08

WORKING PAPERS 2022

FORGETTING APPROACHES TO IMPROVE FORECASTING

Robert Hill | Paulo M. M. Rodrigues

BANCO DE PORTUGAL
EUROSYSTEM

08
WORKING PAPERS 2022

FORGETTING APPROACHES TO IMPROVE FORECASTING

Robert Hill | Paulo M. M. Rodrigues

JUNE 2022
The analyses, opinions and findings of these papers represent the views of the authors, they are not necessarily those of the Banco de Portugal or the Eurosystem.

Please address correspondence to
Banco de Portugal
Rua do Comércio 148, 1100-150 Lisboa, Portugal
Tel.: +351 213 130 000, email: info@bportugal.pt

BANCO DE PORTUGAL
EUROSYSTEM

Lisboa, 2022 • www.bportugal.pt

Working Papers | Lisboa 2022 • Banco de Portugal Rua do Comércio 148 | 1100-150 Lisboa • www.bportugal.pt •
Edition Banco de Portugal • ISBN (online) 978-989-678-833-9 • ISSN (online) 2182-0422

Forgetting Approaches to Improve Forecasting

Robert Hill
Nova School of Business and Economics

Paulo M. M. Rodrigues
Banco de Portugal
Nova School of Business and Economics

June 2022

Abstract
There is widespread evidence of parameter instability in the literature. One way to account for this feature is through the use of time-varying parameter (TVP) models that discount older data in favour of more recent data. This practise is often known as forgetting and can be applied in several different ways. This paper introduces and examines the performance of different (flexible) forgetting methodologies in the context of the Kalman filter. We review and develop the theoretical background and investigate the performance of each methodology in simulations as well as in two empirical forecast exercises using dynamic model averaging (DMA). Specifically, out-of-sample DMA forecasts of CPI inflation and S&P500 returns obtained using different forgetting approaches are compared. Results show that basing the amount of forgetting on the forecast error does not perform as well as avoiding instability by placing bounds on the parameter covariance matrix.

JEL: C22, C51, C53

Keywords: Forgetting, Kalman Filter, Dynamic Model Averaging, In ation, Stock Returns.

Acknowledgements: The authors thank Tassos Magdalinos for many useful discussions on this work.

The analyses, opinions and findings of this paper represents the views of the authors, they are not necessarily those of the Banco de Portugal or the Eurosystem

E-mail: robert.hill@novasbe.pt; pmrodrigues@bportugal.pt

1. Introduction

Time varying parameter (TVP) models offer insights into system dynamics that traditional constant coefficient models are not designed to capture. Over the past two decades TVP models have made their way into mainstream macroeconomics and forecasting. For interesting overviews of work regarding TVPs in the context of state space models see e.g. Harvey (1990), Durbin and Koopman (2012) and references therein. Applications to macroeconomics include Primiceri (2005), Cogley and Sargent (2005), and Koop *et al.* (2009). In these examples, TVPs are usually considered as state variables and the observable data as measurement variables in a state space format. The Kalman filter (KF) offers a feasible closed form method to estimate the parameters in this setting and is frequently used in place of Markov Chain Monte Carlo (MCMC) methods that can be computationally demanding.

The use of KF to track state variables has been well studied in the control theory literature for several decades and has been extended to a number of applications in science and engineering. This is in large part due to KF's simplicity in design and adaptive interpretation (it can, for example, have a recursive least squares (RLS) formulation as well as an appealing Bayesian interpretation (see e.g. Durbin and Koopman, 2000)).

For KF to track TVPs it must avoid convergence to a set distribution. For instance, in a simple autoregressive time series model, if the autoregressive parameter is estimated at each iteration as in RLS then the variation in parameter estimates will decrease as $t \to \infty$. This happens not only in the case of a fixed parameter, but also in the TVP case. Similarly, parameter estimates obtained with KF will also undergo convergence unless some adaptation is made. Forecasting techniques using, for instance, a rolling window or adjusting estimation through some exponential weighting of the data address this issue. The same concept has been applied to KF in a large number of studies in order to avoid convergence to a set distribution and to track TVPs (Jazwinski, 1970). One common way to avoid this convergence is through what is known as forgetting.

Generally, forgetting in KF allows the system to react to changes in the system's dynamics, such as changes in the marginal effects of inputs on outputs. Usually in economics, these changes are unknown *a priori* so the problem of identifying where these occur is challenging. Applying forgetting to limit the convergence of parameter estimates is one way to potentially identify these dynamics. Forgetting was initially applied in an ad hoc fashion, however starting in the early 90s formal properties of different forgetting methods have been analysed; see, for instance, Kulhavỳ and Zarrop (1993) and Parkum *et al.* (1992).

In its simplest form forgetting can be achieved by a single parameter that acts as a discount factor on the data used in RLS, which is known as scalar exponential forgetting (EF). EF is a means by which KF avoids convergence, by essentially flattening the parameter covariance matrix at each iteration and thus stopping the filter from learning the state variables too well. This method of filter tracking is employed in Raftery *et al.* (2010) and in an economic context by Koop and Korobilis (2012). In both cases Bayesian model averaging (BMA) is applied to a set of TVP models, which are estimated through KF using EF in what is known as dynamic model averaging (DMA).

Although EF is easily applied and interpreted, it may not always offer the most efficient means by which to estimate TVPs as the same rate of forgetting is applied to all measurement variables regardless of how much information is contained in each (Pollock, 2003). In a linear KF model, the parameters corresponding to certain measurement variables or predictors may shift over time at different rates. If not enough forgetting is applied, then we may not track these parameters quickly enough, however if too much is applied KF becomes unstable. This motivates the introduction of alternative more flexible forgetting approaches. Several methods have been developed in the systems control literature. For instance, Fortescue *et al.* (1981), Saelid and Foss (1983), Hägglund (1984) and Kulhavỳ and Kárnỳ (1984) introduce approaches which allow the forgetting factor to be dynamic or to vary across predictors depending on the information delivered by each predictor. However, to date, their potential in economics and finance has not yet been evaluated.

The first contribution of this paper is to introduce and classify several methods of forgetting and to demonstrate the usefulness of these approaches through an in depth Monte Carlo analysis. The second contribution consists of forecast applications of these methods within a DMA framework to US inflation and S&P 500 returns. We opt for a DMA approach since it is a commonly used methodology that has been shown to have an interesting performance in terms of forecasting and model analysis (Koop and Korobilis, 2012). Results suggest that the more flexible forgetting approaches outperform the standard EF method. In particular, when predictors undergo a decrease in variance there is a tendency for the information content in the state space system to decrease and EF causes over-correction by essentially fitting noise. Results also show that more accurate forecasts, lower forecast variances and better state variable estimation is achieved by using alternative forgetting schemes.

The remainder of the paper is organized as follows. Section 2 presents some background theory on the KF and the concept of forgetting; Section 3 discusses several dynamic forgetting methods and briefly introduces the DMA methodology. Section 4 presents the Monte Carlo simulation results and Section 5 the empirical forecasting exercise. Lastly, Section 6 concludes.

2. The Concept of Forgetting

This section provides a brief description of the application of the forgetting concept in a simple RLS context and formalizes the idea for KF. Consider the state space TVP model,

$$y_t = \mathbf{x}_t'\boldsymbol{\theta}_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, V_t) \tag{1}$$

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \eta_t, \qquad \eta_t \sim N(0, W_t) \tag{2}$$

where $\boldsymbol{\theta}_t$ is the state variable, y_t is the variable of interest, and $\mathbf{x}_t$ is an $m \times 1$ vector of predictors. The random walk representation of $\boldsymbol{\theta}_t$ in (2) is frequently used in macroeconomic studies; see *e.g.*, Cogley and Sargent (2005), Groen *et al.* (2013), Koop *et al.* (2009), Korobilis (2013) and Primiceri (2005).

2.1. Recursive Least Squares Estimation

In what follows we provide an outline of the RLS methodology and show why it is not always appropriate when parameter values are changing. Ignoring initial values whose relevance decreases over time, the RLS corresponds to the minimization at each t of the loss function,

$$J_t = \frac{1}{2}\sum_{i=1}^{t}(y_i - \mathbf{x}_i'\boldsymbol{\theta}_{\boldsymbol{t}})'(y_i - \mathbf{x}_i'\boldsymbol{\theta}_{\boldsymbol{t}}),\ t = 1, ..., T \tag{3}$$

producing parameter estimates, $\hat{\boldsymbol{\theta}}_t = \left(\sum_{i=1}^{t}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\sum_{i=1}^{t}\mathbf{x}_i y_i$, which can be written recursively as,

$$\hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\theta}}_{t-1} + \left(\sum_{i=1}^{t}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\mathbf{x}_t(y_t - \mathbf{x}_t'\hat{\boldsymbol{\theta}}_{t-1}). \tag{4}$$

If $\left(\sum_{i=1}^{t}\mathbf{x}_i\mathbf{x}_i'\right)^{-1} = \mathbf{P}_t$ the usual expression for the Kalman gain matrix, $\mathbf{K}_t = \mathbf{P}_t\mathbf{x}_t'$, is obtained which can be recursively updated. From $\mathbf{P}_t^{-1} = \mathbf{P}_{t-1}^{-1} + \mathbf{x}_t\mathbf{x}_t'$, the Woodbury matrix identity can be used to arrive at the parameter covariance matrix updating equation,

$$\mathbf{P}_t = \mathbf{P}_{t-1} - \frac{\mathbf{P}_{t-1}\mathbf{x}_t\mathbf{x}_t'\mathbf{P}_{t-1}}{1 + \mathbf{x}_t'\mathbf{P}_{t-1}\mathbf{x}_t}. \tag{5}$$

