EUI WORKING PAPERS

Dealing with Financial Instability under a DSGE modeling approach with Banking Intermediation: a forecastability analysis versus TVP-VARs

Stelios Bekiros, Roberta Cardani, Alessia Paccagnini,
Stefania Villa

European University Institute
Department of Economics

Dealing with Financial Instability under a DSGE modeling approach with Banking Intermediation: a forecastability analysis versus TVP-VARs

Stelios Bekiros, Roberta Cardani, Alessia Paccagnini, Stefania Villa

Dealing with Financial Instability under a DSGE modeling approach with Banking Intermediation: a forecastability analysis versus TVP-VARs

Stelios Bekiros[*]   Roberta Cardani[†]   Alessia Paccagnini[‡]   Stefania Villa[§]

Abstract

Recently there has been an increasing awareness on the role that the banking sector can play in macroeconomic activity, especially within the context of the DSGE literature. In this work, we present a DSGE model with financial intermediation as in Gertler and Karadi (2011). The estimation of the shocks and of the structural parameters shows that time-variation can be crucial in the empirical analysis. As DSGE modelling fails to take into account inherent nonlinearities of the economy, we introduce a novel time-varying coefficient state-space estimation method for VAR processes, for homoskedastic and heteroskedastic error structures (TVP-VAR). We conduct an extensive empirical exercise to compare the out-of-sample forecastability of the DSGE model versus standard ARs, VARs, Bayesian VARs and TVP-VARs. We find that the TVP-VAR provides the best forecasting performance for the series of GDP and net worth of financial intermediaries for all steps-ahead, while the DSGE model with the incorporation of a banking sector outperforms the other specifications in forecasting inflation and the federal funds rate at shorter horizons.

JEL Classification: C11, C13, E37

Keywords: DSGE, Extended Kalman filter, Financial frictions, Banking sector

[*]Corresponding author; European University Institute, Department of Economics, Via della Piazzuola 43, I-50133, Florence, Italy; Tel.: +39 055 4685 916; Fax: +39 055 4685 902; E-mail: stelios.bekiros@eui.eu. The first author has received funding from the EU Horizon2020 Research and Innovation Programme under the MS-C Grant No. 656136.

[†]Department of Economics, Università degli Studi di Milano - Bicocca - Milan 20126, Italy; Tel.: +39 02 6448 3237; Email: roberta.cardani@unimib.it.

[‡]University College Dublin - UCD School of Economics, Newman Building, Belfield, Dublin 4, Ireland; Email: alessia.paccagnini@ucd.ie.

[§]Department of Economics, University of Foggia, 71100 Foggia, Italy; Center for Economic Studies, KU Leuven, Leuven, Belgium; Tel:. +39-0881-753713; Fax: +39-0881-781771; E-mail stefania.villa@unifg.it.

1   Introduction

The recent financial crisis made it clear that disruptions in financial markets could have considerable effects on both the dynamics of the business cycle and on the underlying equilibrium growth path. In the dynamic stochastic general equilibrium (DSGE) literature the links between financial and real sectors have mostly been neglected until recently, when there has been an increasing awareness on the role that financial frictions and the banking sector can play in macroeconomic activity (Bean, 2010 among others). Much of the macroeconomics literature with financial frictions stemming from Bernanke *et al.* (1999) (BGG) emphasizes credit market constraints on non-financial borrowers and treats financial intermediaries largely as a veil.

The literature offers different contributions on DSGE models featuring a banking sectors, such as Goodfriend and McCallum (2007), Curdia and Woodford (2010) and Gerali *et al.* (2010), which are interesting but require a relatively high degree of computational tractability. The former introduce different short-term interest rates. The authors show that ignoring the differences between short-term interest rates could lead to policy mistakes. Curdia and Woodford (2010) consider a DSGE model in which two sources of "purely financial" disturbances: i) originating loan is a costly activity; ii) intermediaries are unable to distinguish between borrowers who will default from those who will repay, and so must offer loans to both on the same terms. In this framework IS curve will depend on measure of the inefficiency of the intra-temporal allocation of resources as a consequence of imperfect financial and on a weighted average of two interest rates. Gerali *et al.* (2010) include in a DSGE model with borrowing constraints *à la* Iacoviello (2005) and real and nominal frictions *à la* Smets and Wouters (2007) an imperfectly competitive banking sector that collects deposits and then, subject to the requirement of using banking capital as an input, supplies loans to the private sector.

Gertler and Karadi (2011) present a model with some real and nominal frictions *à la* Smets and Wouters (2007), where financial intermediaries (not exposed to the possibility of runs) are the source of financial frictions. This feature is particularly interesting since in the recent crisis financial frictions mainly originated from and within the financial intermediation sector, as shown by Galati and Moessner (2011). The agency problem creates: (i) a wedge between lending rates and risk free rates; and (ii) a limit to the financial intermediaries ability to acquire assets, and hence, to lend to the private sector. Therefore, financial intermediaries play an active role in the transmission mechanism of the shocks hitting the economy. Overall, their model is fairly elegant and computationally fairly tractable (Cole, 2011). In addition, the introduction of the financial intermediation sector as a source of shocks, helps understand a number of amplification and propagation mechanisms deriving from the endogeneity of credit spreads, and/or of banks' balance sheets. Notwithstanding these contributions, research is still ongoing to fully account for financial instability in DSGE models.

During the financial crisis, conditions in financial markets deteriorated sharply. Figure 1(a) shows the growth rate of bank real estate loans. Data on commercial real estate loans are available only from 2004 onwards. It is evident that in the aftermath of the financial crisis there has been a severe contraction of bank real estate loans. Figure 1(b) shows the evolution of three alternative measure of credit spreads: (i) Baa minus 10-year Treasury constant maturity rate (TCM); (ii) Moody's Baa minus Aaa yield; and (iii) the bank prime loan rate minus the quarterly Treasury bill rate as in Melina and Villa (2014). All the proxies are clearly countercyclical. In particular, during the financial turbulence the credit spreads increased reaching the peak in the financial crisis.

Figure 1: Financial data

Note: Both data on real estate and commercial real estate loans of all commercial banks and on spreads are extracted from the FRED database. Baa corporate bond yield relative to yield on 10-Year Treasury Constant Maturity (Baa-10yTCM) and Baa minus Aaa corporate bond yield are reported on the left axis, while the bank prime loan rate minus 3-month Treasury bill is reported on the left axis.

Furthermore, the evolution of estimated shocks and parameters show that the time variation of the parameters should be crucial in any attempted empirical analysis (Cardani *et al.*, 2015). However, even rolling-window DSGE estimation and modelling usually fails to take into account inherent nonlinearities of the economy, especially in crisis time periods. Del Negro and Schorfheide (2009; 2012) and Wolters (2013) provide rationale as to whether DSGEs fail to demonstrate a good forecasting behavior in crisis regimes. Basically, the DSGEs lack a good calibration outside "*normal*" times. This could be an after-effect of the imposition of tight restrictions on the data by the simple DSGEs. If the data rejects these restrictions, large stochastic shocks are needed to fit the model to the dataset which results in high shock uncertainty. Under average exogenous shocks the DSGE models return back to a steady state, albeit they could not predict recessions and booms as significantly larger exogenous shocks are required to capture these. Even though hybrid models relax these restrictions and the estimated variance of shocks is lower - a fact that provides more accurate predictions - nevertheless, they do not seem to outperform time varying VAR models as demonstrated in our study. Obviously, the use of time-varying parameters seems to be an attractive alternative as well as in terms of capturing nonlinear economic relationships. Primiceri (2005) uses them extensively in analyzing macroeconomic policy issues. The time-varying parameter autoregressive model (TVP-VAR) enables capturing a possible time-varying nature of underlying structure in the economy in a flexible and robust manner. In this paper, a novel time-varying multivariate state-space estimation method for TVP-VAR processes is proposed both for homoskedastic and heteroskedastic error structures. As an alternative to the homoskedastic TVP-VAR we assume that the error structure of the state space Kalman filter is dependent on state variables, which are unobserved discrete-time, discrete-state Markov process, thus providing a Markov-switching heteroskedasticity. While a simple Markov-switching variance model fails to incorporate the learning process of agents, the classic TPV-VAR model fails to incorporate uncertainty that changes due to future asymmetric random shocks. In this work we consider a more general model in which both types of uncertainty are incorporated. For the TVP-VAR models, the parameters are estimated using a multivariate specification of the standard Kalman set-up (Harvey, 1990) with extended quasi-optimal filtering in particular for the TVP-VAR with Markov-switching heteroskedasticity. The likelihood estimation of the TVP-VAR is performed with a suitable multivariate extension of Kim (1993) and Kim and Nelson (1999a, 1999b) method.

In this work, we consider a DSGE model for the US economy with standard frictions *à la* Smets and Wouters (2007) augmented with financial intermediaries as in Gertler and Karadi (2011). We use rolling-window DSGE estimation and modelling as well as TVP-VARs to account for parameter instabilities especially in crisis time periods. Moreover, we conduct an exhaustive empirical exercise that includes the comparison of the out-of-sample predictive performance of the estimated DSGE model with that of standard ARs, VARs, Bayesian VARs as well as of two time-varying parameter autoregressive models (TVP-VAR) models with homoskedastic and heteroskedastic errors in an attempt to investigate inherent nonlinearities of the economy that cannot be captured by the VAR and DSGE class models. Our main goal is to compare different econometrics strategies in evaluating a DSGE economy, but mainly to stress the importance of considering financial variables in particular for the US economy during and after the recent financial crisis, and their incorporation in DSGE and TVP-VAR models. We use time series data in the form of 20-year rolling windows for the period 1984Q1-2013Q4 in order to capture changes in parameters (regime shifts) as discussed in Gürkaynak *et al.* (2013). The DSGE model is estimated for the US using financial observable variables in addition to standard macroeconomic variables.

The remainder of this paper is organized as follows: Section 2 describes the proposed DSGE model with financial frictions and banking intermediation. In Section 3 the standard (benchmark) VAR and Bayesian VAR models are presented. Section 4 presents the time-varying multivariate state-space homoskedastic TVP-VAR model as well as a Markov-switching heteroskedastic set-up. In Section 5 the data and the DSGE estimation procedure are described. In addition, the evolution of estimated shocks and parameters is displayed. Next, the empirical results of the comparative forecasting evaluation are illustrated and analyzed in Section 6. Finally, Section 7 concludes.

