Adaptive state space models with applications to the business cycle and financial stress[*]

Davide Delle Monache[†]   Ivan Petrella[‡]   Fabrizio Venditti[§]

Banca d'Italia   Warwick Business School   Banca d'Italia

& CEPR

November 1, 2016

Abstract

In this paper we develop a new framework for the analysis of state space models with time-varying parameters. We let the driver of the time variation be the score of the predictive likelihood and derive a new filter that allows us to estimate simultaneously the state vector and the time-varying parameters. In this setup the model remains Gaussian, the likelihood function can be evaluated using the Kalman filter and the model parameters can be estimated via maximum likelihood, without requiring the use of computationally intensive methods. Using a Monte Carlo exercise we show that the proposed method works well for a number of different data generating processes. We also present two empirical applications. In the former we improve the measurement of GDP growth by combining alternative noisy measures, in the latter we construct an index of financial stress and evaluate its usefulness in nowcasting GDP growth in real time. Given that a variety of time series models have a state space representation, the proposed methodology is of wide interest in econometrics and applied macroeconomics.

JEL codes: C22, C32, C51, C53, E31.

Keywords: State space models, time-varying parameters, score-driven models, business cycle, financial stress.

1 Introduction

Following two decades of low macroeconomic volatility and stable correlations across macroeconomic time series, the Great Recession has brought the issue of structural breaks back in the spotlight, spurring a wealth of new research on modeling and forecasting economic time series in the presence of parameter instability. Recent work in the field builds on two separate research agendas. In the former parameter instability manifests itself through large, infrequent breaks. A second view is based on the idea that economic agents adapt to shocks through a slow adjustment of their behavior, which is reflected in a gradual change of the parameters of empirical models.[1] The flexibility of the econometric tools designed to capture these different forms of time variation does not come without costs as the nonlinearity introduced in the models often requires the use of computationally intensive methods and sophisticated filtering techniques. These complications can be particularly burdensome for unobserved component models where also latent states variables need to be inferred on the basis of the observed data.

Motivated by these challenges, in this paper we develop a framework for the estimation of state space models with time-varying parameters (TVP). The building block of the method consists in positing a law of motion for the parameters that is a (linear) function of the score of the predictive likelihood, in line with Creal at al. (2013) and Harvey (2013). In this setup, the model remains Gaussian and the likelihood is evaluated through the Kalman filter (KF). Intuitively, at each point in time, the score of the predictive likelihood determines both the size and the sign of the adjustment of the model parameters that is needed to attain a local maximum. Such mechanism, which forces changes in the parameters to adapt to the local likelihood, justifies the definition of *adaptive state space models* adopted in the title of the paper.

In a state space model with constant parameters the accrual of new information (i.e. a data release) generates a prediction error that, through the KF recursions, is the basis for updating and forecasting the unobserved states. In other words, conditional on the model parameters, new information allows us to refine our view on the most likely location of the state vector and to form the best possible guess of where the data will be in the future. The time variation in the parameters introduces an additional margin of adjustment to this process and the new information calls for a simultaneous update of both the parameters and the latent states. The main contribution of the paper consists in deriving the analytical expressions for a new set of recursions that, running in parallel with the KF, update at each point in time both the vector of TVP and the latent states. Therefore, within this framework, the likelihood of any Gaussian state space model with TVP is available in closed form and the model can be even estimated by maximum likelihood (ML).

After presenting the main results we discuss a number of extensions of the baseline framework: e.g. we show how to incorporate general restrictions on the model parameters, and we

---

[1]Unfortunately the question of which view of the world is more appropriate can not really be settled with an econometric test, as tests designed for detecting large breaks typically perform poorly when the data generating process features slow, continuous changes in the model parameters, see Benati (2007).

illustrate how the model has to be modified to deal with data at mixed frequencies and with missing observations. We also work through two simple analytical examples to show how the proposed approach works in practice.

We assess the small sample properties of the new method in a Monte Carlo exercise. We find that it does a very good job at tracking a wide range of sources of time variation in the parameters. Furthermore, when the data are generated by a model with constant parameters our adaptive filter collapses the parameters to a constant, i.e. the method does not generate spurious time variation in the parameters when this feature is not supported by the data.

Given that a wide range of time series models have a state space representation, the method that we propose is of interest for a broad spectrum of applications in econometrics and applied macroeconomics. We present two empirical exercises that illustrate its usefulness. In the first application we focus on business cycle measurement. In particular, we take as a starting point the model developed by Aruoba et al. (2016) to estimate GDP growth on the basis of underlying (noisy) measures and extend it to account for time variation in the model parameters. We show that most of the parameters in the model of Aruoba et al. (2016) are subject to frequent breaks, which are well captured by our score driven method. Our approach can then retrieve some well known stylized facts of the U.S. business cycle that are overlooked in a constant parameter framework, namely a permanent slowdown in the long-run growth of the US economy, as well as the decline of GDP growth volatility during the Great Moderation and the subsequent leap due to the Great Recession. In a second application we derive from a panel of business cycle and financial variables an indicator of financial stress for the euro area. Our modeling approach allows us to address three related challenges in the construction of measures of financial stress: model instability, irregularities in the pattern of most financial data and the frequency mismatch between financial and real data. We assess the performance of our model in a nowcasting exercise in which financial variables are used to complement standard business cycle indicators in monitoring GDP developments in real time. The model detects significant shifts in the parameters and delivers GDP forecasts that are more accurate than those obtained on the basis of simple univariate benchmarks. Moreover, we find that financial stress indicators provide some additional (with respect to standard business cycle indicators) real time predictive content for the 2008 recession and that the presence of financial data improves the accuracy of density forecasts.

The rest of the paper is structured as follows. Section 2 constitutes the theoretical body of the paper, where we present the main results and discuss how to modify the estimation algorithm to incorporate parameter restrictions and to deal with missing and mixed frequency data. In Section 3 we describe the Monte Carlo exercise. In Section 4 we present the empirical applications and Section 5 concludes the paper.

2 Adaptive state space models

Let us assume that a given time series model has the following state space representation:

$$
\begin{aligned}
y_t &= Z_t \alpha_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{N}(0, H_t), \\
\alpha_{t+1} &= T_t \alpha_t + \eta_t, & \eta_t &\sim \mathcal{N}(0, Q_t), \quad t = 1, ..., n.
\end{aligned}
\tag{1}
$$

y_t is the $N \times 1$ vector of observed variables, ε_t is the $N \times 1$ vector of measurement errors, α_t is the $m \times 1$ vector of state variables and η_t is the corresponding $m \times 1$ vector of disturbances. The two disturbances are assumed to be Gaussian distributed and uncorrelated for all time period, that is $E(\varepsilon_t \eta'_s) = 0$ for $\forall t, s$.[2] The initial value of the state vector is also assumed to be Gaussian $\alpha_1 \sim \mathcal{N}(a_1, P_1)$ and uncorrelated $\forall t$ with ε_t and η_t.

Following Harvey (1989, sec. 3.1), the *system matrices* Z_t, H_t, T_t and Q_t are assumed to be non-stochastic and as a result the system is *linear*. Conditional on the information set $Y_{t-1} = \{y_{t-1}, ..., y_1\}$ and on the vector of parameters θ, the observations and the state vector are Gaussian; i.e. $y_t | Y_{t-1}; \theta \sim \mathcal{N}(Z_t a_t, F_t)$ and $\alpha_t | Y_{t-1}; \theta \sim \mathcal{N}(a_t, P_t)$. It follows that the log-likelihood function at time t is:

$$
\ell_t = \log p(y_t | Y_{t-1}, \theta) \propto -\frac{1}{2} \left(\log |F_t| + v'_t F_t^{-1} v_t \right).
\tag{2}
$$

The prediction error v_t, its covariance matrix F_t, the state vector conditional mean a_t, and its mean square error (MSE) matrix P_t, are estimated optimally[3] by means of the KF:

$$
\begin{aligned}
v_t &= y_t - Z_t a_t, & L_t &= T_t - K_t Z_t, \\
F_t &= Z_t P_t Z'_t + H_t, & a_{t+1} &= T_t a_t + K_t v_t, \\
K_t &= T_t P_t Z'_t F_t^{-1}, & P_{t+1} &= T_t P_t L'_t + Q_t, \quad t = 1, ..., n.
\end{aligned}
\tag{3}
$$

We have that $a_t = E(\alpha_t | Y_{t-1}, \theta)$ is the so-called *predictive* filter and $P_t = E[(a_t - \alpha_t)(a_t - \alpha_t)']$ is the associated MSE. The *real-time* filter is equal to $a_{t|t} = E(\alpha_t | Y_t, \theta) = a_t + P_t Z'_t F_t^{-1} v_t$, and its MSE is $P_{t|t} = E[(a_{t|t} - \alpha_t)(a_{t|t} - \alpha_t)'] = P_t - P_t Z'_t F_t^{-1} Z_t P_t$. It is worth stressing that in this linear model the MSEs are independent from the observations, thus they are also the *unconditional* covariance matrices associated with the conditional mean estimators; see Harvey (1989, sec. 3.2.3).

If the system matrices depend on past observations the model is still conditionally Gaussian since, given Y_{t-1}, the system matrices can be regarded as fixed. In this case they are usually refer to be *predetermined*; see Harvey (1989, sec. 3.7.1). In this framework a_t is the mean of the conditional distribution of α_t and it will no longer be a linear function of the observations. Thus, when a_t is viewed as an estimator of α_t, P_t represents its *conditional* MSE in the sense that it will depend on the particular realization of the observations in the sample. The attractive feature of this setup is that the expression of likelihood function can be still obtained as in (2),

---

[2]This assumption can be relaxed at the cost of a slight complication in some of the filtering formulae (see Remark 2).

[3]It produces the minimum mean square linear estimator (MMSLE) of the state vector.

and the KF is still the MMSLE of the state vector.

A conditionally Gaussian setup of this sort is used in several papers. However, no unified framework to analyze such models is available in the literature. As a result, for each type of time variation considered in these papers, an ad hoc strategy is devised. Harvey et al. (1992) provide approximate filtering and *quasi* ML estimation for unobserved components models with ARCH disturbances. Specifically, time-varying volatilities are driven by the squared disturbances estimated within the system as additional state variables.[4] In the same spirit, Koopman et al. (2010) modify the dynamic Nelson–Siegel model to include time-varying volatility, which is modeled through a GARCH dynamics. Their model also features time-varying loadings, implying a non-linearity that is handled by using the extended KF. Eickmeier et al. (2015), on the other hand, resort to a two-step strategy to estimate the time-varying loading in a factor model. They work under the assumption that principal components are consistent estimators for the factors even if the loadings vary mildly over time.[5] By treating the estimated factors as observable, the relation between the variables and the factors is given by a set of univariate regression models with time-varying coefficients, which evolve as independent random walks. The model can therefore be analyzed equation-wise by state space methods, i.e. estimating the hyperparameters by ML, and applying the KF to back out the TVP paths.[6] Koop and Korobilis (2014) use yet another strategy to estimate a factor model with changing loadings and volatilities. Their estimation procedure iterates until convergence over two steps: (i) given the TVP of the model, the KF is used to estimate the state vector and (ii) given the state vector, the *forgetting factor algorithm* updates the time-varying loadings and volatilities.[7]

Different problems arise when the elements of the system matrices Z_t, H_t, T_t and Q_t are driven by additional stochastic processes. Under this scenario the KF loses its optimality, the likelihood function is not available in closed form and Bayesian simulation techniques are needed; see Durbin and Koopman (2012). Several papers in the literature indeed use this approach. Stock and Watson (2007), for instance study the evolution of the relative importance of permanent and transitory components of US inflation through a trend-noise decomposition with stochastic volatilities. Del Negro and Otrok (2008) analyze the evolution of international business cycles using a dynamic factor model with time-varying loadings and stochastic volatility. Bianchi et al. (2009) introduce parameters' variation in the latent factors of the Nelson Siegel yield curve model. Marcellino et al. (2016) study the nowcasting properties of a small scale dynamic factor model in which volatility changes over time. The common theme in these papers is the use of computationally intensive Monte Carlo Markov Chain methods.

In this paper we contribute to the literature by proposing a new method for analyzing state

---

[4]Since the past values of the disturbances are not directly observable the model is not conditional Gaussian. Nevertheless the model is treated as if it was conditionally Gaussian and the KF is considered to be *quasi-optimal.*

[5]The assumption that the state vector can be consistently estimated in a first step can be valid for factor models but it does not generalize to other state space models.

[6]Mikkelsen et al. (2015) derive the asymptotic theory for this two-step estimator.

[7]The use of the forgetting factor algorithm was proposed in Koop and Korobilis (2013) to deal with TVP in large VAR models. This algorithm is obtained by applying the KF to a restricted state space model in which only two parameters regulate the time variation in the coefficients and volatilities.

space models with TVP. Our work falls in the first of the two lines of research discussed in the Introduction, in that we assume that the parameters evolve continuously and that are driven by past information. More in details, in line with Creal at al. (2013) and Harvey (2013), we assume that the matrices Z_t, H_t, T_t and Q_t are driven by the score of the predictive likelihood. Since the score depends on past information, the model is still conditional Gaussian, and consequently, we do not need to resort to simulation methods for estimation. We analytically derive a new auxiliary filter that runs in parallel to the KF and gives rise to an algorithm that enables the estimation of the parameters and of the state vector in a unified framework.

The key analytical challenge is represented by the joint updating, at each point in time, of both the system matrices and of the state vector. In the existing literature this challenge is typically solved by assuming that the unobserved state vector can be somehow estimated *out of the model*. This is done in papers that use classical methods (Eickmeier et al., 2015 and Koop and Korobilis, 2014) and implicitly also in studies that use Bayesian methods, as cycling through the Gibbs sampler implies conditioning on a given estimate of the whole state vector.

In contrast with existing approaches, we propose a new set of recursions that run in parallel with the KF in order to jointly update (at each point in time) the parameters and the state vector. Therefore, the KF still retains the usual optimality properties, the likelihood function can be evaluated in closed form, and it is even possible to estimate the model by ML. Our method is not tailored to a single application, but it is valid for all state space models. Since a variety of time series models can be cast in state space form, the proposed approach is of interest for a wide spectrum of empirical applications in econometrics and applied macroeconomics.[8]

2.1 Score-driven time-varying parameters

First, the TVP of model (1) are collected in the vector f_t. Second, as in Creal et al. (2013) and Harvey (2013), we posit the following law of motion for the TVP:

$$f_{t+1} = \omega + \Phi f_t + \Omega s_t, \quad s_t = \mathcal{S}_t \nabla_t, \quad t = 1, ..., n, \tag{4}$$

where

$$\nabla_t = \frac{\partial \ell_t}{\partial f_t}, \quad \mathcal{S}_t = \left[-E_t \left(\frac{\partial \ell_t^2}{\partial f_t \partial f_t'} \right) \right]^{-k}, \quad k = 0, 1/2, 1, \tag{5}$$

and ℓ_t is the conditional log-likelihood function of the model.

In the vector θ_f we collect all the elements of ω, Φ, and Ω, regulating the dynamics of the TVP. The system matrices in (1) may possibly contain some static parameters that are collected in the vector θ_m. Thus, the full vector of constant parameters is $\theta = (\theta_m', \theta_f')'$. At each point in time the system matrices depend upon f_t and θ_m, namely $Z_t = Z(f_t, \theta_m)$, $T_t = T(f_t, \theta_m)$, $H_t = H(f_t, \theta_m)$, and $Q_t = Q(f_t, \theta_m)$. In turn, the dynamics of f_t depends on Y_{t-1} and θ_f. Therefore, the model is observation-driven, i.e. for each point in time t, it is

---

[8]Creal et al. (2008), for instance, analyze parameters' variation for the local level model using the score-driven approach. Their filter is nested as a special case by our method, see sub-section 2.4.1.

entirely determined by $\{Y_{t-1}, \theta\}$, the KF retains its optimal properties, and its log-likelihood function is equal to (2).

The driving mechanism for the law of motion (4) is represented by the scaled-score of the conditional likelihood. Specifically, for $k = 0$, s_t equals the score ∇_t with zero mean and variance equal to the Fisher information matrix $\mathcal{I}_t$. Alternatively, for $k = 1$, the scaling matrix $\mathcal{S}_t$ is equal to the inverse of the information matrix $\mathcal{I}_t^{-1}$ and s_t will have variance equal to $\mathcal{I}_t^{-1}$. Finally, for $k = 1/2$, the variance of s_t is the identity matrix.[9]

Let us focus briefly on the updating rule (4), at each point in time the vector f_t is updated so as to maximize the local fit of the model. Specifically, the magnitude of the update depends on the slope and on the curvature of the likelihood function. It follows the score-driven updating rule can be rationalized as a stochastic analogue of the Gauss–Newton search direction for the TVP.[10]

Our theoretical contribution is organized in two Results: in the first one, we show analytically the formulae for the score and the scaling matrix; in the second, we present the filter to compute recursively such scaled-score. It is worth mentioning that the strategy adopted by Koop and Korobilis (2014) can be obtained as special case of the general algorithm presented below.[11]

Before going into the main Results let us introduce some notation (see e.g. Abadir and Magnus, 2005, ch. 11). Given a $N \times m$ matrix X, $vec(X)$ is the vector obtained by stacking the columns of X one underneath the other, while $vech(X)$ eliminates all supradiagonal elements of X from $vec(X)$. The $Nm \times Nm$ commutation matrix $C_{N,m}$ is such that $C_{N,m}vec(X) = vec(X')$. For $N = m$ the $m^2 \times m^2$ commutation matrix is denoted by C_m. Given a square matrix A, the symmetrizer matrix is $N_n = \frac{1}{2}(I_{n^2} + C_n)$ and $N_n vec(A) = vec\left[\frac{1}{2}(A + A')\right]$. For A symmetric $N_n vec(A) = vec(A)$. The identity matrix of order N is denoted by I_N, and '$\otimes$' is the Kronecker product.

Result 1 *Given the conditional Gaussian model (1)-(2), the score and the information matrix are:*

$$
\begin{aligned}
\nabla_t &= \frac{1}{2}\left[\dot{F}_t'(F_t^{-1} \otimes F_t^{-1})vec(v_t v_t' - F_t) - 2\dot{V}_t' F_t^{-1} v_t\right] \\
\mathcal{I}_t &= \frac{1}{2}\left[\dot{F}_t'(F_t^{-1} \otimes F_t^{-1})\dot{F}_t + 2\dot{V}_t' F_t^{-1}\dot{V}_t\right], \quad t = 1, ..., n,
\end{aligned}
\tag{6}
$$

where $\dot{V}_t = \frac{\partial v_t}{\partial f_t'}$ and $\dot{F}_t = \frac{\partial vec(F_t)}{\partial f_t'}$. Proofs in Appendix A.1.

Let us first have a closer look at the expression for the score ∇_t in (6). The score is a linear

---

function of two terms, namely $(v_t v_t' - F_t)$ and v_t. The first term is the difference between the current estimate of the second moment of y_t and the past estimate based on information at $t-1$. Loosely speaking, when F_t 'correctly' estimates the current value of the second moment (i.e. F_t coincides with $v_t v_t'$), this term is zero and it does not contribute to changes in f_t. Such term is weighed by two matrices: $(F_t^{-1} \otimes F_t^{-1})$, which is a scaling factor measuring the degree of uncertainty on the estimated second moment, and $\overset{\bullet}{F}_t$, which measures the sensitivity of the second moment with respect to f_t. The second term in the score is the prediction error, which makes changes in f_t proportional to the forecast error on y_t. This term is also pre-multiplied by two matrices: F_t^{-1}, which scales the prediction error by its variance, and $\overset{\bullet}{V}_t$, which accounts for the sensitivity of the prediction error with respect to f_t. In sum, the stronger the impact new information has on the estimated first and second moments of the observable y_t, the larger the revision of the parameters f_t. Furthermore, the score depends on how sensitive the estimated first and second moments of y_t are to changes in f_t.

The scaling matrix $\mathcal{I}_t$ in (6) is also composed of two terms, namely $\overset{\bullet}{F}_t'(F_t^{-1} \otimes F_t^{-1})\overset{\bullet}{F}_t$ and $\overset{\bullet}{V}_t' F_t^{-1} \overset{\bullet}{V}_t$. Those are the Hessian matrices with respect to f_t of the prediction error and of its variance, both scaled by their respective variances.

Notice that v_t and F_t are recursively computed by means of the KF (3), while their Jacobian counterparts, $\overset{\bullet}{V}_t$ and $\overset{\bullet}{F}_t$, are recursively computed through the new filter presented below.

Result 2 *For $t = 1, ..., n$, the Jacobian counterpart of the KF (3) leads to*

$$
\begin{aligned}
\overset{\bullet}{V}_t &= -[(a_t' \otimes I_N)\overset{\bullet}{Z}_t + Z_t \overset{\bullet}{A}_t], \\
\overset{\bullet}{F}_t &= 2N_N(Z_t P_t \otimes I_N)\overset{\bullet}{Z}_t + (Z_t \otimes Z_t)\overset{\bullet}{P}_t + \overset{\bullet}{H}_t, \\
\overset{\bullet}{K}_t &= (F_t^{-1} Z_t P_t \otimes I_m)\overset{\bullet}{T}_t + (F_t^{-1} Z_t \otimes T_t)\overset{\bullet}{P}_t + (F_t^{-1} \otimes T_t P_t)C_{Nm}\overset{\bullet}{Z}_t - (F_t^{-1} \otimes K_t)\overset{\bullet}{F}_t, \quad (7) \\
\overset{\bullet}{A}_{t+1} &= (a_t' \otimes I_m)\overset{\bullet}{T}_t + T_t \overset{\bullet}{A}_t + (v_t' \otimes I_m)\overset{\bullet}{K}_t + K_t \overset{\bullet}{V}_t \\
\overset{\bullet}{P}_{t+1} &= (T_t \otimes T_t)\overset{\bullet}{P}_t - (K_t \otimes K_t)\overset{\bullet}{F}_t + \overset{\bullet}{Q}_t + 2N_m[(T_t P_t \otimes I_m)\overset{\bullet}{T}_t - (K_t F_t \otimes I_m)\overset{\bullet}{K}_t],
\end{aligned}
$$

where $\overset{\bullet}{Z}_t = \frac{\partial vec(Z_t)}{\partial f_t'}$, $\overset{\bullet}{H}_t = \frac{\partial vec(H_t)}{\partial f_t'}$, $\overset{\bullet}{T}_t = \frac{\partial vec(T_t)}{\partial f_t'}$ and $\overset{\bullet}{Q}_t = \frac{\partial vec(Q_t)}{\partial f_t'}$. Proofs in Appendix A.