Expressions (4) and (5) are derived from RLS, however they are also analogous to the updating equations of the KF. Here, $\mathbf{P}_t$ is the estimated parameter covariance matrix at time t using information at time t. At this point, it becomes apparent why the RLS is not appropriate for estimating TVPs. As $t \to \infty$, $\mathbf{P}_t$ will approach 0 and $\hat{\boldsymbol{\theta}}_t$ will converge to a particular value as long as $\mathbf{x}_t$ is, in some sense, bounded away from 0. Thus, to handle TVPs some discounting must be introduced to stop $\mathbf{P}_t$ from shrinking. This can be achieved by exponentially decreasing the weight of older data through the use of a scalar forgetting factor $\lambda \in [0, 1]$. Hence, the solution of the RLS minimization problem is,

$$\hat{\boldsymbol{\theta}}_t = \left(\sum_{i=1}^{t}\lambda^{t-i}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\sum_{i=1}^{t}\lambda^{t-i}\mathbf{x}_i y_i \tag{6}$$

and the corresponding updating equations are,

$$\hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\theta}}_{t-1} + \mathbf{P}_{t-1}\mathbf{x}_t(y_t - \mathbf{x}_t'\hat{\boldsymbol{\theta}}_{t-1}); \tag{7}$$

$$\mathbf{P}_t = \lambda^{-1}\Big(\mathbf{P}_{t-1} - \frac{\mathbf{P}_{t-1}\mathbf{x}_t\mathbf{x}_t'\mathbf{P}_{t-1}}{1+\mathbf{x}_t'\mathbf{P}_{t-1}\mathbf{x}_t}\Big). \tag{8}$$

Using the forgetting factor, λ, an exponential information decay is forced into the RLS solution, thereby keeping $\mathbf{P}_t$ from reaching zero. This is known as exponential forgetting (EF). Inversely this is the same as preventing the information matrix $\mathbf{P}_t^{-1}$ from growing without bound. This can be shown by rearranging the covariance updating equation in (8) based on the Woodbury identity as,

$$\mathbf{P}_t^{-1} = \lambda\mathbf{P}_{t-1}^{-1} + \mathbf{x}_t\mathbf{x}_t'. \tag{9}$$

2.2. The Kalman Filter

KF works in two steps per iteration. It links $\boldsymbol{\theta}_{t|t-1}$ to $\boldsymbol{\theta}_{t|t}$, by updating the previous parameter estimates using data in period t, and predicts or propagates the parameter distribution for time $t+1$ using the specified model structure. In the TVP framework the updating equations are, $\boldsymbol{\theta}_{t|t} = \boldsymbol{\theta}_{t|t-1} + \mathbf{K}_t e_t$; with $\mathbf{K}_t = \frac{\mathbf{P}_{t|t-1}\mathbf{x}_t}{1+\mathbf{x}_t'\mathbf{P}_{t|t-1}\mathbf{x}_t} = \mathbf{P}_{t|t}\mathbf{x}_t$, where $\boldsymbol{\theta}_{t|t}$ is an estimate of the first moment of the parameter distribution at time t given information at time t, $\mathbf{P}_{t|t}$ is the estimate of its variance, $\mathbf{K}_t$ is the Kalman gain and e_t is the predictive error, $e_t = y_t - \mathbf{x}_t'\boldsymbol{\theta}_{t|t-1}$. The propagating steps are,

$$\boldsymbol{\theta}_{t+1|t} = \boldsymbol{\theta}_{t|t}; \tag{10}$$

$$\mathbf{P}_{t+1|t} = \mathbf{P}_{t|t} + \mathbf{W}_t. \tag{11}$$

Since specifying and estimating $\mathbf{W}_t$ is computationally challenging Raftery *et al.* (2010) opt for a scalar forgetting factor approach using $\lambda \in [0,1]$ to specify $\mathbf{W}_t$ as $(\lambda^{-1}-1)\mathbf{P}_{t|t}$. This is equivalent to writing the propagating equation as $\mathbf{P}_{t+1|t} = \lambda^{-1}\mathbf{P}_{t|t}$ (which is the same as EF). Generally, a forgetting scheme will act as a function of $\mathbf{P}_{t|t}$, i.e., $\mathbf{P}_{t+1|t} = F(\mathbf{P}_{t|t}) \geq \mathbf{P}_{t|t}$. This flattens the estimated parameter distribution, adds uncertainty to the KF output, provides some means by which KF does not converge, and allows for the tracking of the TVPs.

3. Dynamic Model Averaging and Forgetting Methods

Forgetting approaches are applied within the context of KF in order to efficiently manage information. These methods allow for parameter tracking while avoiding instabilities resulting from too much forgetting. This has potential benefits for TVP models, in particular, for those which are based on the state space framework. One application of TVP models is DMA. This forecasting methodology was introduced by Raftery *et al.* (2010) and subsequently brought into economics by Koop and Korobilis (2012), and has since been successfully employed in a number of studies; see, e.g., Aye *et al.* (2015) and Drachal (2016).

The DMA forecast is a weighted average of forecasts from individual models, which are assigned weights according to how well they are performing. In the canonical version of the DMA, TVP models are state space models, estimated with KF, using EF as discussed above. EF offers a simple means to allow for TVP in the state space set up. However, Park and Jun (1992) note that there are drawbacks connected with the use of scalar EF. Firstly, it employs the same rate of discounting on all variables. This may not be optimal if parameters move at different rates. Secondly, a fixed forgetting factor may result in a phenomenon known as parameter blow-up, where more information from the system is discarded via forgetting than added through new measurement variables (see (9)). If $\mathbf{x}_t\mathbf{x}_t'$ is small over several periods, the eigenvalues of $\mathbf{P}_{t+1|t}^{-1}$ will decrease and $\mathbf{K}_t$ will increase making KF unstable and overly sensitive to noise. A number of studies have attempted to address this issue; see *e.g.* Milek (1995), Parkum *et al.* (1992) and Kulhavý and Zarrop (1993).

The reason why a scalar forgetting factor can cause instability is that old information is discarded regardless of whether new information is added into the system. The main goal behind the proposed solutions to the blow-up problem is to better manage the information content of the filter in order to ease the bias-variance trade off between fast tracking and instability. In the following section we present three easy to implement forgetting schemes that potentially permit more efficient parameter tracking by managing the information content in the system, bounding the parameter covariance matrix, or through the use of a reference matrix.

Other approaches that allow for dynamics in the forgetting scheme have been explored by Dangl and Halling (2012), in which models with different scalar forgetting factors are included in the model averaging set, and Koop and Korobilis (2013), where a self-perturbed KF method of dynamic forgetting, based on the approach of Park and Jun (1992), is used. The idea behind methods such as Koop and Korobilis (2013) and Bork and Møller (2015) is that dynamic forgetting can make models more efficient with their information content, and thus mitigate the trade off between tracking and instability. Thus, in a flexible forgetting scheme, the variation of information content between competing models will likely be large which may lead to gains from forecast averaging. In what follows we provide a brief overview of several dynamic flexible forgetting methods which we will implement in the DMA framework. In Table 1 we briefly summarise the main characteristics of these forgetting methods for which a more detailed description is provided in Section 2.1 and in Sections 3.1 - 3.3.

[**Table 1 about here**]

3.1. Selective Forgetting

One way to manage the information in a filter is to change the forgetting factor to keep the trace of $\mathbf{P}_{t|t-1}$ constant over time. This is referred to as selective forgetting and was formalized by Fortescue *et al.* (1981). The idea is that instead of a fixed forgetting factor, λ, this parameter is allowed to vary over time based on the amount of information brought into the system by errors and predictors, i.e.,

$$\lambda_t = 1 - \frac{e_t^2}{\mathbf{V}_t(1 + \mathbf{x}_t' \mathbf{P}_{t|t} \mathbf{x}_t)} \tag{12}$$

where the second term on the right-hand side of (12) is the squared forecast error scaled by the amount of new information entering the system ($\mathbf{V}_t$ is the variance of the innovations); see Saelid and Foss (1983). Thus, low forecast errors or information arriving in the form of predictor variation will result in low forgetting, that is, λ_t will take values close to 1. In macroeconomics, it is often the case that the variance changes over time, which makes selective forgetting factors appealing both theoretically and empirically.

Recently Bork and Møller (2015) developed a system in which a time varying forgetting factor is produced based on the value of the forecast error and which appears to work quite well in their DMA application. Other studies, such as Grassi *et al.* (2017) and Park and Jun (1992) also select forgetting factors based on the ratio of the actual and expected error variance, however the style of forgetting used is somewhat different.

3.2. Directional Selective Forgetting

One drawback of the selective forgetting method described in (12) in managing the information content is that instability can still arise if variation shifts across predictors. When the flow of new information is non-uniform, gains can be made in the bias-variance trade off if information is more quickly forgotten in the dimensions of the predictor space that carry more information. Early algorithms that consider directional selective forgetting (DSF) have been developed by Saelid and Foss (1983) [henceforth DSF_{SF}] and Parkum *et al.* (1992) [henceforth DSF_{PPH}]. This is an interesting idea since it is often the case that different predictors have different variances, and that these variances may change over time.

DSF can be thought of as updating $\boldsymbol{P}_{t|t}$ by multiplication with a diagonal matrix, $\boldsymbol{\Lambda}$, i.e.,

$$\boldsymbol{P}_{t+1|t} = \boldsymbol{\Lambda} \boldsymbol{P}_{t|t} \boldsymbol{\Lambda}, \tag{13}$$

where each diagonal element of $\boldsymbol{\Lambda}$ is $\lambda_m^{-1/2}$, $m = 1, 2, ..., M$, which corresponds to the specific forgetting factor for predictor x_{mt}.

If $\lambda_1 = \lambda_2 = ... = \lambda_M$ this corresponds to a common scalar in the dispersion propagation step of KF. However, if the forgetting factors are different, then each coefficient will weigh past observations differently.