2   The DSGE model with Banking Intermediation

This section briefly describes the linearized version of the DSGE model, which features financial intermediaries as in Gertler and Karadi (2011) in an otherwise setup of Smets and Wouters (2007). The economy is composed by households, labor unions, labor packers, financial intermediaries, a productive sector and a monetary authority. Households consume, accumulate government bonds and supply labor. A labor union differentiates labor and sets wages in a monopolistically competitive market. Competitive labor packers buy labor services from the union, package and sell them to intermediate goods firms. The presence of an agency problem limits the ability of financial intermediaries to obtain deposits from households. This, in turn, affects the leverage ratio of financial intermediaries. Output is produced in several steps, including a monopolistically competitive sector with producers facing price rigidities. The monetary authority sets the short-term nominal interest rate according to a Taylor rule. We report below the list of equilibrium conditions in their log-linear form[1]

$$
\begin{aligned}
c_t = & \frac{h/\gamma}{1+h/\gamma} c_{t-1} + \left(1 - \frac{h/\gamma}{1+h/\gamma}\right) E_t c_{t+1} + \\
& \frac{(\sigma_c - 1)(w_* l_*/c_*)}{\sigma_c(1+h/\gamma)}(l_t - E_t l_{t+1}) - \frac{1-h/\gamma}{\sigma_c(1+h/\gamma)}\left(r_t - E_t \pi_{t+1} + e_t^b\right)
\end{aligned}
\tag{1}
$$

---

[1]All variables are log-linearized around their steady state balanced growth path and starred variables represent steady state values.

$$w_t = \frac{1}{1 + \beta\gamma^{1-\sigma_c}}w_{t-1} + \left(1 - \frac{1}{1 + \beta\gamma^{1-\sigma_c}}\right)(E_t w_{t+1} + E_t \pi_{t+1}) - \frac{1 + \beta\gamma^{1-\sigma_c}\iota_w}{1 + \beta\gamma^{1-\sigma_c}}\pi_t$$

$$+ \frac{\iota_w}{1 + \beta\gamma^{1-\sigma_c}}\pi_{t-1} - \frac{\left(1 + \xi_w \beta\gamma^{1-\sigma_c}\right)(1 - \xi_w)}{(1 + \beta\gamma^{1-\sigma_c})\xi_w\left[(\phi_w - 1)e^w + 1\right]}\mu_t^w + e_t^w \tag{2}$$

$$\mu_w = w_t - \left[\sigma_l l_t + \frac{1}{1-h}(c_t - hc_{t-1})\right] \tag{3}$$

$$y_t = \phi_p\left[\alpha(k_{t-1} + u_t) + (1 - \alpha)l_t\right] + e_t^a \tag{4}$$

$$u_t = \frac{(1 - \psi)}{\psi}z_t^k \tag{5}$$

$$\pi_t = \frac{\iota_p}{1 + \beta\gamma^{1-\sigma_c}\iota_p}\pi_{t-1} + \frac{\beta\gamma^{1-\sigma_c}}{1 + \beta\gamma^{1-\sigma_c}\iota_p}\pi_{t+1} -$$

$$\frac{\left(1 - \beta\gamma^{1-\sigma_c}\xi_p\right)(1 - \xi_p)}{(1 + \beta\gamma^{1-\sigma_c}\iota_p)\xi_p\left[(\phi_p - 1)e^p + 1\right]}\mu_t^p + e_t^p \tag{6}$$

$$\mu_p = \alpha(k_{t-1} + u_t - l_t) + e_t^a - w_t \tag{7}$$

$$z_t^k = -(k_{t-1} + u_t - l_t) + w_t \tag{8}$$

$$i_t = \frac{1}{(1 + \beta\gamma^{1-\sigma_c})}i_{t-1} + \left[1 - \frac{1}{(1 + \beta\gamma^{1-\sigma_c})}\right]E_t i_{t+1} + \frac{1}{\gamma^2\varphi(1 + \beta\gamma^{1-\sigma_c})}q_t + e_t^x \tag{9}$$

$$E_t r_{t+1}^k = \frac{z_*^k}{r_*^k}E_t z_{t+1}^k + \frac{(1 - \delta)}{r_*^k + (1 - \delta)}E_t q_{t+1} - q_t \tag{10}$$

$$k_t = \frac{(1 - \delta)}{\gamma}k_{t-1} + \left[1 - \frac{(1 - \delta)}{\gamma}\right]i_t + \left[1 - \frac{(1 - \delta)}{\gamma}\right](1 + \beta\gamma^{1-\sigma_c})\gamma^2\varphi e_t^x \tag{11}$$

$$r_t^{ep} = E_t r_{t+1}^k - (r_t - E_t \pi_{t+1}) \tag{12}$$

$$q_t + k_t = lev_t + n_t \tag{13}$$

$$lev_t = \eta_t + \frac{v}{\phi - v}v_t \tag{14}$$

$$\eta_t = \frac{\varpi\beta}{\gamma^{\sigma_c}}z_*\left(E_t\Lambda_{t+1} - \Lambda_t + z_t + E_t\eta_{t+1}\right) \tag{15}$$

$$z_t = \frac{lev_* r_*^k}{z_*}r_t^k + r_*(1 - lev_*)(r_{t-1} - \pi_t) + lev_*\left(r_*^k - r_*\right)lev_{t-1} \tag{16}$$

$$v_t = \frac{\varpi\beta}{\gamma^{\sigma_c}}z_*\left(E_t\Lambda_{t+1} - \Lambda_t + x_t + E_t v_{t+1}\right) + \frac{(1 - \varpi)\beta}{\gamma^{\sigma_c}\nu_*}$$

$$\left[r_*^k r_t^k - r_*(r_{t-1} - \pi_t)\right] + \frac{(1 - \varpi)\beta}{\gamma^{\sigma_c}\nu_*}\left(r_*^k - r_*\right)(E_t\Lambda_{t+1} - \Lambda_t) \tag{17}$$

$$x_t = lev_t - lev_{t-1} + z_t \tag{18}$$

$$n_t = \frac{n_*^e}{n}n_t^e + \frac{n_*^n}{n}n_t^n \tag{19}$$

$$n_t^e = n_{t-1}^e + z_t + e_t^n \tag{20}$$

$$n_t^n = \xi lev_*(q_t + k_t) \tag{21}$$

$$r_t = \rho_r r_{t-1} + (1 - \rho_r) \left[\rho_\pi \pi_t + \rho_y \left(y_t - y_t^p \right) \right] + \rho_{\Delta y} \left[\left(y_t - y_t^p \right) - \left(y_{t-1} - y_{t-1}^p \right) \right] + e_t^i \tag{22}$$

$$y_t = (1 - g_y - i_y) c_t + i_y i_t + \frac{z_*^k k_*}{y_*} z_t^k + e_g \tag{23}$$

Equation (1) is the Euler consumption equation: h measures the degree of habits in consumption, γ is the steady state growth rate and σ^c is the relative risk aversion coefficient. Equation (2) represents the Calvo staggered wage setting, β is the households discount factor, ξ_w measures wage stickiness, while ι_w denotes the degree of wage indexation. The wage mark-up, μ_t^w, defined in equation (3), is determined as the difference between the real wage and the marginal rate of substitution between working and consuming.

Equation (4) captures the production technology with fixed costs ϕ_p and capital share α. Capital is augmented by the capital utilization rate, u_t, whose optimality condition is given by equation (5), where ψ represents the positive function of elasticity of the capital utilization adjustment cost and z_t^k is the marginal product of capital. Staggered price stickiness is incorporated into the model through limiting the ability of firms to reset their prices every period with a probability equal to ξ_p, as shown by equation (6), where ι_p governs the degree of price indexation. The price markup dynamics obtained under monopolistic competition is described by (7). Cost minimization by firms, equation (8), implies that the marginal product of capital is negatively related to the capital-labor ratio and positively to real wages. Investment dynamics is described by equation (9), where q_t is the current value of capital stock and φ is the elasticity of the investment adjustment cost. The arbitrage condition for the value of capital is given by equation (10), where $E_t r_{t+1}^k$ is the external cost of funding. The law of motion of installed capital is given by equation (11).

Financial intermediaries raise funds from households and grant loans to intermediate firms producers. Due to a *moral-hazard* costly enforcement problem,[2] the presence of financial intermediation leads to an endogenous credit spread, captured by equation (12), as a difference between the cost of funding state contingent asset of non-financial firms, $E_t r_{t+1}^k$, and gross nominal interest rate paid on deposits to households, r_t. As described by equation (13), the maximum amount of lending by firms depends on the total net worth, n_t and on the ratio of loan asset to equity capital, lev_t. The leverage is endogenously determined by equation (14) and depends on the the gain of increasing one unit of net worth, η_t, on the gain of expanding assets, v_t, and on the fraction ϕ that bankers could divert from the project and transfer it back to their household. The gain of having net worth, equation (15), hinges on the stochastic discount factor, Λ_t, associated to the household problem, the probability of bankers' surviving in the next period, ϖ, and on the gross growth rate of net worth, z_t, whose law of motion is given by equation (16). The gain of expanding assets, equation (17), is mainly affected by the gross growth rate in assets, x_t, which evolves as in equation (18). Total net worth is given by the sum of net worth of existing bankers, n_t^e and of new bankers, n_t^n, equation (19). Net worth of existing bankers equals earnings on assets held in the previous period and the growth of the net worth, as specified by equation (20), while net worth of new banks takes into account the "start-up" funds from the households to which they belong to, equal to the fraction ξ of total assets, as indicated by equation (21).[3]

---

[2] In this model at the beginning of each period the banker can choose to divert the fraction ϕ of available funds from the project and instead transfer them back to the household. Depositors can force the intermediary into bankruptcy and recover the remaining fraction $1 - \phi$ of total assets. However, costly enforcement implies that it is too costly for the depositors to recover the diverted fraction of funds by the banker.

[3] As argued by Cole (2011), a critical assumption of the model by Gertler and Karadi is that financial intermediaries do not efficiently hedge their risk. This is a general assumption in the mainstream DSGE literature on financial frictions (see, for example, the seminal contribution of Bernanke et al. (1999) and can be an avenue for future research. In addition, the model can get closer to the data if firms borrow to finance some fraction of their labor as well as their capital.

The monetary authority follows a Taylor rule, equation (22): r_t ρ_r, ρ_π, ρ_y and $\rho_{\Delta y}$ are policy parameters referring to interest-rate smoothing, and the responsiveness of the nominal interest rate to inflation deviations, to the output gap and to changes in the output gap, respectively. Finally, the resource constraint, equation (23), completes the model.

The model features eight exogenous disturbances: total factor productivity, e_t^a; price mark-up, e_t^p; wage mark-up, e_t^w; investment-specific technology, e_t^x; risk premium, e_t^b; net worth of financial intermediaries, e_t^n; exogenous spending, e_t^g; and monetary policy shocks, e_t^i. All the shocks follow an AR(1) process, but the price and wage mark-up shocks following an ARMA(1,1) process.

Our general estimation and calibration strategy follows the procedure proposed by Smets and Wouters (2007) adapted to a model with financial intermediation. In particular, the parameters that cannot be identified in the data and/or are related to steady state values of variables are calibrated as follows: the discount factor, β, is set equal to 0.99. The depreciation rate of physical capital, δ, is set equal to 0.025. The Kimball aggregators in the goods and labor market are equal to 10, and the steady state gross wage mark-up is set to 1.5. The share of government to GDP is equal to 0.18. Similarly to Villa (2014), the financial parameters – ϖ, ϕ and ξ – are calibrated to target an average working life of bankers of almost a decade, a steady state spread of 150 basis points and a steady state leverage ratio of financial intermediaries equal to 4. The remaining parameters governing the dynamics of the model are estimated using Bayesian techniques. The locations of the prior mean correspond to those in Smets and Wouters (2007).