Results 1 and 2 generalize the results in Harvey (1989, sec. 3.4.6) to multivariate models and to models with TVP.

In order to understand the logic behind the filter presented in (7), let us recall how the recursions in the standard KF work. In the KF, given past estimates of the state vector moments (a_t and P_t) and current information y_t, the prediction error v_t and its variance F_t are obtained. Based on these, new estimates of the state vector moments (a_{t+1} and P_{t+1}) are computed. The link between past and current estimates is given by the Kalman gain K_t whose role is to weigh the current observation. The filter presented in Theorem 2 works much in the same way. Given past Jacobians of the state vector moments $\overset{\bullet}{A}_t$ and $\overset{\bullet}{P}_t$, current Jacobians of the prediction error ($\overset{\bullet}{V}_t$) and of the prediction error variance ($\overset{\bullet}{F}_t$) are computed. The Jacobians $\overset{\bullet}{A}_{t+1}$ and $\overset{\bullet}{P}_{t+1}$ can then be updated through the Jacobian of the Kalman gain $\overset{\bullet}{K}_t$. Notice that

also the Jacobians of the system matrices appear; i.e. $\overset{\bullet}{Z}_t$, $\overset{\bullet}{H}_t$, $\overset{\bullet}{T}_t$ and $\overset{\bullet}{Q}_t$. Those depend on the specific model under consideration; in simple models they result in selection matrices as will be clearer from the examples that we provide below.

Remark 1 *It is sometimes convenient to include mean adjustment terms in (1) resulting in*

$$
\begin{aligned}
y_t &= Z_t \alpha_t + d_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, H_t), \\
\alpha_{t+1} &= T_t \alpha_t + c_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, Q_t),
\end{aligned}
\tag{8}
$$

where d_t and c_t are known vectors possibly time-varying. The filtering (3) is amended in the following elements

$$
\begin{aligned}
v_t &= y_t - Z_t a_t - d_t, \\
a_{t+1} &= T_t a_t + c_t + K_t v_t,
\end{aligned}
\tag{9}
$$

and two expressions of (7) are modified as follows

$$
\begin{aligned}
\overset{\bullet}{V}_t &= -[(a_t' \otimes I_N)\overset{\bullet}{Z}_t + \overset{\bullet}{d}_t + Z_t \overset{\bullet}{A}_t], \\
\overset{\bullet}{A}_{t+1} &= (a_t' \otimes I_m)\overset{\bullet}{T}_t \overset{\bullet}{c}_t + T_t \overset{\bullet}{A}_t + (v_t' \otimes I_m)\overset{\bullet}{K}_t + K_t \overset{\bullet}{V}_t,
\end{aligned}
\tag{10}
$$

where $\overset{\bullet}{d}_t = \frac{\partial d_t}{\partial f_t'}$ and $\overset{\bullet}{c}_t = \frac{\partial c_t}{\partial f_t'}$ are model specific.

Remark 2 *In the case the two disturbances in (1) are correlated, that is $E(\varepsilon_t \eta_t') = G_t$, the Kalman gain in (3) is equal to $K_t = (T_t P_t Z_t' + G_t)F_t^{-1}$, and the third expression of (7) becomes*

$$
\begin{aligned}
\overset{\bullet}{K}_t &= (F_t^{-1} Z_t P_t \otimes I_m)\overset{\bullet}{T}_t + (F_t^{-1} Z_t \otimes T_t)\overset{\bullet}{P}_t + (F_t^{-1} \otimes T_t P_t)C_{Nm}\overset{\bullet}{Z}_t \\
&\quad + [F_t^{-1} \otimes (G_t F_t^{-1} - K_t)]\overset{\bullet}{F}_t + (F_t^{-1} \otimes I)\overset{\bullet}{G}_t,
\end{aligned}
\tag{11}
$$

where $\overset{\bullet}{G}_t = \frac{\partial vec(G_t)}{\partial f_t'}$ is model specific.

Putting together Results 1 and 2, we obtain a new filter that enables to compute the scaled score $s_t = \mathcal{S}_t \nabla_t$ and therefore to estimate the TVP vector using the the score-driven filter (4). Such *auxiliary filter* runs in parallel with the standard KF (3) and gives rise to the Algorithm described below.[12]

The vector of static parameters θ can be estimated by ML: $\widehat{\theta} = \arg\max \sum_{t=1}^n \ell_t(\theta)$. Given the above algorithm, the evaluation of the log-likelihood function is straightforward and the maximization can be obtained numerically. As in Creal et al. (2013, sec. 2.3), we can conjecture that the usual ML results hold and this is backed up by the Monte Carlo experiment in Section 3. Specifically, we have that $\sqrt{n}(\widehat{\theta} - \theta) \to \mathcal{N}(0, \Xi)$, where the asymptotic variance Ξ is evaluated by numerical derivative at the optimum.

---

[12] A simpler version of this algorithm is used by Delle Monache et al. (2016) to analyze inflation rates in the euro area within a dynamic factor model with TVP.

Algorithm

Initialize the following elements a_1, P_1, f_1, $\overset{\bullet}{A}_1$, $\overset{\bullet}{P}_1$.

For $t = 1, ..., n$:

    i. given f_t, evaluate Z_t, H_t, T_t, Q_t, $\overset{\bullet}{Z}_t$, $\overset{\bullet}{H}_t$, $\overset{\bullet}{T}_t$, $\overset{\bullet}{Q}_t$

    ii. compute v_t, F_t, K_t, and ℓ_t using (2)-(3)

    iii. compute $\overset{\bullet}{V}_t$, $\overset{\bullet}{F}_t$, $\overset{\bullet}{K}_t$ using (7)

    iv. compute ∇_t, $\mathcal{I}_t$, (and s_t) using (6)

    v. update a_{t+1}, P_{t+1} using (3)

    vi. update $\overset{\bullet}{A}_{t+1}$, $\overset{\bullet}{P}_{t+1}$ using (7)

    vii. update f_{t+1} using (4)

For $t = n$ we have all the elements to estimate the model (1).

2.2 Parameter restrictions

Applications of TVP models often require imposing restrictions on the parameters' space. For instance, stationarity restrictions or positivity of the volatilities may be desirable. When restrictions are implemented within our setup, the resulting model can still be estimated by ML without the need for demanding simulation methods. This requires the re-parameterization of the TVP vector:

$$\tilde{f}_t = \psi(f_t), \tag{12}$$

where f_t is the unrestricted vector of parameters, $\tilde{f}_t$ is the restricted vector of interest with respect to which the likelihood function (2) is expressed, and $\psi(\cdot)$ is assumed to be a time-invariant, continuous, invertible and twice differentiable function (also known as link function). In particular, we model $f_t = h(\tilde{f}_t)$, where $h(\cdot)$ is the inverse function of $\psi(\cdot)$. The vector f_t continues to follow the updating rule (4), and the algorithm easily takes into account such transformation via the Jacobian $\Psi_t = \frac{\partial \tilde{f}_t}{\partial f_t'}$ ensuring that the restrictions are satisfied at each point in time.

In our modeling strategy the transformation (12) will affect the Jacobian matrices $\overset{\bullet}{Z}_t$, $\overset{\bullet}{H}_t$, $\overset{\bullet}{T}_t$ and $\overset{\bullet}{Q}_t$ above defined. Generally, these matrices contain both constant and TVP, some of which are restricted, some are left unrestricted. To deal with the presence of restrictions, given a generic time-varying system matrix M_t, we propose the following decomposition:

$$vec\left(M_t\right) = c_M + S_{M1}\psi_M\left(S_{M2}f_t\right), \tag{13}$$

where M_t denotes a $n_r \times n_c$ matrix containing both constant and time-varying parameters, c_M is a $n_r n_c \times 1$ vector with constant elements, and S_{M1} and S_{M2} are selection matrices: S_{M1} selects the time-varying elements of M_t and it is obtained from $I_{n_r n_c}$ by retaining the columns associated with the time varying elements of $vec\left(M_t\right)$, while S_{M2} selects the sub-vector of f_t belonging to M_t. Finally, $\psi_M\left(\cdot\right)$ denotes the link function used to restrict the elements of M_t

and is the Jacobian of $\psi_M(\cdot)$ is denoted with Ψ_{Mt}. Given the representation (13), the Jacobian of the matrix is computed as

$$\overset{\bullet}{M}_t = S_{M1}\Psi_{Mt}S_{M2}.$$

Note that different sub-vectors of f_t may need different restrictions and thus different link functions may be used. For instance, in the empirical applications we will impose two main restrictions. First, we impose the restriction that the autoregressive coefficients have stable roots at each point in time; this is implemented by re-parameterising the model with respect to the partial autocorrelations as in Delle Monache and Petrella (2014). A second set of restrictions forces a generic covariance matrix Σ_t to be positive definite at each point in time; this is achieved by means of the log-Cholesky transformation. More in details, $\Sigma_t = J_t J_t'$, with J_t being a lower triangular matrix, and we model TVP vector $vech(\tilde{J}_t)$, where $\tilde{J}_{ij,t} = J_{ij,t}$ for $i \neq j$, and $\tilde{J}_{ii,t} = \log J_{ii,t}$. Thus, the Jacobian is computed as follows

$$\frac{\partial vec(\Sigma_t)}{\partial vech(\tilde{J}_t)'} = (I_{N^2} + C_N)(J_t \otimes I_N)S_J S_{\log t},$$

where C_N is the commutation matrix, S_J is a selection matrix[13] such that $S_J vech(J_t) = vec(J_t)$ and $S_{\log t} = diag[\exp(\tilde{J}_{ii,t})]$.

2.3 Missing observations and mixed frequencies

Assume to have a data set containing missing observations. The observed vector is represented by $W_t y_t$, where W_t is an $N_t \times N$ selection matrix with $1 \leq N_t \leq N$, meaning that at least one observation is available at time t. Note that W_t is obtained by eliminating the $i-th$ row from I_N when the $i-th$ variable is missing. In this setting, at each time t the likelihood ℓ_t is computed using N_t observations; i.e. $\ell_t = \log p(W_t y_t | Y_{t-1}, \theta)$, that is the *marginal* likelihood. In practice, the score of the marginal likelihood is computed and the updating of f_t is based on the available information. We have two polar cases. If there are no missing observations $W_t = I_N$ and the score is computed on the *joint* likelihood. If, on the other hand, no data is available at time t, we have that $N_t = 0$, the marginal density degenerates to a constant and the score is zero. In this case the future values of f_t are obtained by using (4) with $s_t = 0$. A formal discussion of dealing with missing values in score-driven TVP models can be found in Lucas et. al (2016).

Given this reparameterization for the model, the measurement equation in (1) is modified as follows:

$$W_t y_t = W_t Z_t \alpha_t + W_t \varepsilon_t, \quad W_t \varepsilon_t \sim \mathcal{N}(0, W_t H_t W_t'). \tag{14}$$

---

[13]S_J is obtained form $diag(vec(J_t))$ by dropping the columns containing only 0s and replacing the non-zeros elements with 1s.

The first three expressions of the KF (3) are modified as follows:

$$
\begin{aligned}
v_t &= W_t(y_t - Z_t a_t), \\
F_t &= W_t(Z_t P_t Z_t' + H_t)W_t', \\
K_t &= T_t P_t Z_t' W_t' F_t^{-1},
\end{aligned}
\tag{15}
$$

thus the first three formulae of the new filter (7) become

$$
\begin{aligned}
\overset{\bullet}{V}_t &= -[(a_t' \otimes W_t)\overset{\bullet}{Z}_t + W_t Z_t \overset{\bullet}{A}_t], \\
\overset{\bullet}{F}_t &= 2N_{N_t}(W_t Z_t P_t \otimes W_t)\overset{\bullet}{Z}_t + (W_t Z_t \otimes W_t Z_t)\overset{\bullet}{P}_t + (W_t \otimes W_t)\overset{\bullet}{H}_t, \\
\overset{\bullet}{K}_t &= (F_t^{-1} W_t Z_t P_t \otimes I_m)\overset{\bullet}{T}_t + (F_t^{-1} W_t Z_t \otimes T_t)\overset{\bullet}{P}_t + (F_t^{-1} W_t \otimes T_t P_t)C_{Nm}\overset{\bullet}{Z}_t - (F_t^{-1} \otimes K_t)\overset{\bullet}{F}_t,
\end{aligned}
\tag{16}
$$

The derivation of (15) can be found in Durbin and Koopman (2012), while the formulae (16) can be derived following the steps in Appendix A.

The case of mixed frequencies is of particular interest for a number of applications, like for instance forecasting low frequency variables using higher frequency predictors (nowcasting). Mixed frequencies typically involve missing observations and temporal aggregation. Indeed low frequency indicators can be modeled as a latent process that is observed at regular low frequency intervals and missing at higher frequency dates. The relation between the observed low frequency variable and the corresponding (latent) higher frequency indicator depends on whether the variable is a flow or a stock and on how the variable is transformed before entering the model. In all cases, the variable can be rewritten as a weighted average of the unobserved high frequency indicator. Therefore, temporal aggregation only requires a modification of the state space representation leaving the filtering algorithms unchanged, for details see Appendix B.2.

2.4 Examples

In this section we provide some more intuition on how the proposed algorithm works by looking at two specific examples: the unobserved components local level model and the reduced form autoregressive model of order one.

2.4.1 Local level model

Let us consider a simple local level model with time-varying volatilities:

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{N}(0, \sigma_{\varepsilon,t}^2), \\
\mu_{t+1} &= \mu_t + \eta_t, & \eta_t &\sim \mathcal{N}(0, \sigma_{\eta,t}^2).
\end{aligned}
\tag{17}
$$

This model has been proposed by Stock and Watson (2007) to model US inflation. The estimation of (17) using the score-driven approach was initially proposed by Creal et al. (2008, sec. 4.4). Here we show how our general algorithm collapses to the one therein proposed.

Model (17) can be easily cast in the state space form (1) with

$$\alpha_t = \mu_t, \quad Z_t = T_t = 1, \quad H_t = \sigma^2_{\varepsilon,t}, \quad Q_t = \sigma^2_{\eta,t}.$$

The vector of TVP and the corresponding Jacobian matrices are

$$f_t = (\log \sigma_{\varepsilon,t}, \log \sigma_{\eta,t})', \quad \dot{H}_t = \left(2\sigma^2_{\varepsilon,t}, 0\right), \quad \dot{Q}_t = \left(0, 2\sigma^2_{\eta,t}\right), \quad \dot{Z}_t = \dot{T}_t = 0.$$

In order to ensure positive volatility we use the exponential link function, this implies that f_t contains the log standard deviations, $\sigma^2_{\varepsilon,t} = \exp(2f_{1,t})$, and $\sigma^2_{\eta,t} = \exp(2f_{2,t})$. As previously anticipated, the Jacobian of the system matrices incorporate the parameter restrictions used for the specific model. Application of the KF (3) leads to the following recursions:

$$
\begin{aligned}
v_t &= y_t - a_t, & a_{t+1} &= a_t + k_t v_t, \\
d_t &= p_t + \sigma^2_{\varepsilon,t}, & p_{t+1} &= (1 - k_t)p_t + \sigma^2_{\eta,t}, \\
k_t &= p_t/d_t, & t &= 1, ..., n.
\end{aligned}
\tag{18}
$$

The estimator of the state vector conditional mean and variance (a_{t+1} and p_{t+1}) depend on f_t, which is recursively estimated based on the score-driven filter (4). First, the log-likelihood of model (17) is equal to $\ell_t \propto -\frac{1}{2}(\log d_t + v_t^2/d_t)$. Secondly, the elements of the gradient and information matrix in (6) are:

$$\nabla_{i,t} = \frac{1}{2}\left[\frac{\dot{d}_{i,t}}{d_t^2}(v_t^2 - d_t) - 2\dot{v}_{i,t}\frac{v_t}{d_t}\right], \quad i = 1, 2, \tag{19}$$

$$\mathcal{I}_{ij,t} = \frac{1}{2}\left[\frac{\dot{d}_{i,t}\dot{d}_{j,t}}{d_t^2} + \frac{2\dot{v}_{i,t}\dot{v}_{j,t}}{d_t}\right], \quad i, j = 1, 2, \tag{20}$$

where the indexing $i = 1, 2$ refers to the two elements of f_t.

We are now ready to apply the recursions of the new filter (7):

$$
\begin{aligned}
\dot{v}_{i,t} &= -\dot{a}_{i,t}, & \dot{a}_{i,t+1} &= (1 - k_t)\dot{a}_{i,t} + \dot{k}_{i,t}v_t, \\
\dot{d}_{i,t} &= \dot{p}_{i,t} + 2\sigma^2_{\varepsilon,t}1(i = 1), & \dot{p}_{i,t+1} &= (1 - k_t)\dot{p}_{i,t} - \dot{k}_{i,t}p_t + 2\sigma^2_{\eta,t}1(i = 2), \\
\dot{k}_{i,t} &= (\dot{p}_{i,t} - k_t\dot{d}_{i,t})/d_t, & t &= 1, ..., n,
\end{aligned}
\tag{21}
$$

where $1(x)$ is an indicator function, taking value one when the condition x is verified and zero otherwise.[14]

As explained in the general case, the score vector (19) is composed of two terms. The first one $(v_t^2 - d_t)$ is the difference between a current estimate of the prediction error variance and its past estimated value. Such term is scaled by its variance d_t^2 and weighed by $\dot{d}_{i,t}$, which is the sensitivity of the prediction error variance to changes in $f_{i,t}$. The second term is the

---

[14]The recursions in 21 are the same as the ones derived in Creal et al. (2008, sec. 4.4).

scaled prediction error, v_t/d_t, weighed by its sensitivity to changes in $f_{i,t}$. Looking at (21), we have that $\overset{\bullet}{d}_{i,t}$ is a linear function of $\overset{\bullet}{p}_{i,t}$, while $\overset{\bullet}{v}_{i,t}$ is a linear function of $\overset{\bullet}{a}_{i,t}$. Thus, for a given prediction error v_t the model updates the two time-varying variances according to the current signal to noise ratio. Specifically, each element of the score is proportional to the past estimated values of the corresponding variance. Therefore, in periods of high signal to noise ratio, for a given prediction error the model is more likely to update more the trend than the noise variance. Conversely, in periods of low signal to noise ratio the algorithm revises relatively more the noise variance.

Putting together the recursions (18)-(21) and after some algebra, we obtain the algorithm shown in the table below.

Algorithm for the local level model

Given the vector of static parameters θ and the initial values a_1, p_1, f_1, $\overset{\bullet}{a}_1$, $\overset{\bullet}{p}_1$.

For $t = 1, ..., n$:

    i. compute the parameters:
$$\sigma_{\varepsilon,t}^2 = \exp(2f_{1,t}),\ \sigma_{\eta,t}^2 = \exp(2f_{2,t});$$

    i. compute the pred. error, the pred. error var., the Kalman gain, and the log-lik:
$$v_t = y_t - a_t, \quad d_t = p_t + \sigma_{\varepsilon,t}^2, \quad k_t = p_t/d_t, \quad \ell_t \propto -\tfrac{1}{2}(\log d_t + v_t^2/d_t);$$

    iii. compute the gradients of d_t and k_t (i.e. their sensitivity to changes in f_t):
$$\overset{\bullet}{d}_{i,t} = \overset{\bullet}{p}_{i,t} + 2\sigma_{\varepsilon,t}^2 1(i = 1), \quad \overset{\bullet}{k}_{i,t} = (\overset{\bullet}{p}_{i,t}d_t - p_t\overset{\bullet}{d}_{i,t})/d_t^2;$$

    iv. compute the gradient and the Hessian of ℓ_t, and the score $s_t = \mathcal{I}_t^{-1}\nabla_t$:
$$\nabla_{i,t} = \tfrac{1}{2d_t}\left[\overset{\bullet}{d}_{i,t}\left(\tfrac{v_t^2}{d_t} - 1\right) + 2\overset{\bullet}{a}_{i,t}v_t\right], \quad \mathcal{I}_{ij,t} = \tfrac{1}{2d_t}\left[\tfrac{\overset{\bullet}{d}_{i,t}\overset{\bullet}{d}_{j,t}}{d_t} + 2\overset{\bullet}{a}_{i,t}\overset{\bullet}{a}_{j,t}\right];$$

    v. compute the next period estimate of the state vector and its conditional variance:
$$a_{t+1} = a_t + k_t v_t, \quad p_{t+1} = (1 - k_t)p_t + \sigma_{\eta,t}^2;$$

    vi. compute the gradients of a_{t+1} and p_{t+1} (i.e. their sensitivity to changes in f_t):
$$\overset{\bullet}{a}_{i,t+1} = (1 - k_t)\overset{\bullet}{a}_{i,t} + \overset{\bullet}{k}_{i,t}v_t, \quad \overset{\bullet}{p}_{i,t+1} = (1 - k_t)\overset{\bullet}{p}_{i,t} - \overset{\bullet}{k}_{i,t}p_t + 2\sigma_{\eta,t}^2 1(i = 2);$$

    vii. compute the next period estimate of the TVP vector:
$$f_{t+1} = \omega + \Phi f_t + \Omega s_t.$$

2.4.2 Autoregressive model

Let us consider the following autoregressive model of order one:

$$y_{t+1} = \phi_t y_t + \xi_t, \quad \xi_t \sim \mathcal{N}(0, \sigma_t^2). \tag{22}$$

Despite its simplicity, the above model (generalized to p lags) has been successfully used by Delle Monache and Petrella (2014) for modeling US inflation. Blasques et al. (2014) consider the AR(1) model with time-varying coefficient and study the stochastic properties of the implied

non-linear model. The SSF representation (1) of model (22) can be easily obtained setting

$$\alpha_t = y_t, \quad Z_t = 1, \quad \varepsilon_t = H_t = 0, \quad T_t = \phi_t, \quad Q_t = \sigma_t^2.$$

For simplicity here we do not impose parameter restrictions, thus the vector of TVP vector and the corresponding Jacobian matrices are: [15]

$$f_t = (\phi_t, \sigma_t^2)', \quad \overset{\bullet}{T}_t = (1, 0), \quad \overset{\bullet}{Q}_t = (0, 1), \quad \overset{\bullet}{Z}_t = \overset{\bullet}{H}_t = 0.$$

The KF (3) results in the following formulation:

$$
\begin{aligned}
v_t &= \xi_t, & a_{t+1} &= \phi_t y_t, \\
F_t &= P_t, & P_{t+1} &= Q_t = \sigma_t^2, \\
K_t &= \phi_t, & t &= 2, ..., n.
\end{aligned}
\tag{23}
$$

Given the Jacobian matrices, the new filter (7) collapses to:

$$
\begin{aligned}
\overset{\bullet}{V}_t &= -\overset{\bullet}{A}_t = -(y_{t-1}, 0), & \overset{\bullet}{P}_{t+1} &= \overset{\bullet}{Q}_t = (0, 1), \\
\overset{\bullet}{F}_t &= \overset{\bullet}{P}_t = (0, 1), & \overset{\bullet}{A}_{t+1} &= y_t \overset{\bullet}{T}_t = (y_t, 0), \\
\overset{\bullet}{K}_t &= \overset{\bullet}{T}_t = (1, 0), & t &= 2, ..., n.
\end{aligned}
\tag{24}
$$

Given (23) and (24), the gradient and scaling matrix in (6) are equal to:

$$
\begin{aligned}
\nabla_t &= \frac{1}{2}\left[\frac{1}{\sigma_t^4}\begin{pmatrix} 0 \\ \xi_t^2 - \sigma_t^2 \end{pmatrix} + \frac{2}{\sigma_t^2}\begin{pmatrix} y_{t-1}\xi_t \\ 0 \end{pmatrix}\right], \\
\mathcal{I}_t &= \frac{1}{2}\left[\frac{1}{\sigma_t^4}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \frac{2}{\sigma_t^2}\begin{pmatrix} y_{t-1}^2 & 0 \\ 0 & 0 \end{pmatrix}\right].
\end{aligned}
\tag{25}
$$

Thus the scaled-score vector has a very simple formulation:

$$
s_t = \begin{bmatrix} \frac{1}{y_{t-1}^2}(y_{t-1}\xi_t) \\ \xi_t^2 - \sigma_t^2 \end{bmatrix}.
\tag{26}
$$

Specifically, the driving process for the coefficient ϕ_t is the prediction error scaled by the squared regressor, while the time-varying volatility σ_t^2 is driven by the squared prediction error, consistently with similar results in Delle Monache and Petrella (2014) and Blasques et al (2014).[16] This result can be easily extended to vector autoregressive models (see Appendix B.1) as well as to ARMA models.

