BANCA D'ITALIA

EUROSISTEMA

Temi di discussione

(Working Papers)

Nearly exact Bayesian estimation of non-linear
no-arbitrage term structure models

by Marcello Pericoli and Marco Taboga

Number 1189

Temi di discussione

(Working Papers)

Nearly exact Bayesian estimation of non-linear
no-arbitrage term structure models

by Marcello Pericoli and Marco Taboga

The papers published in the Temi di discussione *series describe preliminary results and are made available to the public to encourage discussion and elicit comments.*

The views expressed in the articles are those of the authors and do not involve the responsibility of the Bank.

NEARLY EXACT BAYESIAN ESTIMATION OF NON-LINEAR NO-ARBITRAGE TERM STRUCTURE MODELS

by Marcello Pericoli* and Marco Taboga*

Abstract

We propose a general method for the Bayesian estimation of nonlinear no-arbitrage term structure models. The main innovations we introduce are: 1) a computationally efficient method, based on deep learning techniques, for approximating no-arbitrage model-implied bond yields to any desired degree of accuracy; and 2) computational graph optimizations for accelerating the MCMC sampling of the model parameters and of the unobservable state variables that drive the short-term interest rate. We apply the proposed techniques for estimating a shadow rate model with a time-varying lower bound, in which the shadow rate can be driven by both spanned unobservable factors and unspanned macroeconomic factors.

Contents

---

* Bank of Italy, Directorate General for Economics, Statistics and Research.

1 Introduction

In recent years no-arbitrage term structure models have become an essential tool in the toolbox of financial economists and central bankers. These models are routinely used to gauge market participants' views about future interest rate developments and to understand how the risk premia embedded in long-term interest rates evolve and interact with macroeconomic variables (see, e.g., Adrian et al. 2013, Ang and Piazzesi 2003, Hamilton and Wu 2012a and 2012b, Joslin et al 2014, Rudebusch and Wu 2008). However, unprecedented developments in the conduct of monetary policy and in bond markets across the world have created the need for increasingly complex models, whose estimation often poses econometric challenges that are difficult to surmount. For example, it has become evident that Gaussian affine term structure models, which were widely used in applied work thanks to their tractability and ease of estimation, can provide misleading indications[1] when policy rates are near their lower bound, as they have been for many currencies in recent years (e.g., Kim and Singleton 2012, Krippner 2013). Alternatives, such as shadow rate models (first introduced by Black 1995), have been proposed to deal with the problems posed by near-zero interest rates (e.g., Christensen and Rudebusch 2014). However, these alternative models are often difficult to estimate because they are highly non-linear and, in general, they lack analytic formulae for computing bond prices and for recovering pricing factors from the term structure of interest rates. In the specific case of shadow rate models, several approximations of the bond pricing function have been proposed (see Section 3 for a review), but as yet there seems to be no consensus on the best approximation strategy.

In this paper, we propose a general method for the Bayesian estimation of nonlinear no-arbitrage term structure models, aimed at circumventing the above difficulties.

---

[1]Gaussian affine models are typically unable to reproduce the stylized fact that policy rates tend to remain near their lower bound for prolonged periods of time once they reach it. Furthermore, when policy rates are near the lower bound, these models tend to assign high probability to future scenarios where policy rates decrease further and become significantly negative, which contradicts the very existence of a lower bound. Because of these unrealistic features, Gaussian term structure models produce flawed estimates of the expected future policy rates and the term premia embedded in the yield curve.

First, we propose to approximate the bond pricing function of non-linear models with a deep neural network. By appropriately choosing the architecture of the network (number of layers and number of neurons), the root mean squared approximation error (RMSE) can be made arbitrarily small (e.g., less than 1 basis points). The resulting approximation is not only nearly exact but it is also computationally inexpensive. Furthermore, it can be performed in any model in which the trajectories of the pricing factors under the risk-neutral measure can be simulated by Monte Carlo methods so as to obtain consistent estimators of bond prices. All of the most popular non-linear models, including shadow rate, quadratic (e.g., Ahn et al. 2002, Leippold and Wu 2002, Realdon 2006), and autoregressive gamma (e.g., Monfort et al. 2014) models satisfy this requirement.

Second, we propose an algorithm for the acceleration of the MCMC sampling of the model parameters and of the pricing factors that drive the short-term rate. The latter are treated as unobservable variables to be estimated. The only requirement is that the pricing factors follow a Markov process with known transition probability density. Also this requirement is satisfied by most existing models. The main intuition behind the MCMC acceleration technique, which falls within the broader class of computational graph optimization techniques (see, e.g., Theano Development Team 2016), is that at every transition of the Markov Chain only few of the calculations needed to compute the posterior density are actually performed, while the majority of computations can be "recycled" from previous transitions.

We apply the proposed techniques to the estimation of a shadow rate model with time-varying lower bound. The model, which is estimated with data on euro-denominated risk-free rates, has three latent pricing factors and two unspanned observable factors, namely Consensus Economics forecasts of inflation and GDP growth over a 1-year horizon.

We use the model to: 1) estimate how the perceived lower bound changed between the 2008 crisis and today; 2) carry out a decomposition of 10-year rates into expected rates and term premia, in order to understand how these components evolved in recent years; 3) predict future developments of the short-term (policy rate), by also performing scenario

analyses in which predictions are made conditional on inflation scenarios.

While a detailed analysis of the empirical results can be found in Section 10, the main highlights are as follows: 1) we find evidence of a marked decrease in the perceived lower bound, from 0.6% before the financial crisis to about -0.5% at the end of 2017; 2) the estimated term premium has been on a decreasing trend since the financial crisis; its value was negative (around -0.5%) in 2017, in line with the evidence provided by other researchers; 3) the predictions about lift-off probabilities are consistent with survey evidence at the end of the sample; the timing of the rise would anyway be highly dependent on the evolution of expected inflation and GDP growth.

The rest of the paper is organized as follows: Section 2 describes the class of non-linear models that can be estimated with our methodology; Section 3 explains how the bond pricing function can be approximated with deep neural networks; Section 4 derives formulae for the posterior distribution of model parameters; Section 5 briefly describes the MCMC algorithm used to perform draws from the posterior distribution; Section 6 explains how to optimize the computational graph of the MCMC sampler; Section 7 introduces the shadow rate model; Section 8 presents the data used in the estimation of the model; Section 9 provides more details about the estimation methodology; Section 10 comments on the empirical evidence from the model; Section 11 concludes.

2   The model

This section describes the class of term structure models that can be estimated with the methodology proposed in this paper.

Time is discrete and is indexed by $t = 1, \ldots, T$. We assume that the short-term rate r_t (i.e., the interest rate on 1-period loans) is a known function

$$r_t = g\left(X_t, \theta_r\right) \tag{1}$$

of a vector X_t of pricing factors and of a vector of parameters θ_r.

The factors X_t follow a first-order discrete-time Markov process whose transition probability density

$$f\left(X_t \,|\, X_{t-1}, \theta_{\mathrm{P}}\right) \tag{2}$$

under the physical (or real-world) measure P is also a known function, depending on a vector of parameters θ_{P}.

The vast majority of term-structure models satisfies assumptions (1) and (2). For example,

- in the classical GATSM (Gaussian Affine Term Structure Model), the short-term rate is an affine function of the factors:

$$r_t = a + bX_t \tag{3}$$

with parameter $\theta_r = (a, b)$, and X_t follows a first-order vector autoregression

$$X_t = \mu + \rho X_{t-1} + \Sigma \varepsilon_t \tag{4}$$

where ε_t is Gaussian noise and $\theta_{\mathrm{P}} = (\mu, \rho, \Sigma)$.

- in QTSM (Quadratic Term Structure Models), the law of motion of the factors is specified by (4), but the short term rate is a quadratic function of the factors:

$$r_t = a + bX_t + X_t^\top C X_t \tag{5}$$

with parameter $\theta_r = (a, b, C)$

- in SRTSM (Shadow Rate Term Structure Models), the law of motion of the factors is

specified by (4), but the short term rate is a non-linear function of the factors:

$$r_t = \max \left(\underline{r}, a + bX_t + X_t^\top C X_t \right) \tag{6}$$

where $\underline{r}$ is a lower bound for the short-term rate and $\theta_r = (\underline{r}, a, b, C)$.

- in autoregressive Gamma and extended autoregressive Gamma models, the conditional distribution of X_t is a mixture of Gamma distributions whose parameters depend on X_{t-1}.

The third assumption we make is that X_t is a first-order Markov process also under the risk-neutral pricing measure Q. Its transition probability density under Q is denoted by

$$h \left(X_t \, | X_{t-1}, \theta_Q \right) \tag{7}$$

where θ_Q is a vector of unknown parameters.

Furthermore, we assume discrete compounding, so that the price of a zero-coupon bond expiring in n periods is

$$P_t^n = \mathrm{E}^Q \left[\frac{1}{\prod_{j=0}^{n-1} (1 + r_{t+j})} \, \middle| \, X_t \right] \tag{8}$$

and its yield is

$$y_t^n = \left(\frac{1}{P_t^n} \right)^{1/n} - 1 \tag{9}$$

Finally, we assume that trajectories of X_t under Q can be simulated by Monte Carlo methods so as to obtain consistent estimators of bond prices. In particular, we assume that it is possible to generate double sequences $\{X_{t,s}\}$ such that, for every $n \in \mathbb{N}$

$$\operatorname*{plim}_{S \to \infty} \frac{1}{S} \sum_{s=1}^{S} \frac{1}{\prod_{j=0}^{n-1} (1 + g\left(X_{t+j,s}, \theta_r \right))} = P_t^n \left(\theta_r, \theta_Q, X_t \right) \tag{10}$$

where plim denotes a limit in probability and P_t^n depends not only on θ_r, but also on

the parameter θ_Q used to generate $\{X_{t,s}\}$ and on the initial value X_t of the state variables, which is used to recursively generate the trajectories $\{X_{t+j,s}\}$, by starting from the transition density

$$h\left(X_{t+1} \,|\, X_t, \theta_Q\right) \tag{11}$$

The above assumptions about the dynamics of Q are satisfied by virtually all existing dynamic term structure models. In particular, affine, quadratic and shadow rate models where X_t follows the dynamics (4) under P usually have a pricing kernel such that X_t follows a first-order VAR also under Q:

$$X_t = \bar{\mu} + \bar{\rho} X_{t-1} + \Sigma \eta_t \tag{12}$$

where the parameters $\bar{\mu}$ and $\bar{\rho}$ can be different from the parameters μ and ρ in eq. (4) and η_t is multivariate normal. In these cases, the trajectories of the pricing factors are computed by extracting double sequences of Gaussian error terms $\{\eta_{t,s}\}$ and computing $\{X_{t+j,s}\}$ recursively.