3   Benchmark models

The classical unrestricted VAR, as suggested by Sims (1980), has the following form

$$\mathbf{Y}_t = \mathbf{Z}_t \mathbf{B} + u_t \tag{24}$$

where $\mathbf{Y}_t$ is a $(T \times n)$ matrix, $\mathbf{Z}$ is a $(T \times k)$ matrix ($k = 1 + np$, p =number of lags) with rows $Z_t' = [1, Y_{t-1}', ..., Y_{t-p}']$, $\mathbf{B}$ is a $(k \times n) = [\mathbf{B}_0, \mathbf{B}_1, ..., \mathbf{B}_p]'$, while the one-step ahead forecast errors u_t have a multivariate $N(0, \Sigma_u)$ conditional on past observations of Y. The Bayesian VAR, as described in Litterman (1981), Doan et al. (1984), Todd (1984), Litterman (1986) and Spencer (1993) has become a widely popular approach to overcoming overparameterization. One of main problems in using VAR models is that many parameters need to be estimated, although some of them may be insignificant. This overparameterization problem, resulting in multicollinearity and a loss of degrees of freedom, leads to inefficient estimates. Instead of eliminating longer lags, the BVAR imposes restrictions on these coefficients by assuming that they are more likely to be near zero than the coefficients on shorter lags. Obviously, if there are strong effects from less important variables, the data can counter this assumption. Usually, the restrictions are imposed by specifying normal prior distributions with zero means and small standard deviations for all coefficients, with a decreasing standard deviation as the lags increase. The only exception is the coefficient on a variable's first lag, which has a mean of unity. Litterman (1981) used a diffuse prior for the constant. The means of the prior are popularly called the "Minnesota Priors". The basic principle behind the "Minnesota" prior is that all equations are centered around a random walk with drift. This idea has been modified by Kadiyala and Karlsson (1997) and Sims and Zha (1998). In Ingram and Whiteman (1994), a real business cycle model is used to generate a prior for a reduced form VAR, as a development of the "Minnesota" priors procedure. Formally speaking, these prior means can be written as $\mathbf{B}_i \sim N(1, \sigma_{B_i}^2)$ and $\mathbf{B}_j \sim N(0, \sigma_{B_j}^2)$, where $\mathbf{B}_i$ denotes the coefficients associated with the lagged dependent variables in each equation of the VAR, while

$\mathbf{B}_j$ represents any other coefficient. The prior variances $\sigma^2_{B_i}$ and $\sigma^2_{B_j}$ specify the uncertainty of the prior means, namely $\mathbf{B}_i = 1$ and $\mathbf{B}_j = 0$. The specification of the standard deviation of the distribution of the prior imposed on variable j in equation i at lag m, for all i, j and m, denoted by $s(i, j, m)$, is specified as $s(i, j, m) = (w \times \xi(m) \times T(i, j))(\hat{\sigma}_i/\hat{\sigma}_j)$. The tightness $T(i, j)$ of variable j in equation i relative to variable i is $T(i, j) = 1$ if $i = j$ and $T(i, j) = k_{ij}$ otherwise ($0 \leq k_{ij} \leq 1$). By increasing the interaction it is possible for the value of k_{ij} to loosen the prior (Dua and Ray, 1995). The ratio $\hat{\sigma}_i/\hat{\sigma}_j$ consists of estimated standard errors of the univariate autoregression, for variables i and j. This ratio scales the variables to account for differences in the units of measurement, without taking into account the magnitudes of the variables. The term w measures the standard deviation on the first lag, and also indicates the overall tightness. A decrease in the value of w results in a tighter prior. The function $\xi(m) = m^{-d}$, $d > 0$ is the measurement of the tightness on lag m relative to lag 1, and is assumed to have a harmonic shape with a decay of d, which tightens the prior on increasing lags. Following the standard Minnesota prior settings, the overall tightness (w) is set equal to 0.1, while the lag decay (d) is 0.5 and the interaction parameter (k_{ij}) is set equal to 0.1. This Bayesian setting is very similar to the model used in Liu et al (2009) and Gupta and Kabundi (2010).

4 Multivariate State-Space Time-Varying Parameter VAR models

Time varying parameter autoregression could easily form a state space model with the parameters of the TVP-VAR as state variables. The state space model has been well studied by Harvey (1990) and Durbin and Koopman (2002). According to Kalman (1960, 1963), in a state-space representation the signal extraction is implemented through a model that links the unobserved and observed variables of the system. Kalman filtering involves sequentially updating a linear projection on the vector of interest. For the standard homoskedastic TVP-VAR models, the parameters are estimated using a multivariate specification of the standard Kalman filter (Harvey, 1990; Bekiros and Paccagnini, 2013). The likelihood estimation requires repeating the filtering many times in order to evaluate the likelihood for each set of the time-varying parameters until the maximum is reached. This is performed with a suitable multivariate extension of the Kim and Nelson (1999a, 1999b) method. The calculation of the Hessian for the estimation of the variance-covariance matrix is done with the Broyden-Fletcher-Goldfarb-Shano (BFGS) optimization algorithm.[4]

4.1 MVSS-TVP-VAR model with Homoscedastic errors

The standard homoskedastic TVP-VAR (MVSS-TVP-VAR) can be expressed as

$$\mathbf{y}_t = \mathbf{B}_{0,t} + \mathbf{B}_{1,t}\mathbf{y}_{t-1} + \cdots + \mathbf{B}_{p,t}\mathbf{y}_{t-p} + u_t \tag{25}$$

in which $\mathbf{B}_{0,t}$ is a $k \times 1$ vector of time-varying intercepts, $\mathbf{B}_{i,t}$ ($i = 1, \ldots, p$) are $k \times k$ matrices of time-varying coefficients and u_t are homoscedastic reduced-form residuals with a covariance matrix $\mathbf{\Omega}_t$. This could be transformed into a multivariate state-space form. Consider the following state-space system

---

[4]Other algorithms can also be used with the same results, e.g., the DFP and the Levenberg-Marquardt. The parameters could be also estimated with the use of the Zellner g-prior both for homoskedastic and heteroskedastic TVP-VARs and in this case the numerical evaluation of the posterior distributions is performed with Gibbs sampling (Kim and Nelson, 1999b).

$$y_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \varepsilon_t \tag{26}$$

$$\boldsymbol{\alpha}_t = \mathbf{T}_t \boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t \tag{27}$$

The first equation is known as the *measurement* or *observation* equation and presents that part of the system than can physically be measured, while the second is the *state* equation $\boldsymbol{\alpha}_t$ the vector of state variables. $\mathbf{Z}_t$ is a matrix of known or unknown time varying coefficients and matrix $\mathbf{T}_t$, the state transition matrix. Finally, ε_t is $N\left(0, \sigma^2\right)$ while $\boldsymbol{\eta}_t$ in multivariate normal with an expected value of zero and a homoscedastic covariance matrix of $\mathbf{Q}$. The unknown parameters (hyperparameters) are the elements of the matrices and the variances of the noise processes to be estimated. This is accomplished by maximizing the likelihood function which is presented below for one time period

$$L_t = -\frac{1}{2}\sum_{t=1}^{T}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln f_t - \frac{1}{2}\sum_{t=1}^{T}\frac{\eta_t^2}{f_t} \tag{28}$$

where η_t is the one–step ahead residual at time t and f_t is its variance. It is calculated recursively using the following equations

$$\boldsymbol{\alpha}_{t|t-1} = \mathbf{T}_t \boldsymbol{\alpha}_t \tag{29}$$

$$\mathbf{P}_{t|t-1} = \mathbf{T}_t \mathbf{P}_t \mathbf{T}_t' + \mathbf{Q} \tag{30}$$

$$\eta_{t|t-1} = y_t - \mathbf{Z}_t \boldsymbol{\alpha}_{t|t-1} \tag{31}$$

$$f_{t|t-1} = \mathbf{Z}_t \mathbf{P}_{t|t-1} \mathbf{Z}_t' + \sigma^2 \tag{32}$$

$$\boldsymbol{\alpha}_{t|t} = \boldsymbol{\alpha}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{Z}_t' \eta_t / f_t \tag{33}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{Z}_t' \mathbf{Z}_t \mathbf{P}_{t|t-1} / f_t \tag{34}$$

Hence, Equations (29) to (34) that generate an estimate of the state vector and its covariance matrix $\mathbf{P}_t$ are known as the Kalman filter. Given starting values, an estimate of the unknown regression coefficients is obtained. Then using this information in the likelihood function, one may then estimate the hyperparameters of the model. Once these estimates have been obtained, an estimate of the state vector, the recursive residuals and their variance is obtained, and also an estimate of the updated residual vector $\eta_{t|t-1} = y_t - \mathbf{Z}_t \boldsymbol{\alpha}_{t|t-1}$ is generated.

The framework for a multivariate version of the Kalman filter is provided by Harvey (1990), based on a time series analogue of the seemingly unrelated regression equation (SURE) model introduced into econometrics by Zellner (1963). Harvey (1990) refers to it as a system of seemingly unrelated time series equations (SUTSE) model. The simplest SUTSE model is the multivariate random walk plus noise process

$$\mathbf{y}_t = \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T \tag{35}$$

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t \tag{36}$$

where $\boldsymbol{\alpha}_t$ is an $N \times 1$ vector of local level components and $\boldsymbol{\varepsilon}_t$, and $\boldsymbol{\eta}_t$ are vectors of multivariate white noise with mean zero and covariance matrices $\boldsymbol{\Sigma}_\varepsilon$ and $\boldsymbol{\Sigma}_\eta$ respectively. As in the univariate model, $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ are assumed to be uncorrelated with each other in all time periods. The variables are linked via the off-diagonal elements $\boldsymbol{\Sigma}_\varepsilon$ and $\boldsymbol{\Sigma}_\eta$. A linear time-invariant univariate structural model can be written in the SUTSE state space form for N variables

$$\mathbf{y}_t = (\mathbf{z}' \otimes \mathbf{I}_N)\, \boldsymbol{\alpha}_t + \varepsilon_t \tag{37}$$

$$\boldsymbol{\alpha}_t = (\mathbf{T} \otimes \mathbf{I}_N)\, \boldsymbol{\alpha}_{t-1} + (\mathbf{R} \otimes \mathbf{I}_N)\, \boldsymbol{\eta}_t \tag{38}$$

with $Var(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma}_\varepsilon$ and $Var(\boldsymbol{\eta}_t)$ a block diagonal matrix with the blocks being $\boldsymbol{\Sigma}_k$, $k = 1, ..., g$. For example, in the four-variate case (as in this study) the variance of the error component in the state equation is

$$Var\left(\boldsymbol{\eta}_t\right) = \begin{bmatrix} \boldsymbol{\Sigma}_\eta & 0 & 0 & 0 \\ 0 & \boldsymbol{\Sigma}_\zeta & 0 & 0 \\ 0 & 0 & \boldsymbol{\Sigma}_\omega & 0 \\ 0 & 0 & 0 & \boldsymbol{\Sigma}_\gamma \end{bmatrix} \tag{39}$$

A more general formulation of the SUTSE model does not constrain $Var\left(\boldsymbol{\eta}_t\right)$ to be diagonal, hence $Var\left(\boldsymbol{\eta}_t\right)$ need not be block diagonal. Indeed the SUTSE formulation can be generalized further to allow quantities such as $\mathbf{z}, \boldsymbol{\Sigma}_\varepsilon, \mathbf{T}, \mathbf{R}$ and $Var\left(\boldsymbol{\eta}_t\right)$ to change deterministically over time. As shown in Harvey (1986), the time-domain treatment still goes through. The Kalman filter may be applied to (37) and (38), the number of sets of observations needed to form an estimator of $\boldsymbol{\alpha}_t$, with finite MSE matrix being the same as in the univariate case. The conditions for the filter to converge to a steady state are an obvious generalization of the conditions in the univariate case. Given normality of the disturbances, the log-likelihood function is of the prediction error decomposition form.