---

[15]In case we restrict the model to have stable roots as in Delle Monache and Petrella (2014) and positive variance as in the previous example, we re-parameterize the model as follows $f_t = (\operatorname{arctanh}\phi_t, \log\sigma_t)'$, which implies $\overset{\bullet}{T}_t = [(1 - \phi_t^2), 0]$, and $\overset{\bullet}{Q}_t = [0, 2\sigma_t^2]$.

[16]Note that the recursion starts at time $t = 2$ because the likelihood is conditional on the initial observation. However, the use of the SSF methods allows to have the exact likelihood (also for $t = 1$), thus the score can also be estimated for the initial observation.

3   A Monte Carlo analysis

Before testing our approach on actual data we assess through a Monte Carlo exercise its ability to replicate the salient features of a number of data generating processes (DGPs). Since our claim is that the estimation technique that we propose is 'adaptive', we expect it to work reasonably well in capturing different sources of time variation, whether coming for instance from the loadings of the unobserved components or from the autoregressive coefficients of the transition equations or from the volatilities of the measurement or transition equation errors. More in detail, we design the following DGPs:

DGP1 - Time-Varying loadings

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda_t \end{bmatrix} \mu_t + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \quad \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \sim \mathcal{N}\left(0, I\right),$$
$$\mu_t = 0.8\mu_{t-1} + u_t \qquad u_t, \sim \mathcal{N}\left(0, 1\right).$$

DGP2 - Time-Varying AR coefficient

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \mu_t + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \quad \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \sim \mathcal{N}(0, I),$$
$$\mu_t = \rho_t \mu_{t-1} + u_t, \qquad u_t \sim \mathcal{N}(0, 1).$$

DGP3 - Time-Varying Volatility in the measurement equation

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_{\varepsilon,t}^2),$$
$$\mu_{t+1} = 0.8\mu_t + u_t, \quad u_t \sim \mathcal{N}(0, 1).$$

DGP4 - Time-Varying Volatility in the transition equation

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1),$$
$$\mu_{t+1} = 0.8\mu_t + u_t, \quad u_t \sim \mathcal{N}(0, \sigma_{\eta,t}^2).$$

In DGP1 we design a bivariate factor model and let the loading of the second variable on the common factor μ_t vary over time. In DGP2 we lay out a similar model but this time we keep both factor loadings constant while introducing time variation in the law of motion of the common factor, which evolves as an AR(1) model with time-varying coefficient ρ_t. In DGP3 and DGP4 we experiment with time varying volatilities, which can appear either in the measurement or in the transition equation. We look at two different sample sizes $n = 250$ and $n = 500$. We experiment with six different laws of motions for the TVPs entering DGP1 to DGP4:

CONSTANT $f_t = a_1, \forall t;$

SINE $f_t = a_2 + b_2 \sin\left(\frac{2\pi t}{T/2}\right);$

SINGLE STEP $\quad f_t = a_3 + b_3 (t \geq \tau);$

DOUBLE STEP $\quad f_t = a_4 + b_4 I (t \geq \tau_1) + c_4 I (t \geq \tau_2);$

RAMP $\qquad\qquad f_t = a_5 + \left(\frac{b_5}{T/c_5} \right) \mod (t);$

AR(1) MODEL $\quad f_t = a_6(1 - b_6) + b_6 f_{t-1} + \xi_t, \quad \xi_t \sim \mathcal{N}(0, c_6);$

where $f_t = \lambda_t$ in DGP1, $f_t = \rho_t$ in DGP2, $f_t = \sigma^2_{\varepsilon,t}$ in DGP3, and $f_t = \sigma^2_{u,t}$ in DGP4. Moreover, f_t in the AR(1) model is transformed to be within the unit circle for DGP2, and to be positive for DGP3 and DGP4. The calibration of the parameters a_1, a_2, b_2, ..., b_6 is described in Appendix D.2.

Case 1 is a baseline exercise in which we keep the parameters constant over time. We then move to four cases where the parameters change over time according to a deterministic process. In case 2 the parameters follow a cyclical pattern determined by a sine function. In cases 3 and 4 we let the parameters break at discrete points in time, allowing for either one or two breaks. We set the location of the discrete breaks at given point in the sample. In the case of a single break τ corresponds to half of the sample, while when we consider two breaks τ_1 and τ_2 are located at 1/3 and 2/3 of the sample. Case 5 (RAMP) is a rather challenging case, whereby the parameters increase for some time before returning abruptly to their starting levels. Finally, case 6 is the only one in which we let the parameters vary stochastically, following a very persistent AR(1) model. The DGPs that we design are simple, in that time variation is introduced in all the channels in which it can manifest itself, but only one at the time. By focusing on a single channel at the time allows us to better discriminate the situations in which our model either succeeds or fails at identifying and tracking time variation.

We present the results of the Monte Carlo exercise in Table 1. The table is organized in four panels corresponding to each of the four DGPs under analysis. On the left hand side of the table we show the results for a sample size of n=250, on the right hand side those obtained when setting n=500. In each panel we report the results obtained for the six alternative laws of motion described above. We base our assessment on five different statistics, namely the Root Mean Squared Error (RMSEs), the Mean Absolute Error (MAE), the correlation between actual and estimated coefficients, the Coverage (i.e. percentage of times that the actual parameters fall in a given estimated confidence interval) and the number of cases in which a pile-up occurs (#Pile-up). The last statistics consists of the number of simulations in which the static coefficients that pre-multiply the score end up being lower than 10^{-6}, which we take as sufficient evidence that the estimated parameters are effectively zero, i.e. that there is no time variation.

For each DGP we target 300 simulations. However, the actual number of samples changes depending on the specifications. In the case of constant coefficients, where we would like to see our estimator to end up in a pile-up situation as often as possible, we perform 300 simulations and compute all the statistics on these samples. For the remaining specifications, on the other hand, we keep on drawing artificial samples until we obtain 300 simulations in which

the estimated parameters are different from zero and compute RMSEs, MAEs, correlations and coverage ratios on these 300 artificial samples. At the same time we also keep track of the number of times the pile-up problem arises. To better understand how we proceed let us take a concrete example, that is the top-left panel of Table 1 (DGP1-Time Varying loadings, n=250). In the first row we report the results for the constant coefficient case. As explained, for this case we simulate 300 artificial samples and estimate the model using our algorithm. It turns out that in 215 out of 300 simulations our estimation method ends up in a pile-up. The RMSEs, MAEs, Correlations and Coverages, are estimated on all the 300 simulations. Now let us take in the same panel the last line, referring to the AR(1) specification. In this case we need to draw up to 319 samples to obtain 300 simulations in which the estimation algorithm does not end being stuck in a region of the likelihood where the model loading is zero. Now, in this case all the remaining statistics are computed on the 300 'good' samples. We proceed in this way because we want to appraise two different issues. The former is the percentage of cases in which the algorithm ends up in the pile-up even if the true DGP implies time variation. The second is how well it estimates the parameters *conditional on the model correctly detecting time variation.* The two points are of independent interest because, if we were to find that the model often ends up in the pile-up but it is very precise when it does not, one could decide to force the algorithm to stay away from zero, for instance by using a grid-based estimation method. This is the choice made, for instance Koop and Korobilis (2013), and a solution that we also adopt in one of the two empirical applications.[17]

There are four key takeaways from the Monte Carlo exercise. First, for all the DGPs, when the true parameters are constant the model performs extremely well. This means that the adaptive filter correctly collapses the parameters to a constant. As a result, RMSEs and MAEs are virtually nil, the actual coverage extremely precise and a pile-up at zero occurs in more than two thirds of the cases. This result implies that our estimation method passes an essential prerequisite, i.e. it does not generate spurious time variation in the coefficients when this is not supported by the data. Second, for DGPs 1, 3 and 4 and across all the specifications for the parameters we obtain extremely good coverage. For DGP3 (time-varying volatility in the measurement equation) the results are slightly less favorable in the presence of occasionally braking coefficients (Single Step, Double Step and Ramp) and when n=250. Third, across all DGPs the RAMP specification is the one that the model finds more challenging to track. This specification generally leads to low correlations between actual and estimated parameters, higher RMSEs and MAEs and lower coverage. Fourth, the adaptive filter is very effective in estimating time-varying loadings and auto-regressive coefficients, while it is rather conservative in the estimation of the time-varying variances. In other words, when the variance of either the measurement or the transition equation are subject to occasional breaks the results appear to be satisfactory. Yet if time variation is slow, as in the case of a persistent AR(1) process, the correlation between estimates and actual parameters is markedly lower and in many cases the filter ends in a pile-up. For instance, in the case of DGP4 we need to perform 618 simulations

---

[17]Similarly, in their Monte Carlo Markov Chain estimation, Stock and Watson (2007) reject draws in which the variances are very close to zero.

to accumulate 300 samples where the algorithm estimates parameters different from zero, while in the remaining 318 the estimation does not pick up any change in the model parameters. However, notice that for the cases in which time variation *is* detected, the algorithm performs quite well, i.e. it yields relatively low RMSEs and MAEs and a satisfactory coverage. Hence we take this results as evidence that, in the case of time-varying variances, the algorithm needs substantial evidence of breaks in the parameters to move away from zero, i.e. it is relatively conservative.

To convey a visual idea of the results, in Appendix D.2 we report the graphs of the actual parameters together with the 68% and 90% confidence bands.[18] Most of the features that we have just described are also visually apparent in these graphs. In particular the 68% confidence bands appear to be tightly squeezed around the actual values in the case of constant coefficients, while it is evident the fact that the RAMP specification appears more challenging.

4 Empirical applications

In the following sub-sections we present two empirical exercises showing the usefulness of our method: the former deals with measuring GDP dynamics combining heterogeneous sources of information; in the latter we compile a financial stress indicator and evaluate its relevance for forecasting GDP growth.

4.1 *GDPplus* revisited

Our first empirical application consists of extending the model proposed by Aruoba et al. (2016) in the context of GDP measurement. Before going into the application let us briefly provide a short background to this exercise. Given national account identities, GDP can be measured either from the expenditure or from the income side. In the U.S. the expenditure-side version, which we denote as GDP_E, is more widely-used than the income-side version, which we denote as GDP_I. Even though the two measures are supposed to convey the same information it is often the case that discrepancy between the two concepts is non negligible and can at time give a contrasting picture about the state of the economy. For instance, there was an average difference of two percentage points on an annualized basis in the year preceding the last recession, with GDP_I dipping into negative territory and GDP_E showing an healthy two percent growth.[19] The question of which GDP measure is more reliable is of particular interest to policy makers who need to monitor business cycle developments in real time, since some researchers argue that GDP_I, which has traditionally received much less attention than GDP_E, has actually done a better job at recognizing the start of recessions, see Nalewaik (2012). Aruoba et al. (2016) combine the two existing measures of GDP growth computed from the expenditure and from the income side (defined, respectively, $y_{E,t}$ and $y_{I,t}$) in a dynamic factor

---

[18]In the case of the AR(1) specification we report the difference between actual and estimated, since the actual parameters are stochastic and therefore change at each simulation.

[19]See charts in Appendix E.

model, and extract through optimal filtering techniques a measure of underlying GDP growth (which they define *GDPplus*) that is relatively cleaner of measurement error. In our empirical exercise we extend their framework by allowing for the parameters of the dynamic factor model to change over time, driven by the score of the conditional likelihood.

Formally, the baseline constant parameter specification from which we start is described in the following state space model:[20]

$$
\begin{bmatrix} y_{E,t} \\ y_{I,t} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \alpha_t + \begin{bmatrix} \varepsilon_{E,t} \\ \varepsilon_{I,t} \end{bmatrix}, \quad \begin{bmatrix} \varepsilon_{E,t} \\ \varepsilon_{I,t} \end{bmatrix} \sim \mathcal{N}(0, H),
$$
$$
\alpha_{t+1} = \rho_0 + \rho_1 \alpha_t + \rho_2 \alpha_{t-1} + e_t, \quad e_t \sim \mathcal{N}(0, \sigma^2),
$$

(27)

where H is a full 2×2 matrix. It can be readily recognized that the model is a traditional small-scale dynamic factor model in the spirit of Sargent and Sims (1977). The common factor α_t merges the information from the two GDP indicators, that is *GDPplus*.

We extend model (27) to allow for all the seven parameters, that are ρ_0, ρ_1, ρ_2, σ_t^2, $vech(H)$, to change over time.[21]

We then collect in the vector f_t the three groups of parameters, namely the coefficients of the transition equation, the variance of the transition equation, and the elements of the measurement equation covariance, i.e. $f_t = (\rho_{0,t}, \rho_{1,t}, \rho_{2,t}, \sigma_t^2, vech(\tilde{J}_t)')'$, where $\tilde{J}_t$ is the log-Cholesky transformation of H_t as described earlier. We assume that

$$
f_{t+1} = f_t + \Omega s_t,
$$

where Ω is a diagonal matrix function of only three static coefficients: $\kappa_\rho, \kappa_\sigma, \kappa_l$, one for each group of the TVP. The three static parameters $\kappa_\rho, \kappa_\sigma, \kappa_l$ are estimated by ML.[22] Estimation results are reported in Table 2.

In Figure 1 we report the main estimation results. The first feature that our model unveils, see panel (a), is a secular downward trend in the time-varying intercept $\rho_{0,t}$ coupled with marked falls corresponding to some recessions, namely those of the early 70s and early 2000s. The constant parameter model severely overestimates this term at the end of the sample (dashed green line), with important implications for the estimated long-run growth of the economy.

The evolution of the autoregressive parameters is shown in panel (b). Notice that the second

---

[20]Aruoba et al. (2016) consider up to five alternative specifications, which differ from each other with respect to the correlation structure of the measurement errors. In our application we consider only one of their specifications. Furthermore, in Aruoba et al. (2016) the state variable (i.e. the GDPplus) follows an AR(1) process, we instead prefer the AR(2) specification that is more in line with traditional business cycle models. Notice that the spectrum of an AR(1) model has its peak at zero and most of its mass on the left hand side of the frequency band, while in an AR(2) model the peak of the spectrum can occur at cyclical frequencies.

[21]Several papers provide a motivation for this extension. Antolin-Diaz et al. (2016), for instance, focus on shifts in the intercept ρ_0 as a way to capture different business cycle phases and the secular fall of productivity growth. The importance of changes in GDP growth volatility are stressed by Perez-Quiros and McConnell (2000). The relationship between changes in the conditional mean and in the persistence of US GDP growth is studied by Camacho and Perez Quiros (2007).

[22]The algorithm is initialized fitting an AR(2) model to the average of the two GDP series in the first 4 years. Confidence sets take into account estimation uncertainty following Hamilton (1986).

autoregressive term (red line) fluctuates markedly over time, although it is reasonably centered around zero, the value to which is constrained in the original specification by Aruoba et al. (2016). The first AR term (blue line), also displays non-negligible time variation, especially at the end of the sample. Aruoba et al. (2016) place a lot of emphasis on this parameter, which they estimate at around 0.5 on the whole sample (dashed green line). This value is higher than the one estimated for either $y_{E,t}$ or $y_{I,t}$, which leads them to conclude that the filtered measure is more predictable than its components. Our findings cast some doubt over the possibility of exploiting this auto-correlation in real time, since it seems to be the result of the prolonged fall in macroeconomic activity during the Great Recession rather than a robust, full-sample, feature of $GDPplus$.

Next, in panel (c) we report the estimated long-run growth $\mu_t = \frac{\rho_{0,t}}{(1-\rho_{1,t}-\rho_{1,t})}$. Notice that both the time-varying intercept and the the autoregressive terms contribute to defining μ_t, so that underestimating or overestimating any of this terms can lead to a poor estimate of the long-run GDP growth.[23] According to our estimates, four distinct phases of U.S. long-run growth (measured by the long-run forecast of the estimated model) emerge: the first (in the 1950s and 1960s) characterized by a steady high level of growth, the second (from the mid-1970s) marked by a rapid deceleration, a third period of resurgence starting from mid-1990s and, finally, a sharp decline since the burst of the dotcom bubble in the early 2000s, a development which has been further reinforced by the Great Recession. These results are strikingly in line with the conclusions reached by Antolin-Diaz et al. (2016). As a result, we currently estimate long-run growth in the U.S. to stand at levels between 1.5 and 2 percent, in line with current Congressional Budget Office estimates.

Finally, in panel (d) we show the variance of the error of the common component, a measure of macroeconomic volatility. It is interesting to notice that, after a prolonged period of decrease in the 50s and 60s, volatility reaches a plateau until the mid-80s. From 1984 onwards there is a second marked fall of volatility, corresponding to the Great Moderation, which is interrupted, and partially reversed by the Great Recession in 2008.

In Figure 2 we report the Kalman gain of $y_{E,t}$ relative to that of $y_{I,t}$ over the whole sample, which can be interpreted as the relative weights that the two observed measures receive in the construction of $GDPplus$. Given how this is computed, we have that values below 1 imply that a relatively larger weight is assigned to $y_{I,t}$. From the picture it emerges that, indeed, as claimed by part of the literature, $y_{I,t}$ better captures the behavior of $GDPplus$ in the middle of the sample. However, in recent years, an equal weighting scheme seems warranted. This result squares well with the fact that from July 2015 the BEA started publishing the mean of $y_{E,t}$ and $y_{I,t}$ as an indicator of the business cycle. Our results corroborate the choice of an equal weighting scheme at the current juncture.

In Figure 3 we show the resulting $GDPplus$ estimate (blue line) together with the two

---

[23]It is worth stressing that this is a key variable for both monetary and fiscal policy. Orphanides (2003), for instance, argues that a failure in identifying a permanent slowdown in productivity growth in the first part of the 70s led monetary policy makers to be excessively complacent, resulting in a prolonged period of high inflation. Long-run projections are also at the center of the fiscal policy debate as they determine the solvency of pension systems and, more generally, the sustainability of public debt.

underlying components. From this picture it is clear that the estimated latent factor is considerably smoother than the two observed series, so that in many periods $y_{I,t}$ and $y_{E,t}$ fall outside of the confidence intervals.[24]

Concluding, when applied to the problem of extracting a cleaner measure of output growth from observed GDP series, our model recovers a number of stylized facts in US business cycle dynamics highlighted by a broad number of previous studies but overlooked in the constant parameters framework proposed by Aruoba et al. (2016). In particular, our model delivers a real time/time-varying assessment (i) of the long-run economic growth (ii) of macroeconomic volatility and (iii) of the relative importance of $y_{I,t}$ and $y_{E,t}$. All these three ingredients are of great interest to policy makers and business cycle analysts.

