{t} d_{t+1,i}, \tag{13}$$

where $d_{t+1} = \ln f_{t+1,i}(y_{t+1}) - \ln f_{t+1,RW}(y_{t+1})$. If $CLSD_{t+1,i}$ increases at observation $t + 1$, this indicates that the alternative to the benchmark has a higher log score.

4 Results

4.1 Forecast results

We construct 1-step ahead point and density forecasts for the intraday interbank spread over the sample period from April 2, 2007 to April 30, 2009, for a total of 529 observations.

Table 2 reports point forecast results. We think there are two clear conclusions. First, the ARFIMA models have the higher predictability: several of them produce both the smallest MAPE and RMSPE and the differences with respect to the benchmark RW model are statistically significant in terms of the ECN-t test. Therefore, adding the fractionally integrated parameter has a substantial predictive power, improving the accuracy of forecasts. The ARMA and ARMAX models also provide accurate point forecasts, but not AR and MA models improve forecast accuracy. Baglioni and Monticini (2012) discuss how the intraday rate is affected by the likelihood of a liquidity dry-up in financial markets. The spread at time $t + 1$ is, however, not available when forecasts are made at time t and the variable seems to introduce forecasting errors. A (better) model to forecast the Eonia-Euribor spread might improve results. Furthermore, Robertson and Wright (2012) discuss how in the ARMA class, reduce models excluding exogenous variables can encompass larger models with exogenous variables even if the exogenous variables contain useful information.

Thirdly, LOP is superior to the benchmark model and provides MAPE and RMSPE similar to the best individual models, even if not smaller than them. Our analysis confirms that LOP offers an "insurance" against selecting ex-ante the worst model(s) by also providing accurate results.

The CSPED graph in Figure 3 for a selection of models can give further intuitions on the results. The benchmark model does better at beginning of the turbolent time in August 2007; but alternatives recover from September 2007 and perform marginally better. The shock in August is unexpected and the RW model seems the most adequate; but after some weeks the alternative models learn that there is a new regime and provide more accurate forecasts. During and after the Lehman Brothers collapse on September 15th 2008; the ARMA model, several ARFIMA models and LOP statistically dominate the benchmark. Again, the shock is unexpected, at least the magnitude, but models in the ARMA and ARFIMA class have learnt during the previous year that such shocks imply a period of high volatility.

Complete probability distributions over outcomes provide information which is helpful for making economic decisions. Therefore, we turn our analysis to density forecasting. Table 3 reports

full sample results. We first focus on absolute accuracy. The null hypotheses of correct calibration is not rejected at 5% significance level for the ARMAX, almost all ARFIMA models, and LOP; and a 1% significance level for the ARMA. *PITS* for ARMA, two ARFIMA models, and LOP in Figure 4 span the [0,1] interval quite well, even if they are more concentrated on 0.4-0.6 percentiles, suggesting forecasts are on average a bit too narrow. *PITS* for the AR model, on contrary, span less equally the 10 deciles. Following the discussion in Mitchell and Wallis (2010), we apply log score measures and test on them to discriminate between models. Similar to point forecast evidence, several ARFIMAs give high scores and outperform the benchmark in terms of the log score test; but differently than for the previous forecast metrics not all ARFIMA models, see statistics for the ARFIMA(0,d,1) and ARFIMA(1,d,1). The ARFIMA(0,d,0) and ARFIMA(1,d,0) provide the highest scores. The ARMA model also outperforms the benchmark, such the ARX model but its score is almost 10% lower than the score for the ARFIMA(1,d,0). LOP also in this case provides very accurate forecasts and scores very close to the best two models. LOP weights in Figure 5 show that there is some uncertainty at the beginning of the sample, but after September 2007 and, in particular, September 2008, weights converge to the ARFIMA(0,d,0) and ARFIMA(1,d,0) models, with the latter one dominant. Our ex-post analysis shows that these two models give the more accurate predictive densities and LOP exploits in real-time this information. More generally, looking to the full sample series we could identify two types of shocks: transitory shocks such as in August and September 2007 and September 2008 with very high volatility, and a permanent shock such as after 15 October 2008, when ECB switched from the variable rate tender format to a fixed rate full allotment policy, see Abbassi and Linzert (2011), where the mean of the series shifts upward and volatility is high but less than for the previous period. We find as in Bhardwaj and Swanson (2006) that ARFIMA models produce accurate forecasts when there are several stochastic and unknown structural breaks. Which ARFIMA model to use is however unclear ex-ante; LOP can cope with such uncertainty and mixing predictive density can also approximate different regimes over time caused by different types of shocks.