The decoupling of the Kalman filter is related to the result which arises in a SUTSE system where OLS applied to each equation in turn is fully efficient if each equation contains the same regressors. Hence, all the information needed for estimation, prediction and smoothing can be obtained by applying the same univariate filter to each series in turn. If we consider the multivariate random walk plus noise model and a signal-to-noise ratio q (i.e., $\boldsymbol{\Sigma}_\eta / \boldsymbol{\Sigma}_\varepsilon = q$), the Kalman filter for this model is written as

$$\boldsymbol{\alpha}_{t+1|t} = \boldsymbol{\alpha}_{t|t-1} + \mathbf{K}_t \left(\mathbf{y}_t - \boldsymbol{\alpha}_{t|t-1}\right), t = 2, \ldots, T \tag{40}$$

and

$$\mathbf{P}_{t+1|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{F}_t^{-1}\mathbf{P}_{t|t-1} + q\boldsymbol{\Sigma}_\varepsilon \tag{41}$$

where

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{F}_t^{-1} \tag{42}$$

and

$$\mathbf{F}_t = \mathbf{P}_{t|t-1} + \boldsymbol{\Sigma}_\varepsilon \tag{43}$$

Let w_t denote a positive scalar for $t = 2, \ldots, T$ and suppose that $\mathbf{P}_{t|t-1}$, the MSE matrix of the $N \times 1$ vector $\boldsymbol{\alpha}_{t|t-1}$, is proportional to $\boldsymbol{\Sigma}_\varepsilon$, i.e. $\mathbf{P}_{t|t-1} = w_t \boldsymbol{\Sigma}_\varepsilon$. It then follows from (41) that $\mathbf{P}_{t+1|t}$ is of the same form, that is, $\mathbf{P}_{t+1|t} = w_{t+1} \boldsymbol{\Sigma}_\varepsilon$ with $w_{t+1} = (w_t + w_t q + q) / (w_t + 1)$. Furthermore if $\mathbf{P}_{t|t-1} = w_t \boldsymbol{\Sigma}_\varepsilon$ the gain matrix in (40) is diagonal, that is

$$\mathbf{K}_t = w_t \boldsymbol{\Sigma}_\varepsilon \left(w_t \boldsymbol{\Sigma}_\varepsilon + \boldsymbol{\Sigma}_\varepsilon \right)^{-1} = \left[w_t / (w_{t+1}) \right] \mathbf{I}_N \tag{44}$$

Consider that the above Kalman filter is started off in such a way that $\mathbf{P}_{2|1}$ is proportional to $\boldsymbol{\Sigma}_\varepsilon$; that is $\mathbf{P}_{2|1} = p_{2|1} \boldsymbol{\Sigma}_\varepsilon$, where $p_{2|1}$ is a scalar. As $\mathbf{P}_{t|t-1}$ must continue to be proportional to $\boldsymbol{\Sigma}_\varepsilon$, it follows from (44) that the elements of $\boldsymbol{\alpha}_{t+1|t}$, can be computed from the univariate recursions. It also follows that w_t, must be equal to $p_{t|t-1}$ for all $t = 2, \ldots, T$. The starting values $\boldsymbol{\alpha}_{2|1} = \mathbf{y}_1$ and $\mathbf{P}_{2|1} = \boldsymbol{\Sigma}_\eta + \boldsymbol{\Sigma}_\varepsilon = (1 + q) \boldsymbol{\Sigma}_\varepsilon$ equally correspond to the use of a diffuse prior, and the use of these starting values leads to the exact likelihood function for $\mathbf{y}_2, \ldots, \mathbf{y}_T$ in the prediction error decomposition form

$$\log L = -\frac{(T-1)N}{2} \log 2\pi - \frac{1}{2} \sum_{t=2}^T \log |\mathbf{F}_t| - \frac{1}{2} \sum_{t=2}^T \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t \tag{45}$$

where $\mathbf{v}_t = \mathbf{y}_t - \widehat{\mathbf{y}}_{t|t-1}, t = 1, ..., T$. The decoupling of the Kalman filter allows the elements of $\mathbf{v}_t$ to be computed from the univariate recursions. Furthermore

$$\mathbf{P}_{t|t-1} = p_{t|t-1} \boldsymbol{\Sigma}_\varepsilon \tag{46}$$

and so

$$\mathbf{F}_t = \mathbf{P}_{t|t-1} + \boldsymbol{\Sigma}_\varepsilon = f_t \boldsymbol{\Sigma}_\varepsilon, \ t = 3, \ldots, T \tag{47}$$

where $f_t = \left(p_{t|t-1} + 1 \right)$. Substituting from (47) into (45) gives

$$\log L = -\frac{(T-1)N}{2} \log 2\pi + \frac{(T-1)}{2} \log \left| \boldsymbol{\Sigma}_\varepsilon^{-1} \right| - \frac{N}{2} \sum_{t=2}^T \log f_t - \frac{1}{2} \sum_{t=2}^T \frac{1}{f_t} \mathbf{v}_t' \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{v}_t \tag{48}$$

Differentiating (48) with respect to the distinct elements of $\boldsymbol{\Sigma}_\varepsilon^{-1}$ leads to the ML estimator of $\boldsymbol{\Sigma}_\varepsilon$ being

$$\tilde{\boldsymbol{\Sigma}}_\varepsilon = (T-1)^{-1} \sum_{t=2}^T f_t^{-1} \mathbf{v}_t \mathbf{v}_t' \tag{49}$$

for any given value of q. The ML estimators of q and $\boldsymbol{\Sigma}_\varepsilon$ can therefore be obtained by maximizing the concentrated likelihood function

$$\log L_c = -\frac{(T-1)N}{2} \log 2\pi - \frac{(T-1)}{2} \log \left| \tilde{\boldsymbol{\Sigma}}_\varepsilon \right| - \frac{N}{2} \sum_{t=2}^T \log f_t \tag{50}$$

with respect to q. Once the parameters have been estimated, prediction and smoothing can be carried out. The predictions of future observations are obtained from the univariate recursions

$$\mathrm{MSE}\left(\widehat{\mathbf{y}}_{T+l|T} \right) = f_{T+l|T} \boldsymbol{\Sigma}_\varepsilon, \ l = 1, 2, \ldots \tag{51}$$

where

$$f_{T+l|T} = p_{T+l|T} + 1 \tag{52}$$

The decoupling of the Kalman filter can be shown in a similar way for the time-varying system, as in Bekiros and Paccagnini (2013)

$$\mathbf{y}_t = (\mathbf{z}_t' \otimes \mathbf{I}_N) \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, Var(\boldsymbol{\varepsilon}_t) = h_t \boldsymbol{\Sigma}_* \tag{53}$$

$$\boldsymbol{\alpha}_t = (\mathbf{T}_t \otimes \mathbf{I}_N) \boldsymbol{\alpha}_{t-1} + (\mathbf{R}_t \otimes \mathbf{I}_N) \boldsymbol{\eta}_t, \ Var(\boldsymbol{\eta}_t) = \mathbf{Q}_t \otimes \boldsymbol{\Sigma}_* \tag{54}$$

where h_t and $\mathbf{Q}_t = \text{diag}(q_1, \ldots, q_k)$, with h_t and $q_1, \ldots, q_k$ non-negative scalars. The more general formulation does not constrain $\mathbf{Q}_t$ to be diagonal, although, as in the univariate model, restrictions are needed on $\mathbf{Q}_t$ for the model to be identifiable. All the results on estimation and prediction carry through, with $\mathbf{P}_{t+1|t} = \mathbf{P}_{t+1|t}^* \otimes \boldsymbol{\Sigma}_*$, where $\mathbf{P}_{t+1|t}^*$ is the MSE matrix for the univariate model (Harvey, 1986, 1990).

4.2 MVSS-TVP-VAR model with Markov-Switching Heteroscedasticity

As an alternative to the homoskedastic TVP-VAR we assume that ε_t and $\boldsymbol{\eta}_t$ (i.e., σ^2 and $\mathbf{Q}$) are dependent on Hamilton's (1988) state variable (S_t), which is an outcome of an unobserved discrete-time, discrete-state Markov process. While the Markov-switching variance model fails to incorporate the learning process of agents, the TPV model fails to incorporate uncertainty that changes due to future random shocks. In this study we consider a more general approach in which both types of uncertainty are incorporated. An important motivation for considering a state-space model with Markov-switching heteroskedasticity is due to Lastrapes (1989), Lamoureux and Lastrapes (1990) and Kim (1993) who showed that failure to allow for regime shifts leads to an overstatement of the persistence of the variance of a series. Moreover, in this way we could incorporate different regimes in crisis periods.

Consider the following first-order, ω-state Markov-switching model of heteroskedasticity

$$Q_t = Q^{S_t} = Q_1 \Theta_{1t} + Q_2 \Theta_{2t} + \cdots + Q_\omega \Theta_{\omega t} \tag{55}$$

$$h_t = h^{S_t} = h_1 \Theta_{1t} + h_2 \Theta_{2t} + \cdots + h_\omega \Theta_{\omega t} \tag{56}$$

where $\Theta_{jt} = 1$ if $S_t = j$ and $\Theta_{jt} = 0$ if $S_t \neq j$ ($j = 1, 2, \ldots, \omega$). The unobserved-state variable S_t evolves according to a Markov process with the following transition probability matrix

$$p = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1\omega} \\ p_{21} & p_{22} & \cdots & p_{2\omega} \\ \vdots & \vdots & \ddots & \vdots \\ p_{\omega 1} & p_{\omega 2} & \cdots & p_{\omega\omega} \end{pmatrix} \tag{57}$$

where $p_{ij} = \Pr[S_t = j|S_{t-1} = i]$, for $i, j = 1, 2, \ldots, \omega$, and $\sum_{j=1}^{\omega} p_{ij} = 1$. Under a state-space form we consider the time-varying-parameter model in Eqs. (27)-(28), in this case with a 2-state Markov-switching model of heteroskedasticity

$$\Pr\left[S_t = 1 | S_{t-1} = 1\right] = P_{11}; \ \Pr\left[S_t = 0 | S_{t-1} = 1\right] = 1 - P_{11};$$

$$\Pr\left[S_t = 1 | S_{t-1} = 0\right] = 1 - P_{00}; \ \Pr\left[S_t = 0 | S_{t-1} = 0\right] = P_{00}$$

4.2.1 A Quasi-Optimal Filter for Approximation

Suppose that the parameters Q^j, h^j ($j = 1, 2$), $\mathbf{Z}_t$, $\mathbf{T}_t$, and p are known for the model with a 2-state Markov-switching heteroskedasticity that consists of Equations (26)-(27) and (55)-(57). Based on the assumption that $S_{t-1} = i$ and $S_t = j$ ($i, j = 1, 2$), the Kalman filter can be written as follows

$$\boldsymbol{\alpha}^i_{t|t-1} = \mathbf{T}_t \boldsymbol{\alpha}^i_{t-1|t-1} \tag{58}$$

$$\mathbf{P}^{(i,j)}_{t|t-1} = \mathbf{T}_t \mathbf{P}^i_{t-1|t-1} \mathbf{T}'_t + Q^j \tag{59}$$

$$\eta^i_{t|t-1} = y_t - \mathbf{Z}_t \boldsymbol{\alpha}^i_{t|t-1} \tag{60}$$

$$f^{(i,j)}_{t|t-1} = \mathbf{Z}_t \mathbf{P}^{(i,j)}_{t|t-1} \mathbf{Z}'_t + h^j \tag{61}$$

$$\boldsymbol{\alpha}^{(i,j)}_{t|t} = \mathbf{T}_t \boldsymbol{\alpha}^i_{t|t-1} + K^{(i,j)}_t \eta^j_{t|t-1} \tag{62}$$

$$\mathbf{P}^{(i,j)}_{t|t} = \left(I - K^{(i,j)}_t \mathbf{Z}_t\right) \mathbf{P}^{(i,j)}_{t|t-1} \tag{63}$$

$$K^{(i,j)}_t = \mathbf{P}^{(i,j)}_{t|t-1} \mathbf{Z}_t' \left(f^{(i,j)}_{t|t-1}\right)^{-1} \tag{64}$$

