Estimation of a TVP-VAR with MSV via Efficient Importance Sampling

Project Proposal

Aishameriane Schmidt Advisor Prof. Guilherme Valle Moura

> UFSC - PPGECO Seminários de Dissertação

November, 2017.



Using VAR models to describe macroeconomic relations

"It should be feasible to estimate large-scale macromodels as unrestricted reduced forms, treating all variables as endogenous."

Christopher Sims, Macroeconomics and Reality (1980).

Using VAR models to describe macroeconomic relations

"It should be feasible to estimate large-scale macromodels as unrestricted reduced forms, treating all variables as endogenous."

Christopher Sims, Macroeconomics and Reality (1980).

► RBC models x VAR approach

Using VAR models to describe macroeconomic relations

"It should be feasible to estimate large-scale macromodels as unrestricted reduced forms, treating all variables as endogenous."

Christopher Sims, Macroeconomics and Reality (1980).

- ► RBC models x VAR approach
- Strong evidence of parameter instability in macroeconomic series:
 [Sims, 1999], [Bernanke and Mihov, 1998a], [Bernanke and Mihov, 1998b],
 [Mumtaz and Zanetti, 2015], [Galí and Gambetti, 2015],
 [Mumtaz and Theophilopoulou, 2015],
 [Mumtaz and Theophilopoulou, 2017], [Cogley and Sargent, 2001],
 [Cogley and Sargent, 2005], etc.

Cogley and Sargent (2005)

WHAT EXPLAINS THE WORSENING IN US WORSENING IN US INFLATION-UNEMPLOYMENT OUTCOMES IN 19705?



Cogley and Sargent (2005)

WHAT EXPLAINS THE WORSENING IN US WORSENING IN US WORSENING IN 1970S? INFLATION-UNEMPLOYMENT OUTCOMES IN 1970S?

Bad Luck vs Bad Policy

Changes in the inflation persistence

Changes in volatility structure

Bad luck:

It was not monetary strategy but increase/decrease of the shocks' volatility

Changes in monetary policy rules

Drifting coefficients

Bad policy:

The FED's view negates the natural rate theory and believes in an exploitable trade-off between unemployment and inflation

Based on Cogley and Sargent, 2005.

Cogley and Sargent (2005)

WHAT EXPLAINS THE WORSENING IN US WORSENING IN US WORSENING IN 1970S?

Bad Luck vs Bad Policy

Changes in the inflation persistence

Changes in volatility structure

Bad luck:

It was not monetary strategy but increase/decrease of the shocks' volatility

Changes in monetary policy rules

Drifting coefficients

Bad policy:

The FED's view negates the natural rate theory and believes in an exploitable trade-off between unemployment and inflation

Based on Cogley and Sargent, 2005.

Cogley and Sargent (2005)

WHAT EXPLAINS THE WORSENING IN US WORSENING IN US WORSENING IN 1970S? INFLATION-UNEMPLOYMENT OUTCOMES IN 1970S?

Bad Luck vs Bad Policy

Changes in the inflation persistence

Changes in volatility structure

Bad luck:

It was not monetary strategy but increase/decrease of the shocks' volatility

Changes in monetary policy rules

Drifting coefficients

Bad policy:

The FED's view negates the natural rate theory and believes in an exploitable trade-off between unemployment and inflation

Based on Cogley and Sargent, 2005.

Contents



Contents

Economic Motivation

TVP-VAR + MSV

Cogley and Sargent (2001, 2005)

Primiceri (2005)

Phillipov and Glickman (2006a, 2006b)