The fan chart in Figure 6 shows that there is an increase in uncertainty after August 2007, and an upward level shift and larger volatility after Lehman Brothers collapse. The initial months of the financial crisis in August and September 2007 and days after Bear Stearns acquisitions are characterized by a different pattern: mean forecasts move largely, but variances relative less compared to September 2008. The level shift in Septmber 2008 is absorbed in the following months, but the large volatility does not reduce over the sample. The predictive density is also less symmetric after

Lehmann Brothers collapse. Evidence is similar for the best ARFIMA models. The ARMA model, several ARFIMA models, and LOP cannot anticipate unexpected shocks, but they cope relatively better with volatile observations just after them, giving higher log scores than the benchmark RW, the AR and MA models, as the right panel in Figure 3 shows.

4.2 Investment strategies

An investor with excess collateral or liquidity could take advantage by having a long position in the intraday market. This could be done by, for example, lending at the morning rate R1 in the morning and borrowing at the rate R2 in the afternoon. If the intraday rate, difference between R1 and R2, is positive the investor has a gain. The strategy requires that the investor forecasts the intraday rate and it has enough collateral or liquidity to close the position. Therefore, using $\mu_{t+1,i}$ from the individual models and the LOP to forecast of the intraday rate $t+1$, the realized return at day $t+1$ for the portfolio based on model i is:

$$R_{t+1,i} = W_t(1 + I(\mu_{t+1,i} > 0)y_{t+1}) \tag{14}$$

where W_t is the wealth invested at time t; $I(\cdot)$ is an indicator function which takes value 1 if it is satisfied; and y_{t+1} is the realized intraday rate. We assume the investor has initial wealth equal to 1\$, $W_{\underline{t}} = 1$; she repeats the exercise every day; and she reinvests all the return, $W_{t+1} = R_{t+1}$, up to the end of our sample. The exercise is similar to the strategy presented in Pesaran and Timmermann (1995) to evaluate stock return predictability. Panel 1 in Table 4 report mean portfolio returns, standard deviation and Sharpe Ratio for the different models applied to forecast the intraday rate. The strategy produces sizeable gains; ARFIMA(0,d,0), ARFIMA(1,d,0), and LOP gives the highest returns and Sharpe Ratios.

The previous exercise ignores counterparty risks. A first justification for this is that we use the average intraday rate which depends on transactions with high and low counterparty risk; therefore avering such risk. However, we also implement a different strategy which takes full advantage of the density forecasts for the intraday rate and allows the investor not to trade when the distribution of the predicted rate is too high, assuming this is related to high counterparty risk. Precisely, the investor trades if the predicted mean return is positive and the 95% quantile is below 30 basis points. We choose 30 basis points arbitrarily; findings are qualitatively similar for 20 and 25 basis points. If at least one of the two restrictions is not satisfied at day t, the investor exits the market.

When the investor predicts wrongly, and she invests and realization is above 30 basis points we penalize the investment by a loss equal to the realization. The realized return at day $t+1$ for the portfolio based on model i is

$$
\begin{aligned}
R_{t+1,i} \;=\; & W_t[1 + (1 - I(y_{t+1} > 0.3))(I(\mu_{t+1,i} > 0)I(Q95_{t+1,i} < 0.3)y_{t+1} \\
& - I(y_{t+1} > 0.3)(I(\mu_{t+1,i} > 0)I(Q95_{t+1,i} < 0.3)y_{t+1}]
\end{aligned}
\tag{15}
$$

where $Q95_{t+1,i}$ is the predicted 95th quantile for the intraday rate at day $t+1$ given by model i. The strategies still produce positive returns, but density forecast accuracy play a more relevant role than in the previous exercise. The ARMA model, several ARFIMA models and LOP give sizeable gains with respect to other strategies.