where $\boldsymbol{\alpha}^i_{t|t-1}$ is an inference of $\boldsymbol{\alpha}_t$ based on information up to time $t - 1$, given $S_{t-1} = i$, $\mathbf{P}^{(i,j)}_{t|t-1}$ is the covariance matrix of $\boldsymbol{\alpha}^i_{t|t-1}$, $\eta^i_{t|t-1}$ is the conditional forecast error of y_t based on information up to $t - 1$, given $S_{t-1} = i$, and $f^{(i,j)}_{t|t-1}$ is the conditional variance of forecast error $\eta^i_{t|t-1}$, given $S_{t-1} = i$ and $S_t = j$. Finally, $K^{(i,j)}_t$ is the Kalman gain given $S_{t-1} = i$ and $S_t = j$. If we iterate the preceding Kalman filter from $t = 1$ to $t = T$, the inferences on $\boldsymbol{\alpha}_T$ and its covariance matrix $\left(\boldsymbol{\alpha}_{T|T} \text{ and } \mathbf{P}_{T|T}\right)$ for example, would depend on the whole history of current and past states, $S_0, S_1, \ldots, S_T$. Overall, we would have ω^T cases to consider, which could be quite impossible to deal even with relatively few observations, especially in case when we have an ω-state Markov-switching model. That is because in each iteration of the preceding filter an ω-fold increase is produced in the number of cases to consider. In our model, we have $\omega^T = 2^T$ cases to consider. However, we would like to reduce the dimension of the posteriors in (62) and (63) into (2×1) at the end of each iteration, hence only (2×2) cases to consider for each iteration. The following approximations are employed for this purpose; if $\boldsymbol{\alpha}^{(i,j)}_{t|t}$ in (62) is represented as $E\left[\boldsymbol{\alpha}_t | S_{t-1} = i, S_t = j, \psi_t\right]$, and ψ_t represents information up to time t, it is straightforward to show that

$$\boldsymbol{\alpha}^j_{t|t} = \sum_{i=1}^2 \zeta_t \boldsymbol{\alpha}^{(i,j)}_{t|t} \tag{65}$$

where $\zeta_t = \Pr\left[S_{t-1} = i, S_t = j | \psi_t\right] / \Pr\left[S_t = j | \psi_t\right]$ and $\boldsymbol{\alpha}^j_{t|t}$ would reflect $E\left[\boldsymbol{\alpha}_t | S_t = j, \psi_t\right]$. Hence, the

covariance matrix of $\boldsymbol{\alpha}_t$ conditional on ψ_t and on $S_t = j$ could be derived as follows

$$
\begin{aligned}
\mathbf{P}_{t|t}^{j} &= E\left[(\boldsymbol{\alpha}_t - E[\boldsymbol{\alpha}_t|S_t = j, \psi_t]) \times (\boldsymbol{\alpha}_t - E[\boldsymbol{\alpha}_t|S_t = j, \psi_t])' |S_t = j, \psi_t\right] = E\left[\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}^{j}\right)\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}^{j}\right)'|S_t = j, \psi_t\right] \\
&= \sum_{i=1}^{2} \zeta_t E\left[\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}^{j}\right)\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}^{j}\right)'|S_{t-1} = i, S_t = j, \psi_t\right] \\
&= \sum_{i=1}^{2} \zeta_t E\left[\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}^{(i,j)} + \boldsymbol{\alpha}_{t|t}^{(i,j)} - \boldsymbol{\alpha}_{t|t}^{j}\right)\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}^{(i,j)} + \boldsymbol{\alpha}_{t|t}^{(i,j)} - \boldsymbol{\alpha}_{t|t}^{j}\right)'|S_{t-1} = i, S_t = j, \psi_t\right] \\
&= \sum_{i=1}^{2} \zeta_t \left\{E\left[\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}^{(i,j)}\right)\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}^{(i,j)}\right)'|S_{t-1} = i, S_t = j, \psi_t\right] + \left(\boldsymbol{\alpha}_{t|t}^{j} - \boldsymbol{\alpha}_{t|t}^{(i,j)}\right)\left(\boldsymbol{\alpha}_{t|t}^{j} - \boldsymbol{\alpha}_{t|t}^{(i,j)}\right)'\right\} \\
&\quad + \sum_{i=1}^{2} \zeta_t \left(E[\boldsymbol{\alpha}_t|S_{t-1} = i, S_t = j, \psi_t] - \boldsymbol{\alpha}_{t|t}^{(i,j)}\right)\left(\boldsymbol{\alpha}_{t|t}^{(i,j)} - \boldsymbol{\alpha}_{t|t}^{j}\right)' + \\
&\quad + \sum_{i=1}^{2} \zeta_t \left(\boldsymbol{\alpha}_{t|t}^{(i,j)} - \boldsymbol{\alpha}_{t|t}^{j}\right)\left(E[\boldsymbol{\alpha}_t|S_{t-1} = i, S_t = j, \psi_t] - \boldsymbol{\alpha}_{t|t}^{(i,j)}\right)'
\end{aligned}
$$

$$
= \sum_{i=1}^{2} \zeta_t \left\{E\left[\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}^{(i,j)}\right)\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}^{(i,j)}\right)'|S_{t-1} = i, S_t = j, \psi_t\right] + \left(\boldsymbol{\alpha}_{t|t}^{j} - \boldsymbol{\alpha}_{t|t}^{(i,j)}\right)\left(\boldsymbol{\alpha}_{t|t}^{j} - \boldsymbol{\alpha}_{t|t}^{(i,j)}\right)'\right\} \quad (66)
$$

Here, if $\mathbf{P}_{t|t}^{(i,j)}$ in (63) represents $E\left[\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}^{(i,j)}\right)\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}^{(i,j)}\right)'|S_{t-1} = i, S_t = j, \psi_t\right]$, then (66) could be re-formulated as

$$
\mathbf{P}_{t|t}^{j} = \sum_{i=1}^{\omega} \zeta_t \left\{\mathbf{P}_{t|t}^{(i,j)} + \left(\boldsymbol{\alpha}_{t|t}^{j} - \boldsymbol{\alpha}_{t|t}^{(i,j)}\right)\left(\boldsymbol{\alpha}_{t|t}^{j} - \boldsymbol{\alpha}_{t|t}^{(i,j)}\right)'\right\} \quad (67)
$$

At the end of each iteration, Equations (65) and (67) are employed to "collapse" 2×2 posteriors in (62) and (63) into 2×1 to make the filter appropriate, as in Kim (1993). However these collapsed posteriors involve approximations because $\boldsymbol{\alpha}_{t|t}^{(i,j)}$ and $\mathbf{P}_{t|t}^{(i,j)}$ in (62) and (63) do not calculate $E[\boldsymbol{\alpha}_t|S_{t-1} = i, S_t = j, \psi_t]$ and $E\left[\begin{array}{c}\left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}^{(i,j)}\right) \times \left(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t}^{(i,j)}\right)' | \\ S_{t-1} = i, S_t = j, \psi_t\end{array}\right]$ exactly. This is because $\boldsymbol{\alpha}_t$ conditional on $\psi_{t-1}, S_t = j$ and $S_{t-1} = i$ is a mixture of Normals for $t > 2$. Exactly because this approximation is implemented, the preceding filter is called a *quasi-optimal filter*. Eventually, the last thing that has to be considered to complete the filter is to calculate $\Pr[S_{t-1} = i, S_t = j|\psi_t]$ as well as the other probability terms. Following Hamilton (1988) with a slight modification, we derive the equations below:

$$
\Pr[S_{t-1} = i, S_t = j|\psi_t] = \frac{\Pr[y_t, S_{t-1} = i, S_t = j|\psi_{t-1}]}{\Pr[y_t|\psi_{t-1}]}
$$

$$
= \frac{\Pr[y_t|S_{t-1} = i, S_t = j|\psi_{t-1}] \times \Pr[S_{t-1} = i, S_t = j|\psi_{t-1}]}{\Pr[y_t|\psi_{t-1}]} \quad (68)
$$

where

$$
\Pr[y_t|S_{t-1} = i, S_t = j|\psi_{t-1}] = \frac{1}{\sqrt{2\pi f_{t|t-1}^{(i,j)}}} e^{\left\{-\frac{\left(\eta_{t|t-1}^{i}\right)^2}{2f_{t|t-1}^{(i,j)}}\right\}} \quad (69)
$$

14

$$\Pr\left[y_t|\psi_{t-1}\right] = \sum_{i=1}^{2}\sum_{j=1}^{2}\Pr\left[y_t, S_{t-1}=i, S_t=j|\psi_{t-1}\right] \tag{70}$$

and

$$\Pr\left[S_{t-1}=i, S_t=j|\psi_{t-1}\right] = \Pr\left[S_t=j|S_{t-1}=i\right] \times \Pr\left[S_{t-1}=i|\psi_{t-1}\right] \tag{71}$$

with

$$\Pr\left[S_{t-1}=i|\psi_{t-1}\right] = \sum_{S_{t-2}=1}^{2}\Pr\left[S_{t-2}=s_{t-2}, S_{t-1}=i|\psi_{t-1}\right] \tag{72}$$

Thus Equations (58)-(72) complete the *quasi-optimal filter*. Next, based on the quasi-optimal filter, the *approximated conditional log-likelihood function* can be obtained from (70)

$$\log L = \log\left(\Pr\left[y_1, y_2, \ldots, y_T\right]\right) = \sum_{t=1}^{T}\log\left(\Pr\left[y_t|\psi_{t-1}\right]\right) \tag{73}$$

In order to estimate the parameters of the model, we can maximize the log-likelihood function in Equation (73) with respect to the underlying unknown parameters of the model.

Similarly to the homoskedastic TVP-VAR, the above model can be written in the SUTSE (multivariate) state space form for N variables

$$\mathbf{y}_t = \left(\mathbf{z}' \otimes \mathbf{I}_N\right)\boldsymbol{\alpha}_t + \varepsilon_t \tag{74}$$

$$\boldsymbol{\alpha}_t = \left(\mathbf{T} \otimes \mathbf{I}_N\right)\boldsymbol{\alpha}_{t-1} + \left(\mathbf{R} \otimes \mathbf{I}_N\right)\boldsymbol{\eta}_t \tag{75}$$

yet now $Var(\varepsilon_t) = Var(h_t^j) = \boldsymbol{\Sigma}_{h^j}$ and $Var(\boldsymbol{\eta}_t) = Var(Q_t^j) = \boldsymbol{\Sigma}_{Q^j}$ are block diagonal matrices with the blocks all of them following a 2-state Markov-switching heteroskedastic structure, namely in the four-variate case, the variance of the error components in the state equation is

$$Var\left(h_t^j\right) = \begin{bmatrix} \boldsymbol{\Sigma}_{1h^j} & 0 & 0 & 0 \\ 0 & \boldsymbol{\Sigma}_{2h^j} & 0 & 0 \\ 0 & 0 & \boldsymbol{\Sigma}_{3h^j} & 0 \\ 0 & 0 & 0 & \boldsymbol{\Sigma}_{4h^j} \end{bmatrix} \tag{76}$$

and

$$Var\left(Q_t^j\right) = \begin{bmatrix} \boldsymbol{\Sigma}_{1Q^j} & 0 & 0 & 0 \\ 0 & \boldsymbol{\Sigma}_{2Q^j} & 0 & 0 \\ 0 & 0 & \boldsymbol{\Sigma}_{3Q^j} & 0 \\ 0 & 0 & 0 & \boldsymbol{\Sigma}_{4Q^j} \end{bmatrix} \tag{77}$$

Indeed the SUTSE formulation can be generalized further to allow quantities such as $\mathbf{z}, \mathbf{T}, \mathbf{R}$ and $\boldsymbol{\Sigma}_{h^j}, \boldsymbol{\Sigma}_{Q^j}$ to change over time. The quasi-Kalman filter for the heteroskedastic case may be applied to (74) and (75) with the number of sets of observations needed to form an estimator of $\boldsymbol{\alpha}_t$, and with finite MSE matrix being the same as in the univariate case. Moreover, the conditions for the filter to converge to a steady state define a generalization of the conditions in the univariate case. The decoupling of the Kalman filter is

fully efficient if each equation contains the same regressors. Thus, all the information needed for estimation, prediction and smoothing can be obtained by applying the same univariate filter to each series in turn. If the signal-to-noise ratio is q (i.e., $\mathbf{\Sigma}_{Q^j}/\mathbf{\Sigma}_{h^j} = q$), the Kalman filter for this model is

$$\boldsymbol{\alpha}_{t+1|t}^i = \boldsymbol{\alpha}_{t|t-1}^i + \mathbf{K}_t^{(i,j)}\left(\mathbf{y}_t - \boldsymbol{\alpha}_{t|t-1}^i\right), t = 2,\ldots,T \tag{78}$$

and

$$\mathbf{P}_{t+1|t}^{(i,j)} = \mathbf{P}_{t|t-1}^{(i,j)} - \mathbf{P}_{t|t-1}^{(i,j)}\left(\mathbf{\Sigma}_{Q^j t|t-1}^{(i,j)}\right)^{-1}\mathbf{P}_{t|t-1}^{(i,j)} + q\mathbf{\Sigma}_{h^j} \tag{79}$$

where

$$\mathbf{K}_t^{(i,j)} = \mathbf{P}_{t|t-1}^{(i,j)}\mathbf{Z}_t'\left(\mathbf{\Sigma}_{Q^j t|t-1}^{(i,j)}\right)^{-1} \tag{80}$$

and

$$\mathbf{\Sigma}_{Q^j t|t-1}^{(i,j)} = \mathbf{P}_{t|t-1}^{(i,j)} + \mathbf{\Sigma}_{h^j} \tag{81}$$