Let $n_1, \ldots, n_M$ be M maturities of interest. Assume they are in increasing order, so that n_M is the largest bond maturity. Denote by

$$y_t = \left[\begin{array}{cccc} y_t^{n_1} & y_t^{n_2} & \ldots & y_t^{n_M} \end{array}\right]^{\top} \tag{13}$$

the vector of yields. When the dependence of the vector of yields on model parameters and pricing factors needs to be emphasized, we will write

$$y_t = y\left(\theta_r, \theta_Q, X_t\right) \tag{14}$$

where the function y does not depend explicitly on t because the Markov property implies that all the information at time t is summarized by the value of the vector X_t.

We assume that the yields are observed with error, so that the observed yields are

$$y_t^o = y_t + v_t \tag{15}$$

where $\{v_t\}$ is a sequence of IID pricing errors having known density

$$f\left(v_t \left| \theta_{\mathrm{v}}\right.\right) \tag{16}$$

where θ_{v} is a vector of parameters. In most of the term structure models proposed in the literature, v_t is assumed to have a multivariate normal distribution with zero mean. In this case, the parameter θ_{v} is the covariance matrix of the pricing errors.

3   Pricing function approximation

When an analytical expression for the pricing function $y\left(\theta_r, \theta_{\mathrm{Q}}, X_t\right)$ is not available, it is usually approximated in order to perform model estimation. For example, in the context of SRTSM estimation, the following approximation methods have been proposed: i) lattices (Ichiue and Ueno 2007); ii) finite-difference methods (Kim and Singleton 2012); iii) approximation with a finite state space constructed with antithetic sampling (Bauer and Rudebusch 2016); iv) option pricing approximations (Krippner 2013 and Christensen and Rudebusch 2014); v) approximations that ignore Jensen inequality corrections so as to exploit analytical results on truncated normal distributions (Ichiue and Ueno 2013); vi) perturbations and Taylor series expansions around a deterministic steady state (Andreasen and Meldrum 2013); vi) approximations of forward rates (Wu and Xia 2016).

In this paper, we propose to use neural networks (NN) to approximate the pricing function. As proved, for example, by Hornik (1991), NN are universal approximators, in the sense that a suitably parametrized NN can approximate any given function to any desired degree of accuracy provided that it contains a sufficient amount of neurons. In recent years,

the popularity of NN approximation has increased for several reasons.

First, both theoretical and empirical studies have shown that deep networks (i.e., NN with many layers) are parsimonious approximators of complicated functions (e.g., Delalleau and Bengio 2011). These studies have given rise to so-called deep learning, a set of statistical methods related to the use of deep NN that have been fruitfully applied in disparate fields of engineering and artificial intelligence (LeCun et al. 2015).

Second, computer micro-architectures (both CPUs and GPUs) and numerical software have become increasingly more rapid and efficient in performing the kind of calculations involved in the estimation and computation of NN^2 (e.g., NVIDIA 2016).

Third, research on NN has progressed at a very rapid pace and has substantially improved the optimization algorithms that are used to estimate the parameters of NN (see Ruder 2016 for a review), as well as the symbolic computational engines that are used to implement the algorithms (Bahrampour et al. 2015).

Denote by

$$\overline{\theta} = (\theta_r, \theta_Q, X_t) \tag{17}$$

the triplet of inputs (i.e., arguments) to the pricing function, so that the notation

$$y = y\left(\overline{\theta}\right) \tag{18}$$

can be used interchangeably with the notation in eq. (14).

Abstracting from details that will be further discussed in Section 9, the procedure we use to construct and train a neural network to approximate the pricing function for a given model is as follows:

1. we randomly extract P values $\overline{\theta}_1, \ldots, \overline{\theta}_P$ from the set of all admissible values of $\overline{\theta}$;

---

[2]Roughly speaking, the majority of floating point operations performed by a computer in the process of computing or estimating a neural network are typically done while multiplying matrices (of weights and neuron activations). Matrix multiplication, being highly parallelizable, has become increasingly rapid in recent years thanks to the increase in the number of computational cores in CPUs and GPUs and to the introduction of parallel instructions (such as AVX) in modern CPUs.

2. for each $\overline{\theta}_p$ $(p = 1, \ldots, P)$, we compute an approximation

$$\widetilde{y}_p = \widetilde{y}\left(\overline{\theta}_p\right) \tag{19}$$

of the bond yields by Monte Carlo integration as in equation (10), that is, we compute

$$\widetilde{y}_p^{n_j} = \left(\frac{1}{S}\sum_{s=1}^{S}\frac{1}{\prod_{j=0}^{n_j-1}\left(1 + g\left(X_{t+j,s}, \theta_r\right)\right)}\right)^{-1/n_j} - 1 \tag{20}$$

for $j = 1, \ldots, M$; the number S of random trajectories of $X_{t+j,s}$ is chosen in such a way that the mean squared error of the Monte Carlo approximation is less than a pre-specified threshold T_{MC};

3. we subdivide the sample of couples $\left(\overline{\theta}_p, \widetilde{y}_p\right)$ obtained in step 1) and 2) into a training sample of $P_{train} < P$ couples, a validation sample of $\left\lfloor\frac{1}{2}\left(P - P_{train}\right)\right\rfloor$ couples and a test sample comprising the remaining couples;

4. we use the training set to train a sequence of increasingly complex (i.e., endowed with more neurons or more hidden layers) neural networks

$$\widehat{y} = \widehat{y}\left(\theta\right) \tag{21}$$

that take a vector of parameters $\overline{\theta}$ as input and produce an approximation $\widehat{y}$ to the vector of yields as output; for each network, we use the validation sample to determine when to stop training based on the out-of-sample accuracy of the predictions of the network (by checking the generalization loss; e.g., Prechelt 1998); we stop increasing the complexity of the neural network when the mean squared error of its predictions on the validation sample is less than a pre-specified threshold T_{NN}; finally, we use the test sample to check that the accuracy of the network, as estimated on the validation sample, has not been biased by repeated use of the validation sample.

In other words, we use Monte Carlo integration to construct a sample of examples of the mapping from pricing factors and parameters to bond yields, and then we train a neural network to reproduce the mapping with high accuracy (also for examples it has never seen).

Once the neural network $\widehat{y}\left(\overline{\theta}\right)$ has been estimated (or trained, in machine learning parlance), it can be used to approximate the pricing function $y\left(\overline{\theta}\right)$ whenever the latter needs to be computed. Note that the training of the network (including the Monte Carlo simulations that are used to construct the training sample) is performed only once, before the estimation of the term structure model. Then, the trained network is repeatedly used in the estimation of the term structure model (see Section 6). For example, when new data becomes available, the term-structure model can be re-estimated without re-training the neural network.

4   The posterior distribution of parameters

In this section, we derive the posterior distribution of model parameters and pricing factors, based on the assumptions made in Section 2.

We first deal with the simpler case in which all factors X_t are unobservable and we then analyze the more complex case in which some of the factors in X_t are observable but unspanned.

4.1   Only unobservable factors

Denote the matrix of observed yields by

$$Y^o = \left[\begin{array}{cccc} y_1^o & y_2^o & \ldots & y_T^o \end{array} \right] \tag{22}$$

the matrix of theoretical yields by

$$Y = \left[\begin{array}{cccc} y_1 & y_2 & \ldots & y_T \end{array} \right] \tag{23}$$

and the matrix of pricing factors by

$$X = \begin{bmatrix} X_1 & X_2 & \dots & X_T \end{bmatrix} \tag{24}$$

We estimate quantities of interest by simulating from the posterior distribution of the parameters and of the latent states:

$$f\left(\theta_{\mathrm{P}}, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}, X \mid Y^o\right) \propto f\left(Y^o \mid \theta_{\mathrm{P}}, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}, X\right) f\left(\theta_{\mathrm{P}}, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}, X\right) \tag{25}$$

Prices P_t^n and yields y_t^n depend only on the risk neutral dynamics and on the functional relation between the factors and the short-term rate. Therefore[3],

$$f\left(Y^o \mid \theta_{\mathrm{P}}, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}, X\right) = f\left(Y^o \mid \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}, X\right) \tag{26}$$

Furthermore, we have that

$$
\begin{aligned}
f\left(\theta_{\mathrm{P}}, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}, X\right) &= f\left(X \mid \theta_{\mathrm{P}}, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}\right) f\left(\theta_{\mathrm{P}}, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}\right) \\
&= f\left(X \mid \theta_{\mathrm{P}}\right) f\left(\theta_{\mathrm{P}}, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}\right)
\end{aligned} \tag{27}
$$

because the density of X depends only on the physical dynamics of the factors.

We assume that the prior on the parameters is uniform improper on the set of admissible values

$$f\left(\theta_{\mathrm{P}}, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}\right) \propto 1_A\left(\theta_{\mathrm{P}}, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}\right) \tag{28}$$

where 1_A is the indicator of the set A of admissible values for the parameters (which depends

---

[3]For the sake of simplicity, we are assuming that θ_{P} and θ_{Q} have no shared entries. This assumption can be relaxed, at the cost of significantly increasing the complexity of the notation used in what follows. Furthermore, the restriction that θ_{P} and θ_{Q} have no shared entries can be usually imposed as an identification restriction, in order to uniquely pin down the likelihood of the pricing errors of the term structure model. For example, in Section 7, we present a shadow rate model where the pricing factors follow a VAR both under P and under Q, and the error terms of the two VARs have the same covariance matrix. However, the covariance matrix can be restricted to be equal to the identity matrix, so as to identify the model, and it thus disappears from the vectors of parameters to be estimated.

on the specific model being analyzed).

By putting together equations (25), (26), (27) and (28), we obtain the posterior density

$$f\left(\theta_{\mathrm{P}}, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}, X \,|\, Y^o\right) \propto f\left(Y^o \,|\, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}, X\right) f\left(X \,|\, \theta_{\mathrm{P}}\right) 1_A\left(\theta_{\mathrm{P}}, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}\right) \qquad (29)$$

The three terms in the above product are straightforward to compute. The first term is the likelihood of the pricing errors and can be expanded as follows:

$$f\left(Y^o \,|\, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}, X\right) = \prod_{t=1}^{T} f\left(y_t^o - \widehat{y}\left(\theta_r, \theta_{\mathrm{Q}}, X_t\right) \,|\, \theta_{\mathrm{v}}\right) \qquad (30)$$

The second term is the likelihood of the pricing factors and can be written as

$$f\left(X \,|\, \theta_{\mathrm{P}}\right) = \prod_{t=1}^{T} f\left(X_t \,|\, X_{t-1}, \theta_{\mathrm{P}}\right) \qquad (31)$$

thanks to the Markov property of X_t.

Finally, the third term

$$1_A\left(\theta_{\mathrm{P}}, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}\right) \qquad (32)$$

is an indicator that takes value 1 on the parameter space and 0 elsewhere.