5 Conclusion

The findings presented in this paper point to the conclusion that the dynamic of the intraday interest rate during high volatility periods such as financial crises has a remarkable characteristic: a highly persistent, nonstationary process that nonetheless reverts to the mean. More precisely, we can say that the dependence is typically explained by long memory time series ARFIMA models. We provide evidence in favour of the ARFIMA models compared to random walk, AR and MA models both in terms of in-sample predictability and out-of-sample point and density predictability. Statistical gains are substantially higher when focusing on the full distribution. We also document that predictive densities give substantial economic gains in investment strategies based on lending on the intraday market. Finally, there is, however, potentially large uncertainty on the number of lags, and the use of exogenous variables and we propose to apply a linear opinion pooling to account for it. The method provides insurance against selecting poor models and also gives statistical and economic gains.

Our analysis is unfortunately agnostic on the source of the problems in the intraday money market. The predictability could derive from frictions in the market associated with market participant behaviors (such as unwillingness to borrow, stigma issues, central bank interventions as a source of news), which are difficult to model and above all to forecast.

References

Abbassi, P., Linzert, T., 2011. The effectiveness of monetary policy in steering money market rates during the recent financial crisis. Working Paper Series 1328, European Central Bank.

Amisano, G., Giacomini, R., 2007. Comparing density forecasts via weighted likelihood ratio tests. Journal of Business & Economic Statistics 25 (2), 177–190.

Ang, A., Piazzesi, M., 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics 50, 745–787.

Ángel León, Benito, F., Nave, J., June 2006. Modeling the euro overnight rate. Tech. Rep. 2006-11, Instituto Valenciano de Investigaciones Económicas, S.A. (Ivie).

Angelini, P., 1998. An analysis of competitive externalities in gross settlement systems. Journal of Banking and Finance 22, 1–18.

Angelini, P., 2000. Are banks risk averse? intraday timing of operations in the interbank market. Journal of Money, Credit and Banking 32 (1), 54–73.

Ashley, R., Granger, C. W. J., Schmalensee, R., July 1980. Advertising and aggregate consumption: An analysis of causality. Econometrica 48 (5), 1149–67.

Baglioni, A., Monticini, A., 2008. The intraday price of money: Evidence from the e-mid interbank market. Journal of Money, Credit and Banking 40 (7), 1533–1540.

Baglioni, A., Monticini, A., 2010. The intraday interest rate under a liquidity crisis: The case of august 2007. Economics Letters 107 (2), 198–200.

Baglioni, A., Monticini, A., 2012. Why does the interest rate decline over the day? evidence from the liquidity crisis. Journal of Financial Services Research, 1–12.
URL http://dx.doi.org/10.1007/s10693-012-0139-x

Bech, M., Garratt, R., 2003. The intraday liquidity management game. Journal of Economic Theory 109, 198–219.

Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. Journal of Business & Economic Statistics 19 (4), 465–74.

Bhardwaj, G., Swanson, N. R., 2006. An empirical investigation of the usefulness of arfima models for predicting macroeconomic and financial time series. Journal of Econometrics 131 (1-2), 539–578.

Bhattacharya, J., Haslag, J., Martin, A., 2007. Why does overnight liquidity cost more than intraday liquidity? Tech. Rep. Staff Report no.281, Federal Reserve Bank of New York.

Brunetti, C., di Filippo, M., Harris, J. H., 2011. Effects of central bank intervention on the interbank market during the subprime crisis. Review of Financial Studies 24 (6), 2053–2083.

Byers, D., Davidson, J., Peel, D., 1997. Modelling political popularity: an analysis of long range dependence in opinion poll series. Journal of the Royal Statistical Society Series A 160 (3), 471–490.

Caporin, M., Pres, J., 2010. Modelling and forecasting wind speed intensity for weather risk management. Computational Statistics and Data Analysis forthcoming.

Clark, T., McCracken, M., 2012. Testing for unconditional predictive ability. In: Oxford Handbook on Economic Forecasting,. D. Hendry and M. Clements, eds., Oxford University Press.

de Pooter, M., Ravazzolo, F., van Dijk, D., Mar. 2010. Term structure forecasting using macro factors and forecast combination. Working Paper 2010/01, Norges Bank.

