Adaptive Models and Heavy Tails

Davide Delle Monache[1] and Ivan Petrella[2]

[1]Queen Mary, University of London
[2]Birkbeck, University of London and CEPR

Workshop on Dynamic Models driven by the Score of Predictive Likelihoods.
Tenerife, 9-11 January 2014

Motivation

Since Cogley and Sargent (2002), time-varying parameters (TVP) models have been widely used in macroeconomics for investigating the dynamics of key macroeconomic aggregates:

1. TVP models have been used for policy evaluation (Primiceri, 2006) and it has been shown to provide good forecasts (D'Agostino et al., 2013).
2. The literature on forecasting under parameters instability has been growing fast in the last decades; see Stock and Watson (1996), Pesaran and Pick (2011) and Giraitis et al (2013) $\Rightarrow$ optimal weighting.

The interest in TVP models can be traced back in other fields:

1. In the adaptive control and engineering literature there is a long tradition on the use of discount regression and forgetting factors algorithms (Brown, 1963, Fagin, 1964 and Jazwinski, 1970).
2. Ljung and Soderstrom (1985) derive recursive formulations of a variety of adaptive algorithms for quadratic criterion functions and interpret them as a stochastic approximation of the Gauss-Newton algorithm.

Since Marcet and Sargent (1989), the adaptive algorithms have been used to describe the learning mechanism of the expectations' formation.

Motivation

Since Cogley and Sargent (2002), time-varying parameters (TVP) models have been widely used in macroeconomics for investigating the dynamics of key macroeconomic aggregates:

1. TVP models have been used for policy evaluation (Primiceri, 2006) and it has been shown to provide good forecasts (D'Agostino et al., 2013).
2. The literature on forecasting under parameters instability has been growing fast in the last decades; see Stock and Watson (1996), Pesaran and Pick (2011) and Giraitis et al (2013) ⇒ optimal weighting.

The interest in TVP models can be traced back in other fields:

1. In the adaptive control and engineering literature there is a long tradition on the use of discount regression and forgetting factors algorithms (Brown, 1963, Fagin, 1964 and Jazwinski, 1970).
2. Ljung and Soderstrom (1985) derive recursive formulations of a variety of adaptive algorithms for quadratic criterion functions and interpret them as a stochastic approximation of the Gauss-Newton algorithm.

Since Marcet and Sargent (1989), the adaptive algorithms have been used to describe the learning mechanism of the expectations' formation.

Motivation

Since Cogley and Sargent (2002), time-varying parameters (TVP) models have been widely used in macroeconomics for investigating the dynamics of key macroeconomic aggregates:

1. TVP models have been used for policy evaluation (Primiceri, 2006) and it has been shown to provide good forecasts (D'Agostino et al., 2013).
2. The literature on forecasting under parameters instability has been growing fast in the last decades; see Stock and Watson (1996), Pesaran and Pick (2011) and Giraitis et al (2013) $\Rightarrow$ optimal weighting.

The interest in TVP models can be traced back in other fields:

1. In the adaptive control and engineering literature there is a long tradition on the use of discount regression and forgetting factors algorithms (Brown, 1963, Fagin, 1964 and Jazwinski, 1970).
2. Ljung and Soderstrom (1985) derive recursive formulations of a variety of adaptive algorithms for quadratic criterion functions and interpret them as a stochastic approximation of the Gauss-Newton algorithm.

Since Marcet and Sargent (1989), the adaptive algorithms have been used to describe the learning mechanism of the expectations' formation.

Motivation

Since Cogley and Sargent (2002), time-varying parameters (TVP) models have been widely used in macroeconomics for investigating the dynamics of key macroeconomic aggregates:

1. TVP models have been used for policy evaluation (Primiceri, 2006) and it has been shown to provide good forecasts (D'Agostino et al., 2013).
2. The literature on forecasting under parameters instability has been growing fast in the last decades; see Stock and Watson (1996), Pesaran and Pick (2011) and Giraitis et al (2013) $\Rightarrow$ optimal weighting.

The interest in TVP models can be traced back in other fields:

1. In the adaptive control and engineering literature there is a long tradition on the use of discount regression and forgetting factors algorithms (Brown, 1963, Fagin, 1964 and Jazwinski, 1970).
2. Ljung and Soderstrom (1985) derive recursive formulations of a variety of adaptive algorithms for quadratic criterion functions and interpret them as a stochastic approximation of the Gauss-Newton algorithm.

Since Marcet and Sargent (1989), the adaptive algorithms have been used to describe the learning mechanism of the expectations' formation.

Motivation

Since Cogley and Sargent (2002), time-varying parameters (TVP) models have been widely used in macroeconomics for investigating the dynamics of key macroeconomic aggregates:

1. TVP models have been used for policy evaluation (Primiceri, 2006) and it has been shown to provide good forecasts (D'Agostino et al., 2013).
2. The literature on forecasting under parameters instability has been growing fast in the last decades; see Stock and Watson (1996), Pesaran and Pick (2011) and Giraitis et al (2013) $\Rightarrow$ optimal weighting.

The interest in TVP models can be traced back in other fields:

1. In the adaptive control and engineering literature there is a long tradition on the use of discount regression and forgetting factors algorithms (Brown, 1963, Fagin, 1964 and Jazwinski, 1970).
2. Ljung and Soderstrom (1985) derive recursive formulations of a variety of adaptive algorithms for quadratic criterion functions and interpret them as a stochastic approximation of the Gauss-Newton algorithm.

Since Marcet and Sargent (1989), the adaptive algorithms have been used to describe the learning mechanism of the expectations' formation.

Motivation

Since Cogley and Sargent (2002), time-varying parameters (TVP) models have been widely used in macroeconomics for investigating the dynamics of key macroeconomic aggregates:

1. TVP models have been used for policy evaluation (Primiceri, 2006) and it has been shown to provide good forecasts (D'Agostino et al., 2013).
2. The literature on forecasting under parameters instability has been growing fast in the last decades; see Stock and Watson (1996), Pesaran and Pick (2011) and Giraitis et al (2013) $\Rightarrow$ optimal weighting.

The interest in TVP models can be traced back in other fields:

1. In the adaptive control and engineering literature there is a long tradition on the use of discount regression and forgetting factors algorithms (Brown, 1963, Fagin, 1964 and Jazwinski, 1970).
2. Ljung and Soderstrom (1985) derive recursive formulations of a variety of adaptive algorithms for quadratic criterion functions and interpret them as a stochastic approximation of the Gauss-Newton algorithm.

Since Marcet and Sargent (1989), the adaptive algorithms have been used to describe the learning mechanism of the expectations' formation.

This paper

This paper proposes a new adaptive algorithm that builds on recent developments of the score-driven models ⇐ Creal et al. (2012) and Harvey (2013).

The adaptive algorithm for TVP models extends the traditional ones of Ljung and Soderstrom (1985) along various dimensions:

1. it considers how the existing algorithms are to be modified in the presence of heavy tails (Normal and Student-t are considered) ⇐ Curdia et al (2013)
2. it introduces time-variation in volatility, emphasizing how this interacts with the coefficients' updating rule ⇐ Justiniano and Primiceri (2013)
3. it shows how to impose restrictions so that the model is locally stationary and has a bounded mean ⇐ Projection facility, see Evans and Honkapohja (2001)

Application to the US inflation leads to the following results:

1. allowing for heavy-tails leads to a significant improvement in forecasting ⇒ it is crucial in order to obtain well-calibrated density forecasts
2. the inclusion of bounds on the long-run trend imposes a discipline so that forecasts are (marginally) improved

This paper

This paper proposes a new adaptive algorithm that builds on recent developments of the score-driven models $\Leftarrow$ Creal et al. (2012) and Harvey (2013).

The adaptive algorithm for TVP models extends the traditional ones of Ljung and Soderstrom (1985) along various dimensions:

1. it considers how the existing algorithms are to be modified in the presence of heavy tails (Normal and Student-t are considered) $\Leftarrow$ Curdia et al (2013)
2. it introduces time-variation in volatility, emphasizing how this interacts with the coefficients' updating rule $\Leftarrow$ Justiniano and Primiceri (2013)
3. it shows how to impose restrictions so that the model is locally stationary and has a bounded mean $\Leftarrow$ Projection facility, see Evans and Honkapohja (2001)

Application to the US inflation leads to the following results:

1. allowing for heavy-tails leads to a significant improvement in forecasting $\Rightarrow$ it is crucial in order to obtain well-calibrated density forecasts
2. the inclusion of bounds on the long-run trend imposes a discipline so that forecasts are (marginally) improved

This paper

This paper proposes a new adaptive algorithm that builds on recent developments of the score-driven models $\Leftarrow$ Creal et al. (2012) and Harvey (2013).

The adaptive algorithm for TVP models extends the traditional ones of Ljung and Soderstrom (1985) along various dimensions:

1. it considers how the existing algorithms are to be modified in the presence of heavy tails (Normal and Student-t are considered) $\Leftarrow$ Curdia et al (2013)
2. it introduces time-variation in volatility, emphasizing how this interacts with the coefficients' updating rule $\Leftarrow$ Justiniano and Primiceri (2013)
3. it shows how to impose restrictions so that the model is locally stationary and has a bounded mean $\Leftarrow$ Projection facility, see Evans and Honkapohja (2001)

Application to the US inflation leads to the following results:

1. allowing for heavy-tails leads to a significant improvement in forecasting $\Rightarrow$ it is crucial in order to obtain well-calibrated density forecasts
2. the inclusion of bounds on the long-run trend imposes a discipline so that forecasts are (marginally) improved

This paper

This paper proposes a new adaptive algorithm that builds on recent developments of the score-driven models ⇐ Creal et al. (2012) and Harvey (2013).

The adaptive algorithm for TVP models extends the traditional ones of Ljung and Soderstrom (1985) along various dimensions:

1. it considers how the existing algorithms are to be modified in the presence of heavy tails (Normal and Student-t are considered) ⇐ Curdia et al (2013)
2. it introduces time-variation in volatility, emphasizing how this interacts with the coefficients' updating rule ⇐ Justiniano and Primiceri (2013)
3. it shows how to impose restrictions so that the model is locally stationary and has a bounded mean ⇐ Projection facility, see Evans and Honkapohja (2001)

Application to the US inflation leads to the following results:

1. allowing for heavy-tails leads to a significant improvement in forecasting ⇒ it is crucial in order to obtain well-calibrated density forecasts
2. the inclusion of bounds on the long-run trend imposes a discipline so that forecasts are (marginally) improved

This paper

This paper proposes a new adaptive algorithm that builds on recent developments of the score-driven models ⇐ Creal et al. (2012) and Harvey (2013).