4.2 Measuring financial stress in the euro area

In our second empirical exercise we construct a measure of financial stress for the euro area and evaluate its usefulness in forecasting and nowcasting GDP growth. This application is designed to test our method in a more complex setting, where the information set is characterized by ragged edges, indicators start in different time periods and are sampled at different frequencies. Our index of financial stress is estimated starting from the 15 sub-indices that form the Composite Indicator of Systemic Stress (CISS) for the euro area developed by Hollo' et al. (2012), see Table 3 for details. We model these 15 variables together with 4 business cycle indicators, namely GDP growth, industrial production growth and two survey indicators, the composite Purchasing Manager Index (PMI) and its orders subcomponent. The vector of observable is $y_t = (y_t^{r'}, y_t^{x'})'$, in which y_t^r is a vector of macroeconomic variables including also the quarter on quarter growth rate of GDP and other cyclical indicators, and y_t^x are the financial indicators.[25]

In constructing a measure of financial stress a clear issue is endogeneity, that is we need to disentangle primitive financial shocks from the endogenous response of financial variables to other shocks. In particular, Hatzius et al. (2012) define financial shocks as *"exogenous shifts in financial conditions that influence or otherwise predict future economic activity. True financial shocks should be distinguished from the endogenous reflection or embodiment in financial variables of past economic activity that itself predicts future activity"*. The definition above is implicitly a statement about Granger-causality and motivates the empirical specification of our model. In particular, we look for a common index of financial stress that predicts output beyond the information in the past levels of economic activity.

---

[24]In Figures E and E.3 in Appendix E we zoom these comparisons in two recession periods, namely the 1970s recessions and the Great Recession of 2008. It is evident that during these turbulent periods *GDPplus* gives less volatile indications on output dynamics than the other two measures.

[25]Although financial variables are potentially available at frequencies higher than monthly, we restrict ourselves to the use of monthly averages in order to limit the dimension of the state vector and also because a monthly update of such an indicator seems sufficiently informative for policy makers.

Formally, we link the observed variables to the state equation as follows:

$$\begin{bmatrix} y_t^r \\ y_t^x \end{bmatrix} = \begin{bmatrix} \lambda_t^r & 0 \\ 0 & \lambda_t^x \end{bmatrix} \begin{bmatrix} \alpha_t^r \\ \alpha_t^x \end{bmatrix} + \begin{bmatrix} \varepsilon_t^r \\ \varepsilon_t^x \end{bmatrix}, \qquad \begin{bmatrix} \varepsilon_t^r \\ \varepsilon_t^x \end{bmatrix} \sim \mathcal{N}(0, H_t), \tag{28}$$

where α_t^r and α_t^x summarize the information in the indicators of economic activity and financial stress. Time variation in the loading vector allows for the relevance of the variables included in the system for the factors to vary over time. More generally, the information content of different asset classes could change in different stages of the business cycle. The state vector evolves as follows:

$$\begin{bmatrix} \alpha_{t+1}^r \\ \alpha_{t+1}^x \end{bmatrix} = \begin{bmatrix} \gamma_t^r \\ \gamma_t^x \end{bmatrix} + \Phi_{1t} \begin{bmatrix} \alpha_{t-1}^r \\ \alpha_{t-1}^x \end{bmatrix} + \Phi_{2t} \begin{bmatrix} \alpha_{t-2}^r \\ \alpha_{t-2}^x \end{bmatrix} + u_t, \qquad u_t \sim \mathcal{N}(0, \Sigma_t). \tag{29}$$

The feedback matrices capturing the dynamic relationship between real and financial factors, Φ_{jt}, are restricted to be upper triangular and enforce a Granger-causality restriction between the two factors. Furthermore, time varying volatilities are allowed for both in the measurement error and to the innovations of the transition equations.

Note that the contemporaneous relation between the factors is fully captured by the covariance matrix Σ_t on which we have made no restriction. Therefore, current development in real activity, i.e. in the unpredictable changes in the coincident indicator α_{xt}, affect contemporaneously financial stress and vice versa.[26] Note that even though the measurement equation is block diagonal the fact that the covariance matrix is unrestricted implies that financial variables help to nowcast current economic activity. In fact, it is easy to show that the Kalman gain is not diagonal in this setting. Intuitively, if financial markets react systematically to macroeconomic developments, their reaction helps to pin down the real shocks and therefore to improve the nowcast of the coincident indicator and its components.

Details on the full state space representation of the model and on the estimation method are discussed in Appendix C.

Figure 4 and 5 show the financial and business cycle factors as estimated by our model. We compare our index of financial stress with the Composite Index of Systemic Stress (CISS) produced by Hollo' et al. (2012) and the business cycle factor with Eurocoin, the monthly indicator of medium-term growth proposed by Altissimo et al. (2010) and regularly published by the CEPR and the Bank of Italy.[27] Two observations are in order. First, the estimated financial factor, our measure of financial stress, also captures correctly the two period of financial stress after the Lehman and the Sovereign Debt Crisis. Differently from the CISS, however, it raises warning signals in the early part of the sample, corresponding to the 2000-2001 downturn. This is not surprising, since in our model innovations to the financial factor and to business cycle factor are correlated, but the Granger causality restriction imposed on

---

[26]In principle one can attempt at identifying financial shocks in this context for instance imposing that economic activity reacts the developments in the financial market only with a delay, whereas the news components in the macroeconomic data releases are reflected into changes in asset prices.

[27]See http://eurocoin.cepr.org. All the indicators are rescaled to have zero mean and unit variance.

the model ensures that what we are picking up is not a feedback from the business cycle to financial stress. Second, the business cycle indicator captures the main features of euro-area business cycle, namely the two expansion phases before the Great Recession, as well as the double dip downturn generated by the Great Recession first and by the Sovereign Debt Crisis second.

Next, in Figure 6 we show the time varying elements of the covariance matrix of the unobserved states, i.e. the volatilities of the business cycle and of the financial factor, as well as their time varying covariance. The model detects significant shifts in the volatilities in correspondence with the two recent recessions. Noticeably, the covariance between the disturbances of these two factors appears to be quite stable for most of the sample, but it turns significantly negative after 2007, a period of both severe financial stress and real economy strains.

The distinctive feature of our method is that it can be used to estimate financial stress and evaluate its (real time) predictive content for GDP growth in a unified framework. We therefore set up a real-time out of sample forecast exercise in which we predict GDP 1 to 12 months ahead. We use actual vintages of GDP and industrial production as available from the ECB area Real Time database and align the surveys and financial variables to replicate as closely as possible the actual information set available to forecasters at the end of the reference month. Forecasts are formed for each quarter between 2000q1 and 2014q3. In line with the nowcasting literature we measure steps ahead in terms of months to GDP releases. Given the typical publication lag of GDP (45 days) this means that the 1 step ahead forecast is usually computed with information available in the first month after the reference quarter.[28]

In Figure 7 we report the evolution of the Root Mean Square Forecast Error (RMSE) of our model (blue solid line) and contrast it with that of two different benchmarks. The former is a a simple univariate AR benchmark (red dotted line). The latter is a restricted factor model in which we exclude financial stress variables, so that forecasts are computed only based on business cycle indicators (green dashed line). This latter benchmark is useful in assessing the contribution of financial stress variables to predictive accuracy once timely business cycle indicators are accounted for. Notice that at long horizons (12 months to GDP release) the model that performs best is the restricted model with a single business cycle factor. As the forecast horizon shortens, however, the RMSE of the model that also includes financial stress variables falls more rapidly than the others. At short horizons (1 to 4 months to GDP release), the model with two factors performs better than the other two, indicating that information on financial stress provides valuable coincident information on GDP growth. Not surprisingly both factor models always perform better than the AR benchmark, which means that the presence of monthly information improves predictive accuracy.

In Figure 8 we report the cumulative sum of squared forecast error differentials, computed as

$$CSSED_t = \sum_{j=1}^{t} (e_{j,AR}^2 - e_{j,TVP}^2). \tag{30}$$

---

[28]In the literature this is typically called a *backcast.*

This statistics is very useful in revealing the parts of the forecast sample where the TVP-Factor models accrue their gains relative to the AR benchmark. Positive and increasing values indicate that the TVP models outperforms the AR benchmark, while negative and decreasing values suggest the opposite. For simplicity, we focus on short-term horizons (1 to 4 steps ahead). Two results stand out. First, over most of the sample the CSSED are only mildly upward sloping, while most of the forecasting gains obtained by our TVP-Factor models are obtained during the Great Financial Crisis. Second, the gain in forecast accuracy in 2008 is considerably larger for the model that includes measures of financial stress.

Finally, we look at the relevance of our financial stress indicator for predictive density accuracy. In Figure 9 we compare the results obtained by the factor model that includes financial stress indicators (top panel) with those produced by the model that excludes them (bottom panel). The left-hand panels of Figure 9 show the empirical distribution (p.d.f.) of the Probability Integral Transforms (PITs) together with the 95% confidence interval (broken lines) using a normal approximation to a binomial distribution as in Diebold et al. (1998). If the density forecast produced by the model is satisfactory, the PITs should be distributed uniformly, see Berkowitz (2003). The right-hand panels display the cumulative distribution (c.d.f.) of the PITs together with the critical values based on the Rossi and Sekhposyan (2013) test. Under the null hypothesis the PITs should be uniformly distributed and the c.d.f. of the PITs should therefore be the 45 degrees line.

Clearly, the forecast densities produced by the model that uses information on financial variables are well calibrated. Indeed, the PITs distribution is not significantly different from that of a uniform random variable, and the Rossi and Sekhposyan (2013) test never rejects the null of uniform distribution. In contrast, when financial indicators are excluded from the model the left tail of the PITs distribution, corresponding to periods of low/negative growth, is not well calibrated and the Rossi and Sekhposyan (2013) rejects the null of uniform distribution. This finding is consistent with the results in Alessandri and Mumtaz (2014), Adrian et al. (2016) and Giglio et al. (2016), who find that financial indicators are particularly useful in forecasting economic recessions.

5   Conclusions

In this paper we develop a score-driven approach to estimate state space models with TVP. By letting the dynamics of the parameters be driven by the score of the predictive likelihood we show that the model retains many desirable properties, namely it can be analyzed using an augmented KF procedure. We derive an auxiliary filter that, running in parallel with the standard KF, allows us estimate simultaneously the unobserved state vector and the TVP. Given that a variety of time series models have a state space representation, the proposed methodology has to be considered of wide interest in econometrics and applied macroeconomics.

After presenting the main results we discuss a number of extensions of the theoretical framework. Two of them could be particularly valuable in applied work. First, since in some

applications the researcher would like to impose parameters restrictions, such as stationarity and non-negative variances, we show how to incorporate general restrictions on the model parameters. Second, we illustrate how the model can be extended to deal with data at mixed frequencies and missing observations.

We assess the usefulness of the method in the controlled environment of a Monte Carlo experiment and in two empirical applications. The Monte Carlo exercise shows that the new filter that we propose tracks well a wide range of possible sources of time variation. Moreover, when the true data generating process features constant parameters, it does not generate spurious fluctuations in coefficients and variances. In the empirical section we report the results of two exercises. In the first one we focus on business cycle measurement. In particular, we take as a starting point the model that Aruoba et al. (2016) have developed to estimate GDP growth on the basis of underlying (noisy) measures and extend it to account for time variation in the model parameters. Our model delivers a real time and time-varying assessment of the long-run economic growth and of macroeconomic volatility. In a second application we use our method to extract an indicator of financial stress from a panel of business cycle and financial variables characterized by mixed frequencies and ragged edges. The model picks up significant changes in the co-movement between financial stress and the business cycle and allows us to evaluate the real-time predictive content of financial indicators for GDP growth. We find that financial stress is relevant for very short-term forecasting and relatively more informative for predicting recessions.

References

Abadir K. and J.R. Magnus (2005). *Matrix Algebra*. Cambridge University Press.

Adrian, T., Boyarchenko, N. and D. Giannone (2016). Vulnerable growth. Staff Reports 794, Federal Reserve Bank of New York.

Alessandri P. and H. Mumtaz (2014). Financial indicators and density forecasts for US output and inflation. *Temi di discussione, 977, Bank of Italy.*

Altissimo F., Cristadoro R., Forni M., Lippi M. and G. Veronese (2010). New EuroCOIN: Tracking Economic Growth in Real Time. *The Review of Economics and Statistics, 92(4), 1024-1034.*

Antolin-Diaz J., Drechsel T. and I. Petrella (2016). Tracking the slowdown in long-run GDP growth. *Bank of England working papers 587.*

Aruoba S.B., Diebold F.X., Nalewaik J.J., Schorfheide F. and D. Song (2016). Improving GDP Measurement: A Measurement-Error Perspective. *Journal of Econometrics, forthcoming.*

Bańbura M., Giannone D., Modugno M. and L. Lucrezia (2013). Now-Casting and the Real-Time Data-Flow. *Handbook of Economic Forecasting, 2(A), 195-237.*

Benati L. (2007). Drift and breaks in labor productivity. *Journal of Economic Dynamics and Control, 31(8), 2847-2877.*

Berkowitz J. (2001). Testing Density Forecasts with Applications to Risk Management. *Journal of Business and Economic Statistics, 19, 465-474.*

Bianchi F., Mumtaz H. and P. Surico (2009). The great moderation of the term structure of UK interest rates. *Journal of Monetary Economics, 56(6), 856-871.*

Blasques F., Koopman S.J. and A. Lucas (2014). Optimal Formulations for Nonlinear Autoregressive Processes. *Tinbergen Institute Discussion Papers TI 2014-103/III.*

Blasques F., Koopman S.J. and A. Lucas (2015). Information Theoretic Optimality of Observation Driven Time Series Models. *Biometrika, 102(2), 325-343.*

Camacho M. and G. Perez-Quiros (2007). Jump-and-Rest Effect of U.S. Business Cycles. *Studies in Nonlinear Dynamics and Econometrics, 11(4), 1-39.*

Creal D., Koopman S.J. and A. Lucas (2008). A General Framework for Observation Driven Time-Varying Parameter Models. *Tinbergen Institute Discussion Paper, 108/4.*

Creal D., Kopman S.J. and A. Lucas (2013). Generalised Autoregressive Score Models with Applications. *Journal of Applied Econometrics, 28(5), 777–795.*

Del Negro M. and C. Otrok (2008). Dynamic Factor Models with Time-Varying Parameters: Measuring Changes in International Business Cycles. *Federal Reserve Bank of New York Staff Reports, 326.*

Delle Monache D. and I. Petrella (2014). Adaptive Models and Heavy Tails. *Birkbeck working paper, 1409.*

Delle Monache D., Petrella I. and F. Venditti (2016). Common faith or parting ways? A time varying parameter factor analysis of euro area inflation. *Advances in Econometrics, 35, Dynamic Factor Models. Edited by: Hillebrand E. and S.J. Koopman.*

Diebold F.X., Gunther T.A. and S.A. Tay (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review, 39(4), 863-883.*

Durbin J. and S.J. Koopman (2012). *Time Series Analysis by State Space Methods.* Oxford: Oxford University Press.

Eickmeier S., Lemke W. and M. Massimiliano (2015). A Classical Time Varying FAVAR Model: Estimation, Forecasting, and Structural Analysis. *Journal of the Royal Statistical Society, (Series A), 178, 493-533.*

Giglio, S., Kelly, B. and S. Pruitt (2016). Systemic risk and the macroeconomy: An empirical evaluation, *Journal of Financial Economics, Elsevier, vol. 119(3), 457-471.*

Hamilton J.D. (1986). A standard error for the estimated state vector of a state space model. *Journal of Econometrics, 33(3), 387-397.*

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

Harvey A.C., Ruiz E. and E. Sentana (1992). Unobserved Component Time Series Models With ARCH Disturbances. *Journal of Econometrics, 52, 129-158.*

Harvey A.C. (2013). *Dynamic Models for Volatility and Heavy Tails. With Applications to Financial and Economic Time Series.* Cambridge University Press.

Hatzius J., Hooper P., Mishkin F.S., Schoenholtz L.K., and M.W. Watson (2010). Financial Conditions Indexes: A Fresh Look after the Financial Crisis. *NBER Working Papers 16150.*

Holló, D., Kremer M. and M. Lo Duca (2012). CISS - a composite indicator of systemic stress in the financial system. *European Central Bank Working Paper Series, 1426.*

Koop G. and D. Korobilis (2013). Large Time-Varying Parameter VARs. *Journal of Econometrics, 177, 185-198.*

Koop G. and D. Korobilis (2014). A New Index of Financial Conditions. *European Economic Review, 71, 101-116.*

Koopman S.J., Mallee M.I.P. and M. Van der Wel (2010). Analyzing the Term Structure of Interest Rates Using the Dynamic Nelson-Siegel Model With Time-Varying Parameters. *Journal of Business and Economic Statistics, 28(3), 329-343.*

Lucas A., Opschoor A., and J. Schaumburg (2016). Accounting for missing values in score-driven time-varying parameter models. *Economics Letters, 148, 96-98.*

Illing M. and Y. Liu (2006). Measuring Financial Stress in a Developed Country: An Application to Canada. *Journal of Financial Stability, 2(3), 243–65.*

Marcellino M., Purqueddu M. and F. Venditti (2016). Short-term GDP forecasting with a mixed-frequency dynamic factor model with stochastic volatility. *Journal of Business and Economic Statistics, 34(1), 118-127.*

Mikkelsen J.G., Hillebrand E. and G. Urga (2015). Maximum Likelihood Estimation of Time-Varying Loadings in High-Dimensional Factor Models. *CREATES Research Paper 2015-61.*

Nalewaik J.J. (2012), Estimating Probabilities of Recession in Real Time Using GDP and GDI. *Journal of Money, Credit and Banking, 44, 235-253.*

Orphanides A. (2003). Monetary policy evaluation with noisy information. *Journal of Monetary Economics, Elsevier, 50(3), 605-631.*

Perez-Quiros G. and M.M. McConnell (2000). Output Fluctuations in the United States: What Has Changed since the Early 1980's? *American Economic Review, 90(5), 1464-1476.*

Rossi B. and T. Sekhposyan (2013). Conditional Predictive Density Evaluation in the Presence of Instabilities. *Barcelona Graduate School of Economics, WP688.*

Sargent T.J. and C.A. Sims (1977). Business Cycle Modeling Without Pretending to Have too Much a Priori Theory. *C.A. Sims (ed.), New Methods in Business Cycle Research: Proceedings from a Conference, Federal Reserve Bank of Minneapolis, 45-109.*

Stock J.H. and M.W. Watson (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking, 39(F), 3-33.*

Table 1: MONTE CARLO EXERCISE

	n = 250						n = 500					
	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile-up	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile-up

DGP1: TIME-VARYING LOADING

	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile-up	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile-up
CONSTANT	0.000	0.000	-	0.680	0.896	215	0.00	0.000	-	0.682	0.898	223
SINE	0.493	0.396	0.899	0.640	0.860	0	0.390	0.311	0.939	0.653	0.870	0
SINGLE STEP	0.433	0.294	0.915	0.656	0.872	0	0.362	0.236	0.942	0.660	0.882	0
DOUBLE STEP	0.485	0.348	0.930	0.648	0.868	0	0.390	0.276	0.952	0.652	0.874	0
RAMP	0.431	0.322	0.463	0.680	0.892	0	0.350	0.247	0.641	0.680	0.892	0
AR(1) MODEL	0.207	0.169	0.649	0.672	0.892	19	0.213	0.171	0.727	0.680	0.898	2

DGP2: TIME-VARYING AR COEFF

	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile-up	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile-up
CONSTANT	0.000	0.000	-	0.676	0.896	197	0.000	0.000	-	0.680	0.899	204
SINE	0.361	0.291	0.781	0.676	0.896	0	0.281	0.219	0.870	0.682	0.900	0
SINGLE STEP	0.269	0.193	0.718	0.676	0.896	0	0.225	0.154	0.783	0.678	0.897	0
DOUBLE STEP	0.279	0.220	0.845	0.684	0.900	0	0.235	0.175	0.880	0.682	0.898	0
RAMP	0.193	0.153	0.190	0.680	0.896	0	0.178	0.135	0.379	0.680	0.896	0
AR(1) MODEL	0.243	0.195	0.537	0.678	0.896	48	0.252	0.204	0.634	0.682	0.900	15

DGP3: TIME-VARYING VOLATILITY - MEASUREMENT EQUATION ERROR

	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile-up	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile-up
CONSTANT	0.00	0.00	-	0.674	0.892	229	0.00 0	0.000	-	0.678	0.898	249
SINE	0.582	0.466	0.551	0.688	0.896	24	0.444	0.360	0.725	0.691	0.896	0
SINGLE STEP	0.493	0.359	0.843	0.636	0.856	0	0.382	0.272	0.881	0.652	0.872	0
DOUBLE STEP	0.448	0.355	0.843	0.624	0.844	1	0.355	0.272	0.881	0.650	0.868	0
RAMP	0.603	0.501	0.200	0.648	0.868	163	0.514	0.417	0.368	0.662	0.878	30
AR(1) MODEL	0.341	0.281	0.476	0.676	0.892	300	0.341	0.274	0.513	0.678	0.892	159

DGP4: TIME-VARYING VOLATILITY - TRANSITION EQUATION ERROR

|	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile-up	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile-up
CONSTANT	0.000	0.000	-	0.680	0.896	249	0.000	0.000	-	0.680	0.898	247
SINE	0.571	0.468	0.578	0.680	0.896	105	0.503	0.405	0.702	0.681	0.896	14
SINGLE STEP	0.508	0.374	0.813	0.652	0.876	1	0.421	0.296	0.859	0.664	0.884	0
DOUBLE STEP	0.472	0.375	0.828	0.644	0.866	0	0.369	0.283	0.878	0.660	0.880	0
RAMP	0.616	0.522	0.162	0.656	0.876	39	0.535	0.442	0.319	0.666	0.886	7
AR(1) MODEL	0.375	0.305	0.482	0.672	0.892	318	0.352	0.283	0.532	0.674	0.894	188

Note: The results shown in the first and in the second panel (DGP1 and DGP2) refer to a bivariate factor model in which two variables are driven by a single common factor that evolves as an autoregressive process of order 1. In the first case (DGP1) the loading of the second variable on the common factor varies over time and all the other parameters are kept constant. In the second case (DGP2) the autoregressive component of the common factor varies over time and all the other parameters are kept constant. The results shown in the third and in the fourth panel (DGP3 and DGP4) refer to ARMA(1,1) models that are cast in state space and feature time varying variances of the random disturbance in, respectively, the measurement and the transition equation. The abbreviations *Corr.* and *Cov.* stand, respectively for Correlation and Coverage, while # Pile-up denotes the number of simulations in which the algorithm delivers constant parameters. The different laws of motion of the parameters in the first column (Constant, Sine, Single Step, Double Step, Ramp and AR(1) are described in Section 4).