Diebold, F. X., Gunther, T., Tay, A. S., 1998. Evaluating density forecasts with applications to finance and management. International Economic Review 39, 863–883.

Diebold, F. X., Inoue, A., 2001. Long memory and regime switching. Journal of Econometrics 105 (1), 131–159.

Diebold, F. X., Li, C., 2006. Forecasting the term structure of government bond yields. Journal of Econometrics 130, 337–364.

Doornik, J., Ooms, M., 2006. A package for estimating, forecasting and simulating arfima models: Arfima package 1.04 for ox.

Durré, A., Nardelli, S., 2008. Volatility in the euro area money market: effects from the monetary policy operational framework. International Journal of Finance & Economics 13 (4), 307–322.

Elliott, G., Rothenberg, T. J., Stock, J. H., 1996. Efficient tests for an autoregressive unit root. Econometrica 64 (4), 813–36.

Engle, R. F., Smith, A. D., 1999. Stochastic permanent breaks. The Review of Economics and Statistics 81 (4), 553–574.

FED, 2006. Consultation paper on intraday liquidity management and payment system risk policy. Board of Governors of the Federal Reserve System.

FED, 2007. Comments on the consultation paper. Federal Reserve Bank of Chicago.

Furfine, C. H., 2001. Banks as monitors of other banks: Evidence from the overnight federal funds market. The Journal of Business 74 (1), 33–57.

Geweke, J., Amisano, G., 2010. Comparing and evaluating bayesian predictive distributions of asset returns. International Journal of Forecasting 26 (2), 216–230.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4 (4), 221–238.

Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74, 1545–1578.

Hamilton, J., 1994. Time Series Analysis. Princeton University Press.

Hamilton, J., 2009. Daily changes in fed funds futures prices. Journal of Money, Credit and Banking 41, 567–582.

Heider, F., Hoerova, M., Holthausen, C., Dec. 2009. Liquidity hoarding and interbank market spreads: The role of counterparty risk. Working Paper Series 1126, European Central Bank.

Huurman, C., Ravazzolo, F., Zhou, C., 2010. The power of weather. Computational Statistics and Data Analysis forthcoming.

Inoue, A., Kilian, L., 2004. In-sample or out-of-sample tests of predictability: Which one should we use? Econometric Reviews 23 (4), 371–402.

Jore, A., Mitchell, J., Vahey, S. P., 2010. Combining forecast densities from vars with uncertain instabilities. Journal of Applied Econometrics 25 (4), 621–634.

Jurgilas, M., Zikes, F., 2011. Implicit intraday interest rate in the uk unsecured overnight loan market, unpublished manuscript.

Kascha, C., Ravazzolo, F., 2010. Combining inflation density forecasts. Journal of Forecasting 29 (1-2), 231–250.

Kitamura, Y., 2002. Econometric comparisons of conditional models. Discussion paper, University of Pennsylvania.

Leeb, H., Potscher, B. M., February 2005. Model selection and inference: Facts and fiction. Econometric Theory 21 (01), 21–59.

Lo, A. W., 1991. Long-term memory in stock market prices. Econometrica 59 (5), 1279–313.

Martin, A., 2004. Optimal pricing of intraday liquidity. Journal of Monetary Economics 51, 401–424.

Martin, A., McAndrews, J., 2008. Liquidity-saving mechanisms. Journal of Monetrary Economics 55, 544–557.

Mills, D., Nesmith, T., 2008. Risk and concentration in payment and securities settlement systems. Journal of Monetary Economics 55, 542–553.

Mitchell, J., Hall, S. G., 2005. Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESER "fan" charts of inflation. Oxford Bulletin of Economics and Statistics 67, 995–1033.

Mitchell, J., Wallis, K. F., 2010. Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness. Journal of Applied Econometrics forthcoming.

Pesaran, M., Timmermann, A., 2007. Selection of estimation window in the presence of breaks. Journal of Econometrics 137 (1), 134–161.

Pesaran, M. H., Timmermann, A., 1995. Predictability of stock returns: Robustness and economic significance. The Journal of Finance 50 (4), 1201–1228.

Ravazzolo, F., Vahey, S. P., 2013. Forecast densities for economic aggregates from disaggregate ensembles. Studies of Nonlinear Dynamics and Econometrics forthcoming.