The adaptive algorithm for TVP models extends the traditional ones of Ljung and Soderstrom (1985) along various dimensions:

1. it considers how the existing algorithms are to be modified in the presence of heavy tails (Normal and Student-t are considered) ⇐ Curdia et al (2013)
2. it introduces time-variation in volatility, emphasizing how this interacts with the coefficients' updating rule ⇐ Justiniano and Primiceri (2013)
3. it shows how to impose restrictions so that the model is locally stationary and has a bounded mean ⇐ Projection facility, see Evans and Honkapohja (2001)

Application to the US inflation leads to the following results:

1. allowing for heavy-tails leads to a significant improvement in forecasting ⇒ it is crucial in order to obtain well-calibrated density forecasts
2. the inclusion of bounds on the long-run trend imposes a discipline so that forecasts are (marginally) improved

This paper

This paper proposes a new adaptive algorithm that builds on recent developments of the score-driven models $\Leftarrow$ Creal et al. (2012) and Harvey (2013).

The adaptive algorithm for TVP models extends the traditional ones of Ljung and Soderstrom (1985) along various dimensions:

1. it considers how the existing algorithms are to be modified in the presence of heavy tails (Normal and Student-t are considered) $\Leftarrow$ Curdia et al (2013)
2. it introduces time-variation in volatility, emphasizing how this interacts with the coefficients' updating rule $\Leftarrow$ Justiniano and Primiceri (2013)
3. it shows how to impose restrictions so that the model is locally stationary and has a bounded mean $\Leftarrow$ Projection facility, see Evans and Honkapohja (2001)

Application to the US inflation leads to the following results:

1. allowing for heavy-tails leads to a significant improvement in forecasting $\Rightarrow$ it is crucial in order to obtain well-calibrated density forecasts
2. the inclusion of bounds on the long-run trend imposes a discipline so that forecasts are (marginally) improved

Score-driven model

1. The time-varying parameters model considered here is

$$y_t = \mathbf{x}_t'\phi_{t|t-1} + \epsilon_t, \quad \epsilon_t \sim IID(0, \sigma^2_{t|t-1}), \quad t = 1, ..., n.$$

where $\mathbf{x}_t = (1, y_{t-1}, ..., y_{t-p})'$, $\mathbf{f}_{t|t-1} = (\phi'_{t|t-1}, \sigma^2_{t|t-1})'$.

2. The driving process is represented by the score of the conditional distribution

$$\mathbf{f}_{t+1|t} = \omega + \mathbf{A}\mathbf{f}_{t|t-1} + \mathbf{B}\mathbf{s}_t, \quad \mathbf{s}_t = \mathcal{I}_t^{-1}\nabla_t,$$

$$\nabla_t = \frac{\partial \ell_t\left(y_t|Y_{t-1}, \theta, \mathbf{f}_{t|t-1}\right)}{\partial \mathbf{f}_{t|t-1}}, \quad \mathcal{I}_t = -\mathbf{E}\left[\frac{\partial^2 \ell_t\left(y_t|Y_{t-1}, \theta, \mathbf{f}_{t|t-1}\right)}{\partial \mathbf{f}_{t|t-1}\partial \mathbf{f}'_{t|t-1}}\right].$$

3. In the observation-driven framework, the vector $\mathbf{f}_{t+1}$, although stochastic is perfectly predictable at time t, thus we denote $\mathbf{f}_{t+1|t}$ as the filter estimate
4. The driving mechanism depends upon past observations only ⇒ i.e. a single source of error model
5. Why the score? A stochastic version of the Gauss-Newton searching direction for the parameters' variation ⇒ the score inherits the properties of the distribution of the innovations ϵ_t

Score-driven model

1. The time-varying parameters model considered here is

$$y_t = \mathbf{x}_t' \phi_{t|t-1} + \epsilon_t, \quad \epsilon_t \sim IID(0, \sigma^2_{t|t-1}), \quad t = 1, ..., n.$$

where $\mathbf{x}_t = (1, y_{t-1}, ..., y_{t-p})'$, $\mathbf{f}_{t|t-1} = (\phi'_{t|t-1}, \sigma^2_{t|t-1})'$.

2. The driving process is represented by the score of the conditional distribution

$$\mathbf{f}_{t+1|t} = \omega + \mathbf{A}\mathbf{f}_{t|t-1} + \mathbf{B}\mathbf{s}_t, \quad \mathbf{s}_t = \mathcal{I}_t^{-1} \nabla_t,$$

$$\nabla_t = \frac{\partial \ell_t \left(y_t | Y_{t-1}, \theta, \mathbf{f}_{t|t-1}\right)}{\partial \mathbf{f}_{t|t-1}}, \quad \mathcal{I}_t = -\mathbf{E}\left[\frac{\partial^2 \ell_t \left(y_t | Y_{t-1}, \theta, \mathbf{f}_{t|t-1}\right)}{\partial \mathbf{f}_{t|t-1} \partial \mathbf{f}'_{t|t-1}}\right].$$

3. In the observation-driven framework, the vector $\mathbf{f}_{t+1}$, although stochastic is perfectly predictable at time t, thus we denote $\mathbf{f}_{t+1|t}$ as the filter estimate
4. The driving mechanism depends upon past observations only ⇒ i.e. a single source of error model
5. Why the score? A stochastic version of the Gauss-Newton searching direction for the parameters' variation ⇒ the score inherits the properties of the distribution of the innovations ϵ_t

Score-driven model

1. The time-varying parameters model considered here is

$$y_t = \mathbf{x}_t' \phi_{t|t-1} + \epsilon_t, \quad \epsilon_t \sim IID(0, \sigma^2_{t|t-1}), \quad t = 1, ..., n.$$

   where $\mathbf{x}_t = (1, y_{t-1}, ..., y_{t-p})'$, $\mathbf{f}_{t|t-1} = (\phi'_{t|t-1}, \sigma^2_{t|t-1})'$.

2. The driving process is represented by the score of the conditional distribution

$$\mathbf{f}_{t+1|t} = \omega + \mathbf{A}\mathbf{f}_{t|t-1} + \mathbf{B}\mathbf{s}_t, \quad \mathbf{s}_t = \mathcal{I}_t^{-1} \nabla_t,$$

$$\nabla_t = \frac{\partial \ell_t \left(y_t | Y_{t-1}, \theta, \mathbf{f}_{t|t-1} \right)}{\partial \mathbf{f}_{t|t-1}}, \quad \mathcal{I}_t = -\mathbf{E} \left[\frac{\partial^2 \ell_t \left(y_t | Y_{t-1}, \theta, \mathbf{f}_{t|t-1} \right)}{\partial \mathbf{f}_{t|t-1} \partial \mathbf{f}'_{t|t-1}} \right].$$

3. In the observation-driven framework, the vector $\mathbf{f}_{t+1}$, although stochastic is perfectly predictable at time t, thus we denote $\mathbf{f}_{t+1|t}$ as the filter estimate
4. The driving mechanism depends upon past observations only ⇒ i.e. a single source of error model
5. Why the score? A stochastic version of the Gauss-Newton searching direction for the parameters' variation ⇒ the score inherits the properties of the distribution of the innovations ϵ_t

Score-driven model

1. The time-varying parameters model considered here is

$$y_t = \mathbf{x}_t' \phi_{t|t-1} + \epsilon_t, \quad \epsilon_t \sim IID(0, \sigma^2_{t|t-1}), \quad t = 1, ..., n.$$

   where $\mathbf{x}_t = (1, y_{t-1}, ..., y_{t-p})'$, $\mathbf{f}_{t|t-1} = (\phi'_{t|t-1}, \sigma^2_{t|t-1})'$.

2. The driving process is represented by the score of the conditional distribution

$$\mathbf{f}_{t+1|t} = \omega + \mathbf{A}\mathbf{f}_{t|t-1} + \mathbf{B}\mathbf{s}_t, \quad \mathbf{s}_t = \mathcal{I}_t^{-1} \nabla_t,$$

$$\nabla_t = \frac{\partial \ell_t \left(y_t | Y_{t-1}, \theta, \mathbf{f}_{t|t-1} \right)}{\partial \mathbf{f}_{t|t-1}}, \quad \mathcal{I}_t = -\mathbf{E}\left[\frac{\partial^2 \ell_t \left(y_t | Y_{t-1}, \theta, \mathbf{f}_{t|t-1} \right)}{\partial \mathbf{f}_{t|t-1} \partial \mathbf{f}'_{t|t-1}} \right].$$

3. In the observation-driven framework, the vector $\mathbf{f}_{t+1}$, although stochastic is perfectly predictable at time t, thus we denote $\mathbf{f}_{t+1|t}$ as the filter estimate
4. The driving mechanism depends upon past observations only $\Rightarrow$ i.e. a single source of error model
5. Why the score? A stochastic version of the Gauss-Newton searching direction for the parameters' variation $\Rightarrow$ the score inherits the properties of the distribution of the innovations ϵ_t

Recursions under Gaussian distribution

The conditional log-likelihood of y_t is

$$\ell_t = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma^2_{t|t-1} - \frac{1}{2}\frac{\epsilon_t^2}{\sigma^2_{t|t-1}}.$$

1. We parameterize the model so that $\mathbf{A} = \mathbf{I}$ such that the the parameters follow a random walk-type law of motionand and we restrict $\mathbf{B}$ to depends upon two scalar parameters κ_ϕ and κ_σ
2. We end up with the following recursions:

$$\phi_{t+1|t} = \phi_{t|t-1} + \kappa_\phi \mathbf{R}_t^{-1}\mathbf{x}_t\sigma^{-2}_{t|t-1}(y_t - \mathbf{x}_t'\phi_{t|t-1})$$

$$\mathbf{R}_t = \mathbf{R}_{t-1} + \kappa_h(\mathbf{x}_t\sigma^{-2}_{t|t-1}\mathbf{x}_t' - \mathbf{R}_{t-1})$$

$$\sigma^2_{t+1|t} = \sigma^2_{t|t-1} + \kappa_\sigma(\varepsilon_t^2 - \sigma^2_{t|t-1})$$

   the scaling matrix has been replaced by its smoothed version $\mathbf{R}_t$

3. This generalises the exponential smoothing when regressors are included

Recursions under Gaussian distribution

The conditional log-likelihood of y_t is

$$\ell_t = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma^2_{t|t-1} - \frac{1}{2}\frac{\epsilon_t^2}{\sigma^2_{t|t-1}}.$$

1. We parameterize the model so that $\mathbf{A} = \mathbf{I}$ such that the the parameters follow a random walk-type law of motionand and we restrict $\mathbf{B}$ to depends upon two scalar parameters κ_ϕ and κ_σ
2. We end up with the following recursions:

$$\phi_{t+1|t} = \phi_{t|t-1} + \kappa_\phi \mathbf{R}_t^{-1}\mathbf{x}_t\sigma^{-2}_{t|t-1}(y_t - \mathbf{x}_t'\phi_{t|t-1})$$

$$\mathbf{R}_t = \mathbf{R}_{t-1} + \kappa_h(\mathbf{x}_t\sigma^{-2}_{t|t-1}\mathbf{x}_t' - \mathbf{R}_{t-1})$$

$$\sigma^2_{t+1|t} = \sigma^2_{t|t-1} + \kappa_\sigma(\varepsilon_t^2 - \sigma^2_{t|t-1})$$

the scaling matrix has been replaced by its smoothed version $\mathbf{R}_t$

3. This generalises the exponential smoothing when regressors are included

Recursions under Gaussian distribution

The conditional log-likelihood of y_t is

$$\ell_t = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma^2_{t|t-1} - \frac{1}{2}\frac{\epsilon_t^2}{\sigma^2_{t|t-1}}.$$