Let again w_t denote a positive scalar for $t = 2,\ldots,T$ and suppose that $\mathbf{P}_{t|t-1}^{(i,j)} = w_t\mathbf{\Sigma}_{h^j}$, i.e, the MSE matrix of the $N \times 1$ vector $\boldsymbol{\alpha}_{t|t-1}^i$, is proportional to $\mathbf{\Sigma}_{h^j}$ depending on the whole history of current and past states, $S_0, S_1, \ldots, S_T$. It follows from (79) that $\mathbf{P}_{t+1|t}^{(i,j)}$ is of the same form, that is, $\mathbf{P}_{t+1|t}^{(i,j)} = w_{t+1}\mathbf{\Sigma}_{h^j}$ with $w_{t+1} = (w_t + w_t q + q)/(w_t + 1)$, and if $\mathbf{P}_{t|t-1}^{(i,j)} = w_t\mathbf{\Sigma}_{h^j}$ the gain matrix in (78) is state-dependent diagonal, that is

$$\mathbf{K}_t^{(i,j)} = w_t\mathbf{\Sigma}_{h^j}\left(w_t\mathbf{\Sigma}_{h^j} + \mathbf{\Sigma}_{h^j}\right)^{-1} = \left[w_t/(w_{t+1})\right]\mathbf{I}_N$$

In the heteroskedastic case the above Kalman filter is started off in the same way as in the standard model. However, the use of these starting values now would not lead to an exact likelihood function for $\mathbf{y}_2,\ldots,\mathbf{y}_T$ in the prediction error decomposition form as in (45), but now to a multivariate version of the approximated conditional log-likelihood function of the *quasi-optimal filter*

$$\mathbf{LL} = \log\left(\Pr\left[\mathbf{y}_1, \mathbf{y}_{t2},\ldots,\mathbf{y}_T\right]\right) = \sum_{t=1}^T \log\left(\Pr\left[\mathbf{y}_t|\psi_{t-1}\right]\right) \tag{82}$$

based now on the following probabilities instead of the (69)-(70) for the univariate filter

$$\Pr\left[\mathbf{y}_t|S_{t-1} = i, S_t = j|\psi_{t-1}\right] = \frac{1}{\sqrt{2\pi\mathbf{\Sigma}_{Q^j t|t-1}^{(i,j)}}}e^{\left\{-\frac{\left(f_{t|t-1}^i\right)^2}{2\mathbf{\Sigma}_{Q^j t|t-1}^{(i,j)}}\right\}} \tag{83}$$

$$\Pr\left[\mathbf{y}_t|\psi_{t-1}\right] = \sum_{i=1}^2\sum_{j=1}^2\Pr\left[\mathbf{y}_t, S_{t-1} = i, S_t = j|\psi_{t-1}\right] \tag{84}$$

The predictions of future observations are obtained from the univariate recursions (as in the standard model)

$$\text{MSE}\left(\widehat{\mathbf{y}}_{T+l|T}\right) = f_{T+l|T}\mathbf{\Sigma}_{h^j}, \ l = 1, 2,\ldots \text{ and } p_{ij} = \Pr\left[S_t = j|S_{t-1} = i\right], i, j = 1, 2 \tag{85}$$

Finally, the decoupling of the Kalman filter can be extended in a similar way for a time-varying system

with a 2-state Markov-switching model of heteroskedasticity, as in Bekiros and Paccagnini (2013)

$$\mathbf{y}_t = (\mathbf{z}_t' \otimes \mathbf{I}_N)\,\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \quad \text{with} \quad Var(\varepsilon_t) = Var(h_t^j)\boldsymbol{\Sigma}_* = \boldsymbol{\Sigma}_{h^j}\boldsymbol{\Sigma}_* \tag{86}$$

$$\boldsymbol{\alpha}_t = (\mathbf{T}_t \otimes \mathbf{I}_N)\,\boldsymbol{\alpha}_{t-1} + (\mathbf{R}_t \otimes \mathbf{I}_N)\,\boldsymbol{\eta}_t, \quad \text{with} \quad Var(\boldsymbol{\eta}_t) = Var(Q_t^j) \otimes \boldsymbol{\Sigma}_* = \boldsymbol{\Sigma}_{Q^j} \otimes \boldsymbol{\Sigma}_* \tag{87}$$

where again $Var(\varepsilon_t) = Var(h_t^j) = \boldsymbol{\Sigma}_{h^j}$ and $Var(\boldsymbol{\eta}_t) = Var(Q_t^j) = \boldsymbol{\Sigma}_{Q^j}$ are block diagonal matrices, although a more general formulation does not constrain them to be diagonal. However, as in the univariate model, restrictions are needed on the matrices for the model to be identifiable. The previous results on estimation and prediction apply, with $\mathbf{P}_{t+1|t}^{(i,j)*} = \mathbf{P}_{t+1|t}^{(i,j)} \otimes \boldsymbol{\Sigma}_*$, where $\mathbf{P}_{t+1|t}^{(i,j)}$ is the MSE matrix for the univariate model.

5 DSGE estimation procedure

We estimate the DSGE with banking intermediation model using Bayesian methods as in Smets and Wouters (2007). The data we use in the estimation is in the form of 20-year rolling windows (80 quarter observations). Rolling window estimates may help capturing changes in parameters (regime shifts) as discussed in Gürkaynak et al. (2013). The first estimation period is from 1984Q4 to 2003Q4 and the last one is from 1994Q1 to 2013Q4, for a total of 41 samples. The model is estimated for the United States using the following variables: GDP, investment, consumption, wages, net worth of financial intermediaries, hours of work, GDP deflator inflation and the federal funds rate. We include net worth of financial intermediaries as a financial observable because the model features a net worth shock. Although observations on all variables are available at least from 1973Q2 onward, we concentrate on this period because it is characterized by a single monetary policy regime. Appendix A contains a detailed discussion of data sources, definitions and transformations. The following set of measurement equations shows the link between the observables in the dataset and the endogenous variables of the DSGE model

$$\begin{bmatrix} \Delta Y_t^o \\ \Delta C_t^o \\ \Delta I_t^o \\ \Delta W_t^o \\ \Delta N_t^o \\ L_t^o \\ \pi_t^o \\ r_t^{n,o} \end{bmatrix} = \begin{bmatrix} \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma}^N \\ \bar{l} \\ \bar{\pi} \\ \bar{r}^n \end{bmatrix} + \begin{bmatrix} \hat{Y}_t - \hat{Y}_{t-1} \\ \hat{C}_t - \hat{C}_{t-1} \\ \hat{I}_t - \hat{I}_{t-1} \\ \hat{W}_t - \hat{W}_{t-1} \\ \hat{N}_t - \hat{N}_{t-1} \\ \hat{L}_t \\ \hat{\Pi}_t \\ \hat{R}_t^n \end{bmatrix} \tag{88}$$

where $\bar{\gamma} = 100(\gamma - 1)$ is the common quarterly trend growth rate of GDP, consumption, investment and wages; $\bar{\gamma}^N = 100(\gamma^N - 1)$ is the quarterly trend growth rate of net worth of financial intermediaries, as in Gelain and Ilbas (2014); $\bar{h}$ is the steady-state hours of work; $\bar{\pi}$ is the steady-state quarterly inflation rate; and $\bar{r}^n$ is the steady-state quarterly nominal interest rate. A hat over a variable represents log-deviation from steady state.

5.1 Evolution of estimated shocks and parameters

We report the evolution of the estimated shocks and parameters of the DSGE model based on rolling estimation sample starting from 1984Q1-2003Q4 and ending with the sample 1994Q1- 2013Q4.

Figure 2: Evolution of the shock processes
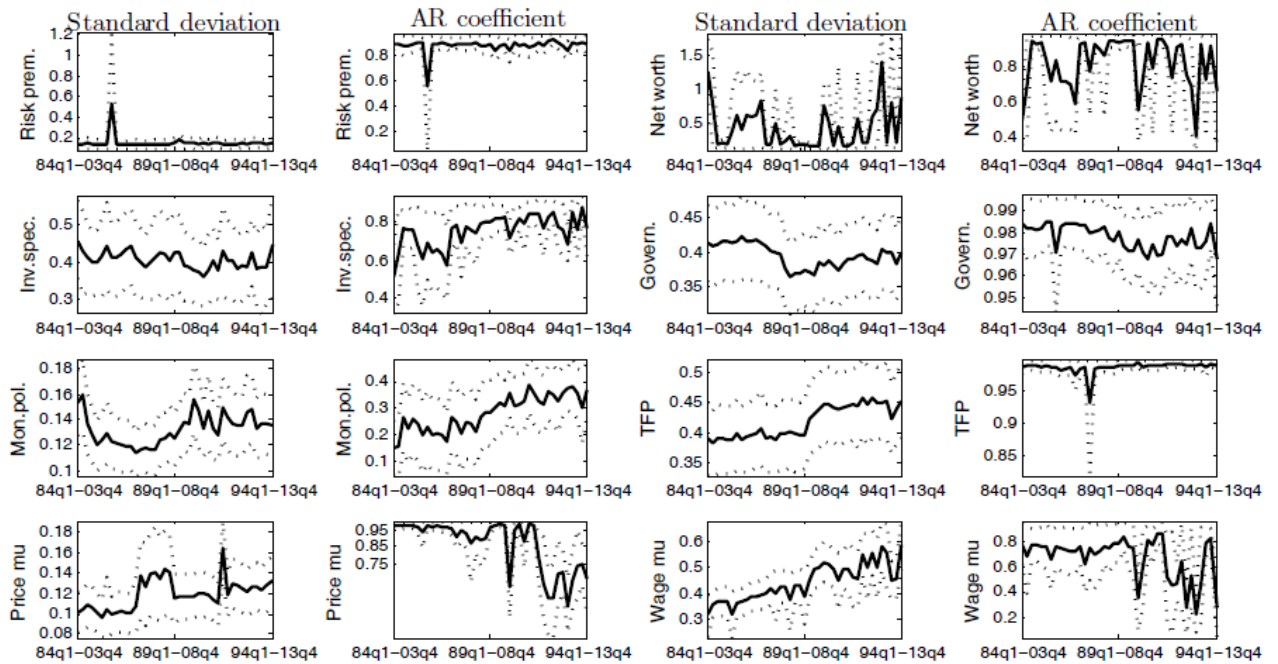
Note: Solid lines represent the posterior mean, while dotted lines confidence intervals. Estimates are computed based on rolling estimation sample starting from 1984Q1-2003Q4 and ending with the sample 1994Q1- 2013Q4.