Table 2: TVP-GDPPLUS, ESTIMATION RESULTS

κ_T	0.030
	(0.007)
κ_Q	0.111
	(0.021)
κ_H	0.002
	(0.002)
Log-lik	-1045.22

Note: The table contains the estimated parameters and their standard errors (in parenthesis) for TVP-GDPplus. κ_T relates to the dynamic of the coefficients of the autoregressive specification of the common factor. κ_Q and κ_H relate to the volatility dynamics in the transition and measurement equation, respectively. The best log-likelihood is indicated with Log-lik.

Figure 1: TIME-VARYING PARAMETERS: *GDPplus* MODEL

(a) Time varying intercept

(b) Time varying AR coefficients

(c) Time varying long-run mean

(d) Time varying error variances
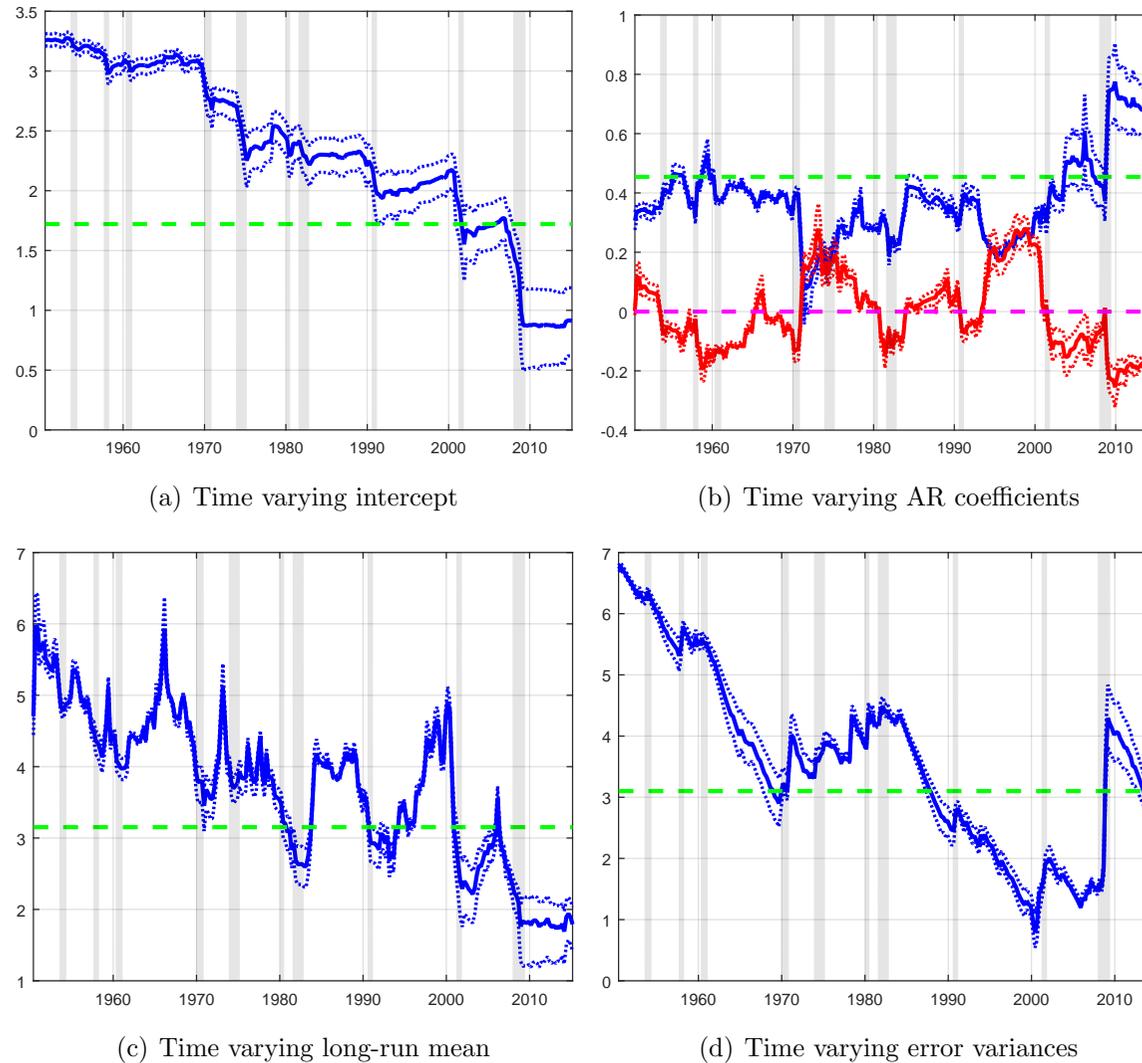
Note: The figure reports median estimates (blue solid line) together with 64% confidence intervals (blue dotted line) of the time-varying coefficient in the time-varying version of the *GDPplus* model. The dashed green line indicates the estimated coefficient in the constant parameter version of the model. Shaded areas indicate recessions as dated by the NBER. Panel (a) reports the intercept, $\rho_{0,t}$. Panel (b) reports the autoregressive coefficients, $\rho_{1,t}$ (blue lines) and $\rho_{2,t}$ (red lines). Panel (c) reports the implied long-run mean, estimated as $\rho_{0,t}/(1 - \rho_{1,t} - \rho_{2,t})$. Panel (d) reports the variance of the innovation to the transition equation, i.e. (σ_t^2).

Figure 2: RELATIVE KALMAN FILTER GAIN
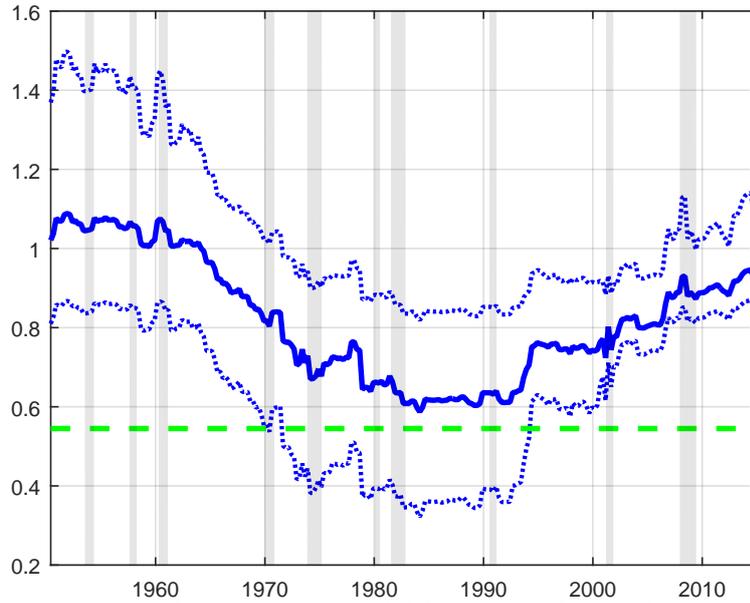
Note: The figure reports median estimates (solid lines) together with 64% confidence intervals (dotted lines) of the time-varying relative Kalman gain (GDP_E over GDP_I) in the time-varying version of the *GDPplus* model. The dashed green line indicates relative Kalman Gain in the constant parameter version of the model. Shaded areas indicate recessions as dated by the NBER.

Figure 3: ESTIMATED COMMON FACTOR, GDP_E AND GDP_I
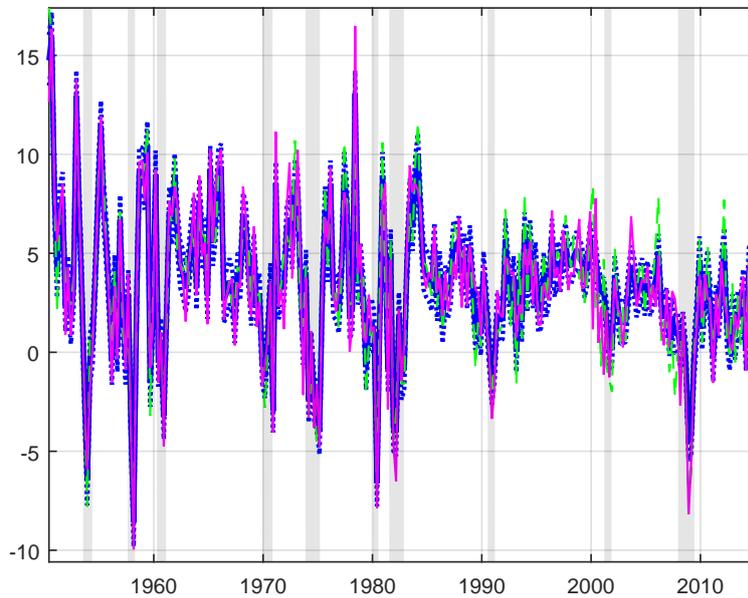
Note: The figure reports the estimated common factor (a_t, green dashed) in the time-varying version of the *GDPplus* model together with $y_{I,t}$ (blue solid line) and $y_{E,t}$ (purple solid line). Shaded areas indicate recessions as dated by the NBER.

33

Table 3: Data used for the Financial Stress Model

	Starting Period	Ending Period	Frequency
GDP	Jan-86	Dec-14	Quarterly
PMI - composite	Jan-98	Jan-15	Monthly
PMI - orders	Jan-98	Jan-15	Monthly
Industrial Production	Feb-87	Dec-14	Monthly
CMAX of non-financial sector stock market	Jan-86	Feb-15	Monthly
CMAX A Rated NFC and Gov. Bonds	Jan-86	Feb-15	Monthly
MFI Emergency Central Bank Lending	Jan-99	Feb-15	Monthly
Stock-Bond Correlation	Jan-86	Feb-15	Monthly
Spread 3 Month Euribor-French T Bill	Dec-98	Feb-15	Monthly
10 year interest swap spread	Mar-98	Feb-15	Monthly
Spread A Rated NFC and Financial Corporations	Mar-98	Feb-15	Monthly
Spread A Rated NFC and Gov. Bonds	Mar-98	Feb-15	Monthly
RV of 10 year Bund	Mar-98	Feb-15	Monthly
RV of the 3-month Euribor rate	Dec-98	Feb-15	Monthly
RV of the idiosyncratic equity return	Jan-86	Feb-15	Monthly
RV Euro/Yen	May-90	Feb-15	Monthly
RV of non-financial sector stock market	Mar-98	Feb-15	Monthly
RV Euro/Pound	Jun-90	Feb-15	Monthly
RV Euro/US$	Oct-89	Feb-15	Monthly

Note: PMI stands for Purchasing Manager Index. CMAX measures the maximum cumulated loss over a moving two-year window. RV stands for realized volatility. For further details see Hollo' et al. (2012).

Figure 4: ESTIMATED FINANCIAL STRESS FACTOR
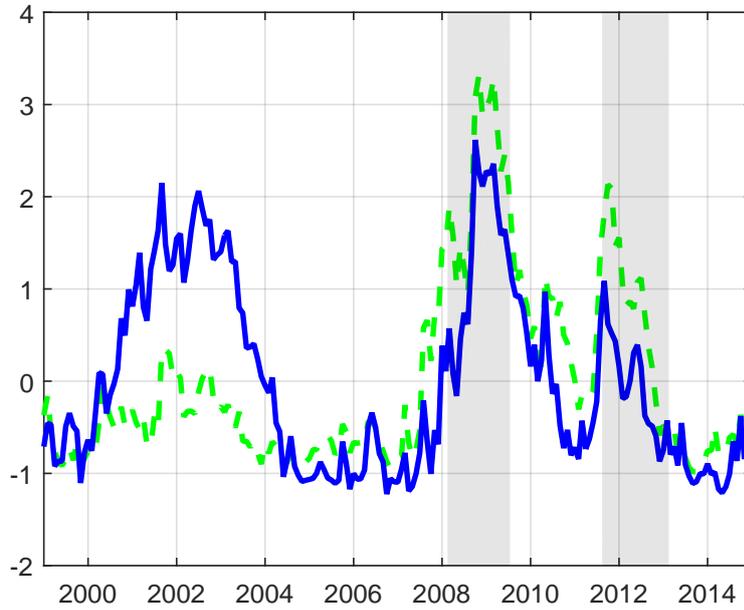
Note: The figure reports the median estimate of the financial stress indicator obtained in our factor model (solid lines) together with the Composite Index of Systemic Stress (CISS) produced by Hollo' et al. (2012). Shaded areas indicate recessions as dated by the CEPR.

Figure 5: ESTIMATED BUSINESS CYCLE FACTOR
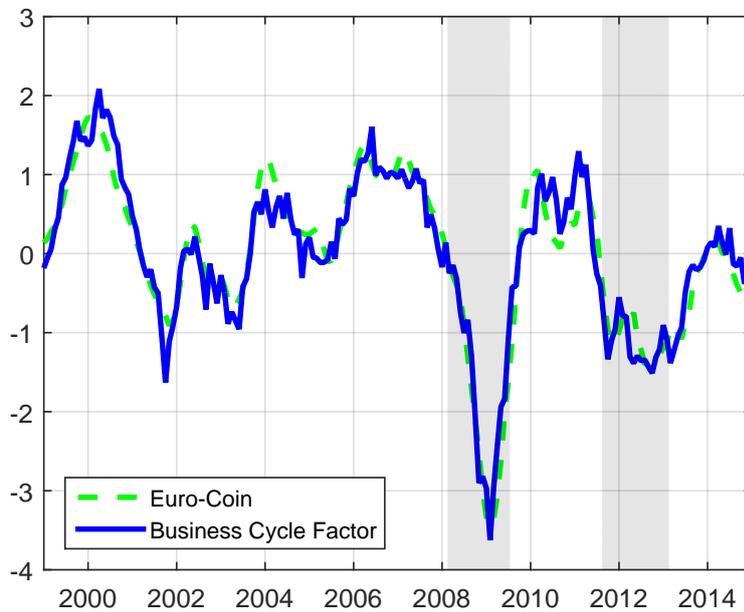
Note: The figure reports the median estimate of the business cycle indicator obtained in our factor model (solid lines) together with EuroCoin, an index of medium-term GDP growth for the euro area published by the CEPR and the Bank of Italy. Shaded areas indicate recessions as dated by the CEPR.

Figure 6: TIME VARYING COVARIANCES
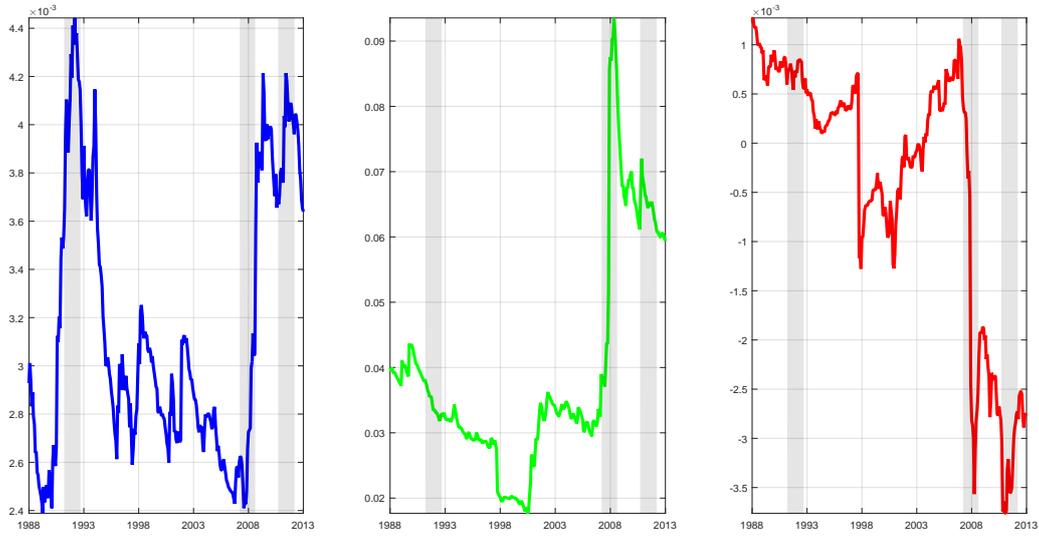
Note: The figure reports (i) in the left panel the estimated time-varying variance of the business cycle factor (ii) in the central panel the estimated time-varying variance of the financial stress factor (iii) in the right panel the estimated time-varying covariance between the business cycle and the financial stress factor. Shaded areas indicate recessions as dated by the CEPR.

Figure 7: ROOT MEAN SQUARE FORECAST ERRORS
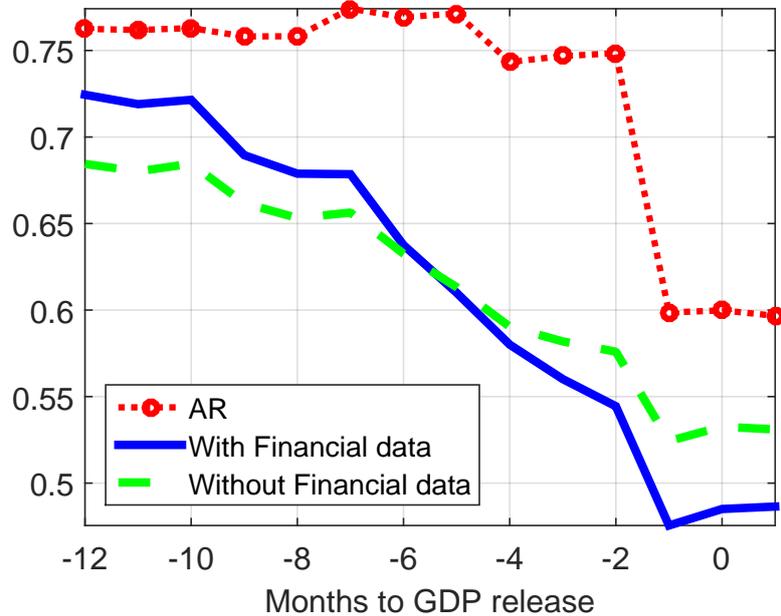
Note: The figure reports the Root Mean Square Forecast Errors (RMSFE) obtained with the Autoregressive model (AR, blue solid line) against the RMSFE attained by the dynamic factor model estimated by equal weight pooling and a restricted version of the factor model that excludes financial data.

Figure 8: CUMULATIVE SUM OF SQUARED ERRORS DIFFERENTIALS
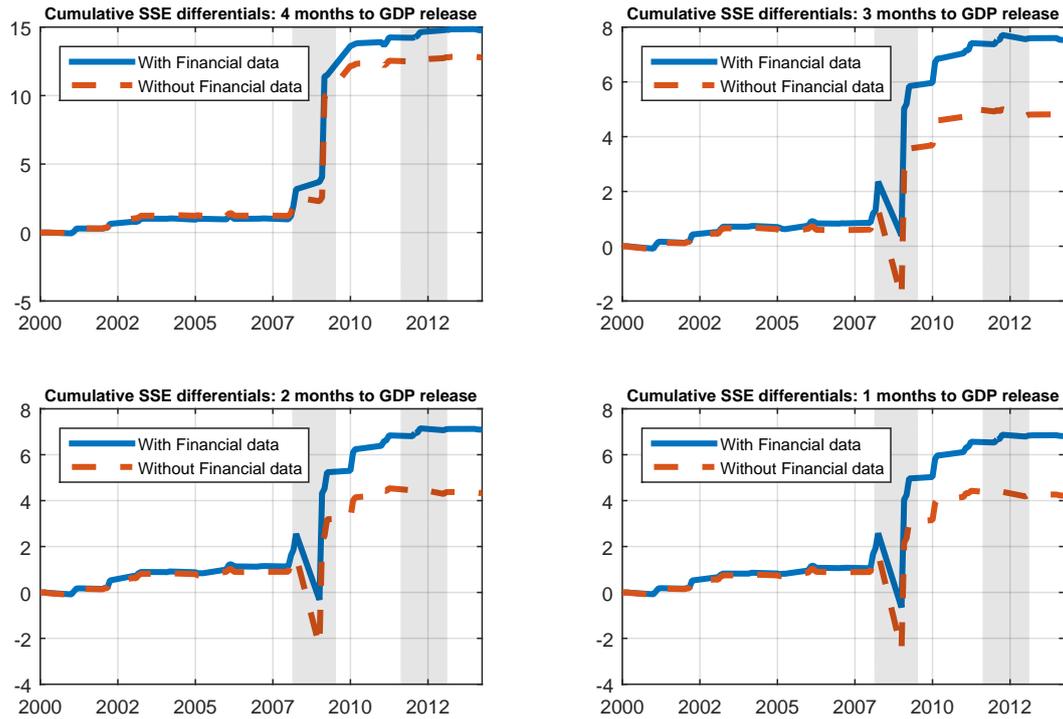
Note: The figure reports Cumulative Sum of Squared Errors Differentials between the dynamic factor model estimated by equal weight pooling and the Autoregressive model (blue solid line) and the restricted version of the factor model that excludes financial data and the Autoregressive model (green dashed line). Shaded areas indicate recessions as dated by the CEPR.