Robertson, D., Wright, S., 2012. The predictive space, or, if x predicts y, what does y tell us about x? Birkbeck Working Papers in Economics and Finance 1210, Birkbeck, Department of Economics, Mathematics & Statistics.

Sowell, F., 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. Journal of Econometrics 53 (1-3), 165–188.

Timmermann, A., 2006. Forecast combinations. Vol. 1 of Handbook of Economic Forecasting. Elsevier, pp. 135 – 196.

VanHoose, D., 1991. Bank behavior, interest rate determination, and monetary policy in a financial system with an intraday federal funds market. Journal of Banking and Finance 15, 343–365.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21 (4), 253–303.

Zhou, R., 2000. Understanding intraday credit in large-value payment systems,. In: Economic Perspectives. Vol. 3. Federal Reserve Bank of Chicago, pp. 29–44.

Table 2: Point forecasting

	MAPE	RMSPE	ECN-t
RW	0.059	0.113	1.000
AR	0.054	0.104	
MA	0.056	0.108	0.999
ARMA	0.051	0.097	**0.004**
RWX	0.061	0.112	0.999
ARX	0.055	0.104	0.627
MAX	0.058	0.108	0.988
ARMAX	0.052	0.097	**0.009**
ARFIMA(0,d,0)	0.051	0.098	**0.000**
ARFIMA(1,d,0)	0.051	0.097	**0.001**
ARFIMA(0,d,1)	0.051	0.097	**0.001**
ARFIMA(1,d,1)	0.051	0.097	**0.002**
ARFIMAX(0,d,0)	0.051	0.098	**0.003**
ARFIMAX(1,d,0)	0.051	0.097	**0.005**
ARFIMAX(0,d,1)	0.051	0.097	**0.006**
ARFIMAX(1,d,1)	0.052	0.097	**0.007**
LOP	0.051	0.097	**0.001**

Note: Table reports results for point forecast accuracy in predicting the 1-day ahead European intraday interbank rate over the period from 2 January 2008 to 30 April 2009. The column MAPE reports the mean absolute prediction error. The column RMSPE gives the root mean square prediction errors. The column ECN-t reports $p-$values for the test for equal forecast accuracy. Bold numbers indicate that the null of equal density predictive accuracy relative to the ARMA benchmark is rejected at 5% significance level. See section 3 for model description.

Table 3: Density forecasting

|	LR3	LS	LS_test
RW	0.000	0.803	0.000
AR	0.000	0.878	
MA	0.000	-0.432	0.000
ARMA	0.026	0.915	**0.007**
RWX	0.000	0.784	0.001
ARX	0.000	0.858	**0.017**
MAX	0.000	-0.473	0.000
ARMAX	**0.071**	0.909	0.093
ARFIMA(0,d,0)	**0.050**	0.932	**0.012**
ARFIMA(1,d,0)	0.024	0.937	**0.024**
ARFIMA(0,d,1)	**0.240**	0.870	0.000
ARFIMA(1,d,1)	**0.160**	0.836	0.000
ARFIMAX(0,d,0)	**0.061**	0.917	**0.003**
ARFIMAX(1,d,0)	**0.073**	0.922	**0.005**
ARFIMAX(0,d,1)	**0.094**	0.918	**0.015**
ARFIMAX(1,d,1)	**0.104**	0.891	**0.004**
LOP	**0.053**	0.915	**0.009**

Note: The column LR2 is the Likelihood Ratio p-value of the test of zero mean, unit variance and independence of the inverse normal cumulative distribution function transformed *PITS*, with a maintained assumption of normality for transformed *PITS* proposed by Berkowitz (2001). A bold number indicates that the null hypothesis of a correctly specified model cannot be rejected at 5% significance level. LS is the average Logarithmic Score over the evaluation period. The number with highest mean LS is reported in bold. The column LS_test is the p-value for the test of equal predictive density accuracy. Bold numbers indicate that the null of the test of equal density predictive accuracy relative to the ARMA benchmark is rejected at 5% significance level.