4.2   Unobservable and observable unspanned factors

We now relax the assumption that all the factors are unobserved and we allow for some observable but unspanned factors (see Joslin et al. 2014 for a definition of unspanned factors). For notational simplicity, we keep denoting the unobservable factors by X_t, while the unspanned observable factors are denoted by X_t^o (by X^o when they are stacked into a matrix).

In this case, the posterior distribution of the parameters and the latent states is

$$f\left(\theta_{\mathrm{P}}, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}, X \,|\, Y^o, X^o\right) \propto f\left(Y^o, X^o \,|\, \theta_{\mathrm{P}}, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}, X\right) f\left(\theta_{\mathrm{P}}, \theta_r, \theta_{\mathrm{Q}}, \theta_{\mathrm{v}}, X\right) \qquad (33)$$

The fact that X^o is unspanned implies that Y^o and X^o are independent conditional on X. Therefore,

$$f\left(Y^o, X^o \,|\, \theta_\mathrm{P}, \theta_r, \theta_\mathrm{Q}, \theta_\mathrm{v}, X\right) = f\left(Y^o \,|\, \theta_\mathrm{P}, \theta_r, \theta_\mathrm{Q}, \theta_\mathrm{v}, X\right) f\left(X^o \,|\, \theta_\mathrm{P}, \theta_r, \theta_\mathrm{Q}, \theta_\mathrm{v}, X\right) \tag{34}$$

and

$$f\left(\theta_\mathrm{P}, \theta_r, \theta_\mathrm{Q}, \theta_\mathrm{v}, X \,|\, Y^o, X^o\right) \propto f\left(Y^o \,|\, \theta_\mathrm{P}, \theta_r, \theta_\mathrm{Q}, \theta_\mathrm{v}, X\right) f\left(\theta_\mathrm{P}, \theta_r, \theta_\mathrm{Q}, \theta_\mathrm{v}, X, X^o\right) \tag{35}$$

From the previous section, we know that

$$f\left(Y^o \,|\, \theta_\mathrm{P}, \theta_r, \theta_\mathrm{Q}, \theta_\mathrm{v}, X\right) = f\left(Y^o \,|\, \theta_r, \theta_\mathrm{Q}, \theta_\mathrm{v}, X\right) \tag{36}$$

Furthermore, we have that

$$\begin{aligned} f\left(\theta_\mathrm{P}, \theta_r, \theta_\mathrm{Q}, \theta_\mathrm{v}, X, X^o\right) &= f\left(X, X^o \,|\, \theta_\mathrm{P}, \theta_r, \theta_\mathrm{Q}, \theta_\mathrm{v}\right) f\left(\theta_\mathrm{P}, \theta_r, \theta_\mathrm{Q}, \theta_\mathrm{v}\right) \\ &= f\left(X, X^o \,|\, \theta_\mathrm{P}\right) f\left(\theta_\mathrm{P}, \theta_r, \theta_\mathrm{Q}, \theta_\mathrm{v}\right) \end{aligned} \tag{37}$$

because the joint density of X and X^o depends only on the physical dynamics of the factors. By putting the above equations together, we have that

$$f\left(\theta_\mathrm{P}, \theta_r, \theta_\mathrm{Q}, \theta_\mathrm{v}, X \,|\, Y^o, X^o\right) \propto f\left(Y^o \,|\, \theta_r, \theta_\mathrm{Q}, \theta_\mathrm{v}, X\right) f\left(X, X^o \,|\, \theta_\mathrm{P}\right) f\left(\theta_\mathrm{P}, \theta_r, \theta_\mathrm{Q}, \theta_\mathrm{v}\right) \tag{38}$$

Therefore, the only difference with respect to (29) is that the likelihood of the pricing factors needs to be substituted with the joint likelihood of the pricing factors and the unspanned observable factors.

5 The MCMC baseline algorithm

The MCMC algorithm used to generate draws from the posterior distributions of the parameters is the random walk Metropolis-Hastings algorithm[4] with block structure (e.g., Bagasheva et al. 2008). Each parameter (including the individual entries of X) is treated as a separate block. While the algorithm is standard, we briefly describe it in this section, in order to set out the notation that will be used in the next section, where we propose a set of computational graph optimizations aimed at making the algorithm more efficient.

Stack all the parameters $\theta_P, \theta_r, \theta_Q, \theta_v$ and X in a vector θ (by merging parameters that are shared by θ_P and θ_Q if necessary), and denote its entries by $\theta_1, \ldots, \theta_N$ (each entry constitutes a block).

The vector of parameters is randomly drawn $J_B + J_K$ times and the first J_B draws are discarded (they are a so-called burn-in sample, also used for tuning the transition density of the chain). The last J_K draws are instead kept and constitute a sample of serially dependent draws (following a Markov chain) from the posterior distribution of θ. The value of θ at the j-th iteration of the Markov Chain is denoted by θ^j and its i-th entry by θ_i^j.

The posterior density of a generic draw θ^j, denoted by $f\left(\theta^j | Y^o\right)$, is known up to a constant of proportionality that does not depend on θ^j.

Define a $N \times 1$ vector κ of standard deviations of the random-walk increments that will be adaptively adjusted during the burn-in phase in order to target a given acceptance rate.

The chain starts from an admissible value θ^0. The j-th iteration is made up of the following steps:

1. set $l = j - N \lfloor j/N \rfloor$ where $N \lfloor j/N \rfloor$ denotes the floor of j/N;

2. draw a random number z_j from a standard normal distribution;

---

[4]Gibbs sampling could be performed on some blocks of parameters, but not on those entering the pricing function, as the latter is non-linear in the parameters (as a consequence their full conditional distributions cannot be derived analytically). As the computational cost of the non-linear part dominates that of the linear part of the model, performing Gibbs sampling with Metropolis steps does not afford any significant computational advantage, also in view of the graph optimizations presented in the next section.

3. build a new $N \times 1$ vector η such that $\eta_i = \theta_i^{j-1}$ for $i \neq l$ and $\eta_i = \theta_i^{j-1} + \kappa_i z_j$ for $i = l$;

4. compute the acceptance probability a_j as follows:

$$a_j = \min\left(1, \frac{f\left(\eta \,|Y^o\right)}{f\left(\theta^j \,|Y^o\right)}\right) \tag{39}$$

5. draw a random number u_j from the uniform distribution on $[0, 1]$;

6. if $u_j \leq a_j$ then set $\theta^j = \eta$; otherwise, set $\theta^j = \theta^{j-1}$;

7. if $j \leq J_B$, adjust k_l [5];

8. if $j = J_B + J_K$ end the algorithm, otherwise go back to step 1.

6 Computational graph optimizations

A naive implementation of the MCMC algorithm described in the previous section is prohibitively computationally expensive, as it requires T evaluations of the pricing function $\widehat{y}$ for each iteration of the chain. However, it is possible to drastically cut down the computational burden by optimizing the computational graph of the algorithm.

At the j-th iteration of the algorithm, the computation of the numerator and denominator of the ratio

$$\frac{f\left(\eta \,|Y^o\right)}{f\left(\theta^j \,|Y^o\right)} \tag{40}$$

can be decomposed into a number of inter-dependent computations each of which may depend only on a subset of the parameter vector. Since only one entry of η (i.e., η_l) is different from

---

[5] For example, in the application presented below, the adjustment targets an acceptance rate between 7.5 and 15 per cent. We initially set $k_i = 0.0001$ for all i. Then, at each iteration, if an exponentially weighted moving average (with forgetting factor equal to 0.99) of past acceptance indicators (1 in case of acceptance and 0 otherwise) is below 7.5 per cent for the parameter θ_i, we decrease κ_i by a factor of 0.99; if the same moving average is above 15 per cent, we increase κ_i by a factor of 1.01. This choice of parameters for adjusting κ, although admittedly arbitrary, maintained the acceptance rate broadly on target in several MCMC simulations for different models.

the corresponding entry of θ^j (i.e., θ^j_l) and all the other entries are equal (i.e., $\eta_i = \theta^{j-1}_i$ for $i \neq l$), we need to perform only the computations that depend directly or indirectly on η_l.

In more formal terms, $f(\eta | Y^o)$ can be written as

$$f(\eta | Y^o) = \lambda_C \left(\lambda_{C-1,1}, \ldots, \lambda_{C-1,R_{C-1}}, \mathrm{sub}_C(\eta), Y^o \right) \tag{41}$$

where there are $C \in \mathbb{N}$ layers of computation, $R_c \in \mathbb{N}$ computations in layer c (for $c = 1, \ldots, C$), and the computations $\lambda_{c,r}$ $(r = 1, \ldots, R_c)$ recursively depend on the computations in the previous layer:

$$\lambda_{c,r} \left(\lambda_{c-1,1}, \ldots, \lambda_{c-1,R_{c-1}}, \mathrm{sub}_{c,r}(\eta), Y^o \right) \tag{42}$$

Furthermore, a given computation $\lambda_{c,r}$ depends on a subset of the entries of η that is denoted by $\mathrm{sub}_{c,r}(\eta)$, and computations in the first layer $(c = 1)$ do not depend on previous computations:

$$\lambda_{1,r} \left(\mathrm{sub}_{1,r}(\eta), Y^o \right) \tag{43}$$

Note that a computation $\lambda_{c,r}$ needs to be performed only if there exist integers $k \leq c$ and $s \leq R_k$ such that $\eta_l \in \mathrm{sub}_{k,s}(\eta)$. On the contrary, if $\lambda_{c,r}$ does not depend directly or indirectly on η_l, then $\lambda_{c,r}$ does not need to be computed again and it takes the same value that it took when $f(\theta^j | Y^o)$ was computed (we call $\lambda_{c,r}$ a "recycled computation").

By avoiding redundant computations, the cost of the MCMC algorithm can decrease by some orders of magnitude (about two in the example presented below).

Which computations can be recycled depends on the specifics of the term structure model[6]. However, there are some computations that can be recycled in all models belonging to the general class of models described in Section 2. The following is a list of the most important ones:

---

[6] Although the recycling strategy outlined above is completely general. In particular, it can be applied to any state space model in which the unobservable states are drawn as separate blocks in a Metropolis step.

1. when η_l is one of the entries of X_t, then only the likelihood

$$f\left(y_t^o - \widehat{y}\left(\theta_r, \theta_Q, X_t\right)|\theta_v\right) \tag{44}$$

of the pricing errors at time t needs to be computed in order to compute (30), while the computation of the likelihoods

$$f\left(y_s^o - \widehat{y}\left(\theta_r, \theta_Q, X_s\right)|\theta_v\right) \tag{45}$$

for $s \neq t$ can be recycled from the previous iteration of the MCMC algorithm. Furthermore, in order to compute (31), we need to re-calculate only the likelihoods

$$f\left(X_t|X_{t-1}, \theta_P\right) \tag{46}$$

and

$$f\left(X_{t+1}|X_t, \theta_P\right) \tag{47}$$

of the pricing factors at times t and $t+1$ while the likelihoods at all other times remain unchanged. Finally, the prior (32) on the parameters does not depend on X_t and does not need to be re-computed.