Figure 9: DENSITY FORECAST EVALUATION
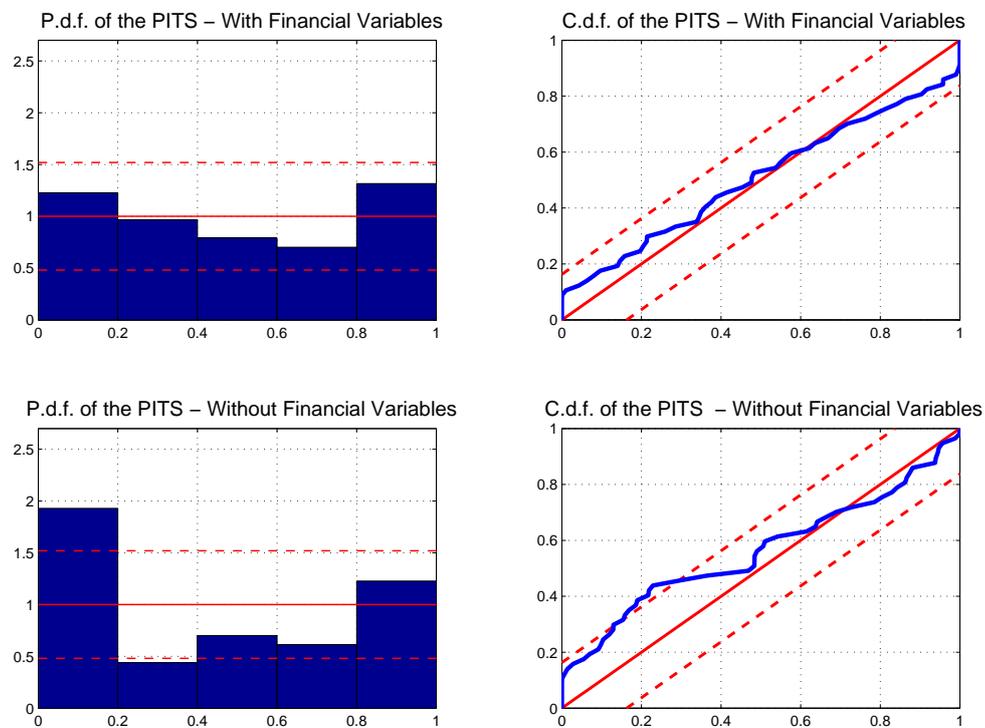
Note. In the left panels we report the p.d.f. of the PITs (normalized) and the 95% critical values (dashed lines) approximated by binomial distribution, constructed using a normal approximation. In the right panels, we show the c.d.f. of the PITs with critical values based on Rossi and Sekhposyan (2013).

A Proofs

We follow the notation and the main results on the matrix differential calculus in Abadir and Magnus (2005, ch 13)

A.1 Result 1

The gradient vector is

$$
\begin{aligned}
\nabla_t &= -\frac{1}{2}\left[\frac{\partial \log|F_t|}{\partial f_t'} + \frac{\partial v_t' F_t^{-1} v_t}{\partial f_t'}\right]' \\
&= -\frac{1}{2}\left[\frac{1}{|F_t|}\frac{\partial |F_t|}{\partial vec(F_t)'}\frac{\partial vec(F_t)}{\partial f_t'} + \frac{\partial v_t' F_t^{-1} v_t}{\partial v_t}\frac{\partial v_t}{\partial f_t'} + \frac{\partial v_t'}{F_t^{-1} v_t}\partial vec(F_t^{-1})'\frac{\partial vec(F_t^{-1})}{\partial vec(F_t)'}\frac{\partial vec(F_t)}{\partial f_t'}\right]' \\
&= -\frac{1}{2}[vec(F_t^{-1})'\dot{F}_t + 2v_t' F_t^{-1}\dot{V}_t - (v_t'\otimes v_t')(F_t^{-1}\otimes F_t^{-1})\dot{F}_t]' \\
&= -\frac{1}{2}\left[\dot{F}_t' vec(F_t^{-1}) - \dot{F}_t'(F_t^{-1}\otimes F_t^{-1})(v_t\otimes v_t) + 2\dot{V}_t' F_t^{-1} v_t\right] \\
&= \frac{1}{2}\left[\dot{F}_t'(F_t^{-1}\otimes F_t^{-1})(v_t\otimes v_t) - \dot{F}_t' vec(F_t^{-1} F_t F_t^{-1}) - 2\dot{V}_t' F_t^{-1} v_t\right] \\
&= \frac{1}{2}\left[\dot{F}_t'(F_t^{-1}\otimes F_t^{-1})(v_t\otimes v_t) - \dot{F}_t'(F_t^{-1}\otimes F_t^{-1})vec(F_t) - 2\dot{V}_t' F_t^{-1} v_t\right] \\
&= \frac{1}{2}\left[\dot{F}_t'(F_t^{-1}\otimes F_t^{-1})[v_t\otimes v_t - vec(F_t)] - 2\dot{V}_t' F_t^{-1} v_t\right].
\end{aligned}
\tag{31}
$$

We compute the information matrix as the expected value of the Hessian

$$
\mathcal{I}_t = -E_t\left[\frac{\partial^2 \ell_t}{\partial f_t \partial f_t'}\right].
\tag{32}
$$

Let re-write the Gradient (31) as follows

$$
\begin{aligned}
\nabla_t &= -\frac{1}{2}\left[\dot{F}_t'(F_t^{-1}\otimes F_t^{-1})[vec(F_t) - vec(v_t v_t')] + 2\dot{V}_t' F_t^{-1} v_t\right] \\
&= -\frac{1}{2}\left[\dot{F}_t'[(F_t^{-1}\otimes F_t^{-1})vec(F_t) - (F_t^{-1}\otimes F_t^{-1})vec(v_t v_t')] - 2\dot{V}_t' F_t^{-1} v_t\right] \\
&= -\frac{1}{2}\left[\dot{F}_t'[vec(F_t^{-1}) - vec(F_t^{-1} v_t v_t' F_t^{-1})] + 2\dot{V}_t' F_t^{-1} v_t\right] \\
&= -\frac{1}{2}\left[\dot{F}_t' vec(F_t^{-1} - F_t^{-1} v_t v_t' F_t^{-1}) + 2\dot{V}_t' F_t^{-1} v_t\right] \\
&= -\frac{1}{2}\left[\dot{F}_t' vec\{F_t^{-1}(I - v_t v_t' F_t^{-1})\} + 2\dot{V}_t' F_t^{-1} v_t\right] \\
&= -\frac{1}{2}\left[\dot{F}_t'(I\otimes F_t^{-1})vec(I - v_t v_t' F_t^{-1}) + 2\dot{V}_t' F_t^{-1} v_t\right].
\end{aligned}
\tag{33}
$$

The negative Hessian is made of two terms:

$$
-\frac{\partial^2 \ell_t}{\partial f_t \partial f_t'} = \underbrace{-\frac{1}{2}\frac{\partial \left[\overset{\bullet}{F}_t{}'(I \otimes F_t^{-1})vec(I - v_t v_t' F_t^{-1})\right]}{\partial f_t'}}_{\Phi_{1t}} + \underbrace{\frac{\partial \left(\overset{\bullet}{V}_t{}' F_t^{-1} v_t\right)}{\partial f_t'}}_{\Phi_{2t}}.
\tag{34}
$$

We first computing the first term of (34)

$$
\begin{aligned}
\Phi_{1t} &= \frac{1}{2}\frac{\partial \left[\overset{\bullet}{F}_t{}'(I \otimes F_t^{-1})vec(I - v_t v_t' F_t^{-1})\right]}{\partial vec(\overset{\bullet}{F}_t{}')'}\frac{\partial vec(\overset{\bullet}{F}_t{}')}{\partial f_t'} \\
&+ \frac{1}{2}\frac{\partial \left[\overset{\bullet}{F}_t{}'(I \otimes F_t^{-1})vec(I - v_t v_t' F_t^{-1})\right]}{\partial vec(I \otimes F_t^{-1})'}\frac{\partial vec(I \otimes F_t^{-1})}{\partial vec(F_t^{-1})'}\frac{\partial vec(F_t^{-1})}{\partial vec(F_t)'}\frac{\partial vec(F_t)}{\partial f_t'} \\
&+ \frac{1}{2}\frac{\partial \left[\overset{\bullet}{F}_t{}'(I \otimes F_t^{-1})vec(I - v_t v_t' F_t^{-1})\right]}{\partial vec(I - v_t v_t' F_t^{-1})'}\frac{\partial vec(I - v_t v_t' F_t^{-1})'}{\partial f_t'} \\
&= \frac{1}{2}\left[vec(I - v_t v_t' F_t^{-1})'(I \otimes F_t^{-1}) \otimes I\right]\overset{\#}{F}_t \\
&- \frac{1}{2}\left[vec(I - v_t v_t' F_t^{-1})' \otimes \overset{\bullet}{F}_t{}'\right]\frac{\partial vec(I \otimes F_t^{-1})}{\partial vec(F_t^{-1})'}(F_t^{-1} \otimes F_t^{-1})\overset{\bullet}{F}_t \\
&- \frac{1}{2}\overset{\bullet}{F}_t{}'(I \otimes F_t^{-1})\underbrace{\frac{\partial vec(v_t v_t' F_t^{-1})}{\partial f_t'}}_{\Gamma_t}
\end{aligned}
\tag{35}
$$

where $\overset{\#}{F}_t = \frac{\partial vec(\overset{\bullet}{F}_t{}')}{\partial f_t'}$ and

$$
\begin{aligned}
\Gamma_t &= \frac{\partial vec(v_t v_t' F_t^{-1})}{\partial vec(v_t v_t')'}\frac{\partial vec(v_t v_t')}{\partial v_t'}\frac{\partial v_t}{\partial f_t'} + \frac{\partial vec(v_t v_t' F_t^{-1})}{\partial vec(F_t^{-1})'}\frac{\partial vec(F_t^{-1})}{\partial vec(F_t)'}\frac{\partial vec(F_t)}{\partial f_t'} \\
&= \left(F_t^{-1} \otimes I\right)(v_t \otimes I + I \otimes v_t)\overset{\bullet}{V}_t - (I \otimes v_t v_t')(F_t^{-1} \otimes F_t^{-1})\overset{\bullet}{F}_t.
\end{aligned}
\tag{36}
$$

Putting together (35) and (36) we obtain

$$
\begin{aligned}
\Phi_{1t} &= \frac{1}{2}[vec(I - v_t v_t' F_t^{-1})'(I \otimes F_t^{-1}) \otimes I]\overset{\#}{F}_t \\
&= -\frac{1}{2}\left[vec(I - v_t v_t' F_t^{-1})' \otimes \overset{\bullet}{F}_t{}'\right]\frac{\partial vec(I \otimes F_t^{-1})}{\partial vec(F_t^{-1})'}(F_t^{-1} \otimes F_t^{-1})\overset{\bullet}{F}_t \\
&- \frac{1}{2}\overset{\bullet}{F}_t{}'(F_t^{-1} \otimes F_t^{-1})(v_t \otimes I + I \otimes v_t)\overset{\bullet}{V}_t + \frac{1}{2}\overset{\bullet}{F}_t{}'(F_t^{-1} \otimes F_t^{-1} v_t v_t' F_t^{-1})\overset{\bullet}{F}_t.
\end{aligned}
\tag{37}
$$

The second term of (34) is equal to

$$
\Phi_{2t} = \frac{\partial\left(\overset{\bullet}{V}{}'_t F_t^{-1} v_t\right)}{\partial vec(\overset{\bullet}{V}_t)'}\frac{\partial vec(\overset{\bullet}{V}_t)}{\partial f'_t} + \frac{\partial\left(\overset{\bullet}{V}{}'_t F_t^{-1} v_t\right)}{\partial vec(F_t^{-1})}\frac{\partial vec(F_t^{-1})}{\partial vec(F_t)'}\frac{\partial vec(F_t)}{\partial f'_t} + \frac{\partial\left(\overset{\bullet}{V}{}'_t F_t^{-1} v_t\right)}{\partial v'_t}\frac{\partial v_t}{\partial f'_t}
$$

$$
= (v'_t F_t^{-1}\otimes I)\overset{\#}{V}_t - (v_t\otimes \overset{\bullet}{V}_t)'(F_t^{-1}\otimes F_t^{-1})\overset{\bullet}{F}_t + \overset{\bullet}{V}{}'_t F_t^{-1}\overset{\bullet}{V}_t, \tag{38}
$$

where $\overset{\#}{V}_t = \frac{\partial vec(\overset{\bullet}{V}{}'_t)}{\partial f'_t}$.

Putting together (37) and (38) we obtain the following expression for (34)

$$
\begin{aligned}
-\frac{\partial^2\ell_t}{\partial f_t f'_t} &= \frac{1}{2}[vec(I - v_t v'_t F_t^{-1})'(I\otimes F_t^{-1})\otimes I]\overset{\#}{F}_t \\
&\quad -\frac{1}{2}\left[vec(I - v_t v'_t F_t^{-1})'\otimes \overset{\bullet}{F}{}'_t\right]\frac{\partial vec(I\otimes F_t^{-1})}{\partial vec(F_t^{-1})'}(F_t^{-1}\otimes F_t^{-1})\overset{\bullet}{F}_t \\
&\quad -\frac{1}{2}\overset{\bullet}{F}{}'_t(F_t^{-1}\otimes F_t^{-1})(v_t\otimes I + I\otimes v_t)\overset{\bullet}{V}_t \\
&\quad +\frac{1}{2}\overset{\bullet}{F}{}'_t\left(F_t^{-1}\otimes F_t^{-1}v_t v'_t F_t^{-1}\right)\overset{\bullet}{F}_t \\
&\quad +(v'_t F_t^{-1}\otimes I)\overset{\#}{V}_t \\
&\quad -(v_t\otimes \overset{\bullet}{V}_t)'(F_t^{-1}\otimes F_t^{-1})\overset{\bullet}{F}_t \\
&\quad +\overset{\bullet}{V}{}'_t F_t^{-1}\overset{\bullet}{V}_t. \tag{39}
\end{aligned}
$$

Following Harvey (1989, p.141), taking the conditional expectation of (39) the only random element is the prediction error v_t and its first and second derivatives are $\overset{\bullet}{V}_t$ and $\overset{\#}{V}_t$. However, such derivatives are fixed given the conditional expectation at time $t-1$. Moreover, we have that $E_t(v_t) = 0$, $E_t(v_t v'_t) = F_t$ and $E_t[vec(I - v_t v'_t F_t^{-1})] = 0$. Therefore, applying the expectations $E_t(.)$, the fourth and the seventh term in (39) are the only nonzero elements and the information matrix is equal to

$$
\mathcal{I}_t = \frac{1}{2}\overset{\bullet}{F}{}'_t\left(F_t^{-1}\otimes F_t^{-1}\right)\overset{\bullet}{F}_t + \overset{\bullet}{V}{}'_t F_t^{-1}\overset{\bullet}{V}_t. \tag{40}
$$

A.2 Result 2

Given the model-specific Jacobian matrices

$$
\underset{Nm\times k}{\overset{\bullet}{Z}_t} = \frac{\partial vec(Z_t)}{\partial f'_t}, \quad \underset{N^2\times k}{\overset{\bullet}{H}_t} = \frac{\partial vec(H_t)}{\partial f'_t}, \quad \underset{m^2\times k}{\overset{\bullet}{T}_t} = \frac{\partial vec(T_t)}{\partial f'_t}, \quad \underset{m^2\times k}{\overset{\bullet}{Q}_t} = \frac{\partial vec(Q_t)}{\partial f'_t}.
$$

We show how compute the following Jacobian matrices

$$
\begin{aligned}
\underset{N\times k}{\overset{\bullet}{V}_t} &= -\left[\frac{\partial Z_t a_t}{\partial vec(Z_t)'}\frac{\partial vec(Z_t)}{\partial f'_t} + \frac{\partial Z_t a_t}{\partial a'_t}\frac{\partial a_t}{\partial f'_t}\right] \\
&= -[(a'_t\otimes I_N)\overset{\bullet}{Z}_t + Z_t\overset{\bullet}{A}_t], \tag{41}
\end{aligned}
$$

$$\underset{N^2 \times k}{\overset{\bullet}{F}_t} = \frac{\partial vec(Z_t P_t Z_t')}{\partial vec(Z_t)'}\frac{\partial vec(Z_t)}{\partial f_t'} + \frac{\partial vec(Z_t P_t Z_t')}{\partial vec(P_t)'}\frac{\partial vec(P_t)}{\partial f_t'} + \overset{\bullet}{H}_t$$

$$= (I_{N^2} + C_{N^2})(Z_t P_t \otimes I_N)\overset{\bullet}{Z}_t + (Z_t \otimes Z_t)\overset{\bullet}{P}_t + \overset{\bullet}{H}_t$$

$$= 2N_N(Z_t P_t \otimes I_N)\overset{\bullet}{Z}_t + (Z_t \otimes Z_t)\overset{\bullet}{P}_t + \overset{\bullet}{H}_t \qquad (42)$$

$$\underset{mN \times k}{\overset{\bullet}{K}_t} = \frac{\partial vec(T_t P_t Z_t' F_t^{-1})}{\partial vec(T_t)'}\frac{\partial vec(T_t)}{\partial f_t'} + \frac{\partial vec(T_t P_t Z_t' F_t^{-1})}{\partial vec(P_t)'}\frac{\partial vec(P_t)}{\partial f_t'}$$

$$+ \frac{\partial vec(T_t P_t Z_t' F_t^{-1})}{\partial vec(Z_t')'}\frac{\partial vec(Z_t')}{\partial vec(Z_t)'}\frac{\partial vec(Z_t)}{\partial f_t'} + \frac{\partial vec(T_t P_t Z_t' F_t^{-1})}{\partial vec(F_t^{-1})'}\frac{\partial vec(F_t^{-1})}{\partial vec(F_t)'}\frac{\partial vec(F_t)}{\partial f_t'}$$

$$= (F_t^{-1} Z_t P_t \otimes I_m)\overset{\bullet}{T}_t + (F_t^{-1} Z_t \otimes T_t)\overset{\bullet}{P}_t + (F_t^{-1} \otimes T_t P_t)C_{Nm}\overset{\bullet}{Z}_t - (I_N \otimes T_t P_t Z_t')(F_t^{-1} \otimes F_t^{-1})\overset{\bullet}{F}_t$$

$$= (F_t^{-1} Z_t P_t \otimes I_m)\overset{\bullet}{T}_t + (F_t^{-1} Z_t \otimes T_t)\overset{\bullet}{P}_t + (F_t^{-1} \otimes T_t P_t)C_{Nm}\overset{\bullet}{Z}_t - (F_t^{-1} \otimes K_t)\overset{\bullet}{F}_t. \qquad (43)$$

$$\underset{m \times k}{\overset{\bullet}{A}_{t+1}} = \frac{\partial T_t a_t}{\partial vec(T_t)'}\frac{\partial vec(T_t)}{\partial f_t'} + \frac{\partial T_t a_t}{\partial a_t'}\frac{\partial a_t}{\partial f_t'} + \frac{\partial K_t v_t}{\partial vec(K_t)'}\frac{\partial vec(K_t)}{\partial f_t'} + \frac{\partial K_t v_t}{\partial v_t'}\frac{\partial v_t}{\partial f_t'}$$

$$= (a_t' \otimes I_m)\overset{\bullet}{T}_t + T_t \overset{\bullet}{A}_t + (v_t' \otimes I_m)\overset{\bullet}{K}_t + K_t \overset{\bullet}{V}_t \qquad (44)$$

$$\underset{m^2 \times k}{\overset{\bullet}{P}_{t+1}} = \frac{\partial vec(T_t P_t T_t')}{\partial vec(T_t)'}\frac{\partial vec(T_t)}{\partial f_t'} + \frac{\partial vec(T_t P_t T_t')}{\partial vec(P_t)'}\frac{\partial vec(P_t)}{\partial f_t'}$$

$$- \frac{\partial vec(K_t F_t K_t')}{\partial vec(K_t)'}\frac{\partial vec(K_t)}{\partial f_t'} - \frac{\partial vec(K_t F_t K_t')}{\partial vec(F_t)'}\frac{\partial vec(F_t)}{\partial f_t'} + \overset{\bullet}{Q}_t$$

$$= (I_{m^2} + C_m)(T_t P_t \otimes I_m)\overset{\bullet}{T}_t + (T_t \otimes T_t)\overset{\bullet}{P}_t - (I_{m^2} + C_m)(K_t F_t \otimes I_m)\overset{\bullet}{K}_t - (K_t \otimes K_t)\overset{\bullet}{F}_t + \overset{\bullet}{Q}_t$$

$$= 2N_m[(T_t P_t \otimes I_m)\overset{\bullet}{T}_t - (K_t F_t \otimes I_m)\overset{\bullet}{K}_t] + (T_t \otimes T_t)\overset{\bullet}{P}_t - (K_t \otimes K_t)\overset{\bullet}{F}_t + \overset{\bullet}{Q}_t \qquad (45)$$

An alternative expression for $\overset{\bullet}{P}_{t+1}$ can be obtain as follows:

$$\underset{m^2 \times k}{\overset{\bullet}{P}_{t+1}} = \frac{\partial vec(T_t P_t T_t')}{\partial vec(T_t)'}\frac{\partial vec(T_t)}{\partial f_t'} + \frac{\partial vec(T_t P_t T_t')}{\partial vec(P_t)'}\frac{\partial vec(P_t)}{\partial f_t'} - \frac{\partial vec(K_t Z_t P_t T_t')}{\partial vec(K_t)'}\frac{\partial vec(K_t)}{\partial f_t'}$$

$$- \frac{\partial vec(K_t Z_t P_t T_t')}{\partial vec(Z_t)'}\frac{\partial vec(Z_t)}{\partial f_t'} - \frac{\partial vec(K_t Z_t P_t T_t')}{\partial vec(P_t)'}\frac{\partial vec(P_t)}{\partial f_t'} - \frac{\partial vec(K_t Z_t P_t T_t')}{\partial vec(T_t)'}\frac{\partial vec(T_t)}{\partial f_t'} + \overset{\bullet}{Q}_t$$

$$= [2N_m(T_t P_t \otimes I_m) - (I_m \otimes K_t Z_t P_t)]\overset{\bullet}{T}_t +$$

$$+ (T_t \otimes L_t)\overset{\bullet}{P}_t - (T_t P_t Z_t' \otimes I_m)\overset{\bullet}{K}_t - (T_t P_t \otimes K_t)\overset{\bullet}{Z}_t + \overset{\bullet}{Q}_t \qquad (46)$$

B   Extensions

B.1   Vector autoregressive model

It is straightforward to generalize the results in Section 2.4.2 to the case of the VAR of order one:

$$y_{t+1} = \Phi_t y_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \Sigma_t).$$

The SSF representation (1) of the above model is

$$\alpha_t = y_t, \quad Z_t = I, \quad T_t = \Phi_t, \quad Q_t = \Sigma_t.$$

The TVP vector is $f_t = (vec(\Phi_t)', vec(\Sigma_t)')'$, and after some algebra we obtain:

$$s_t = \mathcal{I}_t^{-1} \nabla_t = \begin{bmatrix} (X_t' \Sigma_t^{-1} X_t)^{-1} X_t' \Sigma_t^{-1} \epsilon_t \\ vec(\epsilon_t \epsilon_t') - vec(\Sigma_t) \end{bmatrix}$$

$X_t = (y_{t-1}' \otimes I)$. The general VAR of order p is obtained with $X_t = [(y_{t-1}', ..., y_{t-p}') \otimes I]$.