Table 4: Investment startegy

|	Exercise 1			Exercise 2		
	Mean Ret	St Dev	SR	Mean Ret	St Dev	SR
RW	6.966	11.28	0.616	3.210	11.15	0.29
AR	6.943	11.29	0.615	3.301	10.41	0.32
MA	6.975	11.28	0.618	3.280	11.04	0.30
ARMA	6.964	11.29	0.617	3.356	8.54	0.39
RWX	6.490	11.05	0.587	3.162	9.56	0.33
ARX	6.767	11.24	0.602	3.260	9.05	0.36
MAX	6.777	11.26	0.602	3.128	9.85	0.32
ARMAX	6.862	11.25	0.610	3.337	8.25	0.40
ARFIMA(0,d,0)	**6.971**	11.28	**0.618**	3.386	8.57	0.40
ARFIMA(1,d,0)	**6.971**	11.28	**0.618**	3.322	8.53	0.39
ARFIMA(0,d,1)	6.970	11.29	**0.618**	3.326	8.53	0.39
ARFIMA(1,d,1)	6.928	11.30	0.613	3.278	8.52	0.38
ARFIMAX(0,d,0)	6.926	11.20	0.618	**3.459**	8.21	**0.42**
ARFIMAX(1,d,0)	6.927	11.20	**0.618**	3.373	8.18	0.41
ARFIMAX(0,d,1)	6.926	11.20	**0.618**	3.343	**8.17**	0.41
ARFIMAX(1,d,1)	6.910	11.21	0.617	3.327	**8.17**	0.41
LOP	**6.971**	11.28	**0.618**	3.344	8.53	0.39

Note: The columns Mean Ret, St Dev and SR report respectively the mean portfolio returns in basis points, the portfolio standard deviations and the Sharpe ratio of portfolios long on the intraday rate when the portfolio models predict it to be positive.