2. when η_l is one of the entries of θ_P, the likelihood of the pricing errors (30) remains unchanged (its computation is recycled) and only the likelihood of the pricing factors (31) and the prior (32) need to be computed[7].

3. when η_l is one of the entries of θ_Q or of θ_r, the likelihood of the pricing errors (30) and the prior (32) need to be re-computed, while the likelihood of the pricing factors (31) remains unchanged.

4. when η_l is one of the entries of θ_ν, not only the likelihood of the pricing factors (31)

---

[7]But see the comments on shared entries in Section 4.

remains unchanged, but also some of the computations involved in the calculation of the likelihood of the pricing errors (30) can be recycled: for all t, we need to re-compute

$$f\left(y_t^o - \widehat{y}\left(\theta_r, \theta_Q, X_t\right) | \theta_v\right) \tag{48}$$

but the value of $y_t^o - \widehat{y}\left(\theta_r, \theta_Q, X_t\right)$ is left unchanged from the previous MCMC step.

Further recycling may be possible in specific model settings. For example,

1. if X_t follows a VAR under P, i.e.,

$$X_t = \mu + \rho X_{t-1} + \Sigma\varepsilon_t \tag{49}$$

and η_l is one of the entries of μ, then we do not need to re-compute the products ρX_{t-1} in order to calculate (31);

2. if the pricing errors have a Gaussian density and η_l is one of the entries of θ_Q, θ_r or X_t, then the likelihood of the pricing errors (30) depends on the determinant and on the inverse of the covariance matrix of the errors, which do not need to be re-computed because θ_v has not changed from the previous MCMC iteration;

3. if the priors on the parameters $\theta_P, \theta_r, \theta_Q, \theta_v$ are independent, then (32) can be decomposed into four factors, and only one of them needs to be recomputed at each iteration of the MCMC algorithm.

7   A shadow rate model

This section describes the SRTSM that we estimate to exemplify the methodology illustrated in the previous sections. The model has three unobservable pricing factors and two unspanned observable factors.

The short term rate r_t, that is, the interest rate on risk-free bonds expiring in one-period, is equal to the lower bound $\underline{r_t}$ (which can be time-varying) or to the shadow rate s_t, whichever is larger:

$$r_t = \max\left(\underline{r_t}, s_t\right) \tag{50}$$

The shadow rate is an affine function of a 3×1 vector X_t of unobservable pricing factors

$$s_t = a + bX_t \tag{51}$$

where a is a scalar and b is a 1×3 vector of factor loadings.

Denote by X_t^o the 2×1 vector of unspanned observable factors. We assume that X_t and X_t^o jointly follow a first order vector autoregression under the real-world measure P:

$$\begin{bmatrix} X_t \\ X_t^o \end{bmatrix} = \mu + \rho \begin{bmatrix} X_{t-1} \\ X_{t-1}^o \end{bmatrix} + S\varepsilon_t \tag{52}$$

where μ is a 5×1 drift vector, ρ is a 5×5 autoregression matrix, S is a 5×5 volatility matrix and ε_t is a sequence of IID standard multivariate normal 5×1 random vectors.

Under standard assumptions on the functional form of the pricing kernel, the dynamics of X_t under the risk-neutral pricing measure Q are

$$X_t = \overline{\mu} + \overline{\rho}X_{t-1} + \Sigma\eta_t \tag{53}$$

where $\overline{\mu}$ is a 3×1 drift vector and $\overline{\rho}$ is a 3×3 autoregression matrix. Furthermore, the volatility matrix Σ is equal to the 3×3 upper left block of S and the sequence of errors η_t is also IID standard multivariate normal.

We impose the following restrictions on the parameters of the model:

1. $\Sigma = I$;

2. $\overline{\mu} = 0$;

3. $\overline{\rho}_{ii} \geq \overline{\rho}_{jj}$ if $i > j$;

4. $b \geq 0$;

5. $\overline{\rho}$ is diagonal;

While restrictions 1-4 are necessary for identification, restriction 5 is over-identifying and could be replaced by the weaker requirement that $\overline{\rho}$ be an upper diagonal matrix satisfying the real Schur property (see Pericoli and Taboga 2012). Also note that restriction 1 implies that the upper left block of S is equal to the identity matrix and that θ_{P} and θ_{Q} have no elements in common.

The time-varying lower bound $\underline{r}_t$ is modeled as an exogenous smooth function of time. In particular, we assume that $\underline{r}_t$ can be interpolated by a truncated radial basis function network:

$$\underline{r}_t = w_0 + \sum_{i=1}^{D} w_i \varphi \left(\frac{(t - t_i)^+}{q_i} \right) \tag{54}$$

where D is the number of kernels in the network, φ is the probability density function of a standard normal distribution, w_i are the network weights and t_i and q_i are the locations and bandwidths of the kernels.

The assumption that $\underline{r}_t$ is exogenous means that the possibility of future changes in $\underline{r}_t$ is not priced into bond yields. In other words, at any point in time, investors take $\underline{r}_t$ as given and they do not expect it to be revised. Arguably, this is a semantically coherent assumption: a bound is not really a bound if agents expect that it can be moved downwards at any time. In this respect, our point of view is different from that of Wu and Xia (2017), who argue that "*A model with a time-varying lower bound would be internally inconsistent if agents in the model were naive, never updated their beliefs, and always thought the lower bound would stay where it currently was even after seeing it move repeatedly*". See Section 10 for further remarks about the concept of lower bound.

The distribution of the pricing errors is assumed to be multivariate normal with zero mean and diagonal covariance matrix V.

We assume that the prior on the parameters is uniform improper on the set of admissible values

$$f\left(a, b, \mu, \rho, \overline{\rho}, S, V\right) \propto 1_A\left(a, b, \mu, \rho, \overline{\rho}, S, V\right) \tag{55}$$

where 1_A is the indicator of the set A of admissible values for the parameters. The specification of A can be found in the Appendix.

8   The data

We use end-of-quarter data on the term-structure of euro-denominated Overnight Indexed Swap (OIS) rates. OIS are swap contracts where one counterparty receives a variable payment indexed to the interest rate on overnight unsecured interbank deposits between prime banks, and the other counterparty receives the fixed OIS rate. Because overnight interbank deposits between prime banks are considered virtually risk free, OIS rates are deemed a very good proxy for risk free rates[8] (e.g., Morini 2009, Mercurio 2009, Ejsing et al. 2012, Taboga 2014). Given the timing of the payments of OIS contracts, a simple recursive calculation allows to extract a term structure of zero-coupon spot rates from the term-structure of OIS rates. We use the 3-month and 6-month maturities as well as all the yearly maturities from 1 to 10 years. OIS rates, downloaded from Bloomberg, are available for all maturities since 2005. We backfill the dataset back to 1999 with zero-coupon German government bond rates. The last observations refer to the beginning of December 2017 (when we last updated our dataset), which is used in the estimation and treated as if it was the end of the last quarter of 2017.

The two observable unspanned factors in our model are expected inflation and GDP growth 1-year forward derived from Consensus Economics forecasts data. In particular, the

---

[8]By risk free we mean free of credit and liquidity risk. Of course, long-term OIS rates are subject to interest rate risk.

1-year expectation is a weighted average of the median forecast for the current and next years, with weights equal to the proportions of the next 12 months belonging to each of the two years (so, for example, the 1-year forecast made in September 2015 is a weighted average of the forecasts for 2015 and 2016 with weights 3/12 and 9/12 respectively).

9   Estimation

Since only the three spanned factors enter the pricing function of the shadow rate model described in Section 7, the vector of yields is a function

$$y = y\left(a, b, \overline{\rho}, X_t, \underline{r_t}\right) \tag{56}$$

of 11 parameters ($b, \overline{\rho}$ and X_t have dimension 3, while a and $\underline{r_t}$ have dimension 1). Furthermore, the vector y has 12 components, as we use all the maturities in our sample to estimate the models. Thus, the pricing function is $y : \mathbb{R}^{11} \to \mathbb{R}^{12}$.

The training sample for the neural network that approximates the pricing function y is made up of one million uniform draws of the parameters from their support (see the Appendix for details). For each of these draws, a Monte Carlo approximation $\widetilde{y}$ of the no-arbitrage yields is computed by generating $500,000$ trajectories of X_t, which are then doubled by antithetic sampling. This gives rise to an estimated standard deviation of the Monte Carlo approximation error below 1 basis point (across maturities and samples of parameters). The sample of parameters and approximated prices is then split into a training set of $800,000$ units and a validation and test set of $100,000$ units each.

Starting from simpler network architectures and progressively increasing their complexity, we stop when we find an architecture that gives a root mean squared error of less than one basis point on the validation set. This is achieved with a 6-layer network (the input and output layer, and 4 intermediate layers). All the intermediate layers have 30 neurons and the absolute value as the non-linear activation function (i.e., a leaky rectified linear unit with

parameter -1), while the last layer is linear. The layers are fully connected by conformable weight matrices and vectors of biases.

The training of the network is carried out by using stochastic gradient descent with Nesterov momentum, batch size equal to 500, and decreasing learning rate (linearly from 0.01 to 0.0001 over 500,000 iterations). The final mean squared error is 0.7 basis points on the test set (that is, on the sample set that was not used for training and validating the neural network). In separate experiments with affine models where the prices are known analytically, we noted that the error of the network tends to be highly negatively correlated with the approximation error of the Monte Carlo simulations used to generate the sample, thus providing evidence of good generalization and of the ability of the proposed procedure to approximate the true pricing function almost perfectly.

Table 1 compares the computational burden of the neural net approximation thus obtained with that of approximations based on discretization of the state space and antithetic sampling (e.g., Bauer and Rudebusch 2016, Pericoli and Taboga 2015). The neural net is more than 100,000 times faster than an antithetic sampling scheme having a comparable accuracy (around 1 basis point RMSE).

In the MCMC simulation we perform 2,000,000 draws for each entry of the parameter vector (the first 1,000,000 draws are used as a burn-in sample and discarded). Raftery and Lewis' (1995) run length control diagnostic[9] indicates that the sample size is more than 20 times the minimum required size.

At each draw of the parameter vector from the posterior distribution, we compute the value of the shadow rate s_t for each date in the sample ($t = 1 \ldots, T$), as well as simulated trajectories 40 quarters ahead for each date: $s_{t+1}^{(t)}, \ldots, s_{t+40}^{(t)}$ where the superscript (t) indicates a simulation started at time t. We use the simulated trajectories to compute risk-neutral yields[10] and risk premia. The risk-neutral yields are obtained by using the empirical P-

---

[9]The parameters of the diagnostic are set in such a way that the minimum required size allows to estimate the 2.5% quantile of the posterior distribution of each entry of the parameter vector with an error $<1\%$ with probability 95%.