3.2.1. The Saelid and Foss algorithm The DSF_{SF} algorithm updates individual forgetting factors at time t. We start from (13) and use λ_{mt} to represent the forgetting factor that is used to propagate predictor m's uncertainty at time t. The algorithm increases or decreases the value of λ_{mt} based on the forecast error at $t-1$ and predictor values at time t, i.e., $\lambda_{mt} := \max\left(1 - \frac{e_{t-1}^2}{N_0(\mathsf{H}_{t-1} + \mathsf{x}_t' \boldsymbol{P}_{t|t} \mathsf{x}_t)} \gamma_{mt}, \lambda_{min}\right)$

and $\gamma_{mt} := \min\big(\frac{\sum_{j=1}^{M}\sigma_{j,j}\mathbf{x}_{jt}^2}{M\sigma_{m,m}x_{mt}^2}, 1\big)$, where $\sigma_{m,m}$ is the mth diagonal element of $\boldsymbol{P}_{t|t}$, $x_{mt}, m = 1, ..., M$ are the predictors, λ_{min} is the minimum bound for the forgetting parameters, and $\mathbf{H}_t$ is the rolling estimate of the variance of the signal variable. We include a tuning parameter N_0 that dictates the sensitivity of the forgetting algorithm. The part of the algorithm that represents the change in the individual forgetting factor is γ_{mt}, which dictates the predictor specific forgetting factors.

3.2.2. The Parkum, Poulsen and Holst algorithm The DSF_{PPH} algorithm is based on the same concept as DSF_{SF}, however it differs in its application by placing bounds on the eigenvalues of $\boldsymbol{P}_{t+1|t}$. These bounds prevent the propagated covariance matrix from becoming too large and therefore over sensitive to small forecast errors, while the lower bounds prevent $\boldsymbol{P}_{t+1|t}$ from decreasing to a point where it is no longer tracking time varying parameters. The following algorithm selects the individual forgetting factors as,

$$\lambda_{mt} = \begin{cases} \alpha_{t|t}^{(m)}\big[\alpha_{min} + \alpha_{t|t}^{(m)}\frac{\alpha_{max}-\alpha_{min}}{\alpha_{max}}\big]^{-1} & \text{if } \alpha_{t|t}^{(m)} \le \alpha_{max} \\ 1 & \text{if } \alpha_{t|t}^{(m)} > \alpha_{max} \end{cases}, \tag{14}$$

where $\alpha_{t|t}^{(m)}$ is an eigenvalue of $\boldsymbol{P}_{t|t}$. These forgetting factors are then aligned along the main diagonal of $\boldsymbol{\Lambda}$ in (13). DSF_{PPH} essentially sets a range for the eigenvalues of $\boldsymbol{P}_{t|t}$. When the m^{th} eigenvalue of $\boldsymbol{P}_{t|t}$ is greater than α_{max}, indicating that there is a low amount of information being received in the direction of m, the algorithm sets the forgetting factor pertaining to that direction to 1, meaning no forgetting. For eigenvalues between 0 and α_{max} the algorithm gets forgetting factors via an upward sloping concave function. The higher α_{min} the more the filter is able to track time varying parameters.

3.3. Stabilized Linear Forgetting

Stabilized linear forgetting (SLF) is discussed by Milek (1995) and Milek and Kraus (1995) as a RLS based forgetting method that avoids the blow-up of $\mathbf{P}_{t+1|t}$, yet allowing for sufficient parameter tracking. Stabilization comes from a reference matrix, ideally one that includes any prior knowledge regarding the variance of the parameter vector. In this paper, we will focus on linear versions of stabilized forgetting.

The SLF method does not use a fixed λ forgetting factor directly, instead it includes and additive reference matrix, $\mathbf{G}$,

$$\mathbf{P}_{t+1|t} = \mu\mathbf{P}_{t|t} + \mathbf{G} \tag{15}$$

where $\mu \in (0, 1)$ and $\mathbf{G}$ is a positive semi-definite reference matrix that provides an upper bound for the covariance matrix. The idea behind SLF is to forget information proportionally to the difference between $\mathbf{P}_{t|t}$ and some upper bound. Precise details on how the upper bound, and the contraction factor μ stabilize the propagation of the $\mathbf{P}_{t|t}$ matrix and allow parameter tracking while still avoiding instability are given in Milek and Kraus (1995).

One advantage of SLF is its similarity to variance propagation in the familiar KF setting with parameters that follow a random walk and covariance matrix $\mathbf{G}$, i.e., $\mathbf{P}_{t+1|t} = \mathbf{P}_{t|t} + \mathbf{G}$. The difference between this and SLF is that $\mathbf{P}_{t|t}$ in (15) is scaled down by μ producing what in effect are the eigenvalues of $\mathbf{P}_{t+1|t}$ as a linear function of $\mathbf{P}_{t|t}$, that has an intercept greater than 0 and a slope less than 1.

Rearranging (15) it follows that $\mathbf{P}_{t+1|t} = \mathbf{P}_{t|t} + (1-\mu)(\mathbf{P}^* - \mathbf{P}_{t|t})$, where the propagation step is based on the difference between $\mathbf{P}^* = \mathbf{G}/(1-\mu)$ and the updated parameter covariance matrix. This can be seen as a modification to the filter, but also as a model based means of forgetting, similar to that given in (11). One strength of this approach is the reference matrix $\mathbf{G}$, which can incorporate prior information on how parameters are expected to co-move. A diagonal matrix for example will give the eigenvalues of $\mathbf{P}_{t|t}$ a tendency to move towards lines spanned by unit vectors (for example the $(1,0)$ and $(0,1)$ 2 dimensional space). The strength of this tendency will depend on the norm of $\mathbf{G}$.

In a normal setting when adequate information is provided to the system, such that there is no potential for instability, $\mathbf{P}^* - \mathbf{P}_{t|t}$ is a positive semi-definite matrix that increases the variance on at least some axis of the parameter covariance matrix. When not enough information from the predictors is received $\mathbf{P}_{t|t} = (\mathbf{P}_{t|t-1}^{-1} + \boldsymbol{x}_t \boldsymbol{x'}_t)^{-1}$ increases and the gain matrix from the filter becomes unstable. In this case $\mathbf{P}^* - \mathbf{P}_{t|t}$ is negative, and shrinks the propagated covariance matrix relative to the updated one, thus preventing instability in the gain matrix.

4. Monte Carlo Analysis

To better understand where the gains from non-fixed forgetting schemes come from we perform an in depth Monte Carlo investigation. Our simulation study is designed to identify any gains in using different forgetting schemes in terms of the bias-variance trade off. We use a simple KF to identify state parameter variables and conduct forecasts. The results we compare are the average mean square forecast error (MSFE) of the signal variable, as is done in the empirical application, as well as the average mean square prediction error (MSPE) of the actual state variables, and the spread of the 90% confidence interval. Note that we are not simulating the DMA, we are only simulating a state space system in which various forgetting factors are applied to KF. DMA would average across models such as these. However, if individual models are more accurate, overall the DMA is also expected to achieve better estimates of the state variables and better forecasts.

4.1. Simulation Setup

The random walk model of parameter movement is frequently applied in TVP models since little is known *a priori* about the parameter structure. However, other models may also suitably represent parameter variation in the data. In order to examine this more closely, we generate parameters from three different models: i) a random walk, ii) a deterministic setting, and iii) an Ornstein–Uhlenbeck (OU) process. We compare the different forgetting methods across each parameter specification with different forgetting schemes.

Experiment 1: The random walk (RW). In the RW case the data generation process (DGP) is,

$$\begin{aligned} \mathbf{x}_t &= \boldsymbol{\varphi}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t \qquad (16) \\ y_t^* &= 0.1y_{t-1}^* + \mathbf{x}_{t-1}'\boldsymbol{\theta}_t + \nu_t,\ \nu_t \sim N(0, 0.1) \qquad (17) \end{aligned}$$

where the state vector $\boldsymbol{\theta}_t$ is generated as, $\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim N(0, \boldsymbol{\Sigma}_t)$ as in Raftery *et al.* (2010), and $\boldsymbol{\Sigma}_t$ is a diagonal matrix that does not necessarily have common elements on the main diagonal. The predictors $\mathbf{x}_t$ are generated using an AR(1) process as indicated in (16). In this simulation exercise $\mathbf{x}_t$ is a 2×1 vector of predictors, $\boldsymbol{\varphi} :=$ diag$\{0.6, 0.6\}$ and $\boldsymbol{\varepsilon}_t$ is a 2×1 vector of innovations drawn from a mean zero bivariate normal distribution. The variances of the elements of $\boldsymbol{\varepsilon}_t$ are either equal to 1 or 0.1 depending on the particular simulation setup considered (see Table 2).

Experiment 2: The deterministic setting. The deterministic parameter specification includes breaks in the parameter values. The DGP is as in (16) - (17), but in this experiment $\boldsymbol{\theta}_t$ is deterministic, such that $\boldsymbol{\theta}_t = \boldsymbol{\theta} = (\theta_1, \theta_2)'$, with $\theta_1 = 0$ and

$$\theta_2 = \begin{cases} 0 & if\ t < 150 \\ 0.5 & if\ 150 \leq t < 350 \\ -0.5 & if\ 350 \leq t < 450 \\ 0 & if\ t \geq 450 \end{cases}.$$

Experiment 3: The Ornstein-Uhlenbeck (OU) process. When $\boldsymbol{\theta}_t$ is modelled as a RW, it is possible that it may grow in absolute value and cause instabilities in the resulting response variable, y. In order to rectify this concern, our last parameter specification is that of an OU process. Since OU is a mean reverting stochastic process, we avoid having instances in which parameter values increase in absolute value without bound. In this case, the DGP is as in (16) - (17) with the state vector $\boldsymbol{\theta}_t$ generated as $\boldsymbol{\theta_t} = \sigma \int_0^t e^{-\eta(t-s)} dW_s$ where $\sigma = 0.5$, η varies depending on the case considered, and W_s denotes a standard Wiener process. The starting values and the mean of the process are set to zero.