Figure 2: AIC comparisons
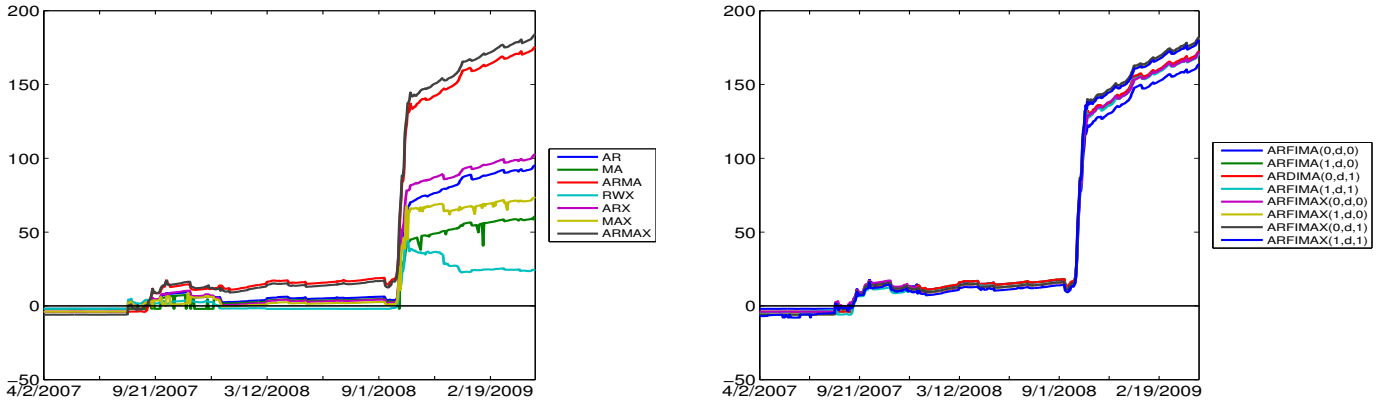
Notes: The graphs show differences in AIC (AIC(benchmark) - AIC(alternative)) over time relative the RW benchmark for the ARMA and ARFIMA models. If the benchmark model generates the better fit, then the AIC differences are negative.
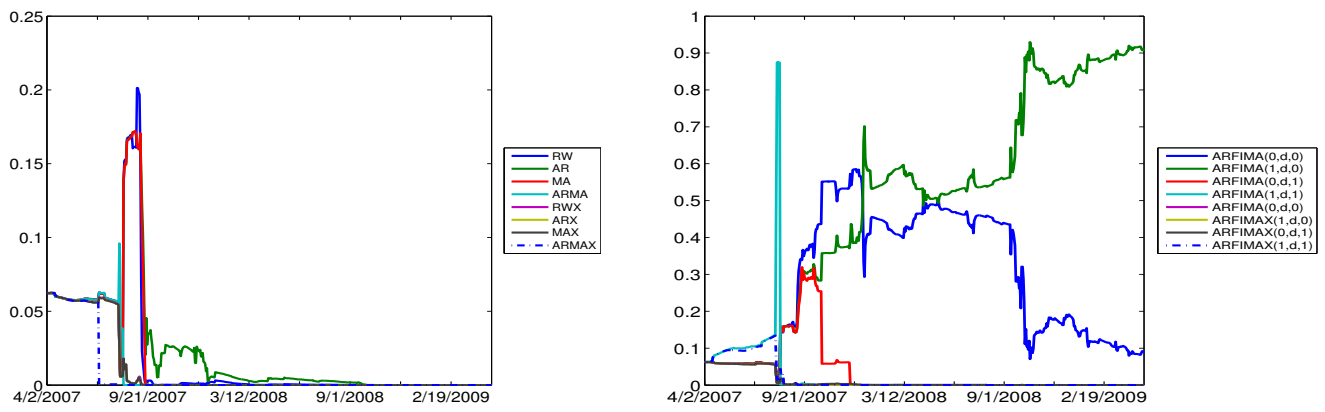
Figure 3
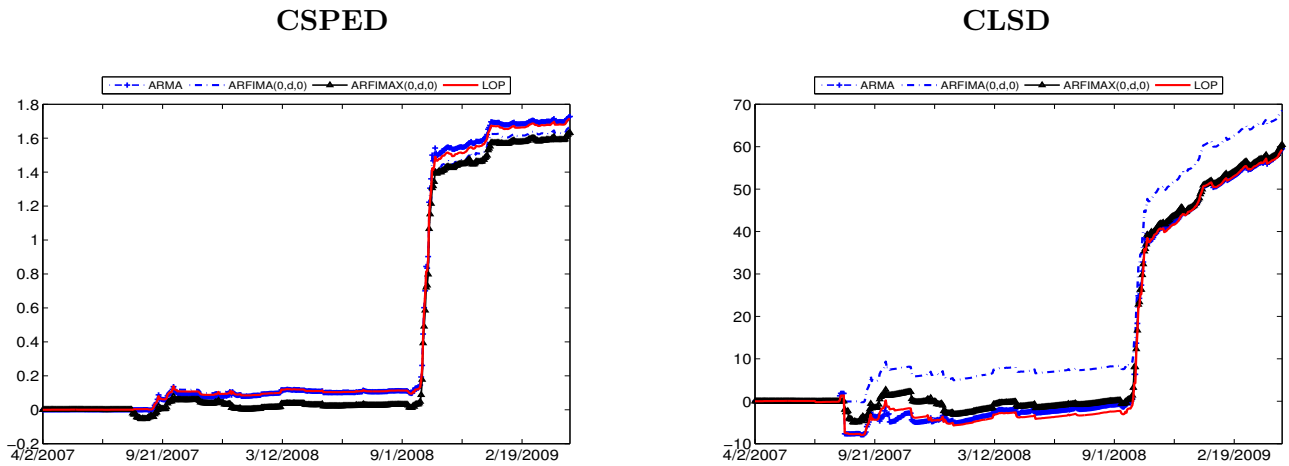
CSPED