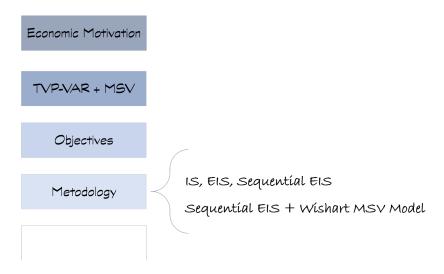
Contents



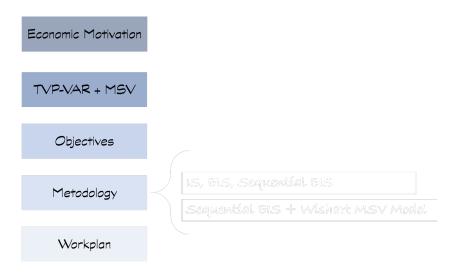


IS, FIS, Sequential FIS Sequential FIS + Wishart MSV Model

Contents



Contents



Literature Review: TVP-VAR w/ MSV

Consider the following state-space representation of a TVP-VAR with MSV:

$$y_t = Z_t \alpha_t + \epsilon_t$$
 $\epsilon_t \sim \mathcal{N}_k(0_k, \Omega_t^{-1})$ (measure eq.) (1)
 $\alpha_t = \alpha_{t-1} + \nu_t$ $\nu_t \sim \mathcal{N}_p(0_p, Q)$ (state transition eq.) (2)

Literature Review: TVP-VAR w/ MSV

Consider the following state-space representation of a TVP-VAR with MSV:

$$y_t = Z_t \alpha_t + \epsilon_t$$
 $\epsilon_t \sim \mathcal{N}_K(0_k, \Omega_t^{-1})$ (measure eq.) (1)
 $\alpha_t = \alpha_{t-1} + \nu_t$ $\nu_t \sim \mathcal{N}_P(0_p, Q)$ (state transition eq.) (2)

Some remarks:

► We don't observe α_t directly (**latent variable**);

Literature Review: TVP-VAR w/ MSV

Consider the following state-space representation of a TVP-VAR with MSV:

$$y_t = Z_t \alpha_t + \epsilon_t$$
 $\epsilon_t \sim \mathcal{N}_K(0_K, \Omega_t^{-1})$ (measure eq.) (1)
 $\alpha_t = \alpha_{t-1} + \nu_t$ $\nu_t \sim \mathcal{N}_P(0_P, Q)$ (state transition eq.) (2)

Some remarks:

- We don't observe α_t directly (**latent variable**);
- ▶ If Ω_t^{-1} were non-stochastic, the Kalman Filter could be used to obtain the ML function;

Literature Review: TVP-VAR w/ MSV

Consider the following state-space representation of a TVP-VAR with MSV:

$$y_t = Z_t \alpha_t + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}_k(0_k, \Omega_t^{-1})$$
 (measure eq.) (1)

$$\alpha_t = \alpha_{t-1} + \nu_t$$
 $\nu_t \sim \mathcal{N}_p(0_p, Q)$ (state transition eq.) (2)

Some remarks:

- We don't observe α_t directly (**latent variable**);
- ▶ If Ω_t^{-1} were non-stochastic, the Kalman Filter could be used to obtain the ML function;
 - Since it is not the case, we have an analytically intractable high-dimensional integral.

Literature Review: Cogley and Sargent (2001, 2005)

Their first work was a TVP-VAR of the form:

$$y_t = Z_t \alpha_t + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}_k(0_k, R)$$
 (3)

$$\alpha_t = \alpha_{t-1} + u_t \qquad u_t \sim \mathcal{N}_p(0_p, Q) \tag{4}$$

Literature Review: Cogley and Sargent (2001, 2005)

Their first work was a TVP-VAR of the form:

$$y_t = Z_t \alpha_t + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}_k(0_k, R)$$

$$\alpha_t = \alpha_{t-1} + u_t \qquad u_t \sim \mathcal{N}_p(0_p, Q)$$
(4)

$$\alpha_t = \alpha_{t-1} + u_t \qquad u_t \sim \mathcal{N}_p(0_p, Q) \tag{4}$$

After being criticized in (2001) for the lack of MVS in their model, [Cogley and Sargent, 2005] proposed changing ϵ_t in (3) for $\epsilon_t = R_t^{1/2} \xi_t$, where ξ_t follows a standard normal distribution and $R_t = B^{-1}H_tB^{-1}$ with:

$$B = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \beta_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1} & \beta_{k2} & \cdots & 1 \end{bmatrix}, \text{ and } H_t = \begin{bmatrix} h_{1t} & 0 & \cdots & 0 \\ 0 & h_{2t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{kt} \end{bmatrix}$$

where
$$ln(h_{it}) = ln(h_{it-1}) + \sigma_i \eta_{it}$$
, $\eta_{it} \sim \mathcal{N}(0, 1)$.