[10]By risk-neutral yields we mean yields that would prevail if risk premia were nil.

measure given by the simulated trajectories instead of the Q-measure in equations (8) and (9)[11]. The risk premia are calculated as differences between observed and risk-neutral yields.

10 Results

Figure 1 displays the posterior median and the first and last deciles of the lower bound $\underline{r}_t$. Note that $\underline{r}_t$ is not interpreted here as a physical limit, depending on the cost of making arbitrages between cash and bonds. It is rather thought of as the market's perception of the level beyond which the central bank is not willing to let the short rate go, also on the basis of considerations about factors linked to strategy, culture and communication (see Grisse et al. 2017 for a review of these factors and a discussion of the perceived lower bound).

We find that until 2007 the posterior median of $\underline{r}_t$ was around 0.6%. Such a number would be justified, for example, if the market expected that the ECB 1) would not let its deposit facility rate (DFR) go below 0.1% (so as to keep it positive), 2) would keep a corridor of about 50 basis points between the deposit facility rate and the rate on the main refinancing operations (MRO), 3) would remain able (by not allowing for excess liquidity) to keep the short-term rate aligned with the MRO rate.

Since 2008 the posterior median of $\underline{r}_t$ started decreasing and kept decreasing until 2017, when it touched a minimum of about -0.5%. We interpret this as a gradual revision of the market's beliefs about the ECB's willingness to lower the minimum level of the short term rate, determined by: 1) the switch to a regime of excess liquidity in which the short-term rate stays closer to the DFR because the ECB is no longer able to keep it in line with the MRO rate; 2) the renouncement of a strictly positive DFR (and the multiple reductions in the DFR below zero carried out by the ECB).

Figure 2 displays the median and the first and last deciles of the posterior distribution of the shadow rate s_t. The standard deviation of the posterior distribution of s_t amounts to

---

[11] The simulated trajectories of s_t are converted into simulated trajectories of r_t, which are then used to compute the plug-in estimator of the expected value in equation (8).

few basis points in the periods when the short-term rate is above the lower bound $\underline{r_t}$ because in these periods r_t and s_t coincide, and the pricing errors and the uncertainty about the exact location of $\underline{r_t}$ are the only sources of randomness that prevent s_t from being exactly observable. On the contrary, when s_t is below the lower bound, the standard deviation of its posterior distribution is in several cases larger than 50 basis points, due to the fact that the value of s_t cannot be recovered from observed factors (not even in the case of zero pricing errors).

The median shadow rate decreased from positive values in 2008 to significantly[12] negative values (below 5 per cent[13]) in 2016. The decrease was not steady, as temporary uptrends were observed in more than one occasion. For example, the shadow rate increased significantly in the first half of 2013. This behavior might reflect the fact that, even if the ECB undertook significant expansionary measures during 2012 (a rate cut, the institution of the Outright Monetary Transactions, a relaxation of collateral rules), the balance sheet of the ECB started to shrink significantly since the beginning of 2013 due to the early repayment of a considerable portion of the funds lent to banks through the 3-year Longer Term Refinancing Operations; this reduction might have been perceived by market participants as tantamount to a monetary restriction. Furthermore, we observe a temporary increase in the shadow rate also in 2015, around the so-called "Bund tantrum" episode, when long-term rates increased significantly. While the causes of the latter episode are still unclear, some commentators have hypothesized that the increase in interest rates might have been triggered by a short-lived uptick in inflation expectations and by expectations that the real effects of the ECB's quantitative easing could be seen sooner than previously thought. Finally, the posterior median of the shadow rate increases markedly at the end of our sample (to about -3.2% in Q4 2017). This movement reflects a moderate increase in the expectations 1-year

---

[12]According to most of the literature on shadow rate term structure models, it seems doubtful that a negative value of the shadow rate can literally be interpreted as the interest rate that would prevail in the absence of a lower bound; it is probably more correct to interpret it as a summary statistic of the stance of monetary policy that provides a measure of the distance from the lift-off (i.e., the date of the first increase in policy rates).

[13]Our estimate of the value of the shadow rate in 2016 is in line with that provided by Wu and Xia (2017).

forward about inflation and GDP growth, accompanied by a pick-up in long-term rates. By forecasting the shadow rate from Q4 2017 onwards, we find that the probability of a significant rise in short-term (policy) rates becomes non-negligible only starting from Q1 2019. This is consistent with survey evidence provided by Bloomberg at the time we estimated our model. The timing of the rise is anyway highly dependent on the evolution of expected inflation and GDP growth. Based on a scenario analysis carried out with the model, if expected inflation 1-year forward were to remain steadily below 2 per cent over the coming years, the first rise in policy rates would likely be carried out after 2020.

Figures 3 and 4 display the estimated decomposition of 10-year rates into expectations and risk premia. According to these estimates, the fall in 10-year rates observed since 2008 has been due both to a decline in expectations about the future path of short-term rates and to a marked decrease in risk premia, which were estimated to be negative at the end of the sample. This is in line with the evidence provided by other studies (see, e.g., Hördahl et al. 2016 for a review)[14]. Some frequently cited explanations for the low or negative level of risk premia are: 1) the low uncertainty about the evolution of policy rates (e.g., Bernanke 2015); 2) the strong negative correlation between bond and stock prices (note that in a CAPM framework, if long-term bonds have a negative beta, they must have a negative risk premium; e.g., Campbell et al. 2016); 3) the increasing relative importance of demand shocks (with respect to supply shocks), which determines a positive correlation between economic growth and inflation shocks (in a consumption-CAPM framework, this implies that if inflation shocks have a negative impact on nominal bond prices, then risk premia must be negative; e.g., Gourio and Ngo 2016). Explanations 1) and 3) would also help to explain why risk premia increased in 2017: as the end of the ECB's asset purchases becomes nearer, there is growing uncertainty about the ECB's next steps; there is evidence that the negative correlation between bond and stock prices is weakening.

---

[14]On the contrary, the estimates of the risk premium from Wu and Xia's (2017) shadow rate model for the same period are mostly positive.

11   Conclusions

We have proposed a general method for the Bayesian estimation of nonlinear no-arbitrage term structure models. There are two main innovations in our method. The first one is the approximation of bond pricing functionals with deep neural networks. The approximation can not only be made very precise (e.g., by requiring an approximation error of less than 1 basis point), but it is also very efficient from a computational viewpoint. The second innovation is a computational recycling technique that dramatically accelerates the MCMC sampling of the model parameters and of the unobservable state variables that drive the short-term rate.

We have used the proposed techniques to estimate a richly specified shadow rate model with data on euro-denominated risk-free rates. The model has a time-varying lower bound for the short-term interest rate and the shadow rate is driven by three spanned unobservable factors and by two unspanned macroeconomic factors (Consensus Economics forecasts of inflation and GDP growth over a 1-year horizon).

We have found evidence of significant time-variation in the perceived lower bound for the short-term rate. This corroborates the evidence provided by other recent studies that advocate the importance of including a time-varying lower bound in no-arbitrage term structure models.

We have used the model to carry out a decomposition of 10-year rates into expectations and term premia. We have found that the estimated term premium has been on a decreasing trend since the financial crisis and its value is now negative (around -1.5%). This in line with the evidence provided by other researchers and can be rationalized in different theoretical frameworks.

We have used the model to predict the evolution of short-term policy rates from Q4 2017 (the last quarter in our sample) onwards. We have found that the probability of a significant rise in these rates becomes non-negligible only starting from Q1 2019, consistently with the survey evidence available when we estimated our model. The timing of the rise would anyway

31

be highly dependent on the evolution of expected inflation and GDP growth.

References

[1] Adrian, T., Crump, R.K. and Moench, E., 2013. Pricing the term structure with linear regressions. Journal of Financial Economics, 110(1), pp.110-138.

[2] Ahn, D.H., Dittmar, R.F. and Gallant, A.R., 2002. Quadratic term structure models: Theory and evidence. Review of Financial Studies, 15(1), pp.243-288.

[3] Andreasen, M.M. and Meldrum, A., 2013. Likelihood inference in non-linear term structure models: the importance of the lower bound. Bank of England Working Paper, no. 481.

[4] Ang, A. and Piazzesi, M., 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary economics, 50(4), pp.745-787.

[5] Bahrampour, S., Ramakrishnan, N., Schott, L. and Shah, M., 2015. Comparative Study of Deep Learning Software Frameworks. arXiv preprint arXiv:1511.06435.

[6] Bauer, M.D. and Rudebusch, G.D., 2016. Monetary policy expectations at the zero lower bound. Journal of Money Credit and Banking, 48(7), pp. 1439-1465.

[7] Bernanke, B., 2015. Why are interest rates so low, part 4: Term premiums. Brookings Institution.

[8] Black, F. 1995. Interest rates as options. Journal of Finance 50(5), 1371-76.

[9] Campbell, J. Y., Sunderam, A. and Viceira, L.M. 2016. Inflation bets or deflation hedges? The changing risks of nominal bonds. Critical Finance Review, forthcoming.

[10] Christensen, J. H. and Rudebusch, G. D. 2014. Estimating shadow-rate term structure models with near-zero yields. Journal of Financial Econometrics, 13(2), pp. 226-259.

[11] Delalleau, O. and Bengio, Y., 2011. Shallow vs. deep sum-product networks. In Advances in Neural Information Processing Systems (pp. 666-674).

[12] Ejsing, J., Grothe, M. and Grothe, O., 2012. Liquidity and credit risk premia in government bond yields. ECB Working Paper Series, 1440, European Central Bank.

[13] Gourio, F. and Ngo, P. 2016. Risk premia at the zlb: a macroeconomic interpretation. Mimeo.

[14] Grisse, C., Krogstrup, S., and Schumacher, S., 2017. Lower bound beliefs and long-term interest rates. Swiss National Bank Working Papers, 5/2017.

[15] Hamilton, J.D. and Wu, J.C., 2012a. The effectiveness of alternative monetary policy tools in a zero lower bound environment. Journal of Money, Credit and Banking, 44(s1), pp.3-46.

[16] Hamilton, J.D. and Wu, J.C., 2012b. Identification and estimation of Gaussian affine term structure models. Journal of Econometrics, 168(2), pp.315-331.

[17] Hördahl, P., J. Sobrun and P. Turner, 2016. Low long-term interest rates as a global phenomenon. BIS Working Papers, No 574.

[18] Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. Neural networks, 4(2), pp.251-257.

[19] Ichiue, H. and Ueno, Y., 2007. Equilibrium Interest Rate and the Yield Curve in a Low Interest Rate Environment. Bank of Japan Working Paper Series 07-E-18, Bank of Japan.

[20] Ichiue, H. and Ueno, Y., 2013. Estimating term premia at the zero bound: An analysis of Japanese, US, and UK yields. Bank of Japan Working Papers No. 13-E-8.