1. We parameterize the model so that $\mathbf{A} = \mathbf{I}$ such that the the parameters follow a random walk-type law of motionand and we restrict $\mathbf{B}$ to depends upon two scalar parameters κ_ϕ and κ_σ
2. We end up with the following recursions:

$$\phi_{t+1|t} = \phi_{t|t-1} + \kappa_\phi \mathbf{R}_t^{-1}\mathbf{x}_t\sigma^{-2}_{t|t-1}(y_t - \mathbf{x}_t'\phi_{t|t-1})$$

$$\mathbf{R}_t = \mathbf{R}_{t-1} + \kappa_h(\mathbf{x}_t\sigma^{-2}_{t|t-1}\mathbf{x}_t' - \mathbf{R}_{t-1})$$

$$\sigma^2_{t+1|t} = \sigma^2_{t|t-1} + \kappa_\sigma(\varepsilon_t^2 - \sigma^2_{t|t-1})$$

the scaling matrix has been replaced by its smoothed version $\mathbf{R}_t$

3. This generalises the exponential smoothing when regressors are included

Recursions under Gaussian distribution

The conditional log-likelihood of y_t is

$$\ell_t = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma^2_{t|t-1} - \frac{1}{2}\frac{\epsilon_t^2}{\sigma^2_{t|t-1}}.$$

1. We parameterize the model so that $\mathbf{A} = \mathbf{I}$ such that the the parameters follow a random walk-type law of motionand and we restrict $\mathbf{B}$ to depends upon two scalar parameters κ_ϕ and κ_σ
2. We end up with the following recursions:

$$\phi_{t+1|t} = \phi_{t|t-1} + \kappa_\phi \mathbf{R}_t^{-1}\mathbf{x}_t\sigma^{-2}_{t|t-1}(y_t - \mathbf{x}_t'\phi_{t|t-1})$$

$$\mathbf{R}_t = \mathbf{R}_{t-1} + \kappa_h(\mathbf{x}_t\sigma^{-2}_{t|t-1}\mathbf{x}_t' - \mathbf{R}_{t-1})$$

$$\sigma^2_{t+1|t} = \sigma^2_{t|t-1} + \kappa_\sigma(\varepsilon_t^2 - \sigma^2_{t|t-1})$$

the scaling matrix has been replaced by its smoothed version $\mathbf{R}_t$

3. This generalises the exponential smoothing when regressors are included

Recursions under Gaussian distribution

The conditional log-likelihood of y_t is

$$\ell_t = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma^2_{t|t-1} - \frac{1}{2}\frac{\epsilon_t^2}{\sigma^2_{t|t-1}}.$$

1. We parameterize the model so that $\mathbf{A} = \mathbf{I}$ such that the the parameters follow a random walk-type law of motionand and we restrict $\mathbf{B}$ to depends upon two scalar parameters κ_ϕ and κ_σ
2. We end up with the following recursions:

$$\phi_{t+1|t} = \phi_{t|t-1} + \kappa_\phi \mathbf{R}_t^{-1}\mathbf{x}_t\sigma^{-2}_{t|t-1}(y_t - \mathbf{x}_t'\phi_{t|t-1})$$

$$\mathbf{R}_t = \mathbf{R}_{t-1} + \kappa_h(\mathbf{x}_t\sigma^{-2}_{t|t-1}\mathbf{x}_t' - \mathbf{R}_{t-1})$$

$$\sigma^2_{t+1|t} = \sigma^2_{t|t-1} + \kappa_\sigma(\varepsilon_t^2 - \sigma^2_{t|t-1})$$

   the scaling matrix has been replaced by its smoothed version $\mathbf{R}_t$

3. This generalises the exponential smoothing when regressors are included

Relation with the learning algorithms

If we assume $\sigma^2_{t|t-1} = \sigma^2$ and $\kappa_\phi = \kappa_h$, the recursions of our model collapse to the Constant Gain Learning (CGL) algorithm of Sargent and William (2005).

1. Lemma 1: the CGL can be obtained from a Kalman filter (KF) by imposing particular restrictions on the state space model $\Rightarrow$ those restrictions imply that the parameter-driven model collapses to an obervation-driven model.

2. Lemma 2: the CGL implies a that past observations are weighted with exponentially decay factor $(1-\kappa)^j$, the engineering literature refers to the constant forgetting factor $\Leftarrow$ Koop and Korrobilis (2012).

3. Lemma 3: other restrictions (on the state space model) has been proposed in the literature $\Rightarrow$ all of them imply that the parameter-driven model collapses to an observation-driven model and the KF converges to a particular score-driven filter $\Leftarrow$ Stock and Watson (1996), Sargent and William (2005), Li (2008) and Evans et al (2010).

4. Remark 4: the CGL can be seen as a recursive solution for quadratic loss function criterion and it leads to a stochastic analog of the recursive Gauss-Newton search direction $\Leftarrow$ Ljung and Soderstrom (1985)

Relation with the learning algorithms

If we assume $\sigma^2_{t|t-1} = \sigma^2$ and $\kappa_\phi = \kappa_h$, the recursions of our model collapse to the Constant Gain Learning (CGL) algorithm of Sargent and William (2005).

1. Lemma 1: the CGL can be obtained from a Kalman filter (KF) by imposing particular restrictions on the state space model $\Rightarrow$ those restrictions imply that the parameter-driven model collapses to an obervation-driven model.

2. Lemma 2: the CGL implies a that past observations are weighted with exponentially decay factor $(1-\kappa)^j$, the engineering literature refers to the constant forgetting factor $\Leftarrow$ Koop and Korrobilis (2012).

3. Lemma 3: other restrictions (on the state space model) has been proposed in the literature $\Rightarrow$ all of them imply that the parameter-driven model collapses to an observation-driven model and the KF converges to a particular score-driven filter $\Leftarrow$ Stock and Watson (1996), Sargent and William (2005), Li (2008) and Evans et al (2010).

4. Remark 4: the CGL can be seen as a recursive solution for quadratic loss function criterion and it leads to a stochastic analog of the recursive Gauss-Newton search direction $\Leftarrow$ Ljung and Soderstrom (1985)

Relation with the learning algorithms

If we assume $\sigma^2_{t|t-1} = \sigma^2$ and $\kappa_\phi = \kappa_h$, the recursions of our model collapse to the Constant Gain Learning (CGL) algorithm of Sargent and William (2005).

1. Lemma 1: the CGL can be obtained from a Kalman filter (KF) by imposing particular restrictions on the state space model $\Rightarrow$ those restrictions imply that the parameter-driven model collapses to an obervation-driven model.

2. Lemma 2: the CGL implies a that past observations are weighted with exponentially decay factor $(1-\kappa)^j$, the engineering literature refers to the constant forgetting factor $\Leftarrow$ Koop and Korrobilis (2012).

3. Lemma 3: other restrictions (on the state space model) has been proposed in the literature $\Rightarrow$ all of them imply that the parameter-driven model collapses to an observation-driven model and the KF converges to a particular score-driven filter $\Leftarrow$ Stock and Watson (1996), Sargent and William (2005), Li (2008) and Evans et al (2010).

4. Remark 4: the CGL can be seen as a recursive solution for quadratic loss function criterion and it leads to a stochastic analog of the recursive Gauss-Newton search direction $\Leftarrow$ Ljung and Soderstrom (1985)

Relation with the learning algorithms

If we assume $\sigma^2_{t|t-1} = \sigma^2$ and $\kappa_\phi = \kappa_h$, the recursions of our model collapse to the Constant Gain Learning (CGL) algorithm of Sargent and William (2005).

1. Lemma 1: the CGL can be obtained from a Kalman filter (KF) by imposing particular restrictions on the state space model $\Rightarrow$ those restrictions imply that the parameter-driven model collapses to an obervation-driven model.

2. Lemma 2: the CGL implies a that past observations are weighted with exponentially decay factor $(1-\kappa)^j$, the engineering literature refers to the constant forgetting factor $\Leftarrow$ Koop and Korrobilis (2012).

3. Lemma 3: other restrictions (on the state space model) has been proposed in the literature $\Rightarrow$ all of them imply that the parameter-driven model collapses to an observation-driven model and the KF converges to a particular score-driven filter $\Leftarrow$ Stock and Watson (1996), Sargent and William (2005), Li (2008) and Evans et al (2010).

4. Remark 4: the CGL can be seen as a recursive solution for quadratic loss function criterion and it leads to a stochastic analog of the recursive Gauss-Newton search direction $\Leftarrow$ Ljung and Soderstrom (1985)

Relation with the learning algorithms

If we assume $\sigma^2_{t|t-1} = \sigma^2$ and $\kappa_\phi = \kappa_h$, the recursions of our model collapse to the Constant Gain Learning (CGL) algorithm of Sargent and William (2005).

1. Lemma 1: the CGL can be obtained from a Kalman filter (KF) by imposing particular restrictions on the state space model $\Rightarrow$ those restrictions imply that the parameter-driven model collapses to an obervation-driven model.

2. Lemma 2: the CGL implies a that past observations are weighted with exponentially decay factor $(1-\kappa)^j$, the engineering literature refers to the constant forgetting factor $\Leftarrow$ Koop and Korrobilis (2012).

3. Lemma 3: other restrictions (on the state space model) has been proposed in the literature $\Rightarrow$ all of them imply that the parameter-driven model collapses to an observation-driven model and the KF converges to a particular score-driven filter $\Leftarrow$ Stock and Watson (1996), Sargent and William (2005), Li (2008) and Evans et al (2010).

4. Remark 4: the CGL can be seen as a recursive solution for quadratic loss function criterion and it leads to a stochastic analog of the recursive Gauss-Newton search direction $\Leftarrow$ Ljung and Soderstrom (1985)

Relation with the learning algorithms

If we assume $\sigma^2_{t|t-1} = \sigma^2$ and $\kappa_\phi = \kappa_h$, the recursions of our model collapse to the Constant Gain Learning (CGL) algorithm of Sargent and William (2005).

1. Lemma 1: the CGL can be obtained from a Kalman filter (KF) by imposing particular restrictions on the state space model $\Rightarrow$ those restrictions imply that the parameter-driven model collapses to an obervation-driven model.

2. Lemma 2: the CGL implies a that past observations are weighted with exponentially decay factor $(1-\kappa)^j$, the engineering literature refers to the constant forgetting factor $\Leftarrow$ Koop and Korrobilis (2012).

3. Lemma 3: other restrictions (on the state space model) has been proposed in the literature $\Rightarrow$ all of them imply that the parameter-driven model collapses to an observation-driven model and the KF converges to a particular score-driven filter $\Leftarrow$ Stock and Watson (1996), Sargent and William (2005), Li (2008) and Evans et al (2010).