Figure 3: Evolution of the most relevant parameters
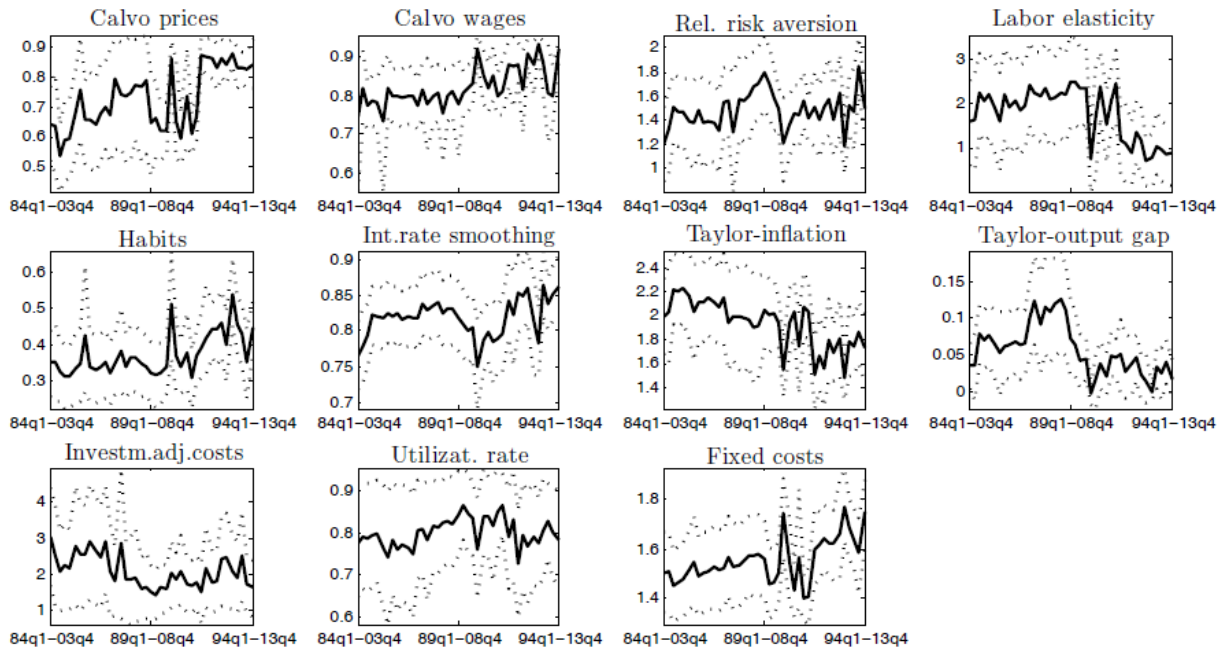
Note: Solid lines represent the posterior mean, while dotted lines confidence intervals. Estimates are computed based on rolling estimation sample starting from 1984Q1-2003Q4 and ending with the sample 1994Q1- 2013Q4.

Figures 2 and 3 show that the time variation of the parameters is crucial in our empirical analysis.

18

As discussed in Cardani *et al.* (2014), the literature offers at least three different approaches to deal with the issue of parameters instability. The first features time-varying coefficients / stochastic volatilities (Fernandez-Villaverde *et al.*, 2010; Caldara *et al.*, 2012; Bekiros and Paccagnini, 2013). Second, Eo (2009) and Foerster *et al.* (2014), among others, propose Markov-switching DSGE modeling. Third, Castelnuovo (2012) presents a rolling-window estimation for detecting instabilities in the structural parameters during a long period with different monetary regimes, i.e. the years between 1965 and 2005. We follow the last methodology, which has the advantage to be applied to a wide set of parameters while it does not force the data to "discretize" the economy (Castelnuovo, 2012) – differently from stochastic volatility approaches.

6   Comparative analysis of predictability

The DSGE model is estimated using a rolling window of 80 observations. We assess forecastability for multi-step horizons $h \in (1, 2, 3, 4, 8)$. The first estimation sample 1984Q1-2003Q4 produces an out-of-sample period starting in 2004Q1 and ending in 2012Q1 for the one-step-ahead, ending in 2012Q2 for two step-ahead forecasts etc., and eventually until 2013Q4 for the eight step-ahead forecasts. We compare the out-of-sample forecasting performance of VAR, BVAR and the homoskedastic and heteroskedastic MVSS-TVP-VARs as well as of the DSGE model in terms of the Root Mean Squared Forecast Error (RMSFE) for the optimal lag specifications (one to four) selected by the Schwartz Bayesian information criterion (SIC). The forecasting investigation for the quarterly US data is performed over the one-, two-, three-, four and eight–quarter-ahead horizon with a rolling estimation sample, based on the works of Marcellino (2004) and Brüggemann *et al.* (2008) for datasets of quarterly frequency. In particular, the models are re-estimated each quarter over the forecast horizon to update the estimate of the coefficients, before producing the quarter-ahead forecasts. The models are comparatively evaluated for the United States using the following observables: GDP, the GDP deflator inflation (INF), the federal funds rate (FFR) and the net worth of financial intermediaries (NWB). We include net worth of financial intermediaries as a financial observable because the model features a net worth shock.

The RMSFE scores for the out-of-sample period are reported in Table 1 for all models and variables. An exhaustive exercise was conducted on VAR and BVAR models with one to four lags based on the Schwartz Bayesian information criterion (SIC). The results provide evidence that in general four lags is the optimal number for these models. Moreover, the SIC for one to four lags (MLE and QMLE estimation) was implemented in order to select the best specification for the MVSS-TVP-VAR and the heteroscedastic MVSS-TVP-MSVAR. In both cases one lag was chosen.

In particular, for the GDP series the MVSS-TVP-VAR clearly outperforms all models for all steps-ahead. Next, the MVSS-TVP-MSVAR model is the best for all steps-ahead except only for the longest horizon where the BVAR presents the lowest RMSFE. In general, the DSGE model provides the worst out-of-sample behaviour compared to the simple VAR and BVAR models. Interestingly, in case of INF the DSGE model with financial frictions and banking intermediation achieves the best score for the first three horizons, namely one-, two- and three-quarters-ahead, thus clearly outperforms all other models. The next lowest RMSFEs are produced by the MVSS-TVP-VAR and standard VAR models. However, considering the longest horizons i.e., four- and eight-steps-ahead the homoskedastic and heteroskedastic TVP-VAR respectively outrank the DSGE model. This result might be an indication of the match of two states/regimes for the MVSS-TVP-VAR models. Possibly the MVSS-TVP-MSVAR because it picks out the crisis period as the "high volatility" regime and the pre-crisis period as the low regime. Hence, the heteroscedastic TVP-VAR attributes the crisis period to the "high volatility" state especially within 2009-

2013 and hence shows a lower RMSFE compared to the "plain vanilla" TVP-VAR in the long forecasting horizon of eight quarters. Consistently, the exact same "picture" emerges from the investigation of the FFR predictability. The novel DSGE model attains the best score for the first three horizons, whilst the two TVP-VARs show the lowest RMSFE for four- and eight-steps-ahead. The next best performer is the BVAR compared to the simple VAR. Similarly, as in the case of INF, the two TVP-VARs produce very close scores when compared to each other. Finally, for the NWB series the plain TVP-VAR is the outranking model for all steps-ahead except for the long eight-quarters-ahead projection where the two-regime TVP-MSVAR emerges as the best performer. Interestingly, the DSGE model with financial frictions generates the highest RMSFEs among the other specifications. The VAR is better than the BVAR model, yet marginally in most forecasting horizons.

Table 1: Root Mean Square Forecast Error (RMSFE) for GDP, INF, FFR and NWB

	VAR	BVAR	DSGE	MVSS-TVP-VAR	MVSS-TVP-MSVAR
GDP					
1	0.7879	0.7891	1.0290	0.5864	0.7087
2	0.7808	0.8056	1.0760	0.6706	0.7083
3	0.9022	0.7932	1.1442	0.5870	0.6416
4	0.8627	0.8116	1.1900	0.5283	0.5746
8	1.5286	1.4556	2.6628	1.0004	1.7117
INF					
1	0.2652	0.3228	0.2030	0.2415	0.3093
2	0.2999	0.3451	0.2096	0.2236	0.2838
3	0.2932	0.3356	0.2063	0.2075	0.2750
4	0.2504	0.3152	0.2123	0.1837	0.2588
8	0.6883	0.5219	0.4967	0.5165	0.4680
FFR					
1	0.6318	0.6147	0.0971	0.2063	0.2216
2	0.7599	0.6094	0.1831	0.2763	0.2880
3	0.9068	0.5975	0.2593	0.3015	0.3329
4	0.7419	0.5799	0.3321	0.3054	0.3386
8	0.9975	0.8696	0.9505	0.6898	0.6716
NWB					
1	2.1596	2.5409	3.4061	2.0952	2.2386
2	2.1398	2.4797	3.1016	2.1315	2.1902
3	2.3212	2.4525	2.8447	2.1665	2.2258
4	2.2072	2.4303	2.6475	2.1936	2.3433
8	6.9893	10.4068	15.9382	7.4460	6.8902

The results for the out-of-sample behavior of the models during the financial crisis period concerning inflation and federal funds rate seem to be in accordance with the works by Del Negro and Schorfheide (2009; 2012) and Wolters (2013). The superiority of the TVP-VAR models for GDP and NWB against basically the DSGEs can be attributed to the fact that the latter lack a good calibration in particular within crisis times. In normal times these models exhibit a balanced forecasting performance. However, in crisis times their predictability appears to weaken. Possibly, this is an after-effect of the imposition of tight restrictions on the data by the simple DSGEs. If the data rejects these restrictions, large stochastic shocks are needed to fit the model to the dataset which results in high shock uncertainty. As mentioned in Wolters (2013), DSGEs provide better results during normal times. Under average exogenous shocks the models return back

to a steady state, albeit they could not predict recessions and booms as significantly larger exogenous shocks are required to capture these. As TVP-VARs relax these restrictions, misspecification can be absorbed by time-varying parameters and the estimated variance of shocks is lower, which results in somewhat tighter predictive distributions. One potential explanation is that the forecasts are generated mostly after the Great Moderation period whilst the estimation sample covers this period or before. According to Del Negro and Schorfheide (2012) the shock standard deviations are estimated to capture the average of the pre- and post-moderation volatility a fact that leads to overprediction of the volatility during the forecast period.

7   Conclusions

In the aftermath of the Great Recession, the DSGE literature has extensively focused on the links between financial and real sectors. Among the different contributions on DSGE models featuring a banking sector, we follow Gertler and Karadi (2011). In such a model financial intermediaries face an endogenous balance sheet constraint due to the presence of asymmetric information.

Rolling-window DSGE estimation and modelling usually fails to take into account inherent nonlinearities of the economy, which are evident also from our estimation of shocks and structural parameters of the DSGE model. The use of time-varying parameters seems to be an attractive alternative as well as in terms of capturing nonlinear economic relationships (Primiceri, 2005). We propose a novel time-varying multivariate state-space estimation method for TVP-VAR processes both for homoskedastic and heteroskedastic error structures. As an alternative to the homoskedastic TVP-VAR we assume that the error structure of the state space Kalman filter is dependent on state variables, which are unobserved discrete-time, discrete-state Markov process, thus providing a Markov-switching heteroskedasticity.

The DSGE model for the US economy features financial intermediaries as in Gertler and Karadi (2011) in an otherwise standard setup à la Smets and Wouters (2007). We employ rolling-window DSGE estimation as well as TVP-VARs to account for parameter instabilities, which can be particularly relevant towards the end of our sample – which spans from 1984Q1 to 2013Q4. In addition, we conduct a comprehensive empirical exercise to compare the out-of-sample predictive performance of the estimated DSGE model with that of standard AR, VARs, Bayesian VARs as well as of two time-varying parameter autoregressive models (TVP-VAR) models with homoskedastic and heteroskedastic errors in an attempt to investigate inherent nonlinearities of the economy that cannot be captured by the VAR and DSGE class models. The main purpose of the paper is to compare different econometric strategies in evaluating a DSGE economy, but mainly to stress the importance of considering the banking intermediation in particular for the US economy during and after the recent financial crisis, and their incorporation in DSGE and TVP-VAR models. Aside from all standard observables, we include the net worth of financial intermediaries as a financial observable because the model features a net worth shock.