In general in each experiment (except in experiment 2), we consider four different cases:

Case 1: Parameters have the same DGP over time;

Case 2: Introduces changes in the parameter generating process. In the RW and OU specifications, half way through the sample, the variance of one of the parameters drops from 0.1 to 0.001 and from 0.5 to 0.1, respectively;

Case 3: The parameter generation process is kept constant, but the predictor generating process changes half way through;

Case 4: The variance of a predictor and a parameter, change half way through the sample.

In all cases predictors are generated as in (16), however, in Cases 2 and 4, the variance of one of the predictors decreases from 1 to 0.1.

To evaluate the simulation results we consider: i) the average MSFE for the y variable (Avy); ii) the average MSPE which measures how well, on average, the KF in each specification tracks the movement of the parameters (Avs); and lastly, iii) the spread of the parameter estimates, i.e., the distance between the 90th and 10th percentile of the estimates. A higher value for the spread suggests a higher dispersion of the computed parameter estimates. The hyper-parameters used in each forgetting method are constant throughout all cases and specifications. All results are based on the sample size $T = 500$ and 1000 Monte Carlo replications.

4.2. The Simulation Results

The competing models used in the simulations are: the RLS, which is equivalent to using a scalar forgetting factor of $\lambda = 1$, i.e. no discounting of older information; a model with a fixed scalar forgetting factor of $\lambda = 0.8$, which corresponds to the value which provided the lowest MSFE in Case 1; and DSF_{SF}, DSF_{PPH} and SLF. DSF_{SF} uses a lower bound of 0.9, DSF_{PPH} includes eigenvalue limits of $\alpha_{max} = 0.1$ and $\alpha_{min} = 0.001$, and SLF is as in (15) with $\mu = 0.99$ and $\mathbf{G} = 0.001\mathbf{I}$, where $\mathbf{I}$ is an identity matrix. Apart from the fixed values, the hyper-parameters chosen were the ones that appeared to work well in the empirical application.

[**Table 2 about here**]

Table 2 shows that RLS presents the worst forecast performance and parameter estimation accuracy. This is not surprising as RLS is not designed for time varying parameters. However, it delivers the lowest spread, which suggests that in cases in which parameters are known not to vary it would perform well.

Fixed forgetting with $\lambda = 0.8$ does relatively well in the random walk specification. In Case 1 fixed forgetting has the lowest forecast and prediction error. With constant variance for both predictors and constant movement of both parameters, we have no need for directional forgetting. Information is constantly arriving into the system and no efficiency gains can be made through the use of a more flexible forgetting method. There is also little chance of a blow up situation.

For changes in parameter variation, predictor variation or both, as in Cases 2, 3 and 4, respectively, we observe the benefits of the other forgetting schemes. In cases where the variance of a predictor drops, forgetting schemes which manage the information more efficiently often have lower parameter prediction errors than the fixed forgetting method. This only translates into lower forecast errors in the case of SLF as the variance of the predictors is quite large compared to the variance of the parameters and thus much of the gains from more accurate parameter predictions is washed out.

In the case of deterministic parameters with fixed break points there is a clear advantage of DSF_{PPH}. The limits placed on the size of the covariance matrix keep the system from becoming unstable after discrete jumps in parameter values. When we introduce variability in the variance of the predictors, we also find that SLF performs quite well. The fixed forgetting methods suffer from a large spread in parameter prediction making them less effective in real world applications when parameters are thought to be mostly stable with large discrete breaks. Interestingly, we notice a marked increase in the spread for all flexible forgetting schemes when we introduce a drop in the variance of a predictor halfway through the sample. SLF appears to cope with this quite well. The decrease in information in one of the predictors does not cause the instability found with EF. Similarly to the DSF setting, SLF also sets a bound for the eigenvalues of the covariance matrix. However, it differs in how information is forgotten. The use of the SLF reference matrix, which in these simulations is set to $0.001\mathbf{I}$, keeps the propagated covariance matrix from becoming too small and losing tracking ability while avoiding the blow up situation.

For parameters following an OU process the SLF and DSF_{PPH} methods perform quite well in terms of forecast and parameter prediction error. DSF_{PPH} suffers however from a large variance of the parameter estimates which could make empirical applications problematic. It is interesting to note that although η, which governs the central tendency of the OU process is 0.5 and drops to 0.1 halfway through the sample depending on the case, SLF and DSF_{PPH} continue to perform well in terms of MSFE, MSPE and spread, with DSF_{PPH} displaying sizable forecasting gains. We use a diagnoal matrix for the reference matrix $\mathbf{G}$ in the SLF method which acts as a prior of independence on $\boldsymbol{\theta}_t$. This may give SLF an advantage as it reinforces the directions of the covariance matrix's eigenvectors towards independence (Milek and Kraus, 1995).

5. Empirical Application of DMA

In this section we explore how the different methods of forgetting compare when applied to DMA in two forecasting applications. Firstly, we consider the US inflation data used in Koop and Korobilis (2012). Secondly, we consider forecasting S&P 500 returns using data from Welch and Goyal (2008). In each case we compare the MSFE and MAFE from different forgetting methods for each TVP model in the DMA and the dynamic model selection (DMS) settings. DMS is a special case of DMA in which the best model receives a weight of 1, and the remaining models are given a weight of 0.

In these two empirical cases, the number of models averaged over in the DMA grows exponentially with the number of predictors. This can make the DMA a time consuming procedure when the predictor set is large. For this reason we reduce the predictor space via factors in the first application and use only 6 predictors in the second application. All models also contain a constant as well as a lagged autoregressive term.

5.1. Forecasting exercises

5.1.1. Inflation forecasting We start by applying DMA using the forgetting methods described above, to forecast US inflation. The predictors used are: 1) Unemployment Rate (UNRATE); 2) Real Personal Consumption Expenditures (PCECC96); 3) Private Residential Fixed Investment (PRFI); 4) Real Gross Domestic Product (GDPC1); 5) Housing Starts: Total: New Privately Owned Housing Units Started (HOUST); 6) Industrial Production Index (INDPRO); 7) All Employees, Total Private (USPRIV); 8) Employment-Population Ratio (EMRATIO); 9) Average Hourly Earnings of Production and Nonsupervisory Employees, Manufacturing (AHEMAN); 10) 3-Month Treasury Bill: Secondary Market Rate (TB3MS); 11) Moody's Seasoned Aaa Corporate Bond Yield (AAA); and 12) M1 Money Stock (M1SL). The predictors are divided into blocks of employment variables (variables 1), 7) and 8)), real economic variables (variables 2), 3), 4), 5), 6), and 9)) and financial variables (variables 10), 11), and 12)). A factor is extracted from each block and used as a predictor.

We forecast two versions of inflation based on personal consumption expenditures: the Chain-type Price Index (PCE) and the Consumer Price Index for All Urban Consumers (CPI). All variables are log first differences. We consider two out-of-sample periods (one (h=1) and four (h=4) quarters ahead), with starting dates 2001:Q1 and 2006:Q1.

Numerous DMA forgetting schemes are tested in our analysis starting with fixed $\lambda = \{0.9, 0.95, 0.97, 0.99, 1\}$. This is followed by the DSF_{SF} and the DSF_{PPH} algorithms with upper and lower bounds set to 1.0E-5 and 1.0E-6, respectively, for the $\mathbf{P}_{t|t}$ matrix; as well as by the variable forgetting method of Bork and Møller (2015) (BM). This method uses a time-varying version of the scalar EF factor λ. Briefly, λ decreases, meaning more forgetting, when the model's forecast errors are large relative to the expanding history of previous forecast errors. A number of quantiles are calculated from the expanding history, and λ shifts based on whether the current forecast error is in a higher quantile than the previous.[1] Lastly, we look

1. This method has been employed in a house price forecasting DMA application in both Bork and Møller (2015) and Risse and Kern (2016).

at a specification of the SLF forgetting scheme of Milek (1995) with $\mu = 0.99$ and reference matrix $G = (1.0e-8)\mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

[**Tables 3 about here**]

Table 3 presents results for CPI.[2] The MSFE and MAFE of the best performing fixed forgetting parameter is included in the table and is $\lambda = 0.97$ and $\alpha = 1$ for one quarter ahead forecasts, and $\lambda = 0.95$ and $\alpha = 1$ for four quarters ahead. The α tuning parameter is used in the DMA procedure to govern how much uncertainty the forecaster attaches to the model weight. α has a similar interpretation to the fixed forgetting parameter λ in the sense that the closer α is to 1, the more memory the model weights will have, meaning the weight reflects not only the model's recent performance, but also how it did in the past. We also present results for different values of α in the tables.

Results show that the DSF_{PPH} and SLF methods work quite well, offering the lowest MSFE and MAFE for both one and four quarters in each out-of-sample period. We also conduct Diebold Mariano (DM) tests for predictive accuracy of the various flexible forgetting schemes against the fixed exponential forgetting scheme with the lowest MSFE. According to the DM test, for the one quarter ahead forecasts for both out-of-sample periods, we find that the DSF_{PPH} and SLF forgetting methods perform significantly better than the best competing exponential forgetting scheme. For four quarters ahead forecasts, we find that SLF has lower forecast errors than the best performing exponential forgetting, however they are only significant for the DMS.

Results for forecasting schemes that increase forgetting upon receiving larger forecasting errors is mixed. Both the BM and DSF_{SF} do not significantly improve upon the best fixed forgetting scheme despite having lower MSFE and MAFE for the one quarter ahead forecasts. This suggests that a large forecast error variance could be causing BM and DSF_{SF} to increase forgetting when it is not required. The idea behind these forgetting schemes is that large forecast errors may be the result of models being misspecified and therefore requiring a larger forgetting factor. This does not appear to be the case in this setting. Large forecast errors appear to be noise, and thus changing parameter values as a result does not improve forecast accuracy.