4. Remark 4: the CGL can be seen as a recursive solution for quadratic loss function criterion and it leads to a stochastic analog of the recursive Gauss-Newton search direction $\Leftarrow$ Ljung and Soderstrom (1985)

Student-t distribution

The conditional log-likelihood is

$$\ell_t = c(\eta) - \frac{1}{2}\ln\sigma^2_{t|t-1} - \left(\frac{\eta+1}{2\eta}\right)\ln\left[1 + \frac{\eta}{1-2\eta}\frac{\epsilon_t^2}{\sigma^2_{t|t-1}}\right],$$

where $c(\eta)$ is a constant and $\upsilon = 1/\eta$ are the degree of freedoms.

1. The distribution has a clear impact on the TVP's dynamics, in fact the two scaled-scores are

$$\mathbf{s}_{\phi t} = \frac{(1-2\eta)(1+3\eta)}{(1+\eta)}(\mathbf{x}_t\sigma^{-2}_{t|t-1}\mathbf{x}'_t)^{-1}\mathbf{x}_t\sigma^{-2}_{t|t-1}w_t\epsilon_t,$$

$$s_{\sigma t} = (1+3\eta)(w_t\epsilon_t^2 - \sigma^2_{t|t-1}).$$

2. The weights $w_t = \frac{(1+\eta)}{(1-2\eta+\eta\zeta_t)}$, where $\zeta_t = \epsilon_t^2/\sigma^2_{t|t-1} \Rightarrow$ the variance affects the conditional mean
3. This leads to a double weighting schema $\Rightarrow$ across obs and across time.

Student-t distribution

The conditional log-likelihood is

$$\ell_t = c(\eta) - \frac{1}{2}\ln\sigma^2_{t|t-1} - \left(\frac{\eta+1}{2\eta}\right)\ln\left[1 + \frac{\eta}{1-2\eta}\frac{\epsilon_t^2}{\sigma^2_{t|t-1}}\right],$$

where $c(\eta)$ is a constant and $\upsilon = 1/\eta$ are the degree of freedoms.

1. The distribution has a clear impact on the TVP's dynamics, in fact the two scaled-scores are

$$\mathbf{s}_{\phi t} = \frac{(1-2\eta)(1+3\eta)}{(1+\eta)}\left(\mathbf{x}_t\sigma^{-2}_{t|t-1}\mathbf{x}_t'\right)^{-1}\mathbf{x}_t\sigma^{-2}_{t|t-1}w_t\epsilon_t,$$

$$s_{\sigma t} = (1+3\eta)\left(w_t\epsilon_t^2 - \sigma^2_{t|t-1}\right).$$

2. The weights $w_t = \frac{(1+\eta)}{(1-2\eta+\eta\zeta_t)}$, where $\zeta_t = \epsilon_t^2/\sigma^2_{t|t-1} \Rightarrow$ the variance affects the conditional mean
3. This leads to a double weighting schema $\Rightarrow$ across obs and across time.

Student-t distribution

The conditional log-likelihood is

$$\ell_t = c(\eta) - \frac{1}{2}\ln\sigma^2_{t|t-1} - \left(\frac{\eta+1}{2\eta}\right)\ln\left[1 + \frac{\eta}{1-2\eta}\frac{\epsilon^2_t}{\sigma^2_{t|t-1}}\right],$$

where $c(\eta)$ is a constant and $\upsilon = 1/\eta$ are the degree of freedoms.

1. The distribution has a clear impact on the TVP's dynamics, in fact the two scaled-scores are

$$\mathbf{s}_{\phi t} = \frac{(1-2\eta)(1+3\eta)}{(1+\eta)}(\mathbf{x}_t\sigma^{-2}_{t|t-1}\mathbf{x}'_t)^{-1}\mathbf{x}_t\sigma^{-2}_{t|t-1}w_t\epsilon_t,$$

$$s_{\sigma t} = (1+3\eta)(w_t\epsilon^2_t - \sigma^2_{t|t-1}).$$

2. The weights $w_t = \frac{(1+\eta)}{(1-2\eta+\eta\zeta_t)}$, where $\zeta_t = \epsilon^2_t/\sigma^2_{t|t-1} \Rightarrow$ the variance affects the conditional mean

3. This leads to a double weighting schema $\Rightarrow$ across obs and across time.

Student-t distribution

The conditional log-likelihood is

$$\ell_t = c(\eta) - \frac{1}{2}\ln\sigma^2_{t|t-1} - \left(\frac{\eta+1}{2\eta}\right)\ln\left[1 + \frac{\eta}{1-2\eta}\frac{\epsilon_t^2}{\sigma^2_{t|t-1}}\right],$$

where $c(\eta)$ is a constant and $\upsilon = 1/\eta$ are the degree of freedoms.

1. The distribution has a clear impact on the TVP's dynamics, in fact the two scaled-scores are

$$\mathbf{s}_{\phi t} = \frac{(1-2\eta)(1+3\eta)}{(1+\eta)}\left(\mathbf{x}_t\sigma^{-2}_{t|t-1}\mathbf{x}'_t\right)^{-1}\mathbf{x}_t\sigma^{-2}_{t|t-1}w_t\epsilon_t,$$

$$s_{\sigma t} = (1+3\eta)\left(w_t\epsilon_t^2 - \sigma^2_{t|t-1}\right).$$

2. The weights $w_t = \frac{(1+\eta)}{(1-2\eta+\eta\zeta_t)}$, where $\zeta_t = \epsilon_t^2/\sigma^2_{t|t-1} \Rightarrow$ the variance affects the conditional mean
3. This leads to a double weighting schema $\Rightarrow$ across obs and across time.

Student-t distribution

The conditional log-likelihood is

$$\ell_t = c(\eta) - \frac{1}{2}\ln\sigma^2_{t|t-1} - \left(\frac{\eta+1}{2\eta}\right)\ln\left[1 + \frac{\eta}{1-2\eta}\frac{\epsilon_t^2}{\sigma^2_{t|t-1}}\right],$$

where $c(\eta)$ is a constant and $\upsilon = 1/\eta$ are the degree of freedoms.

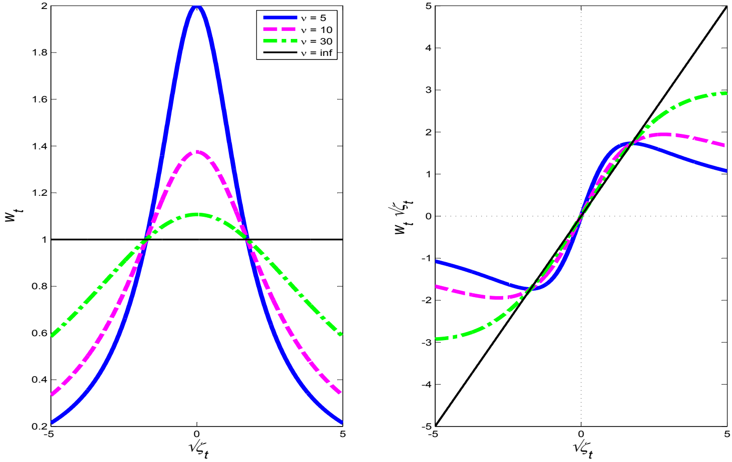
1. The distribution has a clear impact on the TVP's dynamics, in fact the two scaled-scores are

$$\mathbf{s}_{\phi t} = \frac{(1-2\eta)(1+3\eta)}{(1+\eta)}\left(\mathbf{x}_t\sigma^{-2}_{t|t-1}\mathbf{x}_t'\right)^{-1}\mathbf{x}_t\sigma^{-2}_{t|t-1}w_t\epsilon_t,$$

$$s_{\sigma t} = (1+3\eta)\left(w_t\epsilon_t^2 - \sigma^2_{t|t-1}\right).$$

2. The weights $w_t = \frac{(1+\eta)}{(1-2\eta+\eta\zeta_t)}$, where $\zeta_t = \epsilon_t^2/\sigma^2_{t|t-1} \Rightarrow$ the variance affects the conditional mean
3. This leads to a double weighting schema $\Rightarrow$ across obs and across time.

Weighting across realizations



Double weighting schema

As example we consider the time-varying mean and variance only

$$y_t = \mu_{t|t-1} + \epsilon_t, \quad \epsilon_t \sim t_\upsilon(0, \sigma^2_{t|t-1}),$$

$$\mu_{t+1|t} = \mu_{t|t-1} + \kappa_\mu s_{\mu t}, \quad \sigma^2_{t+1|t} = \sigma^2_{t|t-1} + \kappa_\sigma s_{\sigma t}.$$

1. The implied filter for the mean is
$$\mu_{t+1|t} = \frac{\theta}{1-(1-\theta w_t)L}\widetilde{y}_t = \theta \sum_{j=0}^{\infty} \gamma_j \widetilde{y}_{t-j},$$
with $\theta = \frac{\kappa_\mu(1-2\eta)(1+3\eta)}{(1+\eta)}$ and $\widetilde{y}_t = w_t y_t$, providing that $|1-\theta w_t| < 1$.
2. We have a double weighting schema: across realizations regulated by w_t and across time depending on γ_j, where
$$\gamma_0 = 1, \quad \gamma_j = \prod_{k=0}^{j-1}(1-\theta w_{t-k}).$$
3. The estimated variance is $\sigma^2_{t+1|t} = \xi \sum_{j=0}^{\infty}(1-\xi)^j \widetilde{\epsilon}^2_{t-j}$, with $\xi = \kappa_\sigma (1+3\eta)$ and $\widetilde{\epsilon}^2_t = w_t \epsilon^2_t$.
4. We have one-sided low-pass filters on $\widetilde{y}_t = w_t y_t$ and $\widetilde{\epsilon}^2_t = w_t \epsilon^2_t$, respectively. The transfer function for the mean has TVP implying a TV spectral density; see Dahlhaus (2012).

Double weighting schema

As example we consider the time-varying mean and variance only

$$y_t = \mu_{t|t-1} + \epsilon_t, \quad \epsilon_t \sim t_\upsilon(0, \sigma^2_{t|t-1}),$$

$$\mu_{t+1|t} = \mu_{t|t-1} + \kappa_\mu s_{\mu t}, \quad \sigma^2_{t+1|t} = \sigma^2_{t|t-1} + \kappa_\sigma s_{\sigma t}.$$

1. The implied filter for the mean is
$$\mu_{t+1|t} = \frac{\theta}{1-(1-\theta w_t)L}\widetilde{y}_t = \theta \sum_{j=0}^{\infty} \gamma_j \widetilde{y}_{t-j},$$
with $\theta = \frac{\kappa_\mu(1-2\eta)(1+3\eta)}{(1+\eta)}$ and $\widetilde{y}_t = w_t y_t$, providing that $|1-\theta w_t| < 1$.
2. We have a double weighting schema: across realizations regulated by w_t and across time depending on γ_j, where
$$\gamma_0 = 1, \quad \gamma_j = \prod_{k=0}^{j-1}(1-\theta w_{t-k}).$$
3. The estimated variance is $\sigma^2_{t+1|t} = \xi \sum_{j=0}^{\infty}(1-\xi)^j \widetilde{\epsilon}^2_{t-j}$, with $\xi = \kappa_\sigma(1+3\eta)$ and $\widetilde{\epsilon}^2_t = w_t \epsilon^2_t$.
4. We have one-sided low-pass filters on $\widetilde{y}_t = w_t y_t$ and $\widetilde{\epsilon}^2_t = w_t \epsilon^2_t$, respectively. The transfer function for the mean has TVP implying a TV spectral density; see Dahlhaus (2012).