Literature Review: Primiceri (2005)

[Primiceri, 2005] proposed the same specification as [Cogley and Sargent, 2005] but allowed the covariances in *B* to evolve over time:

$$B_{t} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \beta_{21,t} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1,t} & \beta_{k2,t} & \cdots & 1 \end{bmatrix} \quad \text{with} \quad \beta_{t} = \beta_{t-1} + \nu_{t}$$
 (5)

where v_t are mean-zero normal errors with constant covariance matrix.

Literature Review: Primiceri (2005)

[Primiceri, 2005] proposed the same specification as [Cogley and Sargent, 2005] but allowed the covariances in *B* to evolve over time:

$$B_{t} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \beta_{21,t} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1,t} & \beta_{k2,t} & \cdots & 1 \end{bmatrix} \quad \text{with} \quad \beta_{t} = \beta_{t-1} + \nu_{t}$$
 (5)

where v_t are mean-zero normal errors with constant covariance matrix.

Drawback: (5) is not invariant to the observable variables y_t order!

Literature Review: Primiceri (2005)

[Primiceri, 2005] proposed the same specification as [Cogley and Sargent, 2005] but allowed the covariances in *B* to evolve over time:

$$B_{t} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \beta_{21,t} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1,t} & \beta_{k2,t} & \cdots & 1 \end{bmatrix} \quad \text{with} \quad \beta_{t} = \beta_{t-1} + \nu_{t}$$
 (5)

where v_t are mean-zero normal errors with constant covariance matrix.

Drawback: (5) is not invariant to the observable variables y_t order! The problem lies in the MSV structure!

Literature Review: Phillipov and Glickman (2006a, 2006b)

[Philipov and Glickman, 2006b] and [Philipov and Glickman, 2006a] proposes the following model:

$$y_t = \epsilon_t, \qquad \epsilon \sim \mathcal{N}_k(0_k, \Omega_t^{-1})$$
 (6)

$$\Omega_t | \Omega_{t-1} \sim W(\nu, S_{t-1}), \quad \text{with } S_t = \frac{1}{\nu} A^{1/2} \Omega_t^d A^{1/2'}$$
 (7)

where

- ► A is a positive definite symmetric matrix containing parameters determining the intertemporal sensitivity of each element of the precision matrix and
- ▶ $d \in [0, 1)$ is a scalar that accounts for overall persistence of the process.

Literature Review: Phillipov and Glickman (2006a, 2006b)

[Philipov and Glickman, 2006b] and [Philipov and Glickman, 2006a] proposes the following model:

$$y_t = \epsilon_t, \qquad \epsilon \sim \mathcal{N}_k(0_k, \Omega_t^{-1})$$
 (6)

$$\Omega_t | \Omega_{t-1} \sim \mathcal{W}(\nu, S_{t-1}), \quad \text{with } S_t = \frac{1}{\nu} \mathbf{A}^{1/2} \Omega_t^{\mathbf{d}} \mathbf{A}^{1/2'}$$
 (7)

where

- ► A is a positive definite symmetric matrix containing parameters determining the intertemporal sensitivity of each element of the precision matrix and
- ▶ $d \in [0, 1)$ is a scalar that accounts for overall persistence of the process.

This specification is more flexible than Primiceri's one, and it is invariant to ordering!

Literature Review: Phillipov and Glickman (2006a, 2006b)

► [Philipov and Glickman, 2006a] had trouble estimating simultaneously the elements of the matrix A and persistence parameter *d* and proposed to use a fixed *d*;

Literature Review: Phillipov and Glickman (2006a, 2006b)

- ► [Philipov and Glickman, 2006a] had trouble estimating simultaneously the elements of the matrix A and persistence parameter *d* and proposed to use a fixed *d*;
- ► [Asai and McAleer, 2009] reported convergence problems in the Gibbs sampler.

Where are we?