[21] Joslin, S., Priebsch, M. and Singleton, K.J., 2014. Risk premiums in dynamic term structure models with unspanned macro risks. The Journal of Finance, 69(3), pp.1197-1233.

[22] Kim, D.H. and Singleton, K.J., 2012. Term structure models and the zero bound: an empirical investigation of Japanese yields. Journal of Econometrics, 170(1), pp.32-49.

[23] Krippner, L., 2013. A tractable framework for zero lower bound Gaussian term structure models. Reserve Bank of New Zealand Discussion Paper Series DP2013/02. Reserve Bank of New Zealand.

[24] LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. Nature, 521(7553), pp.436-444.

[25] Leippold, M. and Wu, L., 2002. Asset pricing under the quadratic class. Journal of Financial and Quantitative Analysis, 37(02), pp.271-295.

[26] Mercurio, F., 2009. Interest rates and the credit crunch: new formulas and market models. Bloomberg portfolio research paper, (2010-01).

[27] Monfort, A., Pegoraro, F., Renne, J.P. and Roussellet, G., 2014. Staying at zero with affine processes: A new dynamic term structure model. Journal of Econometrics, forthcoming.

[28] Morini, M., 2009. Solving the puzzle in the interest rate market. Available at SSRN 1506046.

[29] NVIDIA (2016). Cuda 8 Peformance Overview, November 2016, available at http://developer.download.nvidia.com/compute/cuda/compute-docs/cuda-performance-report.pdf.

[30] Pericoli, M. and Taboga, M., 2012. Bond risk premia, macroeconomic fundamentals and the exchange rate. International Review of Economics & Finance, 22(1), pp.42-65.

[31] Pericoli, M. and Taboga, M., 2015. Understanding policy rates at the zero lower bound: insights from a Bayesian shadow rate model. Temi di discussione (Economic working papers) 1023, Bank of Italy, Economic Research and International Relations Area.

[32] Prechelt, L., 1998. Early stopping-but when?. In Neural Networks: Tricks of the trade (pp. 55-69). Springer Berlin Heidelberg.

[33] Raftery, A. and Lewis, S., 1995. The number of iterations, convergence diagnostics and generic Metropolis algorithms. Practical Markov Chain Monte Carlo.

[34] Realdon, M., 2006. Quadratic term structure models in discrete time. Finance Research Letters, 3(4), pp.277-289.

[35] Rudebusch, G.D. and Wu, T., 2008. A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy. The Economic Journal, 118(530), pp.906-926.

[36] Ruder, S., 2016. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

[37] Taboga, M., 2014. What is a prime bank? A Euribor–OIS spread perspective. International Finance, 17(1), pp.51-75.

[38] Theano Development Team, 2016. Theano: A Python framework for fast computation of mathematical expressions. arXiv preprint arXiv:1605.02688.

[39] Wu, J.C. and Xia, F.D., 2016. Measuring the macroeconomic impact of monetary policy at the zero lower bound. Journal of Money, Credit and Banking, 48(2-3), pp.253-291.

[40] Wu, J. C., and Xia, F. D., 2017. Time-Varying Lower Bound of Interest Rates in Europe, mimeo, University of Chicago.

Appendix

The admissible set

This section describes the specification of the admissible set.

We impose some restrictions beyond those that are necessary for identification. This is motivated by the observation that, unless some over-identifying constraints are imposed on the parameters, shadow rate models tend to produce implied volatilities of the shadow rate that are an order of magnitude higher than the historical volatilities of the short-term rate, as well as estimated trajectories of the shadow rate that are characterized by implausibly large negative values (the latter drawback has been found in the majority of studies on shadow rate models).

A vector of parameters belongs to the admissible set A (i.e., $1_A\left(a, b, \mu, \rho, \overline{\rho}, V\right) = 1$) if and only if:

- all the identification restrictions are satisfied;

- all the eigenvalues of ρ are between 0 and 0.99 (this corresponds to an upper bound of approximately 17 years on the half-life of the shocks);

- all the diagonal elements of $\overline{\rho}$ are between 0 and 0.99 (as a consequence, we put an upper bound on the half-life of the shocks also under the pricing measure);

- D is such that all the standard deviations of the pricing errors are less than 20 basis points (in other words, we assume that on average all maturities are priced at a reasonable level of accuracy);

- the conditional standard deviation of the one-period-ahead forecast errors of the shadow rate, that is,

$$\sqrt{\mathrm{Var}\left[s_{t+1} \,|\, X_t\right]} = \|a\| \tag{57}$$

is between 25 and 100 basis points[15] (to broadly match the sample standard deviation of the first differences of the short term rate, which is equal to 0.42 for the euro and to 0.50 for the US dollar);

- the unconditional mean of the shadow rate under Q, which is equal to a, is between 1 and 6 per cent;

- the unconditional mean of the shadow rate under P, which is equal to

$$a + b\left(I - \rho\right)^{-1}\mu \tag{58}$$

is between 1 and 4 per cent;

- the unconditional mean of the shadow rate under Q is greater than the unconditional mean of the shadow rate under P (which implies that on average bond risk premia are positive), but the positive difference between the two means cannot exceed 200 basis points (based on the empirical evidence that the expectations hypothesis is, at least unconditionally, difficult to reject and, therefore, large deviations from the expectations hypothesis are, unconditionally, deemed implausible).

---

[15] Note that quarterly compounded rates need to be multiplied by 400 in order to be expressed in percentage points on an annual basis.

Tables and figures

Table 1 - Computational intensity of the neural net approximation and of competing approximations[16]

	Time (ms)	Estimated RMSE (bp)
Discrete state space (500 trajectories + antithetic)	152	9.3
Discrete state space (5,000 trajectories + antithetic)	1,430	3.1
Discrete state space (50,000 trajectories + antithetic)	16,950	1.1
Discrete state space (500,000 trajectories + antithetic)	233,960	0.4
Neural net (6-layers - 30 neurons per intermediate layer)	0.12	0.9

---

[16]The computer employed has an Intel dual core i5-4300U @ 1.9GHz processor and 8GB RAM. The time, in milliseconds, is the time required to compute no-arbitrage bond yields up to the 10-year maturity for 100 different parameter vectors. The estimated RMSE, in basis points, is an average across maturities.

Figure 1 - The short-term rate and the time-varying lower bound[17]

---
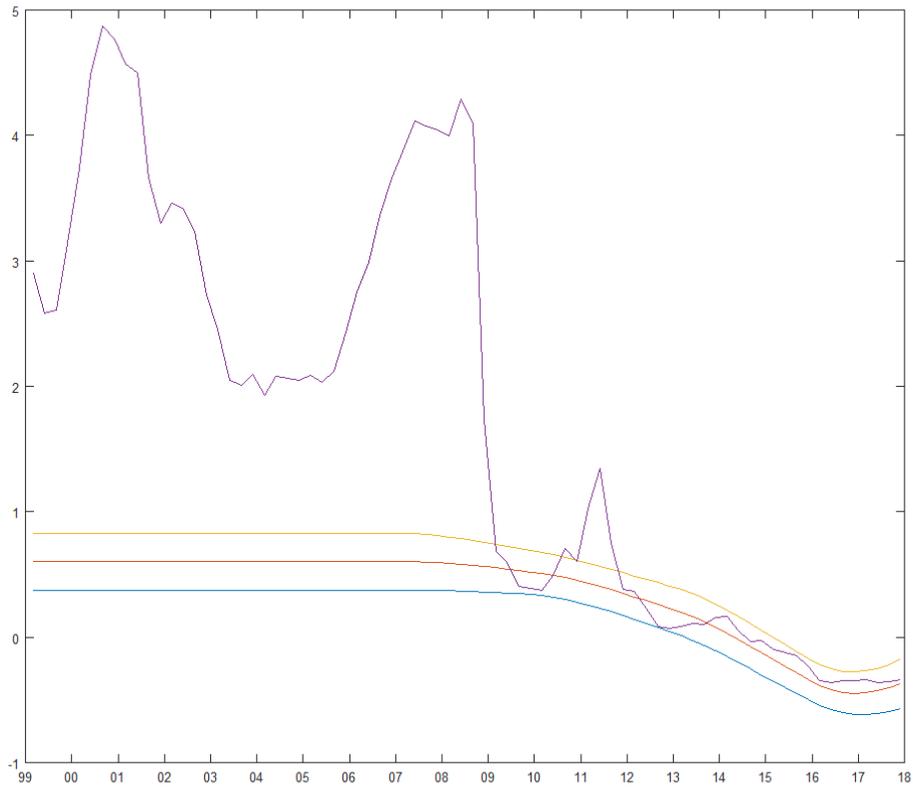
[17]The violet line is the 3-month OIS rate. The red line is the posterior median of the lower bound for interest rates (blue and yellow are the first and last deciles of the posterior distribution).

Figure 2 - The shadow rate[18]

Figure 3 - Observed and risk-adjusted 10-year yields[19]
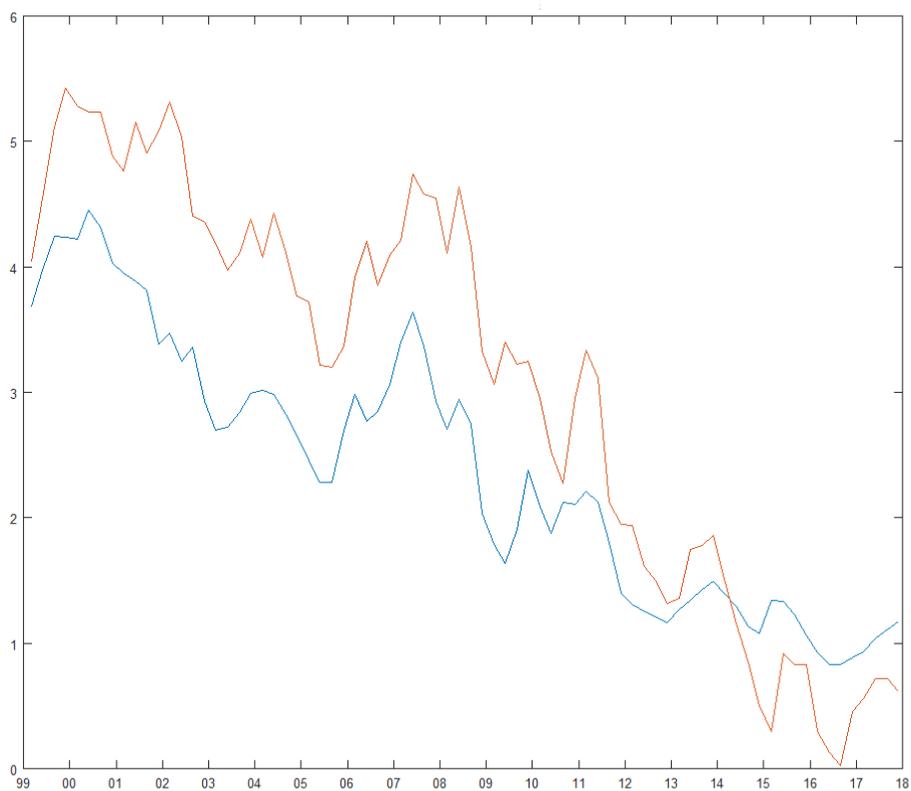
[19]The red line is the 10-year zero-coupon rate. The blue line is the posterior median of the expectations component of the 10-year zero-coupon rate.