Double weighting schema

As example we consider the time-varying mean and variance only

$$y_t = \mu_{t|t-1} + \epsilon_t, \quad \epsilon_t \sim t_\upsilon(0, \sigma^2_{t|t-1}),$$

$$\mu_{t+1|t} = \mu_{t|t-1} + \kappa_\mu s_{\mu t}, \quad \sigma^2_{t+1|t} = \sigma^2_{t|t-1} + \kappa_\sigma s_{\sigma t}.$$

1. The implied filter for the mean is

$$\mu_{t+1|t} = \frac{\theta}{1-(1-\theta w_t)L}\widetilde{y}_t = \theta \sum_{j=0}^{\infty} \gamma_j \widetilde{y}_{t-j},$$

   with $\theta = \frac{\kappa_\mu(1-2\eta)(1+3\eta)}{(1+\eta)}$ and $\widetilde{y}_t = w_t y_t$, providing that $|1-\theta w_t| < 1$.

2. We have a double weighting schema: across realizations regulated by w_t and across time depending on γ_j, where

$$\gamma_0 = 1, \quad \gamma_j = \prod_{k=0}^{j-1}(1-\theta w_{t-k}).$$

3. The estimated variance is $\sigma^2_{t+1|t} = \xi \sum_{j=0}^{\infty}(1-\xi)^j \widetilde{\epsilon}^2_{t-j}$, with $\xi = \kappa_\sigma(1+3\eta)$ and $\widetilde{\epsilon}^2_t = w_t \epsilon^2_t$.

4. We have one-sided low-pass filters on $\widetilde{y}_t = w_t y_t$ and $\widetilde{\epsilon}^2_t = w_t \epsilon^2_t$, respectively. The transfer function for the mean has TVP implying a TV spectral density; see Dahlhaus (2012).

Double weighting schema

As example we consider the time-varying mean and variance only

$$y_t = \mu_{t|t-1} + \epsilon_t, \quad \epsilon_t \sim t_\upsilon(0, \sigma^2_{t|t-1}),$$

$$\mu_{t+1|t} = \mu_{t|t-1} + \kappa_\mu s_{\mu t}, \quad \sigma^2_{t+1|t} = \sigma^2_{t|t-1} + \kappa_\sigma s_{\sigma t}.$$

1. The implied filter for the mean is

$$\mu_{t+1|t} = \frac{\theta}{1-(1-\theta w_t)L}\widetilde{y}_t = \theta \sum_{j=0}^{\infty} \gamma_j \widetilde{y}_{t-j},$$

   with $\theta = \frac{\kappa_\mu(1-2\eta)(1+3\eta)}{(1+\eta)}$ and $\widetilde{y}_t = w_t y_t$, providing that $|1-\theta w_t| < 1$.
2. We have a double weighting schema: across realizations regulated by w_t and across time depending on γ_j, where

$$\gamma_0 = 1, \quad \gamma_j = \prod_{k=0}^{j-1}(1-\theta w_{t-k}).$$

3. The estimated variance is $\sigma^2_{t+1|t} = \xi \sum_{j=0}^{\infty}(1-\xi)^j \widetilde{\epsilon}^2_{t-j}$, with $\xi = \kappa_\sigma(1+3\eta)$ and $\widetilde{\epsilon}^2_t = w_t \epsilon^2_t$.
4. We have one-sided low-pass filters on $\widetilde{y}_t = w_t y_t$ and $\widetilde{\epsilon}^2_t = w_t \epsilon^2_t$, respectively. The transfer function for the mean has TVP implying a TV spectral density; see Dahlhaus (2012).

Double weighting schema

As example we consider the time-varying mean and variance only

$$y_t = \mu_{t|t-1} + \epsilon_t, \quad \epsilon_t \sim t_\upsilon(0, \sigma^2_{t|t-1}),$$

$$\mu_{t+1|t} = \mu_{t|t-1} + \kappa_\mu s_{\mu t}, \quad \sigma^2_{t+1|t} = \sigma^2_{t|t-1} + \kappa_\sigma s_{\sigma t}.$$

1. The implied filter for the mean is

$$\mu_{t+1|t} = \frac{\theta}{1-(1-\theta w_t)L}\widetilde{y}_t = \theta \sum_{j=0}^{\infty} \gamma_j \widetilde{y}_{t-j},$$

   with $\theta = \frac{\kappa_\mu(1-2\eta)(1+3\eta)}{(1+\eta)}$ and $\widetilde{y}_t = w_t y_t$, providing that $|1-\theta w_t| < 1$.
2. We have a double weighting schema: across realizations regulated by w_t and across time depending on γ_j, where

$$\gamma_0 = 1, \quad \gamma_j = \prod_{k=0}^{j-1}(1-\theta w_{t-k}).$$

3. The estimated variance is $\sigma^2_{t+1|t} = \xi \sum_{j=0}^{\infty}(1-\xi)^j \widetilde{\epsilon}^2_{t-j}$, with $\xi = \kappa_\sigma(1+3\eta)$ and $\widetilde{\epsilon}^2_t = w_t \epsilon^2_t$.
4. We have one-sided low-pass filters on $\widetilde{y}_t = w_t y_t$ and $\widetilde{\epsilon}^2_t = w_t \epsilon^2_t$, respectively. The transfer function for the mean has TVP implying a TV spectral density; see Dahlhaus (2012).

Double weighting schema

As example we consider the time-varying mean and variance only

$$y_t = \mu_{t|t-1} + \epsilon_t, \quad \epsilon_t \sim t_\upsilon(0, \sigma^2_{t|t-1}),$$

$$\mu_{t+1|t} = \mu_{t|t-1} + \kappa_\mu s_{\mu t}, \quad \sigma^2_{t+1|t} = \sigma^2_{t|t-1} + \kappa_\sigma s_{\sigma t}.$$

1. The implied filter for the mean is
$$\mu_{t+1|t} = \frac{\theta}{1-(1-\theta w_t)L}\widetilde{y}_t = \theta \sum_{j=0}^{\infty} \gamma_j \widetilde{y}_{t-j},$$
with $\theta = \frac{\kappa_\mu(1-2\eta)(1+3\eta)}{(1+\eta)}$ and $\widetilde{y}_t = w_t y_t$, providing that $|1-\theta w_t| < 1$.
2. We have a double weighting schema: across realizations regulated by w_t and across time depending on γ_j, where
$$\gamma_0 = 1, \quad \gamma_j = \prod_{k=0}^{j-1}(1-\theta w_{t-k}).$$
3. The estimated variance is $\sigma^2_{t+1|t} = \xi \sum_{j=0}^{\infty}(1-\xi)^j \widetilde{\epsilon}^2_{t-j}$, with $\xi = \kappa_\sigma (1+3\eta)$ and $\widetilde{\epsilon}^2_t = w_t \epsilon^2_t$.
4. We have one-sided low-pass filters on $\widetilde{y}_t = w_t y_t$ and $\widetilde{\epsilon}^2_t = w_t \epsilon^2_t$, respectively. The transfer function for the mean has TVP implying a TV spectral density; see Dahlhaus (2012).

Model restrictions

Often we may want to impose restrictions on the parameters space $\Rightarrow$ this is achived by re-parameterizing the model:

1. In a score-driven approach we need to define the transf. fne $\mathbf{f}_t = g(\widetilde{\mathbf{f}}_t)$, where $\mathbf{f}_t$ is the original vector of par. (we want to restrict), $\widetilde{\mathbf{f}}_t$ is the transf. vector (unrestricted) and $g(.)$ is a continous and twice diff. function (link function)
2. As a consequence, the model is expressed wrt $\widetilde{\mathbf{f}}_t$ and the scaled-score is

$$\widetilde{\mathbf{s}}_t = (\Psi_t' \mathcal{I}_t \Psi_t)^{-1} \Psi_t' \nabla_t,$$

where $\Psi_t = \frac{\partial \mathbf{f}_{t|t-1}}{\partial \widetilde{\mathbf{f}}'_{t|t-1}}$ is the Jacobian of $g(.) \Rightarrow$ re-weights the Gauss-Newton search so that restrictions are satisfied $\Rightarrow$ optimal way to implement the projection facility

The transformation function g(.) imposes (possibly) non-linear restrictions. However, the dynamic of the unrestricted parameters is still conditional linear

Proposition 5: under Gaussian dist and with restrictions on the parameter-driven model (Lemma 1) $\Rightarrow$ the Extended KF leads exactely to the score-driven filter.

Model restrictions

Often we may want to impose restrictions on the parameters space $\Rightarrow$ this is achived by re-parameterizing the model:

1. In a score-driven approach we need to define the transf. fne $\mathbf{f}_t = g(\widetilde{\mathbf{f}}_t)$, where $\mathbf{f}_t$ is the original vector of par. (we want to restrict), $\widetilde{\mathbf{f}}_t$ is the transf. vector (unrestricted) and $g(.)$ is a continous and twice diff. function (link function)
2. As a consequence, the model is expressed wrt $\widetilde{\mathbf{f}}_t$ and the scaled-score is

$$\widetilde{\mathbf{s}}_t = (\Psi_t' \mathcal{I}_t \Psi_t)^{-1} \Psi_t' \nabla_t,$$

where $\Psi_t = \frac{\partial \mathbf{f}_{t|t-1}}{\partial \widetilde{\mathbf{f}}_{t|t-1}'}$ is the Jacobian of $g(.) \Rightarrow$ re-weights the Gauss-Newton search so that restrictions are satisfied $\Rightarrow$ optimal way to implement the projection facility

The transformation function g(.) imposes (possibly) non-linear restrictions. However, the dynamic of the unrestricted parameters is still conditional linear

Proposition 5: under Gaussian dist and with restrictions on the parameter-driven model (Lemma 1) $\Rightarrow$ the Extended KF leads exactely to the score-driven filter.