5.1.2. Stock returns forecasting It is well known that returns are difficult to predict and any predictive power of a given model may be short lived (Pesaran and Timmermann, 2002). We investigate the performance of the methods discussed above for forecasting $S\&P$500 end-of-month returns. Results are shown in Table 4. As predictors we use the 12 month moving sum of dividends (D12), the 12 month moving sum of earnings (E12), the book to market (b/m), the treasure bills

2. Results for PCE forecasts lead to similar conclusions and are therefore omitted, but these can be obtained from the authors.

(tbl), and AAA and BAA corporate bond yields[3], which are all log first differences. We use a sample starting in March 1990 and three out of sample periods. The first starts in December 2000, the second in December 2005 and the third in December 2008. The end of all out of sample periods is November 2019. We present results for both $h = 1$ and $h = 4$ (see Table 4). Three model specifications are considered with varying values of the DMA tuning parameter α, namely: i) the DSF_{PPH} method; ii) the BM method; and iii) the SLF. The DSF_{SF} method was not as competitive so we left it out in order to simplify Table 4.

[**Table 4 about here**]

SLF appears to give the strongest forecasting performance in terms of lower MSFE and MAFE, for both forecast horizons, however according to the DM test it is only significantly better for four month ahead forecasts. Other forgetting schemes are not as promising, particularly the SF and BM method which adjust forgetting based on the forecast error. The usage of the reference matrix $\mathbf{G}$ in the SLF propagation step may be what is helping the SLF model. We adjust $\mathbf{G}$ to allow more forgetting for the AR coefficient relative to the constant and other predictor coefficients. It is interesting to note that for four month ahead forecasts in the December 2008 to November 2019 sample, the fixed or EF appears to deliver superior forecasts. Note that the best fixed forgetting scheme uses $\lambda = 1$ for $h = 1$ and $\lambda = 0.99$ for $h = 4$, which suggests that little or no forgetting is applied. This in line with Case 1 in the simulation exercise, where the out of sample period is characterized by constant variance for predictors and parameters thus favouring the same constant forgetting factor applied to all predictors.

6. Conclusion

In this paper we used different means of discounting older information in favour of more recent information in the DMA procedure put forward by Raftery *et al.* (2010). These discounting methods, known as forgetting schemes vary from a simple fixed scalar forgetting factor that discounts uniformly across the predictors, to more complicated methods that either change the amount of discounting based on the magnitude of the squared forecast error, place bounds on eigenvalues of the covariance matrix, or use a reference matrix to guide the evolution of the covariance matrix through the filter. According to our simulated and empirical applications we find the specific context will dictate which method is preferred. For instance, in the simulation, depending on whether we expect the parameters to evolve as a random walk, or act deterministically with discrete breaks, the SLF and DSF_{PPH} methods appear to offer better forecasts. On the empirical side, the DSF_{PPH} and SLF

3. The data used in the returns' prediction analysis was taken from Amit Goyal's website: https://www.hec.unil.ch/agoyal.

methods again appear to do quite well in forecasting CPI and returns over different forecast horizons in comparison to the other competing forgetting schemes.

Our findings suggest that when predicting noisy series, basing the amount of discounting on the magnitude of the forecast error could lead us to discard older information under the incorrect assumption of parameter movement. One interpretation of the empirical work is that this could be the case. On the other hand, placing a bound on the eigenvalues of the covariance matrix appears to work well in simulations, however it offers room for improvement when forecasting returns.

It is important to consider, that as with most forecasting methodologies, tuning parameters play an important role in model performance. Choosing the correct forgetting factor tuning parameter, regardless of whether it is a single scalar, or involves picking bounds of eigenvalues can be challenging and usually involves some trial and error. One advantage of using SLF is the possibility of selecting the reference matrix $\mathbf{G}$ using some prior knowledge about the distribution of the parameters.

References

Aye, Goodness, Rangan Gupta, Shawkat Hammoudeh, and Won Joong Kim (2015). "Forecasting the price of gold using dynamic model averaging." *International Review of Financial Analysis*, 41, 257–266.

Bork, Lasse and Stig V Møller (2015). "Forecasting house prices in the 50 states using Dynamic Model Averaging and Dynamic Model Selection." *International Journal of Forecasting*, 31(1), 63–78.

Cogley, Timothy and Thomas J Sargent (2005). "Drifts and volatilities: monetary policies and outcomes in the post WWII US." *Review of Economic Dynamics*, 8(2), 262–302.

Dangl, Thomas and Michael Halling (2012). "Predictive regressions with time-varying coefficients." *Journal of Financial Economics*, 106(1), 157–181.

Drachal, Krzysztof (2016). "Forecasting spot oil price in a dynamic model averaging framework: Have the determinants changed over time?" *Energy Economics*, 60, 35–46.

Durbin, James and Siem Jan Koopman (2000). "Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1), 3–56.

Durbin, James and Siem Jan Koopman (2012). *Time series analysis by state space methods*. Oxford university press.

Fortescue, TR, Lester S Kershenbaum, and B Erik Ydstie (1981). "Implementation of self-tuning regulators with variable forgetting factors." *Automatica*, 17(6), 831–835.

Grassi, Stefano, Nima Nonejad, and Paolo Santucci De Magistris (2017). "Forecasting With the Standardized Self-Perturbed Kalman Filter." *Journal of Applied Econometrics*, 32(2), 318–341.

Groen, Jan JJ, Richard Paap, and Francesco Ravazzolo (2013). "Real-time inflation forecasting in a changing world." *Journal of Business & Economic Statistics*, 31(1), 29–44.

Hägglund, Tore (1984). "The problem of forgetting old data in recursive estimation." In *Adaptive Systems in Control and Signal Processing 1983*, pp. 213–214. Elsevier.

Harvey, Andrew C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Jazwinski, Andrew H (1970). *Stochastic processes and filtering theory*. Academic Press Inc.

Koop, Gary and Dimitris Korobilis (2012). "Forecasting inflation using dynamic model averaging." *International Economic Review*, 53(3), 867–886.

Koop, Gary and Dimitris Korobilis (2013). "Large time-varying parameter VARs." *Journal of Econometrics*, 177(2), 185–198.

Koop, Gary, Roberto Leon-Gonzalez, and Rodney W Strachan (2009). "On the evolution of the monetary policy transmission mechanism." *Journal of Economic*

Dynamics and Control, 33(4), 997–1017.

Korobilis, Dimitris (2013). "Assessing the transmission of monetary policy using time-varying parameter dynamic factor models." *Oxford Bulletin of Economics and Statistics*, 75(2), 157–179.

Kulhavỳ, R and M Kárnỳ (1984). "Tracking of slowly varying parameters by directional forgetting." *IFAC Proceedings Volumes*, 17(2), 687–692.

Kulhavỳ, R and Martin B Zarrop (1993). "On a general concept of forgetting." *International Journal of Control*, 58(4), 905–924.

Milek, JJ and FJ Kraus (1995). "Time-varying stabilized forgetting for recursive least squares identification." In *Adaptive Systems in Control and Signal Processing 1995*, pp. 137–142. Elsevier.

Milek, JT (1995). *Stabilized adaptive forgetting in recursive parameter estimation*. 4, vdf Hochschulverlag AG.

Park, D-J and B-E Jun (1992). "Selfperturbing recursive least squares algorithm with fast tracking capability." *Electronics Letters*, 28(6), 558–559.

Parkum, JE, Niels Kjølstad Poulsen, and Jan Holst (1992). "Recursive forgetting algorithms." *International Journal of Control*, 55(1), 109–128.

Pesaran, M Hashem and Allan Timmermann (2002). "Market timing and return prediction under model instability." *Journal of Empirical Finance*, 9(5), 495–510.

Pollock, D Stephen G (2003). "Recursive estimation in econometrics." *Computational Statistics & Data Analysis*, 44(1-2), 37–75.

Primiceri, Giorgio E (2005). "Time varying structural vector autoregressions and monetary policy." *The Review of Economic Studies*, 72(3), 821–852.

Raftery, Adrian E, Miroslav Kárnỳ, and Pavel Ettler (2010). "Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill." *Technometrics*, 52(1), 52–66.

Risse, Marian and Martin Kern (2016). "Forecasting house-price growth in the Euro area with dynamic model averaging." *The North American Journal of Economics and Finance*, 38, 70–85.

Saelid, Steinar and Bjarne Foss (1983). "Adaptive controllers with a vector variable forgetting factor." In *The 22nd IEEE Conference on Decision and Control*, pp. 1488–1494, IEEE.

Welch, Ivo and Amit Goyal (2008). "A comprehensive look at the empirical performance of equity premium prediction." *The Review of Financial Studies*, 21(4), 1455–1508.