Phillipov and

Primiceri (2005)

- The order of the variables y_t matters;
- Simple random walk structure for covariances;
- Missing steps in Gibbs sampler.

Estimation problems

Where are we?

Primiceri (2005)

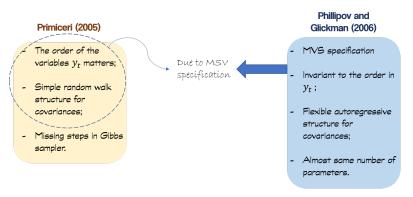
The order of the variables y_t matters;

Simple random walk structure for covariances;

- Missing steps in Gibbs sampler.

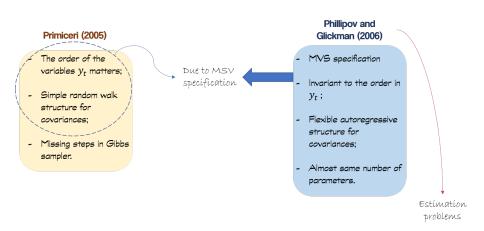
Estimation problems

Where are we?



problems

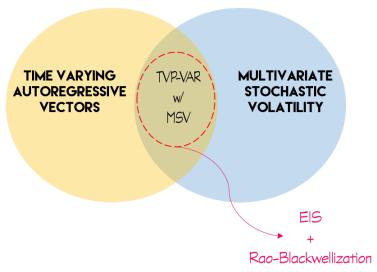
Where are we?



Estimation of a TVP-VAR with MSV via EIS Objective

Develop and implement efficient statistical inference procedures in TVP-VAR with MSV models.

Proposal



Proposal

Build a **TVP-VAR** with **Wishart MSV** model based on [Philipov and Glickman, 2006b] and obtain **maximum likelihood** estimates using **EIS** [Richard and Zhang, 2007] combined with a **Rao-Blackellization** step [Moura and Turatti, 2014].

Proposal

Build a **TVP-VAR** with **Wishart MSV** model based on [Philipov and Glickman, 2006b] and obtain **maximum likelihood** estimates using **EIS** [Richard and Zhang, 2007] combined with a **Rao-Blackellization** step [Moura and Turatti, 2014].

Specific Objectives:

 Obtain the ML estimates for the MSV Wishart model from [Philipov and Glickman, 2006b] using EIS as well develop diagnostic tools to study numerical properties;

Proposal

Build a **TVP-VAR** with **Wishart MSV** model based on [Philipov and Glickman, 2006b] and obtain **maximum likelihood** estimates using **EIS** [Richard and Zhang, 2007] combined with a **Rao-Blackellization** step [Moura and Turatti, 2014].

Specific Objectives:

- Obtain the ML estimates for the MSV Wishart model from [Philipov and Glickman, 2006b] using EIS as well develop diagnostic tools to study numerical properties;
- Generalize the previous model to a TVP-VAR with drifting coefficients plus MSV and implement EIS algorithm combined with a Rao-Blackwellization step, as well as develop diagnostic tools;

Proposal

Build a **TVP-VAR** with **Wishart MSV** model based on [Philipov and Glickman, 2006b] and obtain **maximum likelihood** estimates using **EIS** [Richard and Zhang, 2007] combined with a **Rao-Blackellization** step [Moura and Turatti, 2014].

Specific Objectives:

- Obtain the ML estimates for the MSV Wishart model from [Philipov and Glickman, 2006b] using EIS as well develop diagnostic tools to study numerical properties;
- Generalize the previous model to a TVP-VAR with drifting coefficients plus MSV and implement EIS algorithm combined with a Rao-Blackwellization step, as well as develop diagnostic tools;
- 3. Develop an empirical application using the model developed in (2), in line with [Mumtaz and Theophilopoulou, 2017].

Methodology: Monte Carlo Methods and Importance Sampling

- ► Monte Carlo (MC) Methods are an alternative way to solve complex integrals;
 - Specially used for high-dimension problems, where non-stochastic algorithms are way too slow;
- ► The idea is to resample values from a probability density (thus, is a stochastic method), using a pseudo-random number generator and the inverting theorem.