Figure 4 - 10-year risk premia[20]
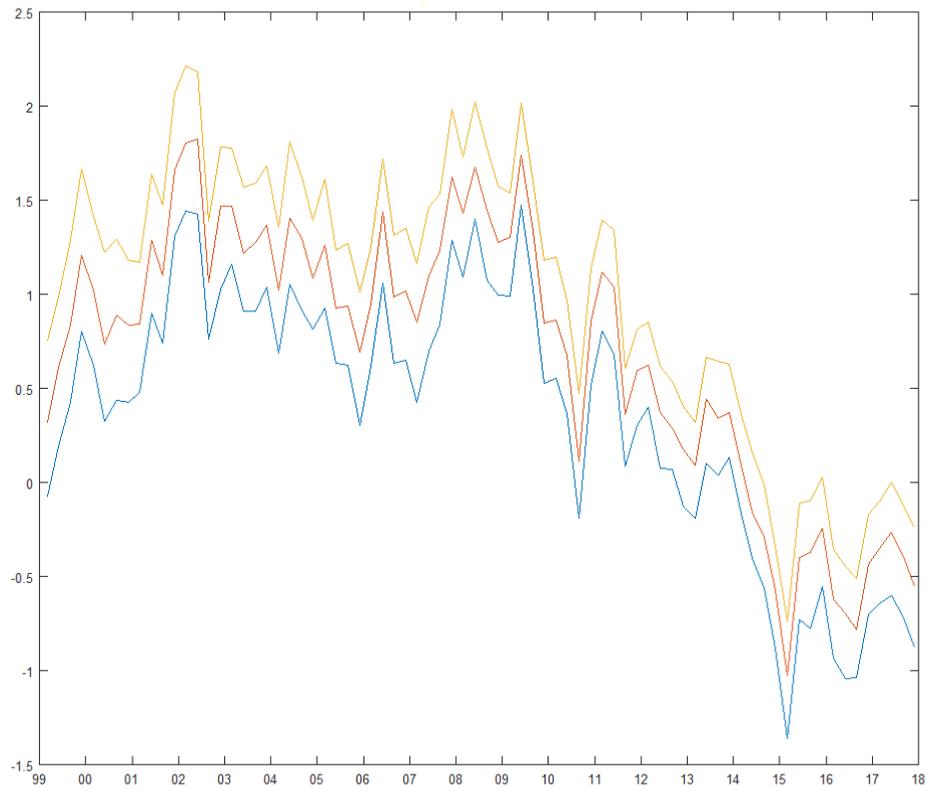
[20]The red line is the posterior median of the risk premium embedded in the 10-year zero coupon rate. The blue and yellow lines are the first and ninth decile of the posterior distribution of the risk premium, respectively.

43

2016

ALBANESE G., G. DE BLASIO and P. SESTITO, *My parents taught me. evidence on the family transmission of values,* Journal of Population Economics, v. 29, 2, pp. 571-592, **TD No. 955 (March 2014).**

ANDINI M. and G. DE BLASIO, *Local development that money cannot buy: Italy's Contratti di Programma,* Journal of Economic Geography, v. 16, 2, pp. 365-393, **TD No. 915 (June 2013).**

BARONE G. and S. MOCETTI, *Inequality and trust: new evidence from panel data,* Economic Inquiry, v. 54, pp. 794-809, **TD No. 973 (October 2014).**

BELTRATTI A., B. BORTOLOTTI and M. CACCAVAIO, *Stock market efficiency in China: evidence from the split-share reform,* Quarterly Review of Economics and Finance, v. 60, pp. 125-137, **TD No. 969 (October 2014).**

BOLATTO S. and M. SBRACIA, *Deconstructing the gains from trade: selection of industries vs reallocation of workers,* Review of International Economics, v. 24, 2, pp. 344-363, **TD No. 1037 (November 2015).**

BOLTON P., X. FREIXAS, L. GAMBACORTA and P. E. MISTRULLI, *Relationship and transaction lending in a crisis*, Review of Financial Studies, v. 29, 10, pp. 2643-2676, **TD No. 917 (July 2013).**

BONACCORSI DI PATTI E. and E. SETTE, *Did the securitization market freeze affect bank lending during the financial crisis? Evidence from a credit register,* Journal of Financial Intermediation , v. 25, 1, pp. 54-76, **TD No. 848 (February 2012).**

BORIN A. and M. MANCINI, *Foreign direct investment and firm performance: an empirical analysis of Italian firms,* Review of World Economics, v. 152, 4, pp. 705-732, **TD No. 1011 (June 2015).**

BRAGOLI D., M. RIGON and F. ZANETTI, *Optimal inflation weights in the euro area,* International Journal of Central Banking, v. 12, 2, pp. 357-383, **TD No. 1045 (January 2016).**

BRANDOLINI A. and E. VIVIANO, *Behind and beyond the (headcount) employment rate,* Journal of the Royal Statistical Society: Series A, v. 179, 3, pp. 657-681, **TD No. 965 (July 2015).**

BRIPI F., *The role of regulation on entry: evidence from the Italian provinces,* World Bank Economic Review, v. 30, 2, pp. 383-411, **TD No. 932 (September 2013).**

BRONZINI R. and P. PISELLI, *The impact of R&D subsidies on firm innovation,* Research Policy, v. 45, 2, pp. 442-457, **TD No. 960 (April 2014).**

BURLON L. and M. VILALTA-BUFI, *A new look at technical progress and early retirement,* IZA Journal of Labor Policy, v. 5, **TD No. 963 (June 2014).**

BUSETTI F. and M. CAIVANO, *The trend–cycle decomposition of output and the Phillips Curve: bayesian estimates for Italy and the Euro Area,* Empirical Economics, V. 50, 4, pp. 1565-1587, **TD No. 941 (November 2013).**

CAIVANO M. and A. HARVEY, *Time-series models with an EGB2 conditional distribution,* Journal of Time Series Analysis, v. 35, 6, pp. 558-571, **TD No. 947 (January 2014).**

CALZA A. and A. ZAGHINI, *Shoe-leather costs in the euro area and the foreign demand for euro banknotes,* International Journal of Central Banking, v. 12, 1, pp. 231-246, **TD No. 1039 (December 2015).**

CESARONI T. and R. DE SANTIS, *Current account "core-periphery dualism" in the EMU,* The World Economy, v. 39, 10, pp. 1514-1538, **TD No. 996 (December 2014).**

CIANI E., *Retirement, Pension eligibility and home production,* Labour Economics, v. 38, pp. 106-120, **TD No. 1056 (March 2016).**

CIARLONE A. and V. MICELI, *Escaping financial crises? Macro evidence from sovereign wealth funds' investment behaviour,* Emerging Markets Review, v. 27, 2, pp. 169-196, **TD No. 972 (October 2014).**

CORNELI F. and E. TARANTINO, *Sovereign debt and reserves with liquidity and productivity crises,* Journal of International Money and Finance, v. 65, pp. 166-194, **TD No. 1012 (June 2015).**

D'AURIZIO L. and D. DEPALO, *An evaluation of the policies on repayment of government's trade debt in Italy,* Italian Economic Journal, v. 2, 2, pp. 167-196, **TD No. 1061 (April 2016).**

DE BLASIO G., G. MAGIO and C. MENON, *Down and out in Italian towns: measuring the impact of economic downturns on crime,* Economics Letters, 146, pp. 99-102, **TD No. 925 (July 2013).**

DOTTORI D. and M. MANNA, *Strategy and tactics in public debt management,* Journal of Policy Modeling, v. 38, 1, pp. 1-25, **TD No. 1005 (March 2015).**

LIBERATI D., M. MARINUCCI and G. M. TANZI, *Science and technology parks in Italy: main features and analysis of their effects on hosted firms*, Journal of Technology Transfer, v. 41, 4, pp. 694-729, **TD No. 983 (November 2014).**

MARCELLINO M., M. PORQUEDDU and F. VENDITTI, *Short-Term GDP forecasting with a mixed frequency dynamic factor model with stochastic volatility,* Journal of Business & Economic Statistics , v. 34, 1, pp. 118-127, **TD No. 896 (January 2013).**

RODANO G., N. SERRANO-VELARDE and E. TARANTINO, *Bankruptcy law and bank financing,* Journal of Financial Economics, v. 120, 2, pp. 363-382, **TD No. 1013 (June 2015).**

ZINNA G., *Price pressures on UK real rates: an empirical investigation,* Review of Finance,v. 20, 4, pp. 1587-1630, **TD No. 968 (July 2014).**

2017

ADAMOPOULOU A. and G.M. TANZI, *Academic dropout and the great recession,* Journal of Human Capital, V. 11, 1, pp. 35–71, **TD No. 970 (October 2014).**

ALBERTAZZI U., M. BOTTERO and G. SENE, *Information externalities in the credit market and the spell of credit rationing,* Journal of Financial Intermediation, v. 30, pp. 61–70, **TD No. 980 (November 2014).**

ALESSANDRI P. and H. MUMTAZ, *Financial indicators and density forecasts for US output and inflation,* Review of Economic Dynamics, v. 24, pp. 66-78, **TD No. 977 (November 2014).**

BARBIERI G., C. ROSSETTI and P. SESTITO, *Teacher motivation and student learning,* Politica economica/Journal of Economic Policy, v. 33, 1, pp.59-72, **TD No. 761 (June 2010).**

BENTIVOGLI C. and M. LITTERIO, *Foreign ownership and performance: evidence from a panel of Italian firms,* International Journal of the Economics of Business, v. 24, 3, pp. 251-273, **TD No. 1085 (October 2016).**

BRONZINI R. and A. D'IGNAZIO, *Bank internationalisation and firm exports: evidence from matched firm-bank data,* Review of International Economics, v. 25, 3, pp. 476-499 **TD No. 1055 (March 2016).**

BRUCHE M. and A. SEGURA, *Debt maturity and the liquidity of secondary debt markets,* Journal of Financial Economics, v. 124, 3, pp. 599-613, **TD No. 1049 (January 2016).**

BURLON L., *Public expenditure distribution, voting, and growth,* Journal of Public Economic Theory,, v. 19, 4, pp. 789–810, **TD No. 961 (April 2014).**

BURLON L., A. GERALI, A. NOTARPIETRO and M. PISANI, *Macroeconomic effectiveness of non-standard monetary policy and early exit. a model-based evaluation,* International Finance, v. 20, 2, pp.155-173, **TD No. 1074 (July 2016).**