Approach	Description	Usage	Forgetting Parameters
Exponential forgetting (EF) See Section 2.1.	Discounts information uniformly across all predictors and over time. Computationally efficient and simple to implement with KF. Potentially unstable if too much information is forgotten relative to that received.	Good to use in large model averaging settings.	Typically, a fixed forgetting parameter, $\lambda \in [0,1]$ is used. $\lambda = 1$ corresponds to no forgetting.
Dynamic Selective Forgetting: Saelid and Foss (1983) (DSF_{SF}). See Section 3.2.1	Forgetting is based on information content of the state space system. Forgetting increases in all directions as forecast error increases and is applied to each component of the parameter covariance matrix based on the relative potential for instability.	When noise and model confidence is expected to be low such that large forecasting errors are interpreted as coming from incorrect models and not large idiosyncratic innovations.	The forgetting parameter λ_{mt} increases or decreases based on the forecast error at time $t-1$ and predictor values at time t, i.e., $$\lambda_{mt} := \max\big(1 - \frac{e_{t-1}^2}{N_0(\mathbf{H}_{t-1} + \mathbf{x}_t' \boldsymbol{P}_{t\|t} \mathbf{x}_t)}\gamma_{mt}, \lambda_{min}\big)$$ $$\gamma_{mt} := \min\big(\frac{\sum_{j=1}^{M} \sigma_{j,j} \mathbf{x}_{jt}^2}{M \sigma_{m,m} x_{mt}^2}, 1\big)$$ where $\sigma_{m,m}$ is the mth diagonal element of $\boldsymbol{P}_{t\|t}$, x_{mt}, $m = 1, ..., M$ are the predictors, λ_{min} is the minimum bound for the forgetting parameters, and $\mathbf{H}_t$ is the rolling estimate of the variance of the signal variable. N_0 is a sensitivity parameter.
Dynamic Selective Forgetting: Parkum et al. (1992) (DSF_{PPH}). See Section 3.2.2	Forgetting factors are specific to each component of the parameter space and are time varying so that the eigenvalues of the parameter covariance matrix are bounded from above and below.	This method works well in most cases, however some experimentation is required to find appropriate bounds for the eigenvalues.	The forgetting parameter is, $$\lambda_{mt} = \begin{cases} \alpha_{t\|t}^{(m)}\big[\alpha_{min} + \alpha_{t\|t}^{(m)} \frac{\alpha_{max} - \alpha_{min}}{\alpha_{max}}\big]^{-1} & \text{if } \alpha_{t\|t}^{(m)} \leq \alpha_{max} \\ 1 & \text{if } \alpha_{t\|t}^{(m)} > \alpha_{max} \end{cases},$$ where $\alpha_{t\|t}^{(m)}$ is an eigenvalue of $\boldsymbol{P}_{t\|t}$, α_{max} is the upper eigenvalue bound on $\boldsymbol{P}_{t+1\|t}$, and α_{min} is the parameter governing the amount of forgetting.
Stabilized Linear Forgetting (SLF): Milek (1995). See Section 3.3	Sets a reference matrix to build an upper bound of the parameter covariance matrix. Forgetting is proportional to the difference between the upper bound of the parameter covariance matrix and the reference matrix.	When prior knowledge of parameters is available it can be incorporated into the reference matrix, allowing for certain parameters to vary more than others within some upper bound.	The SLF method uses, $$\mathbf{P}_{t+1\|t} = \mu \mathbf{P}_{t\|t} + \mathbf{G}$$ where $\mu \in (0,1)$ and $\mathbf{G}$ is a positive semi-definite reference matrix that provides an upper bound for the covariance matrix. Larger elements of reference matrix $\mathbf{G}$ mean more forgetting for the corresponding coefficient. μ dictates how much forgetting will take places, and at what point eigenvalues of $\boldsymbol{P}_{t+1\|t}$ will start to decrease to avoid instability.

Table 1. Summary of the main features of the EF, DSF_{SF}, DSF_{PPH} and SLF Forgetting Methods

		random walk parameter dynamics						deterministic parameter dynamics						OU parameter dynamics					
		$\sigma^2_{T_1}$	$\sigma^2_{T_2}$	Approach	Avy	Avs	Spread	$\sigma^2_{T_1}$	$\sigma^2_{T_2}$	Approach	Avy	Avs	Spread	$\sigma^2_{T_1}$	$\sigma^2_{T_2}$	Approach	Avy	Avs	Spread
CASE 1	$\sigma^2_{\beta_1}$	0.001	0.001	RLS	0.022	0.022	**483.979**	-	-	RLS	0.022	0.021	**27.542**	0.5	0.5	RLS	0.027	0.026	**470.034**
	$\sigma^2_{\beta_2}$	0.001	0.001	$EF_{(\lambda=0.8)}$	**0.008**	**0.006**	760.806	-	-	$EF_{(\lambda=0.8)}$	0.009	0.008	72.334	0.5	0.5	$EF_{(\lambda=0.8)}$	0.011	0.010	698.806
	$\sigma^2_{x_1}$	1	1	DSF_{SF}	0.009	0.007	732.504	1	1	DSF_{SF}	0.013	0.012	54.174	1	1	DSF_{SF}	0.011	0.011	686.539
	$\sigma^2_{x_2}$	1	1	SLF	0.008	0.006	744.440	1	1	SLF	0.009	0.008	62.530	1	1	SLF	0.010	0.010	694.702
				DSF_{PPH}	0.008	0.006	763.775			DSF_{PPH}	**0.008**	**0.007**	93.798			DSF_{PPH}	**0.009**	**0.009**	717.230
CASE 2	$\sigma^2_{\beta_1}$	0.001	0.001	RLS	0.020	0.019	**498.216**							0.5	0.5	RLS	0.023	0.022	**464.653**
	$\sigma^2_{\beta_2}$	0.001	0.0001	$EF_{(\lambda=0.8)}$	**0.008**	**0.005**	734.924							0.5	0.1	$EF_{(\lambda=0.8)}$	0.010	0.009	619.681
	$\sigma^2_{x_1}$	1	1	DSF_{SF}	0.008	0.006	709.578							1	1	DSF_{SF}	0.010	0.009	609.805
	$\sigma^2_{x_2}$	1	1	SLF	0.008	0.005	719.185							1	1	SLF	0.009	0.008	615.429
				DSF_{PPH}	0.008	0.006	737.214									DSF_{PPH}	**0.009**	**0.008**	636.076
CASE 3	$\sigma^2_{\beta_1}$	0.001	0.0001	RLS	0.018	0.023	**450.455**	-	-	RLS	0.025	0.021	**30.649**	0.5	0.5	RLS	0.022	0.027	**441.398**
	$\sigma^2_{\beta_2}$	0.001	0.0001	$EF_{(\lambda=0.8)}$	0.008	0.008	756.195	-	-	$EF_{(\lambda=0.8)}$	0.016	0.015	220.695	0.5	0.5	$EF_{(\lambda=0.8)}$	0.009	0.011	684.773
	$\sigma^2_{x\_1}$	1	1	DSF_{SF}	0.008	0.008	725.973	1	1	DSF_{SF}	0.017	0.013	134.742	1	1	DSF_{SF}	0.010	0.011	671.347
	$\sigma^2_{x\_2}$	1	0.1	SLF	**0.007**	**0.007**	733.880	1	0.1	SLF	0.016	0.010	107.404	1	0.1	SLF	0.009	0.011	676.980
				DSF_{PPH}	0.008	0.007	755.641			DSF_{PPH}	**0.016**	**0.010**	161.459			DSF_{PPH}	**0.008**	**0.010**	704.771
CASE 4	$\sigma^2_{\beta_1}$	0.001	0.001	RLS	0.018	0.020	**468.452**							0.5	0.5	RLS	0.021	0.023	**466.457**
	$\sigma^2_{\beta_2}$	0.001	0.0001	$EF_{(\lambda=0.8)}$	0.007	0.007	751.956							0.5	0.1	$EF_{(\lambda=0.8)}$	0.009	0.009	624.217
	$\sigma^2_{x_1}$	1	1	DSF_{SF}	0.008	0.007	724.789							1	1	DSF_{SF}	0.010	0.010	614.165
	$\sigma^2_{x_2}$	1	0.1	SLF	**0.007**	**0.006**	734.132							1	0.1	SLF	0.009	0.008	620.381
				DSF_{PPH}	0.008	0.007	751.631									DSF_{PPH}	**0.008**	**0.008**	640.545

Note: Results correspond to the MSFE of the simulated measurement series (Avy), and the MSPE of the state parameters' variances (Avs). Spread is the sum of the absolute difference between the 90th and 10th percentiles of the predictions for state parameters over time. σ_{T_1} and σ_{T_2} refer to the sample variances of the first and second half of the simulated sample and $T_1 + T_2 = 500$. DSF_{SF} is computed using a lower bound of 0.9; SLF is as in (15) with $\mu = 0.99$ and $G = 0.001\mathbf{I}$; and DSF_{PPH} is computed with upper and lower eigenvalue bounds $\alpha_{max} = 0.1$ and $\alpha_{min} = 0.001$, respectively. The Table is divided into four cases, which show results for the different predictor (σ_{x_i}) and parameter (σ_{β_i}) variances used.

Table 2. Simulation results of the forecast performance of RLS, EF with $\lambda = 0.8$, DSF_{SF}, DSF_{PPH} and SLF under random walk, deterministic and OU parameter dynamics