Methodology: Monte Carlo Methods and Importance Sampling

Consider the following integral:

$$I = \int_{\mathcal{X}} g(x)f(x)dx \tag{8}$$

Methodology: Monte Carlo Methods and Importance Sampling

Consider the following integral:

$$I = \int_{\mathcal{X}} g(x)f(x)dx \tag{8}$$

If *X* is a random vector with pfd $f_X(\cdot)$, then (8) is the expectation of g(X) with respect to $f(\cdot)$:

$$I = \mathbb{E}_f[g(x)] \tag{9}$$

Methodology: Monte Carlo Methods and Importance Sampling

Consider the following integral:

$$I = \int_{\mathcal{X}} g(x)f(x)dx \tag{8}$$

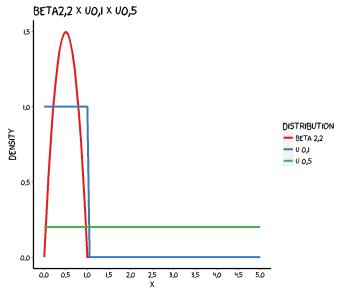
If *X* is a random vector with pfd $f_X(\cdot)$, then (8) is the expectation of g(X) with respect to $f(\cdot)$:

$$I = \mathbb{E}_f[g(x)] \tag{9}$$

In cases where the function f is known, using the law of large numbers we can use the sample mean to approximate (8):

$$I \approx \frac{1}{N} \sum_{i=1}^{N} g(x_i) \tag{10}$$

Methodology: Monte Carlo Methods and Importance Sampling



Methodology: Monte Carlo Methods and Importance Sampling

Consider the following integral:

$$I = \int_{\mathcal{X}} g(x) \frac{f(x)}{m(x)} m(x) dx$$
 (11)

Now we have an expectation with respect to the density $m(\cdot)$:

$$I = \mathbb{E}_{m} \left[\frac{g(x)f(x)}{m(x)} \right]$$
 (12)

In cases where the function m is known, using the law of large numbers we can use the sample mean to approximate (11):

$$I \approx I_N^{IS} = \frac{1}{N} \sum_{i=1}^{N} \omega(x_i) g(x_i), \quad \text{with} \quad \omega_i = \frac{f(x_i)}{m(x_i)}$$
 (13)

Methodology: Efficient Importance Sampling

Two main "problems" arise when dealing with IS:

► The choice of the class of samplers, *M*;

Methodology: Efficient Importance Sampling

Two main "problems" arise when dealing with IS:

- ► The choice of the class of samplers, *M*;
- ► The selection of a sampler in *M* that minimizes the Monte Carlo variance.

Methodology: Efficient Importance Sampling

Two main "problems" arise when dealing with IS:

- ► The choice of the class of samplers, *M*;
- ► The selection of a sampler in *M* that minimizes the Monte Carlo variance.
 - ⇒ EIS resolves the last one!

Methodology: Efficient Importance Sampling

Two main "problems" arise when dealing with IS:

- ► The choice of the class of samplers, *M*;
- ► The selection of a sampler in *M* that minimizes the Monte Carlo variance.
 - ⇒ EIS resolves the last one!



Methodology: Efficient Importance Sampling

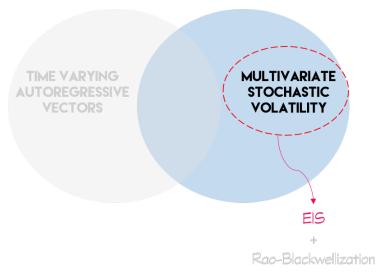
Two main "problems" arise when dealing with IS:

- ► The choice of the class of samplers, *M*;
- ► The selection of a sampler in *M* that minimizes the Monte Carlo variance.
 - ⇒ EIS resolves the last one!



 Sequential EIS is used in high-dimensional problems and consists in "breaking" the EIS algorithm in a sequence of low-dimension optimizations.