CLSD

Note: The figures show cumulative square prediction error differences (CSPED) and cumulative log score differences (CLSD) of various models versus the RW benchmark to predict the intraday interbank rate.
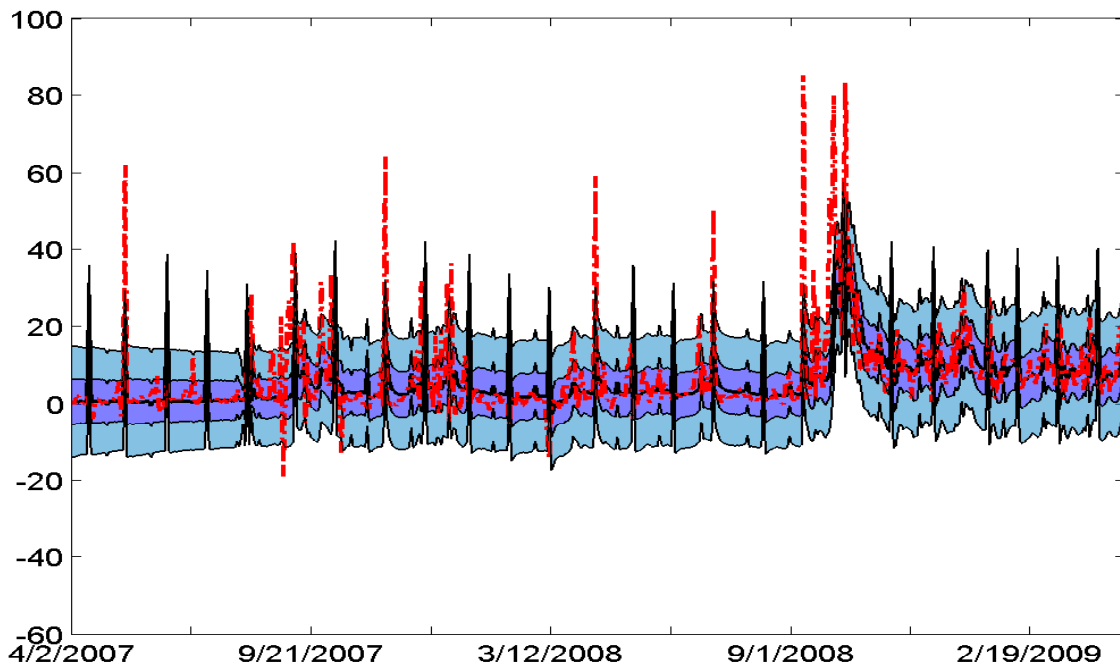
Figure 4: *PITS* histograms
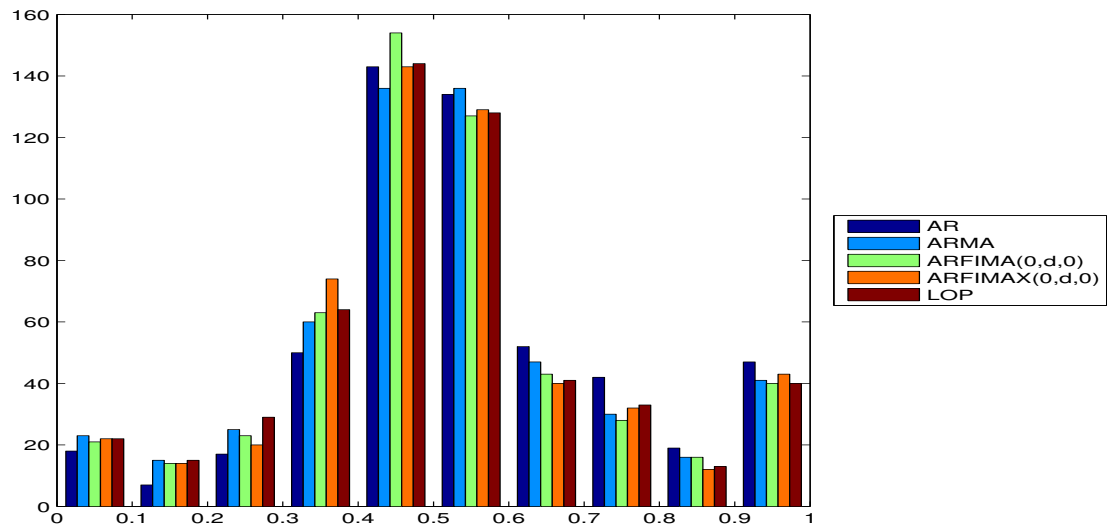
Note: The histograms show the decile counts of the *PITS* transforms for different models.

Figure 5: LOP weights

Note: The figures show LOP weights for different models.

Figure 6: Fan charts

Note: The figure shows the fan chart given by the LOP with 5th, 25th, 50th, 75th and 95th percentiles and the European overnight interbank rate (red dashed line).