Model restrictions

Often we may want to impose restrictions on the parameters space $\Rightarrow$ this is achived by re-parameterizing the model:

1. In a score-driven approach we need to define the transf. fne $\mathbf{f}_t = g(\widetilde{\mathbf{f}}_t)$, where $\mathbf{f}_t$ is the original vector of par. (we want to restrict), $\widetilde{\mathbf{f}}_t$ is the transf. vector (unrestricted) and $g(.)$ is a continous and twice diff. function (link function)
2. As a consequence, the model is expressed wrt $\widetilde{\mathbf{f}}_t$ and the scaled-score is

$$\widetilde{\mathbf{s}}_t = (\Psi_t' \mathcal{I}_t \Psi_t)^{-1} \Psi_t' \nabla_t,$$

where $\Psi_t = \frac{\partial \mathbf{f}_{t|t-1}}{\partial \widetilde{\mathbf{f}}'_{t|t-1}}$ is the Jacobian of $g(.) \Rightarrow$ re-weights the Gauss-Newton search so that restrictions are satisfied $\Rightarrow$ optimal way to implement the projection facility

The transformation function g(.) imposes (possibly) non-linear restrictions. However, the dynamic of the unrestricted parameters is still conditional linear

Proposition 5: under Gaussian dist and with restrictions on the parameter-driven model (Lemma 1) $\Rightarrow$ the Extended KF leads exactely to the score-driven filter.

Model restrictions

Often we may want to impose restrictions on the parameters space $\Rightarrow$ this is achived by re-parameterizing the model:

1. In a score-driven approach we need to define the transf. fne $\mathbf{f}_t = g(\widetilde{\mathbf{f}}_t)$, where $\mathbf{f}_t$ is the original vector of par. (we want to restrict), $\widetilde{\mathbf{f}}_t$ is the transf. vector (unrestricted) and $g(.)$ is a continous and twice diff. function (link function)
2. As a consequence, the model is expressed wrt $\widetilde{\mathbf{f}}_t$ and the scaled-score is

$$\widetilde{\mathbf{s}}_t = (\Psi_t' \mathcal{I}_t \Psi_t)^{-1} \Psi_t' \nabla_t,$$

where $\Psi_t = \frac{\partial \mathbf{f}_{t|t-1}}{\partial \widetilde{\mathbf{f}}_{t|t-1}'}$ is the Jacobian of $g(.) \Rightarrow$ re-weights the Gauss-Newton search so that restrictions are satisfied $\Rightarrow$ optimal way to implement the projection facility

The transformation function g(.) imposes (possibly) non-linear restrictions. However, the dynamic of the unrestricted parameters is still conditional linear

Proposition 5: under Gaussian dist and with restrictions on the parameter-driven model (Lemma 1) $\Rightarrow$ the Extended KF leads exactely to the score-driven filter.

Model restrictions

Often we may want to impose restrictions on the parameters space $\Rightarrow$ this is achived by re-parameterizing the model:

1. In a score-driven approach we need to define the transf. fne $\mathbf{f}_t = g(\widetilde{\mathbf{f}}_t)$, where $\mathbf{f}_t$ is the original vector of par. (we want to restrict), $\widetilde{\mathbf{f}}_t$ is the transf. vector (unrestricted) and $g(.)$ is a continous and twice diff. function (link function)
2. As a consequence, the model is expressed wrt $\widetilde{\mathbf{f}}_t$ and the scaled-score is

$$\widetilde{\mathbf{s}}_t = (\Psi_t' \mathcal{I}_t \Psi_t)^{-1} \Psi_t' \nabla_t,$$

where $\Psi_t = \frac{\partial \mathbf{f}_{t|t-1}}{\partial \widetilde{\mathbf{f}}_{t|t-1}'}$ is the Jacobian of $g(.) \Rightarrow$ re-weights the Gauss-Newton search so that restrictions are satisfied $\Rightarrow$ optimal way to implement the projection facility

The transformation function g(.) imposes (possibly) non-linear restrictions. However, the dynamic of the unrestricted parameters is still conditional linear

Proposition 5: under Gaussian dist and with restrictions on the parameter-driven model (Lemma 1) $\Rightarrow$ the Extended KF leads exactely to the score-driven filter.

Imposing stationary coefficients

1. We re-parametrize the model wrt to the partial autocorrelations (PACs).
2. For each time t, the coefficients $\phi_t \in \mathbf{S}_t^p$, where $\mathbf{S}_t^p$ is the hyperplane with stat roots, i.e. $\phi_t(\mathbf{z}_t) \neq 0$, where $\mathbf{z}_t = (z_{1t}, ..., z_{pt}) \in \mathbf{C}^p$ and $|z_{jt}| < 1$.
3. We consider $\phi_t = (\phi_{1t}, ..., \phi_{pt})'$, the corresponding PACs $\pi_t = (\pi_{1t}, ..., \pi_{pt})'$ and the unrestricted parameters $\alpha_t = (\alpha_{1t}, ..., \alpha_{pt})'$.
4. There is a unique function mapping the AR coeffs to the PACs that is $\phi_t = \Phi(\pi_t) \Rightarrow$ the DL algorithm; see Bandorff-Nielsen and Schou (1973)
5. A stationary AR process has that $\pi_{jt} \in (-1, 1)$, thus we define another function $\pi_t = \Upsilon(\alpha_t)$, such that $|\pi_{jt}| < 1$; e.g. inverse Fisher transf
6. We finally have the mapping between the stationary coeffs and the unrestricted parameters, i.e. $\phi_t = g(\alpha_t) = \Phi[\Upsilon(\alpha_t)]$, where $\alpha_t \in (-\infty, \infty)$, $\pi_t \in (-1, 1)$ and $\phi_t \in \mathbf{S}^p$.

Imposing stationary coefficients

1. We re-parametrize the model wrt to the partial autocorrelations (PACs).
2. For each time t, the coefficients $\phi_t \in \mathbf{S}_t^p$, where $\mathbf{S}_t^p$ is the hyperplane with stat roots, i.e. $\phi_t(\mathbf{z}_t) \neq 0$, where $\mathbf{z}_t = (z_{1t}, ..., z_{pt}) \in \mathbf{C}^p$ and $|z_{jt}| < 1$.
3. We consider $\phi_t = (\phi_{1t}, ..., \phi_{pt})'$, the corresponding PACs $\pi_t = (\pi_{1t}, ..., \pi_{pt})'$ and the unrestricted parameters $\alpha_t = (\alpha_{1t}, ..., \alpha_{pt})'$.
4. There is a unique function mapping the AR coeffs to the PACs that is $\phi_t = \Phi(\pi_t) \Rightarrow$ the DL algorithm; see Bandorff-Nielsen and Schou (1973)
5. A stationary AR process has that $\pi_{jt} \in (-1, 1)$, thus we define another function $\pi_t = \Upsilon(\alpha_t)$, such that $|\pi_{jt}| < 1$; e.g. inverse Fisher transf
6. We finally have the mapping between the stationary coeffs and the unrestricted parameters, i.e. $\phi_t = g(\alpha_t) = \Phi[\Upsilon(\alpha_t)]$, where $\alpha_t \in (-\infty, \infty)$, $\pi_t \in (-1, 1)$ and $\phi_t \in \mathbf{S}^p$.

Imposing stationary coefficients

1. We re-parametrize the model wrt to the partial autocorrelations (PACs).
2. For each time t, the coefficients $\phi_t \in \mathbf{S}_t^p$, where $\mathbf{S}_t^p$ is the hyperplane with stat roots, i.e. $\phi_t(\mathbf{z}_t) \neq 0$, where $\mathbf{z}_t = (z_{1t}, ..., z_{pt}) \in \mathbf{C}^p$ and $|z_{jt}| < 1$.
3. We consider $\phi_t = (\phi_{1t}, ..., \phi_{pt})'$, the corresponding PACs $\pi_t = (\pi_{1t}, ..., \pi_{pt})'$ and the unrestricted parameters $\alpha_t = (\alpha_{1t}, ..., \alpha_{pt})'$.
4. There is a unique function mapping the AR coeffs to the PACs that is $\phi_t = \Phi(\pi_t) \Rightarrow$ the DL algorithm; see Bandorff-Nielsen and Schou (1973)
5. A stationary AR process has that $\pi_{jt} \in (-1, 1)$, thus we define another function $\pi_t = \Upsilon(\alpha_t)$, such that $|\pi_{jt}| < 1$; e.g. inverse Fisher transf
6. We finally have the mapping between the stationary coeffs and the unrestricted parameters, i.e. $\phi_t = g(\alpha_t) = \Phi[\Upsilon(\alpha_t)]$, where $\alpha_t \in (-\infty, \infty)$, $\pi_t \in (-1, 1)$ and $\phi_t \in \mathbf{S}^p$.

Imposing stationary coefficients

1. We re-parametrize the model wrt to the partial autocorrelations (PACs).

2. For each time t, the coefficients $\phi_t \in \mathbf{S}_t^p$, where $\mathbf{S}_t^p$ is the hyperplane with stat roots, i.e. $\phi_t(\mathbf{z}_t) \neq 0$, where $\mathbf{z}_t = (z_{1t}, ..., z_{pt}) \in \mathbf{C}^p$ and $|z_{jt}| < 1$.

3. We consider $\phi_t = (\phi_{1t}, ..., \phi_{pt})'$, the corresponding PACs $\pi_t = (\pi_{1t}, ..., \pi_{pt})'$ and the unrestricted parameters $\alpha_t = (\alpha_{1t}, ..., \alpha_{pt})'$.

4. There is a unique function mapping the AR coeffs to the PACs that is $\phi_t = \Phi(\pi_t) \Rightarrow$ the DL algorithm; see Bandorff-Nielsen and Schou (1973)

5. A stationary AR process has that $\pi_{jt} \in (-1, 1)$, thus we define another function $\pi_t = \Upsilon(\alpha_t)$, such that $|\pi_{jt}| < 1$; e.g. inverse Fisher transf

6. We finally have the mapping between the stationary coeffs and the unrestricted parameters, i.e. $\phi_t = g(\alpha_t) = \Phi[\Upsilon(\alpha_t)]$, where $\alpha_t \in (-\infty, \infty)$, $\pi_t \in (-1, 1)$ and $\phi_t \in \mathbf{S}^p$.

Imposing stationary coefficients

1. We re-parametrize the model wrt to the partial autocorrelations (PACs).
2. For each time t, the coefficients $\phi_t \in \mathbf{S}_t^p$, where $\mathbf{S}_t^p$ is the hyperplane with stat roots, i.e. $\phi_t(\mathbf{z}_t) \neq 0$, where $\mathbf{z}_t = (z_{1t}, ..., z_{pt}) \in \mathbf{C}^p$ and $|z_{jt}| < 1$.
3. We consider $\phi_t = (\phi_{1t}, ..., \phi_{pt})'$, the corresponding PACs $\pi_t = (\pi_{1t}, ..., \pi_{pt})'$ and the unrestricted parameters $\alpha_t = (\alpha_{1t}, ..., \alpha_{pt})'$.
4. There is a unique function mapping the AR coeffs to the PACs that is $\phi_t = \Phi(\pi_t) \Rightarrow$ the DL algorithm; see Bandorff-Nielsen and Schou (1973)
5. A stationary AR process has that $\pi_{jt} \in (-1, 1)$, thus we define another function $\pi_t = \Upsilon(\alpha_t)$, such that $|\pi_{jt}| < 1$; e.g. inverse Fisher transf
6. We finally have the mapping between the stationary coeffs and the unrestricted parameters, i.e. $\phi_t = g(\alpha_t) = \Phi[\Upsilon(\alpha_t)]$, where $\alpha_t \in (-\infty, \infty)$, $\pi_t \in (-1, 1)$ and $\phi_t \in \mathbf{S}^p$.