Methodology: Wishart MSV model



Methodology: Wishart MSV model

Recall the model given in (6)-(7). Defining $f(\underline{y}, \underline{\Omega}, \theta)$ as the joint density of $\underline{y} = \{y_t\}_{t=1}^T$ and $\underline{\Omega} = \{\Omega_t\}_{t=1}^T$, where θ is the vector of unknown parameters, the likelihood associated to \underline{y}_t is given by:

$$L(\theta|\underline{y}) = \int f(\underline{y}.\Omega,\theta)d\Omega$$

$$= \int \cdots \int \prod_{t=1}^{T} f(y_t,\Omega_t|\Omega_{t-1},\underline{y}_{t-1};\theta)d\Omega_1 \ldots d\Omega_T$$
(14)

with \underline{y}_t e $\underline{\Omega}_t$ given by $\{y_t\}_{t=1}^T$ e $\{\Omega_t\}_{t=1}^T$, respectively.

Methodology: Wishart MSV model

The respective space-state representation is:

$$f(y_t, \Omega_t | \underline{\Omega}_{t-1}, \underline{y}_{t-1}; \theta) = g(y_t | \Omega_t, \underline{y}_{t-1}; \theta) \cdot p(\Omega_t | \underline{\Omega}_{t-1}, \underline{y}_{t-1}; \theta)$$
 (15)

where

$$g(y_t|\Omega_t, \underline{y}_{t-1}; \theta) \propto |\Omega_t|^{1/2} \exp\left\{-\frac{1}{2}y_t'\Omega_t y_t\right\}$$
 (16)

$$p(\Omega_t | \underline{\Omega}_{t-1}, \underline{y}_{t-1}; \theta) \propto |\Omega_t|^{\frac{\nu-k-1}{2}} \exp\left\{ -\frac{1}{2} tr \left[S_{t-1}^{-1} \cdot \Omega_t \right] \right\}$$
 (17)

Methodology: Wishart MSV model

Likelihood:

$$L(\theta|\underline{y}) = \int \prod_{t=1}^{T} \left[\frac{f(y_t, \Omega_t | \underline{\Omega}_{t-1}, \underline{y}_{t-1}; \theta)}{m(\Omega_t | \underline{\Omega}_{t-1}; \gamma_t)} \right] \prod_{t=1}^{T} m(\Omega_t | \underline{\Omega}_{t-1}; \gamma_t) d\underline{\Omega}$$

(18)

with $m(\Omega_t | \Omega_{t-1;\gamma_t}) = \frac{k(\Omega_t;\gamma_t)}{\chi(\Omega_{t-1};\gamma_t)}$ where $\chi(\Omega_{t-1};\gamma_t) = \int k(\Omega_{t-1};\gamma_t) d\Omega_t$.

MC estimate:

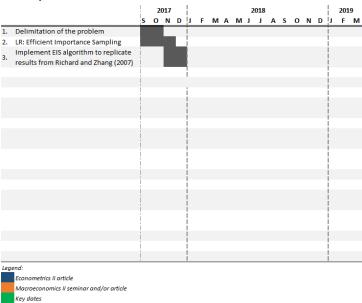
$$\tilde{L}(\theta|\underline{y}) = \frac{1}{N} \sum_{i=1}^{N} \left[\prod_{\ell=1}^{L} \frac{f(y_t, \tilde{\Omega}_t^i | \tilde{\Omega}_{t-1}^i, \underline{y}_{t-1}; \theta) \cdot \mathcal{X}(\tilde{\Omega}_{t-1}^i; \gamma_t)}{k(\tilde{\Omega}_t^i; \gamma_t)} \right]$$

Minimization problem:

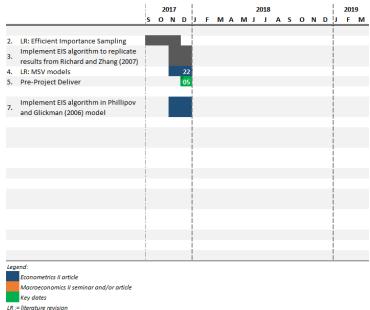
$$\hat{\gamma}_{t}(\theta) = \min_{\gamma_{t}} \sum_{i=1}^{N} \left\{ \ln \left[f(y_{t}, \tilde{\Omega}_{t}^{i} | \tilde{\Omega}_{t-1}^{i}, \underline{y}_{t-1}; \theta) \cdot \mathcal{X}(\tilde{\Omega}_{t-1}^{i}; \gamma_{t}) \right] - c_{t} - \ln k(\underline{\Omega}_{t}; \gamma_{t}) \right\}$$
(20)