Given the score computed above, we can obtain exactly the algorithm proposed by Koop and Korobilis (2013) by the following restrictions: (i) in the law of motion (4), $\Phi = I$ and Ω depends on two scalar parameters driving the coefficients and volatility, respectively; (ii) the scaling matrix $X_t' \Sigma_t^{-1} X_t$ is replaced by its smoothed estimator $\mathcal{S}_t = (1 - \lambda)\mathcal{S}_{t-1} + \lambda(X_t' \Sigma_t^{-1} X_t)$.

B.2   Mixed frequencies and temporal aggregation

Let consider the high frequency variable x_t, which is unobserved, and the corresponding observed low frequency series, x_t^k with $k > 1$. The relation between the observed low frequency variable and the corresponding indicator depends on whether the variable is a flow or a stock variable and on how the variable is transformed before entering the model. In all cases the variable can be rewritten as a weighted average of the unobserved high frequency indicator, specifically

$$x_t = \sum_{j=0}^{2k-2} \omega_j^k x_{t-j}^k \tag{47}$$

Here a summary of the implied weights (see e.g. Bańbura et al. (2013)). If the variable enters:

- in *level* and is a *stock*: $\omega_0^k = 1$, and $\omega_j^k = 0$ for $j > 0$;

- in *level* and is a *flow*: $\omega_j^k = 1$ for $j = 0, ..., k-1$, and $\omega_j^k = 0$ for $j \geq k$;

- in *first diff.* and is a *stock* $\omega_j^k = 1$ for $j = 0, ..., k-1$, and $\omega_j^k = 0$ for $j \geq k$;

- in *first diff.* and is a *flow* $\omega_j^k = j + 1$ for $j = 0, ..., k-1$, and $\omega_j^k = 2k - j - 1$ for $j \geq k$.

Let consider a single indicator (e.g., the GDP) to be aggregated from quarterly to monthly (i.e. $k = 3$); assuming that the unobserved monthly variable follows the state space model

$$y_{i,t} = Z_{i,t}\alpha_t + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}\left(0, \sigma_{i,t}^2\right), \tag{48}$$

at quarterly frequency we would observe

$$y_{i,t}^q = \sum_{j=0}^{4} \omega_j^3 Z_{i,t-j}\alpha_{t-j} + \sum_{j=0}^{4} \omega_j^4 \varepsilon_{i,t-j}^4. \tag{49}$$

Thus, the state space models needs to be accommodated taking into account the aggregation (49) and the implied missing observations.[29]

C   State space representation and estimation of the Financial Stress model

Let first discuss the state space representation of the model (28)-(29) taking into account the mixed frequency and aggregation. For the macro variables y_t^r, we have three types of indicators: the quarterly GDP y_t^q, two monthly business cycle surveys indicators in the vector y_t^s, and the industrial production y_t^i. For the financial variables, we have 15 monthly indicators in the vector y_t^x. Therefore, the vector of observables $y_t = (y_t^q, y_t^{s\prime}, y_t^i, y_t^{x\prime})'$ contains 19 variables. The state vector is

$$\alpha_t = \left(1, 1, \alpha_t^y, \alpha_t^x, \alpha_{t-1}^y, \alpha_{t-1}^x, \alpha_{t-2}^y, \alpha_{t-3}^y, \alpha_{t-4}^y, \varepsilon_t^q, \varepsilon_{t-1}^q, \varepsilon_{t-2}^q, \varepsilon_{t-3}^q, \varepsilon_{t-4}^q\right)'$$

where α_t^y is the real factor α_t^x is the financial factor and ε_t^q is the measurement error of the quarterly macro variable (i.e., GDP). Due to time aggregation we have the moving average terms. The 1 appearing in the first two positions of the state vector allow us to include score-driven breaking intercepts in the transition equations. The corresponding time-varying matrix of factor loading, consistent with the time aggregation, is equal to

$$Z_t = \begin{bmatrix} 0 & 0 & \frac{1}{3}\lambda_t^q & 0 & \frac{2}{3}\lambda_{t-1}^q & 0 & \lambda_{t-2}^q & \frac{2}{3}\lambda_{t-3}^q & \frac{1}{3}\lambda_{t-4}^q & \frac{1}{3} & \frac{2}{3} & 1 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{3}\lambda_t^s & 0 & \frac{1}{3}\lambda_{t-1}^s & 0 & \frac{1}{3}\lambda_{t-2}^s & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_t^m & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_t^x & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

where λ_t^q is scalar, λ_t^s is 2×1 vector, λ_t^m is a scalar, and λ_t^x is 15×1 vector. Note that we impose restrictions on the factor loadings such that the one relative to the first macro

---

[29]Note that although the aggregation implies that the measurement error now follows a moving average process of order related to the dimension of the high frequency, it nevertheless remains white noise when observed at the low frequency frequency (see Aruoba et al., 2011). Hence we treat the measurement error at the low frequency as white noise in what follows.

variable (the GDP) and the one relative to the first financial variable are both normalized to one. The measurement errors are $\varepsilon_t = (0, \varepsilon_t^{s\prime}, \varepsilon_t^i, \varepsilon_t^{x\prime})' \sim \mathcal{N}(0, H_t)$, where H_t is diagonal. We impose positive volatility on non-zero diagonal elements of H_t by log transformation element by element.

Consistently with the previous discussion, the elements of the transition equation are:

$$
T_t = \begin{bmatrix} T_{11,t} & T_{12,t} & T_{13,t} \\ T_{21,t} & T_{22,t} & T_{23,t} \\ T_{31,t} & T_{32,t} & T_{33,t} \end{bmatrix}, \quad T_{21,t} = \begin{bmatrix} \gamma_t^r & 0 \\ 0 & \gamma_t^x \\ 0 & 0 \end{bmatrix} \quad T_{22,t} = \begin{bmatrix} \phi_{11,t} & \phi_{12,t} & \phi_{13,t} & \phi_{14,t} \\ 0 & \phi_{22,t} & 0 & \phi_{24,t} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},
$$

all the other T_{ij} are sub-matrices of 0s and 1s, and $Q_t = blockdiag \left[0_{2\times2}, \Sigma_t, 0_{5\times5}, \sigma_{qt}^2, 0_{4\times4} \right]$. In order to impose stable roots in the sub-matrix $T_{22,t}$, we use the transformation described early on for the two pairs

$$
\phi_{r,t} = (\phi_{11,t}, \phi_{13,t})' \in S^2 \quad \phi_{x,t} = (\phi_{22,t}, \phi_{24,t})' \in S^2
$$

where S^2 is the space with stable roots, while $\phi_{12,t}$ and $\phi_{14,t}$ are left unrestricted. Finally, we restrict the matrix Σ_t to be positive definite using the log-Cholesky transformation. Collecting all the time varying parameters in the vector f_t we specialize their score driven law of motion as

$$
f_{t+1} = f_t + \Omega s_t
$$

The matrix Ω is restricted to be diagonal and to depend on two constants, κ_s and κ_v, where the former drives the amount of time variation in the factor loadings and in the AR coefficients, while κ_v is a smoothing parameter that governs the time variation of the model volatilities.

In the financial stress application, given the irregular behaviour of the financial variables over the considered sample, we find that the maximization of the likelihood is particularly sensitive to the starting parameter values. Hence, rather than using maximum likelihood methods we set up a grid of plausible values for the smoothing parameters κ_s and κ_v, and estimate the model using model combination strategies, namely equal weights averaging, Dynamic Model Selection (DMS) and Dynamic Model Averaging (DMA), see Appendix C for further details. The grid that we use implies the following values for the smoothing parameters: $[0.006, 0.009, 0.012, 0.015]$, which are broadly in line with those used by Koop and Korobilis (2013). Since the different estimation methods turn out to deliver broadly similar result, unless explicitly stated, we refer to the model estimated with equal weights.

Estimation via model averaging and selection proceeds as follows. We specify a grid of values for the parameters κ_s and κ_v. Each point in this grid defines a new model. Weights for each model j (defined $\pi_{t|t-1,j}$) are obtained as a function of the predictive density at time $t-1$

through the following recursions:

$$\pi_{t|t-1,j} = \frac{\pi_{t-1|t-1,j}^{\alpha}}{\sum_{l=1}^{J} \pi_{t-1|t-1,l}^{\alpha}} \tag{50}$$

$$\pi_{t|t,j} = \frac{\pi_{t|t-1,j} p_j(y_t|Y_{t-1})}{\sum_{l=1}^{J} \pi_{t|t-1,l} p_l(y_t|Y_{t-1})} \tag{51}$$

where $p_j(y_t|Y_{t-1})$ is the predictive likelihood of model j. Since this is a function of the prediction errors and of the prediction errors variance, which are part of the output of the KF, the model weights can be computed at no cost along with the model parameters estimation. Note that here a new forgetting factor appears, α, which discounts past predictive likelihoods. We set this parameter to 0.95. At each point in time, forecast are obtained on the basis of the model with the highest weight $\pi_{t|t-1,j}$ or by averaging on the basis of the model weights, see also Koop and Korobilis (2013). In Figures C.1 and C.2 we report a time series plot of the weights assigned by the Dynamic Model Averaging algorithm and the indicator of the best model as selected by Dynamic Model Selection. It can be noticed that the model achieves model sparsity, in the sense that, after a necessary period of learning, it assigns nontrivial weights only to a limited number of models. Figure C.2, on the other hand, reveals that the highest weights are typically assigned to models in the upper part of grid, where time variation is relatively more pronounced.

Figure C.1: MODEL WEIGHTS (DYNAMIC MODEL AVERAGING)
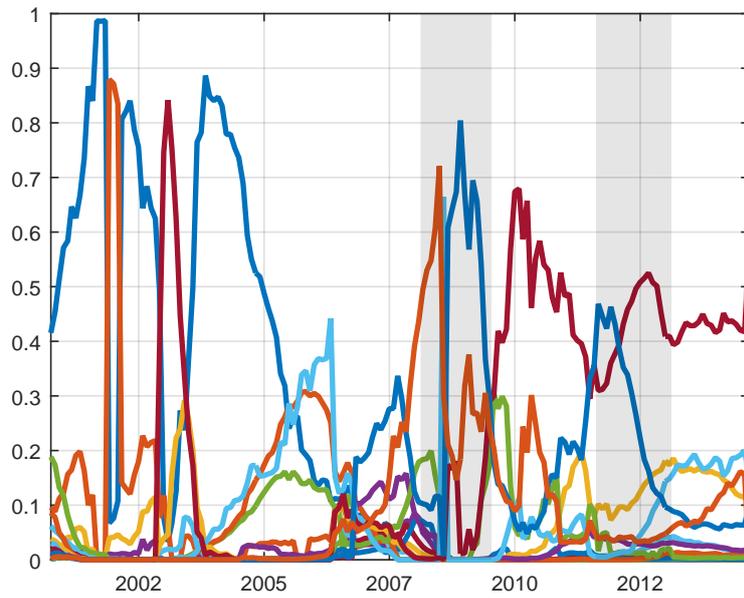
Note. The figure reports the Dynamic Weights computed as in equation 51 for the 16 models estimated using a 4 points grid for the static parameters κ_s and κ_v.

Figure C.2: BEST MODEL (DYNAMIC MODEL SELECTION)
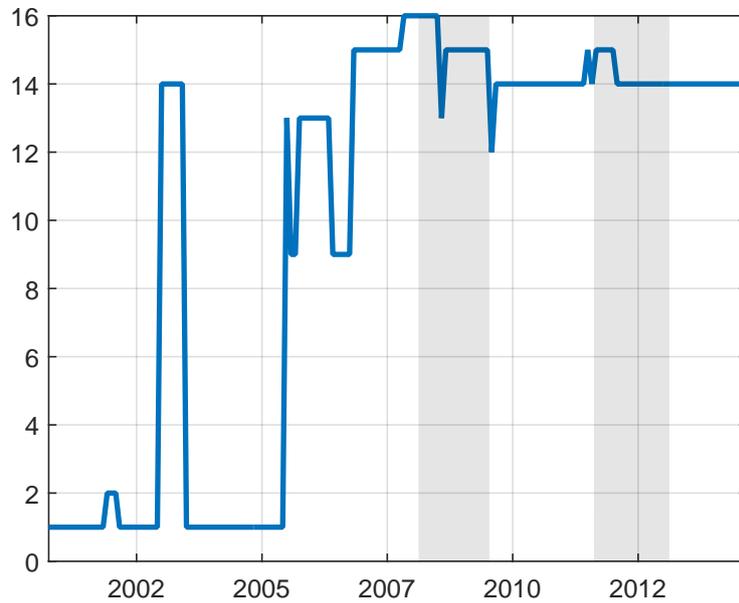
Note. The figure reports the index between 1 and 16 of the model that attains the highest Dynamic Weight, out of the 16 models estimated using a 4 points grid for the static parameters κ_s and κ_v.

D Monte Carlo exercise

## D.1 Calibration

DGP1:Time-varying loadings
CONSTANT: $a_1 = 1$;
SINE: $a_2 = 2$, $b_2 = 1.5$;
SINGLE STEP: $a_3 = 1$, $b_3 = 2$, $\tau = (2/5)n$;
DOUBLE STEP: $a_4 = 1$, $b_4 = c_4 = 1.5$, $\tau_1 = (1/5)n$, $\tau_2 = (3/5)n$;
RAMP: $a_5 = 0.5$, $b_5 = 1.5$, $c_5 = 3$;
MODEL: $a_6 = 1$, $b_6 = 0.99$, $c_6 = 0.05^2$.

DGP2: Time-varying autoregressive coefficient
CONSTANT: $a_1 = 0.7$;
SINE: $a_2 = 0$, $b_2 = 0.7$;
SINGLE STEP $a_3 = 0.8$, $b_3 = -0.6$, $\tau = (2/5)n$;
DOUBLE STEP:$a_4 = 0.8$, $b_4 = c_4 = -0.5$, $\tau_1 = (1/5)n$, $\tau_2 = (3/5)n$;
RAMP: $a_5 = 0.3$, $b_5 = 0.6$. $c_5 = 2$;
MODEL $a_6 = 0.2$, $b_6 = 0.99$, $c_6 = 0.05^2$;
with the restriction that $|\rho_t| < 1$.

DGP3 and DGP4: Time-varying volatilities
CONSTANT: $a_1 = 1$;
SINE: $a_2 = 7$, $b_2 = 5$;
SINGLE STEP: $a_3 = 1$, $b_3 = -4$, $\tau = (2/5)n$;
DOUBLE STEP:$a_4 = 1$, $b_4 = c_4 = -3$, $\tau_1 = (1/5)n$, $\tau_2 = (3/5)n$;
RAMP: $a_5 = 0.5$, $b_5 = 4.5$, $c_5 = 3$;
MODEL $a_6 = 0$, $b_6 = 0.99$, $c_6 = 0.05^2$.

In DGP3 and DGP4, after having simulated the dynamic of the volatility the time-varying volatilities are rescaled so as to have a fixed ratio between the measurement and transition error variances equal to 1.

In figures D.1 - D.8 we report the simulated true process for the time-varying parameters (black line), 16% and 88% (green dotted lines) and 5% and 95% (red broken lines) quantiles of the filtered parameters. In the case of the AR(1) specification we report the difference between actual and estimated parameters. The figures are based on 300 replications.