Imposing stationary coefficients

1. We re-parametrize the model wrt to the partial autocorrelations (PACs).
2. For each time t, the coefficients $\phi_t \in \mathbf{S}_t^p$, where $\mathbf{S}_t^p$ is the hyperplane with stat roots, i.e. $\phi_t(\mathbf{z}_t) \neq 0$, where $\mathbf{z}_t = (z_{1t}, ..., z_{pt}) \in \mathbf{C}^p$ and $|z_{jt}| < 1$.
3. We consider $\phi_t = (\phi_{1t}, ..., \phi_{pt})'$, the corresponding PACs $\pi_t = (\pi_{1t}, ..., \pi_{pt})'$ and the unrestricted parameters $\alpha_t = (\alpha_{1t}, ..., \alpha_{pt})'$.
4. There is a unique function mapping the AR coeffs to the PACs that is $\phi_t = \Phi(\pi_t) \Rightarrow$ the DL algorithm; see Bandorff-Nielsen and Schou (1973)
5. A stationary AR process has that $\pi_{jt} \in (-1, 1)$, thus we define another function $\pi_t = \Upsilon(\alpha_t)$, such that $|\pi_{jt}| < 1$; e.g. inverse Fisher transf
6. We finally have the mapping between the stationary coeffs and the unrestricted parameters, i.e. $\phi_t = g(\alpha_t) = \Phi[\Upsilon(\alpha_t)]$, where $\alpha_t \in (-\infty, \infty)$, $\pi_t \in (-1, 1)$ and $\phi_t \in \mathbf{S}^p$.

Imposing stationary coefficients

1. We re-parametrize the model wrt to the partial autocorrelations (PACs).
2. For each time t, the coefficients $\phi_t \in \mathbf{S}_t^p$, where $\mathbf{S}_t^p$ is the hyperplane with stat roots, i.e. $\phi_t(\mathbf{z}_t) \neq 0$, where $\mathbf{z}_t = (z_{1t}, ..., z_{pt}) \in \mathbf{C}^p$ and $|z_{jt}| < 1$.
3. We consider $\phi_t = (\phi_{1t}, ..., \phi_{pt})'$, the corresponding PACs $\pi_t = (\pi_{1t}, ..., \pi_{pt})'$ and the unrestricted parameters $\alpha_t = (\alpha_{1t}, ..., \alpha_{pt})'$.
4. There is a unique function mapping the AR coeffs to the PACs that is $\phi_t = \Phi(\pi_t) \Rightarrow$ the DL algorithm; see Bandorff-Nielsen and Schou (1973)
5. A stationary AR process has that $\pi_{jt} \in (-1, 1)$, thus we define another function $\pi_t = \Upsilon(\alpha_t)$, such that $|\pi_{jt}| < 1$; e.g. inverse Fisher transf
6. We finally have the mapping between the stationary coeffs and the unrestricted parameters, i.e. $\phi_t = g(\alpha_t) = \Phi[\Upsilon(\alpha_t)]$, where $\alpha_t \in (-\infty, \infty)$, $\pi_t \in (-1, 1)$ and $\phi_t \in \mathbf{S}^p$.

The Jacobian of the DL mapping function

The Jacobian of inverse Fisher transf. is straightforward, while a novel expression for $\Gamma_t = \partial\phi_t/\partial\pi_t'$ is derived $\Rightarrow$ the last iteration of the recursion

$$\Gamma_{k,t} = \begin{bmatrix} \widetilde{\Gamma}_{k-1,t} & \mathbf{b}_{k-1,t} \\ \mathbf{0}_{k-1}' & 1 \end{bmatrix}, \quad \widetilde{\Gamma}_{k-1,t} = \mathbf{J}_{k-1,t}\Gamma_{k-1,t}, \quad k = 2, ..., p,$$

where $\Gamma_{1,t} = 1$, $\mathbf{J}_{1,t} = (1 - \pi_{2t})$ and

$$\mathbf{b}_{k-1,t} = - \begin{bmatrix} \phi_t^{k-1,k-1} \\ \phi_t^{k-2,k-1} \\ \vdots \\ \phi_t^{2,k-1} \\ \phi_t^{1,k-1} \end{bmatrix}, \quad \mathbf{J}_{k-1,t} = \begin{bmatrix} 1 & 0 & \cdots & 0 & -\pi_{kt} \\ 0 & 1 & 0 & -\pi_{kt} & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & -\pi_{kt} & 0 & 1 & 0 \\ -\pi_{kt} & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

if k is even the central element of $\mathbf{J}_{k-1,t}$ is $(1 - \pi_{kt})$.

The Jacobian of the DL mapping function

The Jacobian of inverse Fisher transf. is straightforward, while a novel expression for $\Gamma_t = \partial\phi_t/\partial\pi_t'$ is derived $\Rightarrow$ the last iteration of the recursion

$$\Gamma_{k,t} = \begin{bmatrix} \widetilde{\Gamma}_{k-1,t} & \mathbf{b}_{k-1,t} \\ \mathbf{0}_{k-1}' & 1 \end{bmatrix}, \quad \widetilde{\Gamma}_{k-1,t} = \mathbf{J}_{k-1,t}\Gamma_{k-1,t}, \quad k = 2, ..., p,$$

where $\Gamma_{1,t} = 1$, $\mathbf{J}_{1,t} = (1 - \pi_{2t})$ and

$$\mathbf{b}_{k-1,t} = - \begin{bmatrix} \phi_t^{k-1,k-1} \\ \phi_t^{k-2,k-1} \\ \vdots \\ \phi_t^{2,k-1} \\ \phi_t^{1,k-1} \end{bmatrix}, \quad \mathbf{J}_{k-1,t} = \begin{bmatrix} 1 & 0 & \cdots & 0 & -\pi_{kt} \\ 0 & 1 & 0 & -\pi_{kt} & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & -\pi_{kt} & 0 & 1 & 0 \\ -\pi_{kt} & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

if k is even the central element of $\mathbf{J}_{k-1,t}$ is $(1 - \pi_{kt})$.

The Jacobian of the DL mapping function

The Jacobian of inverse Fisher transf. is straightforward, while a novel expression for $\Gamma_t = \partial\phi_t/\partial\pi'_t$ is derived $\Rightarrow$ the last iteration of the recursion

$$\Gamma_{k,t} = \begin{bmatrix} \widetilde{\Gamma}_{k-1,t} & \mathbf{b}_{k-1,t} \\ \mathbf{0}'_{k-1} & 1 \end{bmatrix}, \quad \widetilde{\Gamma}_{k-1,t} = \mathbf{J}_{k-1,t}\Gamma_{k-1,t}, \quad k = 2, ..., p,$$

where $\Gamma_{1,t} = 1$, $\mathbf{J}_{1,t} = (1 - \pi_{2t})$ and

$$\mathbf{b}_{k-1,t} = -\begin{bmatrix} \phi_t^{k-1,k-1} \\ \phi_t^{k-2,k-1} \\ \vdots \\ \phi_t^{2,k-1} \\ \phi_t^{1,k-1} \end{bmatrix}, \quad \mathbf{J}_{k-1,t} = \begin{bmatrix} 1 & 0 & \cdots & 0 & -\pi_{kt} \\ 0 & 1 & 0 & -\pi_{kt} & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & -\pi_{kt} & 0 & 1 & 0 \\ -\pi_{kt} & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

if k is even the central element of $\mathbf{J}_{k-1,t}$ is $(1 - \pi_{kt})$.

Imposing bounded mean

We may want to discipline the algorithm so to have a bounded mean:

$$\mu_t = \frac{\phi_{0t}}{1 - \sum_{j=1}^{p} \phi_{jt}} \in (a; b).$$

We impose the following transformation

$$\phi_{0t} = \frac{a + b\exp(\alpha_{0t})}{1 + \exp(\alpha_{0t})} \left(1 - \sum_{j=1}^{p} \phi_{j,t}\right),$$

and the Jacobian of this transformation is also derived.

Imposing bounded mean

We may want to discipline the algorithm so to have a bounded mean:

$$\mu_t = \frac{\phi_{0t}}{1 - \sum_{j=1}^{p} \phi_{jt}} \in (a; b).$$

We impose the following transformation

$$\phi_{0t} = \frac{a + b\exp(\alpha_{0t})}{1 + \exp(\alpha_{0t})} \left(1 - \sum_{j=1}^{p} \phi_{j,t}\right),$$

and the Jacobian of this transformation is also derived.

Application to the US inflation

We model the quarterly US CPI inflation 1955q1:2012q4

$$\pi_t = \phi_{0,t} + \phi_{1,t}\pi_{t-1} + \ldots + \phi_{p,t}\pi_{t-p} + \epsilon_t, \quad \epsilon_t \sim IID(0, \sigma_t^2)$$

Normal	Trend	Trend-B	AR(1)	AR(1)-B	AR(2)	AR(2)-B	AR(4)	AR(4)-B
κ_c	0.3367	0.3547	0.0407	0.0387	0.0479	0.0286	0.0325	0.0239
	(0.0480)	(0.0190)	(0.0048)	(0.0039)	(0.0027)	(0.0029)	(0.0028)	(0.0021)
κ_σ	0.1479	0.1910	0.1127	0.1341	0.1044	0.1484	0.0806	0.1036
	(0.0277)	(0.0189)	(0.0180)	(0.0225)	(0.0128)	(0.0293)	(0.0064)	(0.0142)
LogLik	-561.9535	-573.3571	-546.0505	-535.0638	-554.6153	-535.9258	-555.9525	-539.2061
AIC	1131.9071	1154.7142	1102.1009	1080.1277	1121.2306	1083.8516	1127.9050	1094.4123
BIC	1146.1498	1168.9570	1119.9044	1097.9311	1142.5946	1105.2156	1156.3905	1122.8977

Student-t	Trend	Trend-B	AR(1)	AR(1)-B	AR(2)	AR(2)-B	AR(4)	AR(4)-B
κ_c	0.5415	0.1841	0.0452	0.0367	0.0310	0.0366	0.0413	0.0286
	(0.0115)	(0.0058)	(0.0035)	(0.0038)	(0.0013)	(0.0043)	(0.0043)	(0.0025)
κ_σ	0.1632	0.2461	0.1445	0.2046	0.1620	0.1804	0.1520	0.1789
	(0.0105)	(0.1131)	(0.0096)	(0.0633)	(0.0116)	(0.0497)	(0.0505)	(0.0122)
υ	5.7577	4.8640	4.5656	5.3994	5.3317	5.0520	4.6766	4.8111
	(0.0876)	(0.4650)	(0.0827)	(0.4189)	(0.0675)	(0.4380)	(0.4650)	(0.0808)
LogLik	-521.8818	-548.3469	-513.6383	-512.7556	-519.5567	-514.3814	-515.5358	-508.2651
AIC	1053.7637	1106.6937	1039.2766	1037.5111	1053.1134	1042.7628	1049.0716	1034.5302
BIC	1071.5671	1124.4972	1060.6407	1058.8752	1078.0381	1067.6876	1081.1177	1066.5763