	h=1				h=4			
2000Q4-2019Q3	$MSFE_{DMA}$	$MAFE_{DMA}$	$MSFE_{DMS}$	$MAFE_{DMS}$	$MSFE_{DMA}$	$MAFE_{DMA}$	$MSFE_{DMS}$	$MAFE_{DMS}$
Bork Møller								
$\alpha = 90$	1.0019	1.0096	0.9784	0.9927	1.2001	1.2331	1.1886	1.2500
$\alpha = 95$	1.0017	1.0102	0.9921	1.0007	1.1983	1.2328	1.1536	1.2097
$\alpha = 97$	0.9997	1.0096	0.9872	1.0009	1.1991	1.2342	1.1913	1.2195
$\alpha = 99$	0.9888	1.0034	0.9600	0.9778	1.2100	1.2447	1.1935	1.2418
$\alpha = 1$	0.9781	0.9983	0.9432	0.9700	1.2405	1.2712	1.2543	1.2884
DSF_{SF}								
$\alpha = 90$	1.0056	1.0126	1.0021	0.9960	1.1563	1.1814	1.1403	1.1621
$\alpha = 95$	1.0091	1.0147	1.0042	1.0000	1.1588	1.1841	1.1360	1.1526
$\alpha = 97$	1.0094	1.0151	0.9998	0.9989	1.1598	1.1852	1.1444	1.1781
$\alpha = 99$	1.0037	1.0114	0.9973	0.9972	1.1619	1.1865	1.1429	1.1752
$\alpha = 1$	0.9983	1.0106	0.9975	1.0248	1.1783	1.2012	1.1825	1.2128
DSF_{PPH}								
$\alpha = 90$	0.9347**	0.9506**	0.9320**	0.9365**	1.0128	1.0518	1.0354	1.0671
$\alpha = 95$	0.9362**	0.9520**	0.9623*	0.9541*	1.0091	1.0493	1.0587	1.0708
$\alpha = 97$	0.9363**	0.9524**	0.9528*	0.9414*	1.0069	1.0468	1.0002	1.0064
$\alpha = 99$	0.9339**	0.951**	0.9366**	0.9508**	1.0124	1.0438	1.0297	1.0523
$\alpha = 1$	0.9301**	0.9506**	0.9265**	0.9454**	1.0295	1.0512	1.0297	1.0523
SLF								
$\alpha = 90$	0.8854**	0.9068**	0.8992**	0.9052**	0.9586	0.9784	0.9474	0.9757
$\alpha = 95$	0.8878**	0.9079**	0.9081*	0.9133*	0.9622	0.9803	0.9229**	0.9263**
$\alpha = 97$	0.8877**	0.9086**	0.9144*	0.9042*	0.9649	0.9823	0.931*	0.9454*
$\alpha = 99$	0.8828**	0.9096**	0.8938**	0.8940**	0.9774	0.9894	1.0024	1.0167
$\alpha = 1$	0.8825**	0.9121**	0.8866**	0.9096**	1.0052	1.0141	1.0101	1.0202

2005Q4-2019Q3	$MSFE_{DMA}$	$MAFE_{DMA}$	$MSFE_{DMS}$	$MAFE_{DMS}$	$MSFE_{DMA}$	$MAFE_{DMA}$	$MSFE_{DMS}$	$MAFE_{DMS}$
Bork Møller								
$\alpha = 90$	0.9908	1.0078	0.9573	0.9825	1.1436	1.1882	1.1079	1.1952
$\alpha = 95$	0.9908	1.0086	0.9661	0.9911	1.1414	1.1874	1.0816	1.1526
$\alpha = 97$	0.9887	1.0081	0.9617	0.9910	1.1418	1.1885	1.1280	1.1621
$\alpha = 99$	0.9762	1.0015	0.9306	0.9657	1.1520	1.1991	1.1475	1.1973
$\alpha = 1$	0.9636	0.9958	0.9238	0.9639	1.1839	1.2295	1.1971	1.2471
DSF_{SF}								
$\alpha = 90$	1.0011	1.0132	0.9951	0.9918	1.1019	1.1365	1.0701	1.1065
$\alpha = 95$	1.0058	1.0156	0.9952	0.9959	1.1050	1.1395	1.0582	1.0828
$\alpha = 97$	1.0063	1.0159	0.9892	0.9946	1.1065	1.1410	1.0694	1.1025
$\alpha = 99$	0.9989	1.0110	0.9830	0.9875	1.1099	1.1438	1.0972	1.1292
$\alpha = 1$	0.9926	1.0091	0.9691	0.9835	1.1278	1.1622	1.1299	1.1733
DSF_{PPH}								
$\alpha = 90$	0.9367*	0.953*	0.9383**	0.9365**	0.9754	1.0296	0.9970	1.0430
$\alpha = 95$	0.9388*	0.9547*	0.9706*	0.9540*	0.9713	1.0269	1.0108	1.0487
$\alpha = 97$	0.9393*	0.9552*	0.9603*	0.9404*	0.9688	1.0242	0.9578	0.9778
$\alpha = 99$	0.9373*	0.9535*	0.9370*	0.9496*	0.9755	1.0222	0.9964	1.0329
$\alpha = 1$	0.9343*	0.9530*	0.9293**	0.9443**	0.9966	1.0324	0.9964	1.0329
SLF								
$\alpha = 90$	0.8970*	0.9100*	0.9060*	0.9047*	0.9539	0.9800	0.9169	0.9695
$\alpha = 95$	0.9002*	0.9113*	0.9190*	0.9133*	0.9580	0.9816	0.8950**	0.9148**
$\alpha = 97$	0.9005*	0.9122*	0.9273	0.9027	0.9610	0.9837	0.9054**	0.9369**
$\alpha = 99$	0.8946*	0.9132*	0.8926*	0.8870*	0.9794	0.9949	1.0149	1.0283
$\alpha = 1$	0.8947*	0.9155*	0.8944*	0.9088*	1.0190	1.0267	1.0247	1.0322

Note: This Table reports the MSFE and MAFE ratios of different forgetting schemes relative to the lowest MSFE of a model with a fixed forgetting scheme. *, ** and *** indicate significant Diebold-Mariano test results at 10%, 5% and 1% nominal levels, respectively. CPI refers to Consumer Price Index for All Urban Consumers (FRED: CPIAUCSL). Bork and Møller method includes 5 quantiles. DSF_{SF} has tuning parameter $N_0 = 10$ and a lower bound of 0.9. DSF_{PPH} has tuning parameters $\alpha_{max} = 1.0e\text{-}6$ and $\alpha_{min} = \alpha_{max}/10$. For the SLF, we set $\mu = 0.99$ and $G = \{1.0e\text{-}8\}\mathbf{I}$. The fixed forgetting scheme in the benchmark model uses $\lambda = 0.97$ for one quarter ahead ($h = 1$) and $\lambda = 0.99$ for four quarters ahead ($h = 4$) forecast horizons.

Table 3. Out of sample MSFE and MAFE ratios of different forgetting schemes relative to the lowest MSFE fixed forgetting scheme for CPI

	Dec-00				Dec-05				Dec-08			
h=1	$MSFE_{DMA}$	$MAFE_{DMA}$	$MSFE_{DMS}$	$MAFE_{DMS}$	$MSFE_{DMA}$	$MAFE_{DMA}$	$MSFE_{DMS}$	$MAFE_{DMS}$	$MSFE_{DMA}$	$MAFE_{DMA}$	$MSFE_{DMS}$	$MAFE_{DMS}$
DSF_{PPH}												
$\alpha = 90$	1.0093	1.0149	1.0532	1.0942	1.0032	1.0031	1.0435	1.0966	1.0018	1.0117	1.0585	1.1563
$\alpha = 95$	1.0147	1.0265	1.0580	1.1004	1.0096	1.0181	1.0399	1.0999	1.0079	1.0302	1.0584	1.1451
$\alpha = 97$	1.0181	1.0323	1.0626	1.0991	1.0133	1.0252	1.0320	1.0801	1.0122	1.0397	1.0475	1.1147
$\alpha = 99$	1.0222	1.0353	1.0256	1.0362	1.0209	1.0295	1.0214	1.0266	1.0213	1.0488	1.0337	1.0586
$\alpha = 1$	1.0059	1.0115	1.0085	1.0252	1.0132	1.0097	1.0413	1.0571	1.0150	1.0321	1.0333	1.0547
Bork Møller												
$\alpha = 90$	1.0305	1.0404	1.0803	1.1058	1.0126	1.0193	1.0397	1.0695	1.0106	1.0280	1.0425	1.0872
$\alpha = 95$	1.0307	1.0460	1.0719	1.1096	1.0102	1.0245	1.0189	1.0512	1.0082	1.0332	1.0213	1.0629
$\alpha = 97$	1.0280	1.0463	1.0714	1.1036	1.0080	1.0259	1.0246	1.0543	1.0067	1.0362	1.0279	1.0668
$\alpha = 99$	1.0173	1.0333	1.0364	1.0747	1.0063	1.0203	1.0277	1.0800	1.0100	1.0375	1.0305	1.0935
$\alpha = 1$	1.0091	1.0004	0.9936	0.9806	1.0016	0.9874	0.9831	0.9634	1.0081	1.0137	0.9993	0.9979
SLF												
$\alpha = 90$	0.9901	0.9988	0.9727	0.9898	0.9831	0.9843	0.9704	0.9843	0.9703	0.9729	0.9843	1.0048
$\alpha = 95$	0.9911	1.0006	0.9960	1.0115	0.9836	0.9859	0.9706	0.9893	0.9713	0.9765	0.9593	0.9924
$\alpha = 97$	0.9951	1.0040	1.0173	1.0213	0.9863	0.9885	0.9910	0.9965	0.9750	0.9807	0.9905	1.0100
$\alpha = 99$	1.0099	1.0151	1.0361	1.0412	1.0027	0.9978	1.0424	1.0440	0.9940	0.9902	1.0551	1.0724
$\alpha = 1$	1.0092	1.0215	1.0066	1.0147	1.0024	1.0074	1.0242	1.0414	0.9923	0.9837	1.0186	1.0259
h=4	$MSFE_{DMA}$	$MAFE_{DMA}$	$MSFE_{DMS}$	$MAFE_{DMS}$	$MSFE_{DMA}$	$MAFE_{DMA}$	$MSFE_{DMS}$	$MAFE_{DMS}$	$MSFE_{DMA}$	$MAFE_{DMA}$	$MSFE_{DMS}$	$MAFE_{DMS}$
DSF_{PPH}												
$\alpha = 90$	1.0088	1.0083	1.0335	1.0463	1.0116	0.9997	1.0003	1.0000	1.0292	1.0293	1.0274	1.0598
$\alpha = 95$	1.0106	1.0149	1.0285	1.0458	1.0100	1.0039	0.9992	1.0060	1.0281	1.0288	1.0103	1.0446
$\alpha = 97$	1.0096	1.0159	1.0193	1.0441	1.0071	1.0052	0.9997	0.9994	1.0257	1.0281	1.0105	1.0390
$\alpha = 99$	1.0003	1.0007	0.9981	1.0006	1.0005	1.0012	1.0007	1.0016	1.0155	1.0222	1.0061	1.0313
$\alpha = 1$	1.0014	0.9670	1.0151	0.9500	1.0076	0.9731	1.0022	0.9388	1.0010	1.0002	1.0001	1.0001
Bork Møller												
$\alpha = 90$	1.1847	1.3834	1.2355	1.4892	1.2555	1.4755	1.3202	1.6065	1.1830	1.2849	1.2460	1.4245
$\alpha = 95$	1.2080	1.4439	1.2583	1.5287	1.2845	1.5466	1.3417	1.6346	1.2019	1.3247	1.2776	1.5010
$\alpha = 97$	1.2162	1.4535	1.2561	1.5278	1.2906	1.5579	1.3412	1.6331	1.2102	1.3519	1.2696	1.4719
$\alpha = 99$	1.2236	1.4188	1.2839	1.4754	1.2892	1.5158	1.3528	1.5649	1.2114	1.3754	1.2818	1.4882
$\alpha = 1$	1.2055	1.3132	1.2265	1.2998	1.2625	1.3939	1.3034	1.4268	1.1981	1.3255	1.2174	1.3579
SLF												
$\alpha = 90$	0.9565	0.9055	0.9932	0.9431	0.9438	0.8642	0.9473	0.8695	1.0203	1.0101	1.0307	1.0586
$\alpha = 95$	0.9651	0.9157	0.9944	0.9399	0.9486*	0.8697*	0.9627	0.8687	1.0142	1.0124	1.0611	1.0878
$\alpha = 97$	0.9710	0.9164	1.0156	0.9636	0.9527*	0.8686*	0.9895	0.9027	1.0147	1.0115	1.0430	1.0756
$\alpha = 99$	0.9551*	0.8782*	0.9480*	0.8527*	0.9474*	0.8475*	0.9299*	0.8234*	1.0178	1.0087	0.9936	1.0045
$\alpha = 1$	0.9500*	0.8531*	0.9652	0.8402	0.9380	0.8306	0.9355	0.8020	1.0251	1.0150	1.0765	1.1320