The main results are as follows. The best forecasting performance for the GDP series is produced by the TVP-VAR which clearly outperforms all models for all steps-ahead. The DSGE model provides the worst out-of-sample behavior compared to the other models. Instead, in case of inflation and federal funds rate the DSGE model with financial frictions and banking intermediation achieves the best score for the three short-term examined horizons, albeit for the longest horizons of four- and eight-steps-ahead the homoskedastic and heteroskedastic TVP-VAR respectively and marginally outrank the DSGE model. One explanation could be that the Markov-Switching heteroscedastic TVP-VAR attributes the crisis period to the "high volatility" state especially within 2009-2013 and hence shows a lower RMSFE compared to the

plain TVP-VAR and DSGE in the long forecasting horizon of eight quarters. Finally, for the net worth of financial intermediaries series the plain TVP-VAR is the outranking model for all steps-ahead except for the long eight-quarters-ahead projection where the two-regime TVP-VAR emerges again as the best performer. Interestingly, the TVP-VAR outranks the other models, whilst the DSGE model generates the highest RMSFE among the other specifications. As proposed by the recent literature of forecasting with DSGE models (Wolters, 2015; Kolasa and Rubaszek; 2015), an avenue for future research in our comparison exercise is to assess the actual uncertainty of the estimated parameters by means of the density forecast. Eventually, a first attempt is made via our work to find macro-financial micro-founded DSGE models as well as adaptive TVP-VARs, which are able to deal with financial instabilities via incorporating banking intermediation.

References

Bean C (2010) Joseph Schumpeter lecture the great moderation, the great panic, and the great contraction, Journal of the European Economic Association, 8(2-3), 289-325

Bekiros S., Paccagnini A., (2013) On the predictability of time-varying VAR and DSGE models, Empirical Economics, 45 (1), 635-664

Bekiros S., Paccagnini A., (2014) - Macroprudential policy and forecasting using Hybrid DSGE models with financial frictions and State space Markov-Switching TVP-VARs., Macroeconomic Dynamics, doi: 10.1017/S1365100513000953

Bernanke B, Gertler M, Girlstichr S, (1999) The financial accelerator in a quantitative business cycle framework in: J. B. Taylor & M. Woodford (ed.), Handbook of Macroeconomics, edition 1, volume 1, chapter 21, pp. 1341-1393

Brüggemann R, Lütkepohl H, Marcellino M (2008) Forecasting euro area variables with German pre-EMU data, Journal of Forecasting, 27(6), 465–481

Caldara D, Fernandez-Villaverde J, Rubio-Ramirez JF, Yao W (2012) Computing DSGE models with recursive preferences and stochastic volatility, Review of Economic Dynamics, 15(2) 188-206

Canova F (2009) What Explains in the Great Moderation in the US? A Structural Analysis, Journal of the European Economic Association, 7(4), 697–721.

Cardani R, Paccagnini A, Villa S (2015) Forecasting in a DSGE model with banking intermediation: Evidence from the US. University of Milano-Bicocca working paper no. 292.

Castelnuovo E (2012) Fitting U.S. Trend Inflation: A Rolling-Window Approach. In Balke, N., Canova, F., Milani, F., and Wynne, M., editors, Advances in Econometrics: DSGE Models in Macroeconomics - Estimation, Evaluation, and New Developments, 201-252

Cole H (2011) Discussion of Gertler and Karadi: A model of unconventional monetary policy, Journal of Monetary Economics, 58(1), 35-38

Curdia V, Woodford M (2010) Credit Spreads and Monetary Policy , Journal of Money, Credit and Banking, Blackwell Publishing, vol. 42(s1), pages 3-35, 09

Del Negro M, Schorfheide F (2009) Monetary Policy Analysis with Potentially Misspecified Models, American Economic Review, 99, 1415–50.

Del Negro M, Schorfheide F, (2012) DSGE Model-Based Forecasting, Handbook of Economic Forecasting, Volume 2

Doan T, Litterman R, Sims C (1984) Forecasting and Conditional Projections Using Realistic Prior Distributions, Econometric Reviews, 3, 1-100

Dua P, Ray SC (1995) A BVAR model for the Connecticut Economy, Journal of Forecasting, 14(3), 167-180

Durbin J, Koopman SJ (2002) Time Series Analysis by State Space Methods, Oxford University Press

Eo Y (2009) Bayesian Analysis of DSGE Models with Regime Switching. MPRA Paper, 13910

Fernández-Villaverde J, Guerron-Quintana PA, Rubio-Ramírez J (2010) The New Macroeconometrics: A Bayesian Approach, in A. O'Hagan and M. West, eds., Handbook of Applied Bayesian Analysis, Oxford: Oxford University Press

Foerster AT, Rubio-Ramirez JF, Waggoner DF, Zha TA (2014) Perturbation methods for Markov-switching DSGE models. NBER Working Papers, 20390

Galati G, Moessner R (2011) Macroprudential policy-a literature review, BIS Working Paper, 337

Gelain P, Ilbas P (2014) Monetary and macroprudential policies in an estimated model with financial intermediation, NBB Working Paper, 258

Gerali A, Neri S, Sessa L, Signoretti F, (2010) Credit and Banking in a DSGE Model of the Euro Area. Journal of Money, Credit and Banking, 42:107-141

Gertler M, Karadi, P, (2011) A Model of Unconventional Monetary Policy, Journal of Monetary Economics, Elsevier, vol. 58(1), pages 17-34

Giraitis L, Kapetanios G, Theodoridis K, Yates T, (2014) Estimating time-varying DSGE models using minimum distance methods, Bank of England working paper no. 507.

Goodfriend M, McCallum BT, (2007) Banking and interest rates in monetary policy analysis: A quantitative exploration, Journal of Monetary Economics, Elsevier, vol. 54(5), pages 1480-1507

Gupta R, Kabundi A (2010) Forecasting Macroeconomic Variables in a Small Open economy: A Comparison between Small- and Large-Scale Models, Journal of Forecasting, 29, 168-185

Gurkaynak RS, Kisacikoglu B, Rossi B (2013) Do DSGE Models Forecast More Accurately Out-of-Sample than VAR Models? CEPR Discussion Papers, (9576)

Hamilton, J (1988) Rational-Expectation Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates, Journal of Economic Dynamics and Control, 12, 385-423.

Harvey AC (1990) Forecasting, structural time series and the Kalman filter, Cambridge University Press

Harvey AC (1986) Analysis and generalization of a multivariate exponential smoothing model, Management Science, 32, 374-380

Ingram B, Whiteman C (1994) Supplanting the Minnesota Prior - Forecasting Macroeconomics Time Series using Real Business Cycle Model Priors, Journal of Monetary Economics, 34, 497-510

Iacoviello M, (2005) House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle, American Economic Review, American Economic Association, vol. 95(3), pages 739-764, June

Kadiyala KR, Karlsson S (1997) Numerical Methods for Estimation and Inference in Bayesian VAR-Models, Journal of Applied Econometrics, 12(2), 99-132

Kalman RE (1960) A New Approach to Linear Filtering and Prediction. Journal of Basic Engineering, Transactions of the ASME Series D 82, 35-45

Kalman RE (1963) New Methods in Wiener Filtering Theory. In: Bogdanoff J.L., Kozin F. (eds.) Proceedings of the First Symposium of Engineering Applications of Random Function Theory and Probability, Wiley, New York, 270-388

Kim C-J (1993) Unobserved-Component Time Series Models with Markov-Switching Heteroscedasticity: Changes in Regime and the Link between Inflation Rates and Inflation Uncertainty, Journal of Business & Economic Statistics, 11 (3), 341-349

Kim C-J, Nelson CR (1999a) Has The U.S. Economy Become More Stable? A Bayesian Approach Based On A Markov-Switching Model Of The Business Cycle, The Review of Economics and Statistics, MIT Press, 81(4), 608-616

Kim C-J, Nelson CR (1999b) State space models with Regime switching, MIT Press

Kolasa M and Rubaszek M (2015). Forecasting with DSGE models with financial frictions. International Journal of Forecasting, 31(1):1-19.

Lamoureux CG, Lastrapes W (1990) Persistence in Variance, Structural Change, and the GARCH Model, Journal of Business & Economic Statistics, 8, 225- 234

Lastrapes WD (1989) Exchange Rate Volatility and U.S. Monetary Policy: An ARCH Application, Journal of Money, Credit, and Banking, 21, 66-77

Litterman RB (1981) A Bayesian Procedure for Forecasting with Vector Autoregressions, Working Paper, Federal Reserve Bank of Minneapolis

Litterman RB (1986) Forecasting with Bayesian Vector Autoregressions: Five Years of Experience, Journal of Business and Statistics 4(1), 25–38

Liu GD, Gupta R, Schaling E (2009) A New-Keynesian DSGE Model for Forecasting the South African Economy, Journal of Forecasting, 28, 387-404

Marcellino M (2004) Forecasting EMU macroeconomic variables, International Journal of Forecasting, 20, 359–372

Melina G, Villa S (2014) Fiscal policy and lending relationships, Economic Inquiry, 52(2), 696-712

Primiceri GE (2005) Time Varying Structural Vector Autoregressions and Monetary Policy, The Review of Economic Studies, 72, 821-852

Sims CA (1980) Macroeconomics and reality, Econometrica, 48(1), 1-48

Sims CA, Zha T (1998) Bayesian Methods for Dynamic Multivariate Models, International Economic Review, 39, 949-968

Smets F, Wouters R (2007) Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach, American Economic Review 97 (3), 586–606

Spencer DE (1993) Developing a Bayesian Vector Autoregression Forecasting Model, International Journal of Forecasting 9(3), 407–421

Todd RM (1984) Improving Economic Forecasting with Bayesian Vector Autoregression, Quarterly Review, Federal Reserve Bank of Minneapolis

Villa S (2014) Financial frictions in the Euro Area and the United States: a Bayesian assessment, Macro-economic Dynamics. Forthcoming

Wolters MH (2015) Evaluating point and density forecasts of DSGE models, Journal of Applied Econometrics. Forthcoming

Zellner A (1963) Estimation for seemingly unrelated regression equations: some exact finite sample results, Journal of the American Statistical Association, 58, 977-992

8 Appendix

A Appendix: Data sources and transformations

This section discusses the sources of the eight observables used in the estimation and their transformation. GDP, GDP deflator inflation, the federal funds rate, civilian population (CNP160V) and civilian employment (CE160V) are downloaded from the ALFRED database of the Federal Reserve Bank of St. Louis. Private consumption expenditures and fixed private investment are extracted from the NIPA Table 1.1.5 of the Bureau of Economic Analysis. Net worth of banks is downloaded from the FRED database and it is computed as the difference between total assets of all commercial banks (TLAACBW027SBOG) and total liabilities of all commercial banks (TLBACBM027SBOG). Average weekly hours worked (PRS85006023) and compensation per hour (PRS85006103) are downloaded from the Bureau of Labor Statistics.

Data are transformed as in Smets and Wouters (2007). In particular, GDP, consumption, investment and net worth are transformed in real per-capita terms by dividing their nominal values by the GDP deflator and the civilian population. Real wages are computed by dividing compensation per hour by the GDP deflator. As shown in the measurement equations, the observable variables of GDP, consumption, investment, wages and net worth are expressed in first differences. Hours worked are multiplied by civilian employment, expressed in per capita terms and demeaned. The inflation rate is computed as a quarter-on-quarter difference of the log of the GDP deflator. The fed funds rate is expressed in quarterly terms. Remaining variables are expressed in 100 times log. All series are seasonally adjusted.