Workplan - Completed tasks

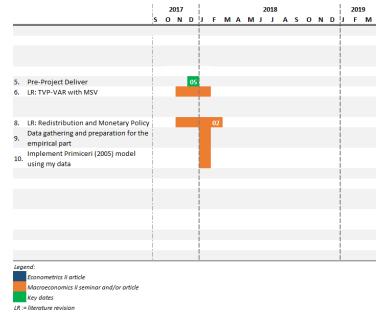
IR := literature revision



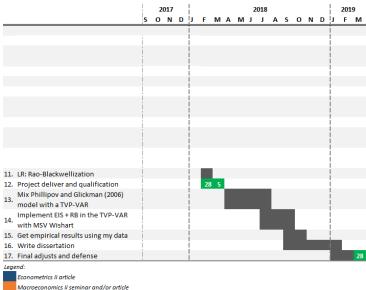
Workplan - Econometrics II related tasks



Workplan - Macroeconomics II related tasks

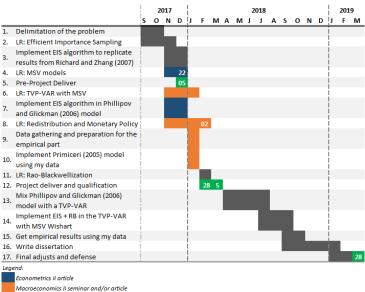


Workplan - What future holds



Key dates

Workplan - General Overview



Key dates

Referências I



Asai, M. and McAleer, M. (2009).

The structure of dynamic correlations in multivariate stochastic volatility models. *Journal of Econometrics*, 150(2):182–192.



Bernanke, B. S. and Mihov, I. (1998a).

The liquidity effect and long-run neutrality.

In Carnegie-Rochester conference series on public policy, volume 49, pages 149–194. Elsevier.



Bernanke, B. S. and Mihov, I. (1998b).

Measuring monetary policy.

The Quarterly Journal of Economics, 113(3):869–902.



Cogley, T. and Sargent, T. J. (2001).

Evolving Post-World War II U.S. Inflation Dynamics.

NBER Macroeconomics Annual, 16(1):331-373.



Cogley, T. and Sargent, T. J. (2005).

Drifts and volatilities: monetary policies and outcomes in the post wwii us.

Review of Economic dynamics, 8(2):262-302.



Galí, J. and Gambetti, L. (2015).

The effects of monetary policy on stock market bubbles: Some evidence.

American Economic Journal: Macroeconomics, 7(1):233-257.

Referências II



Moura, G. V. and Turatti, D. E. (2014).

Efficient estimation of conditionally linear and gaussian state space models. *Economics Letters*, 124(3):494–499.



Mumtaz, H. and Theophilopoulou, A. (2015).

Monetary policy and inequality in the uk.

Technical report, Working Paper, School of Economics and Finance, Queen Mary, University of London.



Mumtaz, H. and Theophilopoulou, A. (2017).

The impact of monetary policy on inequality in the uk. an empirical analysis.

European Economic Review, 98:410-423.



Mumtaz, H. and Zanetti, F. (2015).

Labor market dynamics: A time-varying analysis.

Oxford Bulletin of Economics and Statistics, 77(3):319–338.



Philipov, A. and Glickman, M. E. (2006a).

Factor multivariate stochastic volatility via wishart processes.

Econometric Reviews, 25(2-3):311-334.



Philipov, A. and Glickman, M. E. (2006b).

Multivariate stochastic volatility via wishart processes.

Journal of Business & Economic Statistics, 24(3):313–328.

Referências III



Primiceri, G. E. (2005).

Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852.



Richard, J.-F. and Zhang, W. (2007).

Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141(2):1385–1411.



Sims, C. A. (1999).

Drifts and breaks in monetary policy.
Technical report, Princeton University Press.