Note: This Table reports the MSFE and MAFE ratios of different forgetting schemes relative to the lowest MSFE of a model with a fixed forgetting scheme. *, ** and *** indicate significant Diebold-Mariano test results at 10%, 5% and 1% nominal levels, respectively. Bork Møller forgetting method uses five quantiles. The SLF tuning parameters are $\mu = 0.99$ and $G = \{1.0e\text{-}8\}\mathbf{I}$. The fixed forgetting scheme in the benchmark model uses $\lambda = 1$ for $h = 1$ and $\lambda = 0.99$ for $h = 4$.

Table 4. Out of sample MSFE and MAFE ratios of different forgetting schemes relative to the lowest MSFE fixed forgetting scheme for S&P 500

Working Papers

2020

1|20 On-site inspecting zombie lending
Diana Bonfim | Geraldo Cerqueiro | Hans Degryse | Steven Ongena

2|20 Labor earnings dynamics in a developing economy with a large informal sector
Diego B. P. Gomes | Felipe S. Iachan | Cezar Santos

3|20 Endogenous growth and monetary policy: how do interest-rate feedback rules shape nominal and real transitional dynamics?
Pedro Mazeda Gil | Gustavo Iglésias

4|20 Types of International Traders and the Network of Capital Participations
João Amador | Sónia Cabral | Birgitte Ringstad

5|20 Forecasting tourism with targeted predictors in a data-rich environment
Nuno Lourenço | Carlos Melo Gouveia | António Rua

6|20 The expected time to cross a threshold and its determinants: A simple and flexible framework
Gabriel Zsurkis | João Nicolau | Paulo M. M. Rodrigues

7|20 A non-hierarchical dynamic factor model for three-way data
Francisco Dias | Maximiano Pinheiro | António Rua

8|20 Measuring wage inequality under right censoring
João Nicolau | Pedro Raposo | Paulo M. M. Rodrigues

9|20 Intergenerational wealth inequality: the role of demographics
António Antunes | Valerio Ercolani

10|20 Banks' complexity and risk: agency problems and diversification benefits
Diana Bonfim | Sónia Felix

11|20 The importance of deposit insurance credibility
Diana Bonfim | João A. C. Santos

12|20 Dream jobs
Giordano Mion | Luca David Opromolla | Gianmarco I.P. Ottaviano

13|20 The DEI: tracking economic activity daily during the lockdown
Nuno Lourenço | António Rua

14|20 An economic model of the Covid-19 pandemic with young and old agents: Behavior, testing and policies
Luiz Brotherhood | Philipp Kircher | Cezar Santos | Michèle Tertilt

15|20 Slums and Pandemics
Luiz Brotherhood | Tiago Cavalcanti | Daniel Da Mata | Cezar Santos

16|20 Assessing the Scoreboard of the EU Macroeconomic Imbalances Procedure: (Machine) Learning from Decisions
Tiago Alves | João Amador | Francisco Gonçalves

17|20 Climate Change Mitigation Policies: Aggregate and Distributional Effects
Tiago Cavalcanti | Zeina Hasna | Cezar Santos

18|20 Heterogeneous response of consumers to income shocks throughout a financial assistance program
Nuno Alves | Fátima Cardoso | Manuel Coutinho Pereira

19|20 To change or not to change: the impact of the law on mortgage origination
Ana Isabel Sá

2021

1|21 Optimal Social Insurance: Insights from a Continuous-Time Stochastic Setup
João Amador | Pedro G. Rodrigues

2|21 Multivariate Fractional Integration Tests allowing for Conditional Heteroskedasticity withan Application to Return Volatility and Trading
Marina Balboa | Paulo M. M. Rodrigues Antonio Rubia | A. M. Robert Taylor

3|21 The Role of Macroprudential Policy in Times of Trouble
Jagjit S. Chadha | Germana Corrado | Luisa Corrado | Ivan De Lorenzo Buratta

4|21 Extensions to IVX Methodsnof Inference for Return Predictability
Matei Demetrescu | Iliyan Georgiev | Paulo M. M. Rodrigues | A.M. Robert Taylor

5|21 Spectral decomposition of the information about latent variables in dynamic macroeconomic models
Nikolay Iskrev

6|21 Institutional Arrangements and Inflation Bias: A Dynamic Heterogeneous Panel Approach
Vasco Gabriel | Ioannis Lazopoulos | Diana Lima

7|21 Assessment of the effectiveness of the macroprudential measures implemented in the context of the Covid-19 pandemic
Lucas Avezum | Vítor Oliveiral | Diogo Serra

8|21 Risk shocks, due loans, and policy options: When less is more!
Paulo Júlio | José R. Maria | Sílvia Santos

9|21 Sovereign-Bank Diabolic Loop: The Government Procurement Channel!
Diana Bonfim | Miguel A. Ferreira | Francisco Queiró | Sujiao Zhao

10|21 Assessing the effectiveness of the Portuguese borrower-based measure in the Covid-19 context
Katja Neugebauer | Vítor Oliveira | Ângelo Ramos

11|21 Scrapping, Renewable Technology Adoption, and Growth
Bernardino Adão | Borghan Narajabad | Ted Temzelides

12|21 The Persistence of Wages
Anabela Carneiro | Pedro Portugal | Pedro Raposo | Paulo M.M. Rodrigues

13|21 Serial Entrepreneurs, the Macroeconomy and top income inequality
Sónia Félix | Sudipto Karmakar | Petr Sedláček

14|21 COVID-19, Lockdowns and International Trade: Evidence from Firm-Level Data
João Amador | Carlos Melo Gouveia | Ana Catarina Pimenta

15|21 The sensitivity of SME's investment and employment to the cost of debt financing
Diana Bonfim | Cláudia Custódio | Clara Raposo

16|21 The impact of a macroprudential borrower based measure on households' leverage and housing choices
Daniel Abreu | Sónia Félix | Vítor Oliveira | Fátima Silva

17|21 Permanent and temporary monetary policy shocks and the dynamics of exchange rates
Alexandre Carvalho | João Valle e Azevedo | Pedro Pires Ribeiro

18|21 On the Cleansing Effect of Recessions and Government Policy: Evidence from Covid-19
Nicholas Kozeniauskas | Pedro Moreira | Cezar Santos

19|21 Trade, Misallocation, and Capital Market Integration
Laszlo Tetenyi

20|21 Not All Shocks Are Created Equal: Assessing Heterogeneity in the Bank Lending Channel
Laura Blattner | Luísa Farinha | Gil Nogueira

21|21 Coworker Networks and the Labor Market Outcomes of Displaced Workers: Evidence from Portugal
Jose Garcia-Louzao | Marta Silva

22|21 Markups and Financial Shocks
Philipp Meinen | Ana Cristina Soares

2022

1|22 Business cycle clocks: Time to get circular
Nuno Lourenço | António Rua

2|22 The Augmented Bank Balance-Sheet Channel of Monetary Policy
Christian Bittner | Diana Bonfim | Florian Heider | Farzad Saidi | Glenn Schepens | Carla Soares

3|22 Optimal cooperative taxation in the global economy
V. V. Chari | Juan Pablo Nicolini | Pedro Teles

4|22 How Bad Can Financial Crises Be? A GDP Tail Risk Assessment for Portugal
Ivan De Lorenzo Buratta | Marina Feliciano | Duarte Maia

5|22 Comparing estimated structural models of different complexities: What do we learn?
Paulo Júlio | José R. Maria

6|22 Survival of the fittest: Tourism Exposure and Firm Survival
Filipe B. Caires | Hugo Reis | Paulo M. M. Rodrigues

7|22 Mind the Build-up: Quantifying Tail Risks for Credit Growth in Portugal
Ivan de Lorenzo Buratta | Marina Feliciano | Duarte Maia

8|22 Forgetting Approaches to Improve Forecasting
Robert Hill | Paulo M. M. Rodrigues