BUSETTI F., *Quantile aggregation of density forecasts,* Oxford Bulletin of Economics and Statistics, v. 79, 4, pp. 495-512, **TD No. 979 (November 2014).**

CESARONI T. and S. IEZZI, *The predictive content of business survey indicators: evidence from SIGE,* Journal of Business Cycle Research, v.13, 1, pp 75–104, **TD No. 1031 (October 2015).**

CONTI P., D. MARELLA and A. NERI, *Statistical matching and uncertainty analysis in combining household income and expenditure data,* Statistical Methods & Applications, v. 26, 3, pp 485–505, **TD No. 1018 (July 2015).**

D'AMURI F., *Monitoring and disincentives in containing paid sick leave,* Labour Economics, v. 49, pp. 74-83, **TD No. 787 (January 2011).**

D'AMURI F. and J. MARCUCCI, *The predictive power of google searches in forecasting unemployment,* International Journal of Forecasting, v. 33, 4, pp. 801-816, **TD No. 891 (November 2012).**

DE BLASIO G. and S. POY, *The impact of local minimum wages on employment: evidence from Italy in the 1950s,* Journal of Regional Science, v. 57, 1, pp. 48-74, **TD No. 953 (March 2014).**

DEL GIOVANE P., A. NOBILI and F. M. SIGNORETTI, *Assessing the sources of credit supply tightening: was the sovereign debt crisis different from Lehman?,* International Journal of Central Banking, v. 13, 2, pp. 197-234, **TD No. 942 (November 2013).**

DEL PRETE S., M. PAGNINI, P. ROSSI and V. VACCA, *Lending organization and credit supply during the 2008–2009 crisis,* Economic Notes, v. 46, 2, pp. 207–236, **TD No. 1108 (April 2017).**

DELLE MONACHE D. and I. PETRELLA, *Adaptive models and heavy tails with an application to inflation forecasting,* International Journal of Forecasting, v. 33, 2, pp. 482-501, **TD No. 1052 (March 2016).**

FEDERICO S. and E. TOSTI, *Exporters and importers of services: firm-level evidence on Italy,* The World Economy, v. 40, 10, pp. 2078-2096, **TD No. 877 (September 2012).**

GIACOMELLI S. and C. MENON, *Does weak contract enforcement affect firm size? Evidence from the neighbour's court,* Journal of Economic Geography, v. 17, 6, pp. 1251-1282, **TD No. 898 (January 2013).**

LOBERTO M. and C. PERRICONE, *Does trend inflation make a difference?,* Economic Modelling, v. 61, pp. 351–375, **TD No. 1033 (October 2015).**

MANCINI A.L., C. MONFARDINI and S. PASQUA, *Is a good example the best sermon? Children's imitation of parental reading,* Review of Economics of the Household, v. 15, 3, pp 965–993, **D No. 958 (April 2014).**

MEEKS R., B. NELSON and P. ALESSANDRI, *Shadow banks and macroeconomic instability,* Journal of Money, Credit and Banking, v. 49, 7, pp. 1483–1516, **TD No. 939 (November 2013).**

MICUCCI G. and P. ROSSI, *Debt restructuring and the role of banks' organizational structure and lending technologies,* Journal of Financial Services Research, v. 51, 3, pp 339–361, **TD No. 763 (June 2010).**

MOCETTI S., M. PAGNINI and E. SETTE, *Information technology and banking organization,* Journal of Journal of Financial Services Research, v. 51, pp. 313-338, **TD No. 752 (March 2010).**

MOCETTI S. and E. VIVIANO, *Looking behind mortgage delinquencies,* Journal of Banking & Finance, v. 75, pp. 53-63, **TD No. 999 (January 2015).**

NOBILI A. and F. ZOLLINO, *A structural model for the housing and credit market in Italy,* Journal of Housing Economics, v. 36, pp. 73-87, **TD No. 887 (October 2012).**

PALAZZO F., *Search costs and the severity of adverse selection,* Research in Economics, v. 71, 1, pp. 171-197, **TD No. 1073 (July 2016).**

PATACCHINI E. and E. RAINONE, *Social ties and the demand for financial services,* Journal of Financial Services Research, v. 52, 1–2, pp 35–88, **TD No. 1115 (June 2017).**

PATACCHINI E., E. RAINONE and Y. ZENOU, *Heterogeneous peer effects in education,* Journal of Economic Behavior & Organization, v. 134, pp. 190–227, **TD No. 1048 (January 2016).**

SBRANA G., A. SILVESTRINI and F. VENDITTI, *Short-term inflation forecasting: the M.E.T.A. approach,* International Journal of Forecasting, v. 33, 4, pp. 1065-1081, **TD No. 1016 (June 2015).**

SEGURA A. and J. SUAREZ, *How excessive is banks' maturity transformation?,* Review of Financial Studies, v. 30, 10, pp. 3538–3580, **TD No. 1065 (April 2016).**

VACCA V., *An unexpected crisis? Looking at pricing effectiveness of heterogeneous banks,* Economic Notes, v. 46, 2, pp. 171–206, **TD No. 814 (July 2011).**

VERGARA CAFFARELI F., *One-way flow networks with decreasing returns to linking,* Dynamic Games and Applications, v. 7, 2, pp. 323-345, **TD No. 734 (November 2009).**

ZAGHINI A., *A Tale of fragmentation: corporate funding in the euro-area bond market,* International Review of Financial Analysis, v. 49, pp. 59-68, **TD No. 1104 (February 2017).**

2018

ADAMOPOULOU A. and E. KAYA, *Young Adults living with their parents and the influence of peers,* Oxford Bulletin of Economics and Statistics, v. 80, pp. 689-713, **TD No. 1038 (November 2015).**

BELOTTI F. and G. ILARDI *Consistent inference in fixed-effects stochastic frontier models,* Journal of Econometrics, v. 202, 2, pp. 161-177, **TD No. 1147 (October 2017).**

BRILLI Y. and M. TONELLO, *Does increasing compulsory education reduce or displace adolescent crime? New evidence from administrative and victimization data,* CESifo Economic Studies, v. 64, 1, pp. 15–4, **TD No. 1008 (April 2015).**

BUONO I. and S. FORMAI *The heterogeneous response of domestic sales and exports to bank credit shocks,* Journal of International Economics, v. 113, pp. 55-73, **TD No. 1066 (March 2018).**

CARTA F. and M. DE PHLIPPIS, *You've Come a long way, baby. husbands' commuting time and family labour supply,* Regional Science and Urban Economics, v. 69, pp. 25-37, **TD No. 1003 (March 2015).**

CARTA F. and L. RIZZICA, *Early kindergarten, maternal labor supply and children's outcomes: evidence from Italy,* Journal of Public Economics, v. 158, pp. 79-102, **TD No. 1030 (October 2015).**

CASIRAGHI M., E. GAIOTTI, L. RODANO and A. SECCHI, *A "Reverse Robin Hood"? The distributional implications of non-standard monetary policy for Italian households,* Journal of International Money and Finance, v. 85, pp. 215-235, **TD No. 1077 (July 2016).**

CECCHETTI S., F. NATOLI and L. SIGALOTTI, *Tail co-movement in inflation expectations as an indicator of anchoring,* International Journal of Central Banking, v. 14, 1, pp. 35-71, **TD No. 1025 (July 2015).**

CIPRIANI M., A. GUARINO, G. GUAZZAROTTI, F. TAGLIATI and S. FISHER, *Informational contagion in the laboratory,* Review of Finance, v. 22, 3, pp. 877-904, **TD No. 1063 (April 2016).**

NUCCI F. and M. RIGGI, *Labor force participation, wage rigidities, and inflation,* Journal of Macroeconomics, v. 55, 3 pp. 274-292, **TD No. 1054 (March 2016).**

SEGURA A., *Why did sponsor banks rescue their SIVs?,* Review of Finance, v. 22, 2, pp. 661-697, **TD No. 1100 (February 2017).**

FORTHCOMING

ALBANESE G., G. DE BLASIO and P. SESTITO, *Trust, risk and time preferences: evidence from survey data,* International Review of Economics, **TD No. 911 (April 2013).**

APRIGLIANO V., G. ARDIZZI and L. MONTEFORTE, *Using the payment system data to forecast the economic activity,* International Journal of Central Banking, **TD No. 1098 (February 2017).**

BARONE G., G. DE BLASIO and S. MOCETTI, *The real effects of credit crunch in the great recession: evidence from Italian provinces,* Regional Science and Urban Economics, **TD No. 1057 (March 2016).**

BELOTTI F. and G. ILARDI, *Consistent inference in fixed-effects stochastic frontier models,* Journal of Econometrics, **TD No. 1147 (October 2017).**

BERTON F., S. MOCETTI, A. PRESBITERO and M. RICHIARDI, *Banks, firms, and jobs,* Review of Financial Studies, **TD No. 1097 (February 2017).**

BOFONDI M., L. CARPINELLI and E. SETTE, *Credit supply during a sovereign debt crisis,* Journal of the European Economic Association, **TD No. 909 (April 2013).**

CIANI E. and C. DEIANA, *No Free lunch, buddy: housing transfers and informal care later in life,* Review of Economics of the Household, **TD No. 1117 (June 2017).**

CIANI E. and P. FISHER, *Dif-in-dif estimators of multiplicative treatment effects,* Journal of Econometric Methods, **TD No. 985 (November 2014).**

D'AMURI F., *Monitoring and disincentives in containing paid sick leave,* Labour Economics, **TD No. 787 (January 2011).**

ERCOLANI V. and J. VALLE E AZEVEDO, *How can the government spending multiplier be small at the zero lower bound?,* Macroeconomic Dynamics, **TD No. 1174 (April 2018).**

FEDERICO S. and E. TOSTI, *Exporters and importers of services: firm-level evidence on Italy,* The World Economy, **TD No. 877 (September 2012).**

GERALI A. and S. NERI, *Natural rates across the atlantic,* Journal of Macroeconomics, **TD No. 1140 (September 2017).**

GIACOMELLI S. and C. MENON, *Does weak contract enforcement affect firm size? Evidence from the neighbour's court,* Journal of Economic Geography, **TD No. 898 (January 2013).**

LINARELLO A., *Direct and indirect effects of trade liberalization: evidence from Chile,* Journal of Development Economics, **TD No. 994 (December 2014).**

NATOLI F. and L. SIGALOTTI, *Tail co-movement in inflation expectations as an indicator of anchoring,* International Journal of Central Banking, **TD No. 1025 (July 2015).**

RIGGI M., *Capital destruction, jobless recoveries, and the discipline device role of unemployment,* Macroeconomic Dynamics, **TD No. 871 (July 2012).**

SEGURA A., *Why did sponsor banks rescue their SIVs?,* Review of Finance, **TD No. 1100 (February 2017).**