D.2 Actual and estimated parameters

Figure D.1: TIME VARYING LOADINGS, N=250
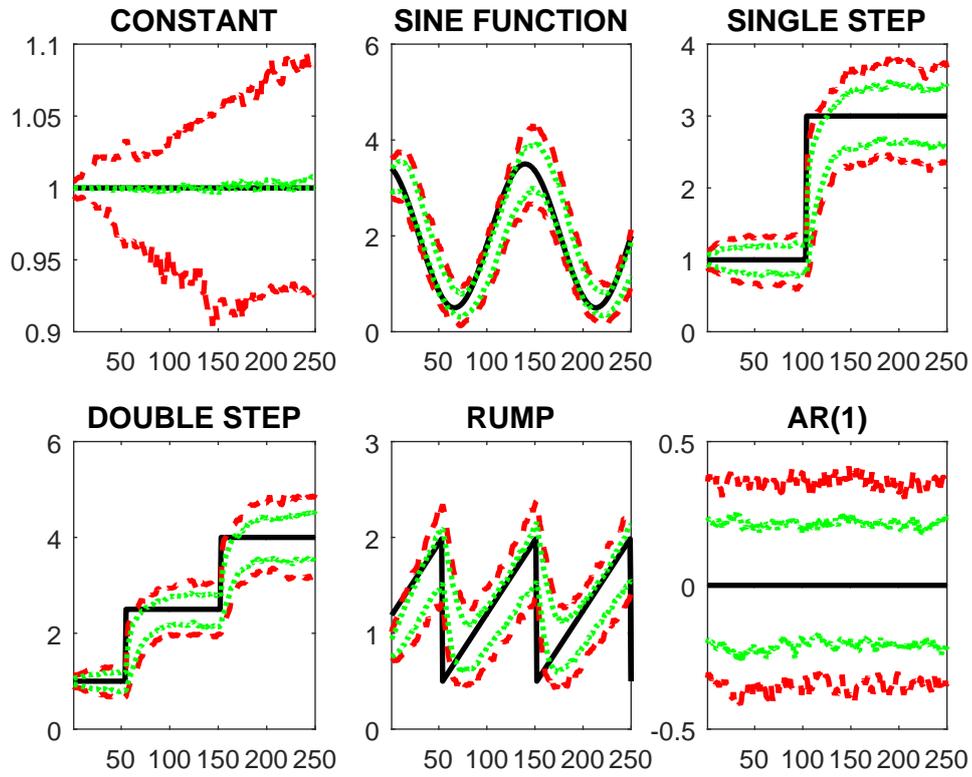
Figure D.2: TIME VARYING LOADINGS, N=500
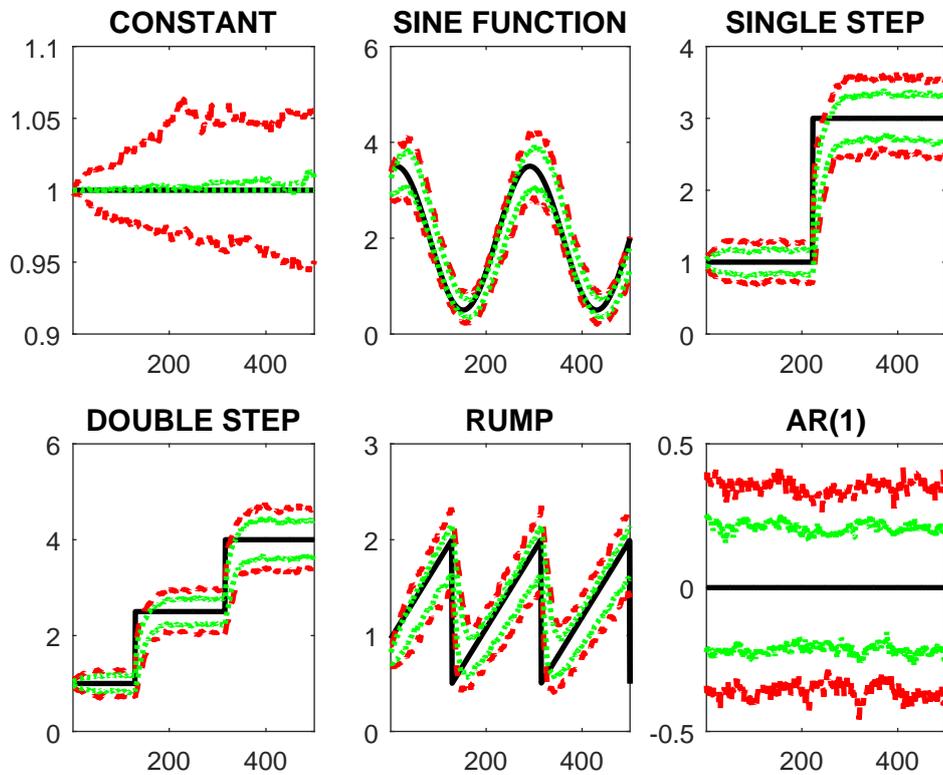
Figure D.3: TIME VARYING AUTOREGRESSIVE COEFFICIENTS, n=250
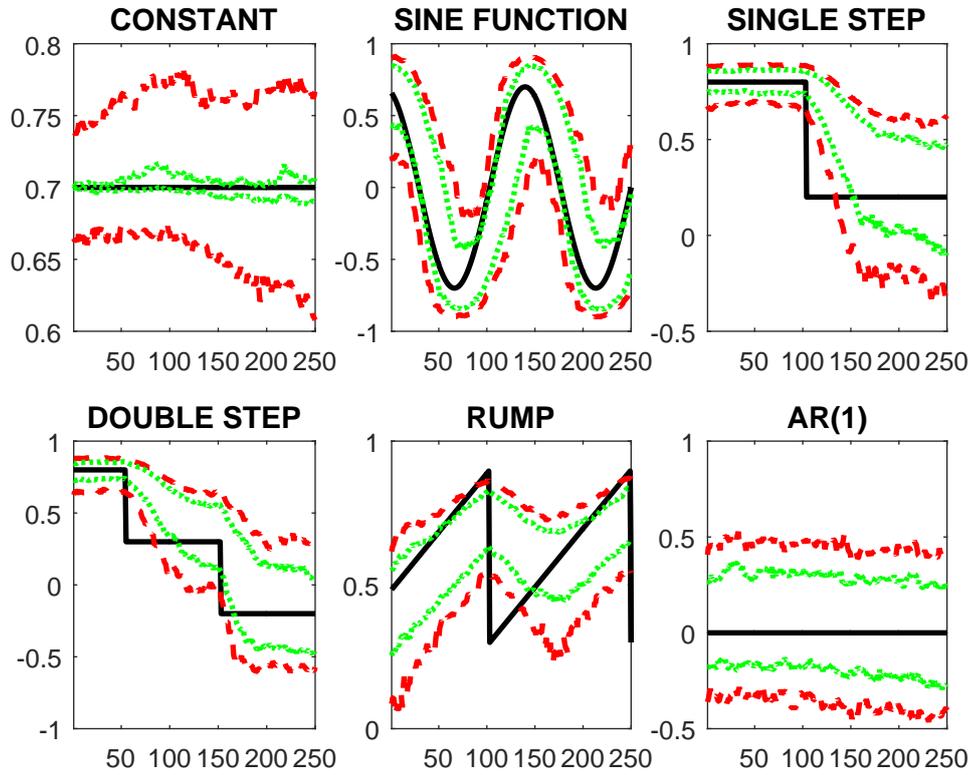
Figure D.4: TIME VARYING AUTOREGRESSIVE COEFFICIENTS, n=500
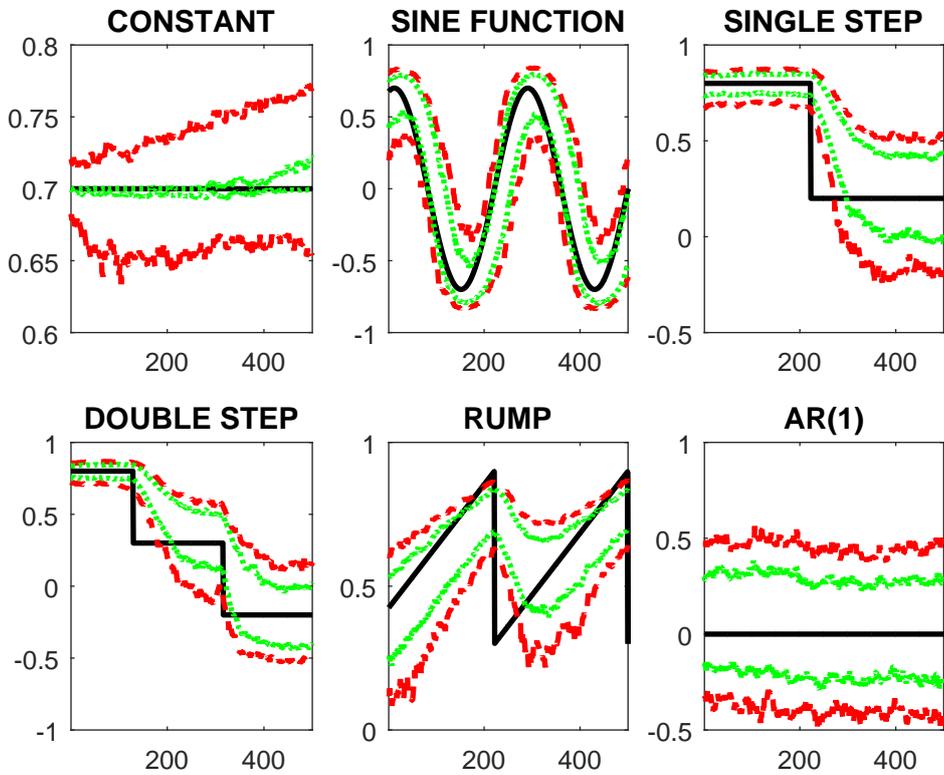
Figure D.5: TIME VARYING MEASUREMENT EQUATION ERROR VARIANCE, n=250
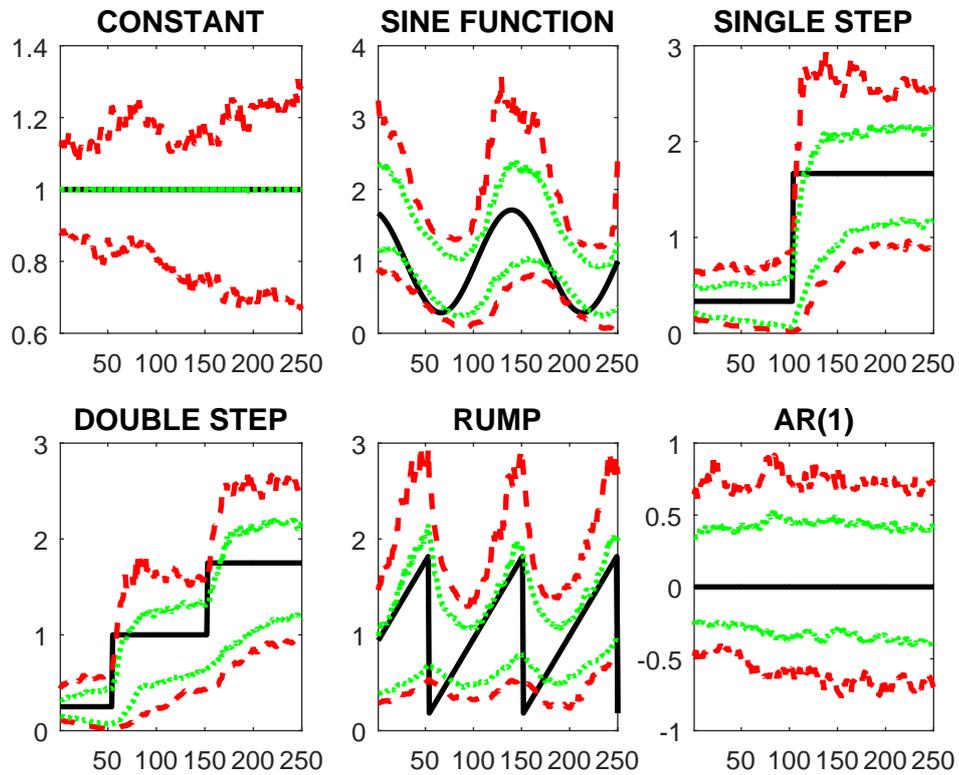
Figure D.6: TIME VARYING MEASUREMENT EQUATION ERROR VARIANCE, n=500
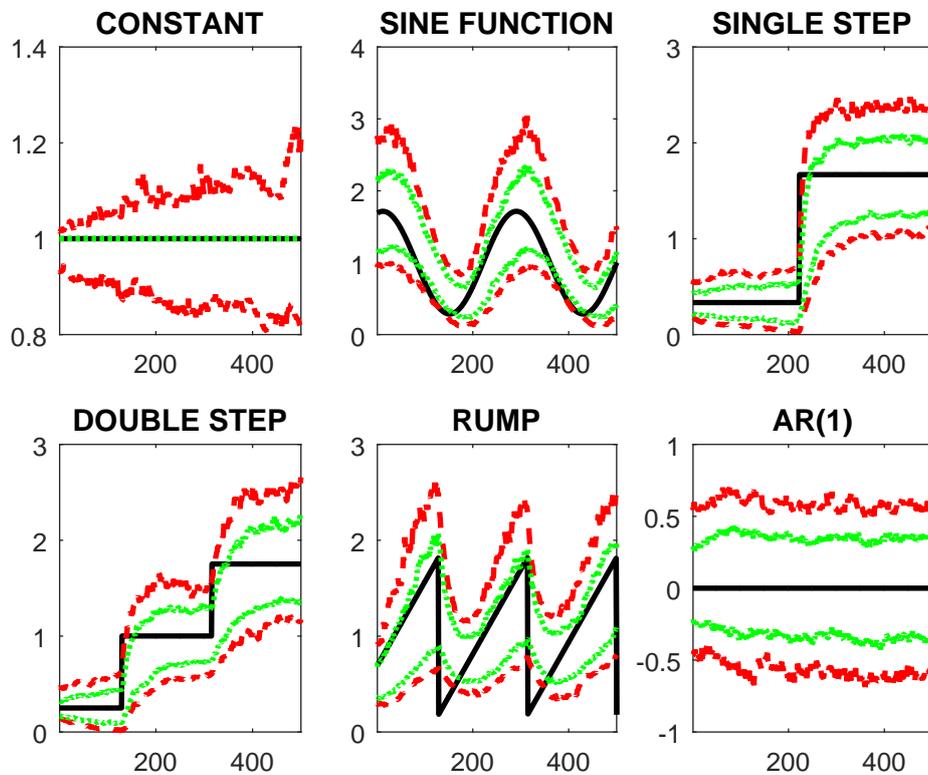
Figure D.7: TIME VARYING TRANSITION EQUATION ERROR VARIANCE, n=250
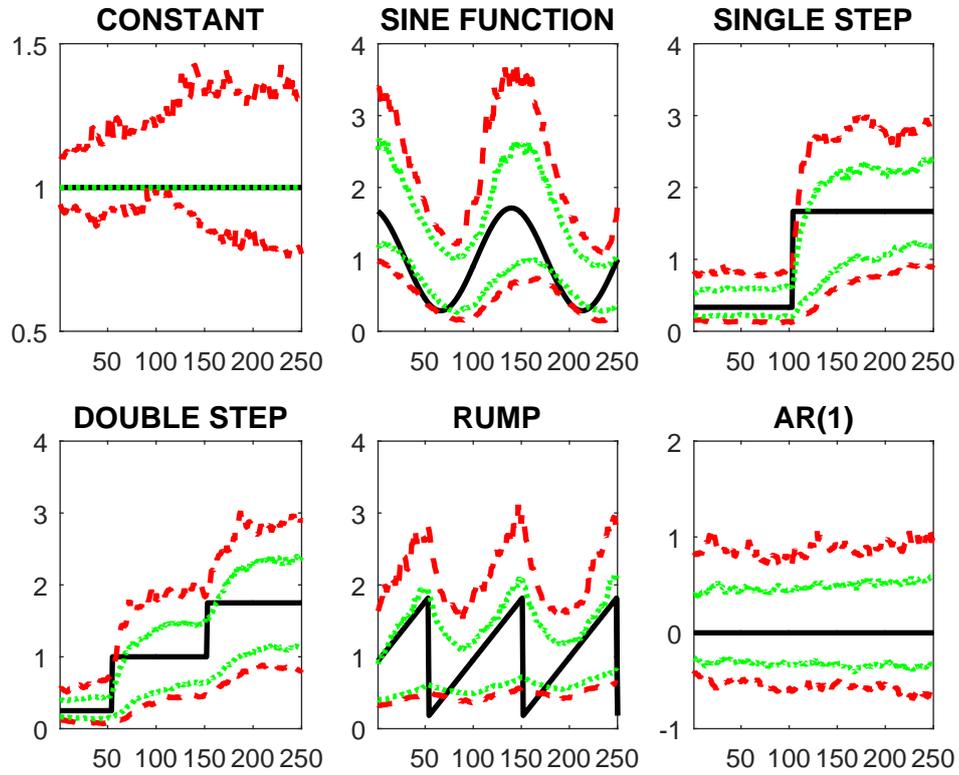
Figure D.8: TIME VARYING TRANSITION EQUATION ERROR VARIANCE, n=500
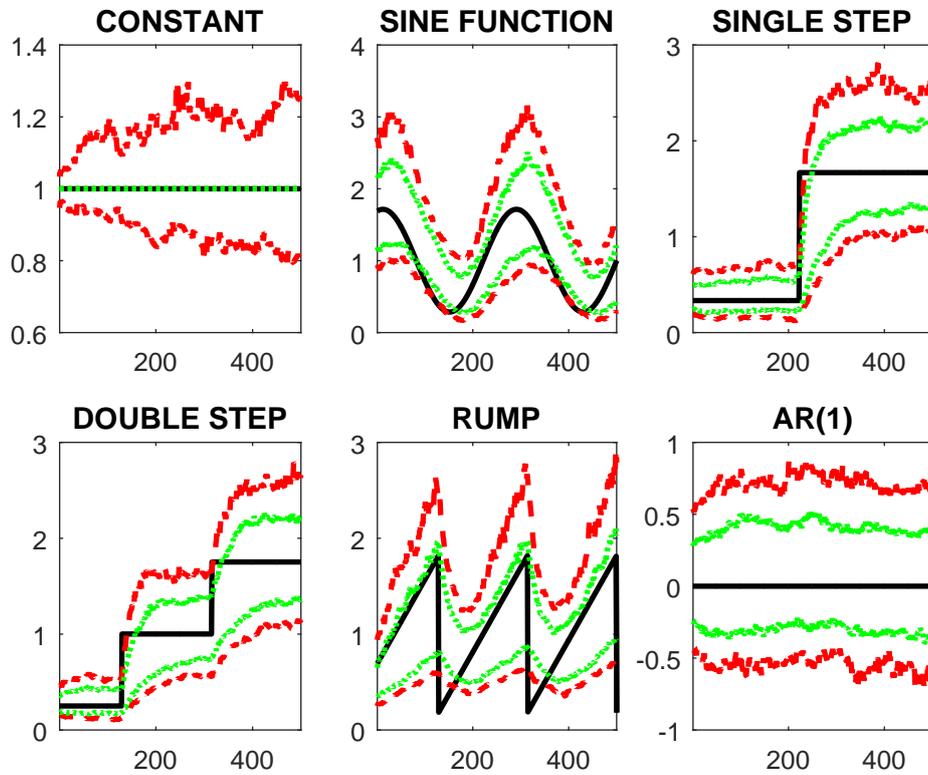
E Additional figures

Figure E.1: GDP(E) vs. GDP(I)
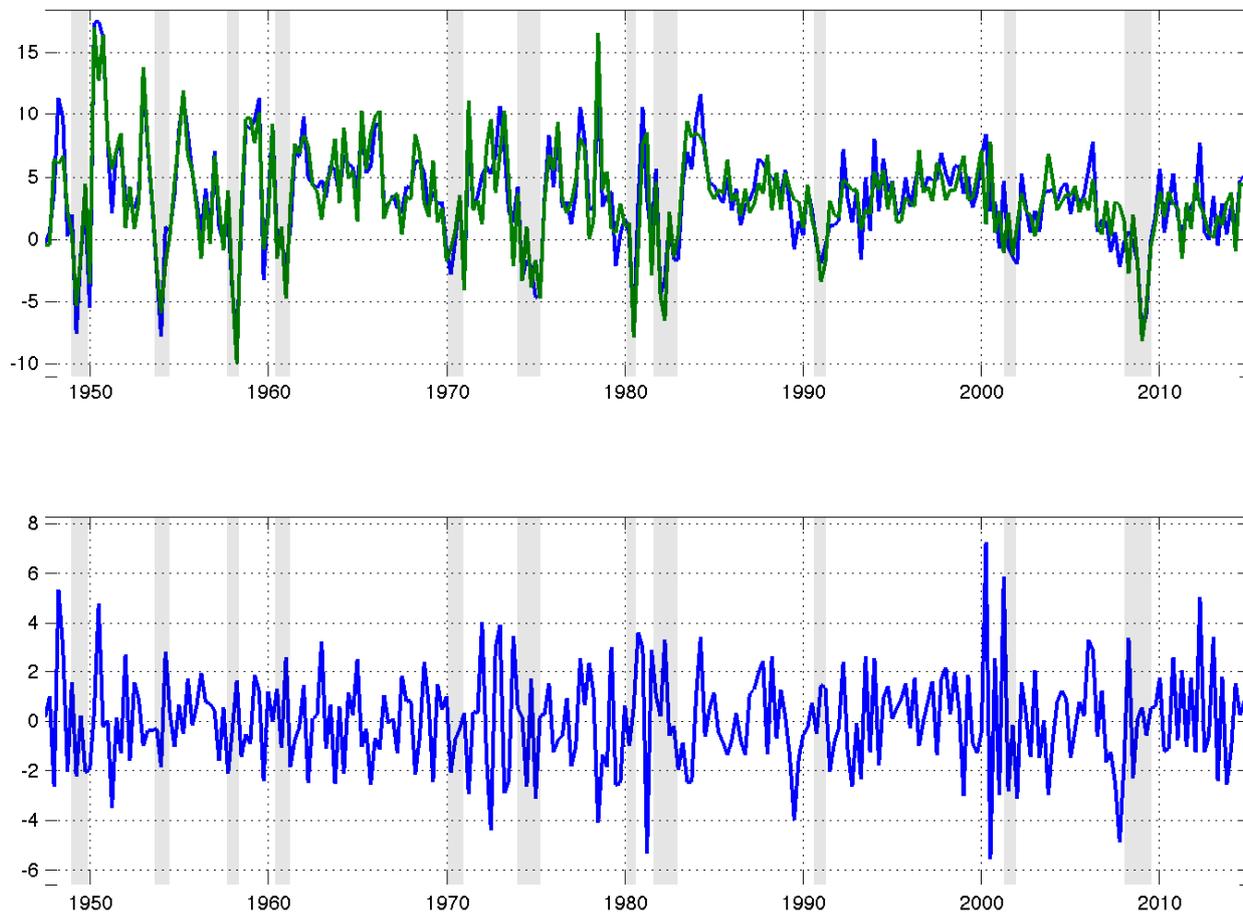
Note: Upper panel: GDP(E) (green line) and GDP(I) (blue line). Lower panel: Difference between the two GDP measures in annualized percentage rate.

Figure E.2: Zooming in the 1970s recessions
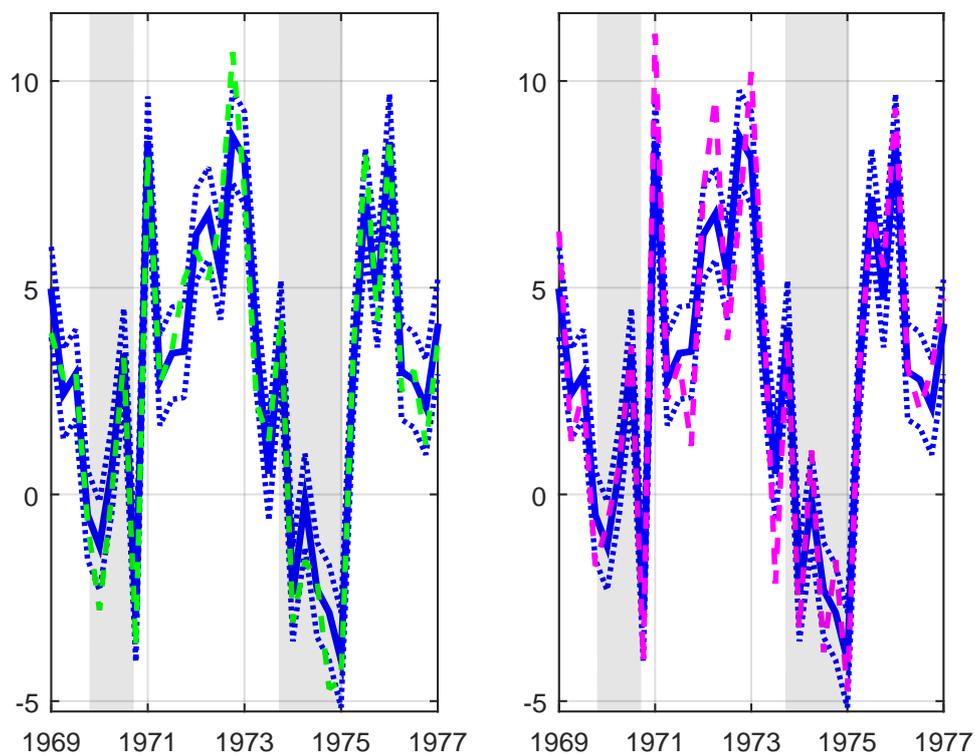
Figure E.3: Zooming in the Great Recession
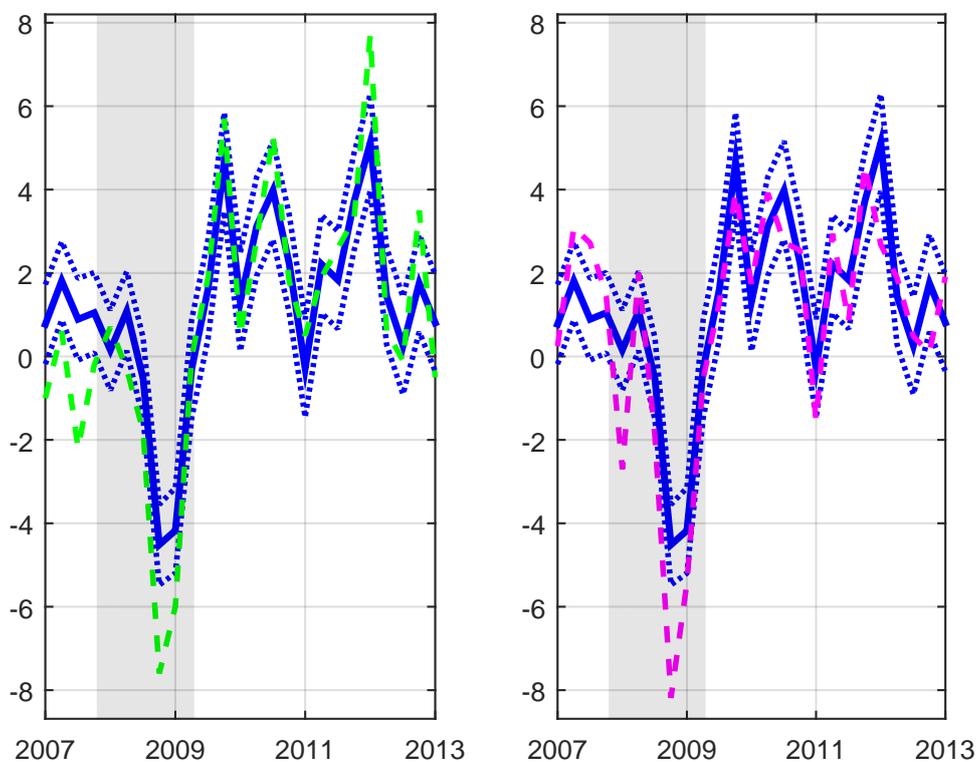
Note: GDP sample paths, 1969 – 1977 (Fig. E.2) and 2007 – 2013 (Fig. E.3). In each panel we show the sample path of the estimated *GDPplus* factor (blue continuous line) with the 68% confidence interval (blue dotted lines), together with GDP(I) (green broken line, left panel) and GDP(E) (pink broken line, right panel).