Application to the US inflation

We model the quarterly US CPI inflation 1955q1:2012q4

$$\pi_t = \phi_{0,t} + \phi_{1,t}\pi_{t-1} + ... + \phi_{p,t}\pi_{t-p} + \epsilon_t, \quad \epsilon_t \sim IID(0, \sigma_t^2)$$

	Normal							
	Trend	Trend-B	AR(1)	AR(1)-B	AR(2)	AR(2)-B	AR(4)	AR(4)-B
κ_c	0.3367	0.3547	0.0407	0.0387	0.0479	0.0286	0.0325	0.0239
	(0.0480)	(0.0190)	(0.0048)	(0.0039)	(0.0027)	(0.0029)	(0.0028)	(0.0021)
κ_σ	0.1479	0.1910	0.1127	0.1341	0.1044	0.1484	0.0806	0.1036
	(0.0277)	(0.0189)	(0.0180)	(0.0225)	(0.0128)	(0.0293)	(0.0064)	(0.0142)
LogLik	-561.9535	-573.3571	-546.0505	-535.0638	-554.6153	-535.9258	-555.9525	-539.2061
AIC	1131.9071	1154.7142	1102.1009	1080.1277	1121.2306	1083.8516	1127.9050	1094.4123
BIC	1146.1498	1168.9570	1119.9044	1097.9311	1142.5946	1105.2156	1156.3905	1122.8977
	Student-t							
	Trend	Trend-B	AR(1)	AR(1)-B	AR(2)	AR(2)-B	AR(4)	AR(4)-B
κ_c	0.5415	0.1841	0.0452	0.0367	0.0310	0.0366	0.0413	0.0286
	(0.0115)	(0.0058)	(0.0035)	(0.0038)	(0.0013)	(0.0043)	(0.0043)	(0.0025)
κ_σ	0.1632	0.2461	0.1445	0.2046	0.1620	0.1804	0.1520	0.1789
	(0.0105)	(0.1131)	(0.0096)	(0.0633)	(0.0116)	(0.0497)	(0.0505)	(0.0122)
υ	5.7577	4.8640	4.5656	5.3994	5.3317	5.0520	4.6766	4.8111
	(0.0876)	(0.4650)	(0.0827)	(0.4189)	(0.0675)	(0.4380)	(0.4650)	(0.0808)
LogLik	-521.8818	-548.3469	-513.6383	-512.7556	-519.5567	-514.3814	-515.5358	-508.2651
AIC	1053.7637	1106.6937	1039.2766	1037.5111	1053.1134	1042.7628	1049.0716	1034.5302
BIC	1071.5671	1124.4972	1060.6407	1058.8752	1078.0381	1067.6876	1081.1177	1066.5763

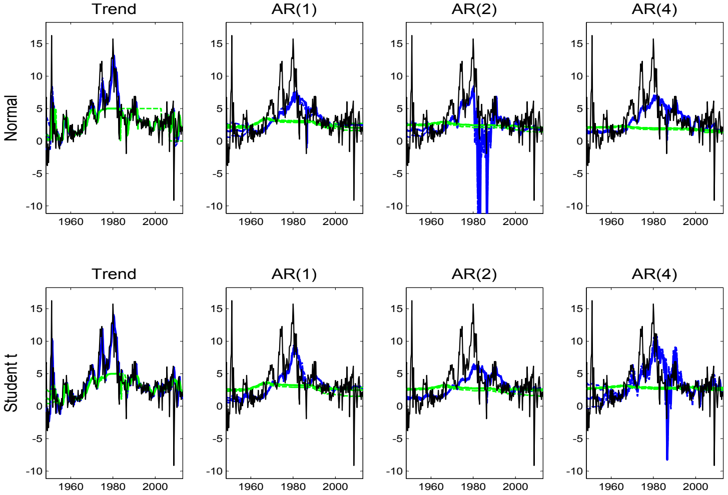
Application to the US inflation

We model the quarterly US CPI inflation 1955q1:2012q4

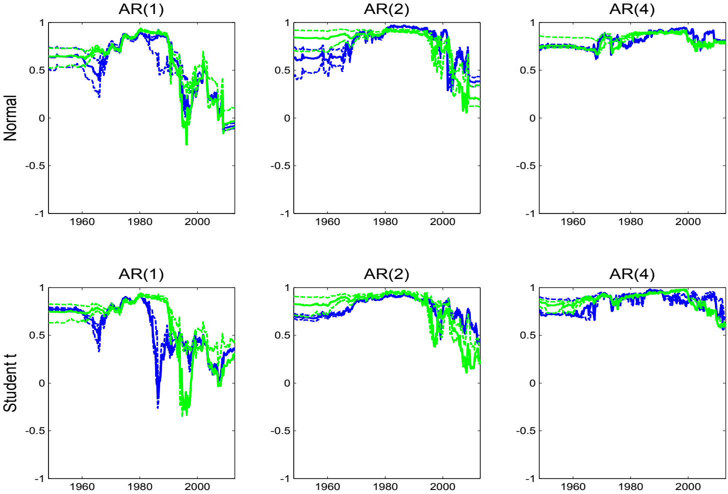
$$\pi_t = \phi_{0,t} + \phi_{1,t}\pi_{t-1} + ... + \phi_{p,t}\pi_{t-p} + \epsilon_t, \quad \epsilon_t \sim IID(0, \sigma_t^2)$$

	Normal							
	Trend	Trend-B	AR(1)	AR(1)-B	AR(2)	AR(2)-B	AR(4)	AR(4)-B
κ_c	0.3367	0.3547	0.0407	0.0387	0.0479	0.0286	0.0325	0.0239
	(0.0480)	(0.0190)	(0.0048)	(0.0039)	(0.0027)	(0.0029)	(0.0028)	(0.0021)
κ_σ	0.1479	0.1910	0.1127	0.1341	0.1044	0.1484	0.0806	0.1036
	(0.0277)	(0.0189)	(0.0180)	(0.0225)	(0.0128)	(0.0293)	(0.0064)	(0.0142)
LogLik	-561.9535	-573.3571	-546.0505	-535.0638	-554.6153	-535.9258	-555.9525	-539.2061
AIC	1131.9071	1154.7142	1102.1009	1080.1277	1121.2306	1083.8516	1127.9050	1094.4123
BIC	1146.1498	1168.9570	1119.9044	1097.9311	1142.5946	1105.2156	1156.3905	1122.8977
	Student-t							
	Trend	Trend-B	AR(1)	AR(1)-B	AR(2)	AR(2)-B	AR(4)	AR(4)-B
κ_c	0.5415	0.1841	0.0452	0.0367	0.0310	0.0366	0.0413	0.0286
	(0.0115)	(0.0058)	(0.0035)	(0.0038)	(0.0013)	(0.0043)	(0.0043)	(0.0025)
κ_σ	0.1632	0.2461	0.1445	0.2046	0.1620	0.1804	0.1520	0.1789
	(0.0105)	(0.1131)	(0.0096)	(0.0633)	(0.0116)	(0.0497)	(0.0505)	(0.0122)
υ	5.7577	4.8640	4.5656	5.3994	5.3317	5.0520	4.6766	4.8111
	(0.0876)	(0.4650)	(0.0827)	(0.4189)	(0.0675)	(0.4380)	(0.4650)	(0.0808)
LogLik	-521.8818	-548.3469	-513.6383	-512.7556	-519.5567	-514.3814	-515.5358	-508.2651
AIC	1053.7637	1106.6937	1039.2766	1037.5111	1053.1134	1042.7628	1049.0716	1034.5302
BIC	1071.5671	1124.4972	1060.6407	1058.8752	1078.0381	1067.6876	1081.1177	1066.5763

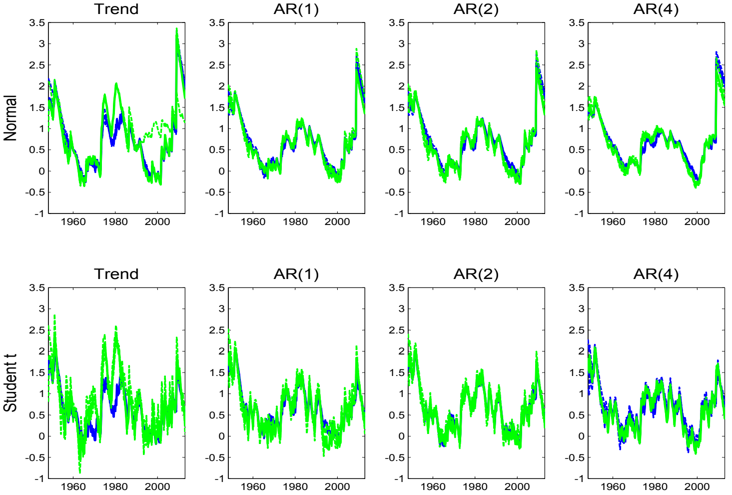
Long-run trend



Inflation Persistence: sum of ARs coeffs



Volatility


Trend
AR(1)
AR(2)
AR(4)
Normal
Student t
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
1960
1980
2000

Point Forecast 1984q1–2012q4

	RMSFE			MAFE		
	h=1	h=4	h=8	h=1	h=4	h=8
			Normal			
Trend	2.2259	2.4902	2.4748	1.3787	1.6961	1.8652
	—	—	—	—	—	—
Trend-B	0.8550	0.8444	0.8665	0.8909	0.8537	0.8391
	(0.0314)	(0.0071)	(0.1668)	(0.0677)	(0.0396)	(0.0649)
AR(1)	0.9294	0.8547	0.9097	0.9648	0.8921	0.8703
	(0.3795)	(0.0423)	(0.4582)	(0.6312)	(0.1934)	(0.2954)
AR(1)-B	0.9131	0.8052	0.8137	0.9413	0.7723	0.7381
	(0.2868)	(0.0048)	(0.1117)	(0.4127)	(0.0014)	(0.0152)
AR(2)	0.9446	0.8224	0.7996	0.9629	0.8040	0.7589
	(0.1138)	(0.0040)	(0.0613)	(0.4258)	(0.0018)	(0.0142)
AR(2)-B	0.9603	0.8426	0.7949	0.9535	0.8031	0.7388
	(0.4839)	(0.0037)	(0.0620)	(0.4545)	(0.0013)	(0.0113)
AR(4)	0.9627	0.8466	0.8116	0.9368	0.8147	0.7536
	(0.4609)	(0.0054)	(0.0539)	(0.2372)	(0.0054)	(0.0073)
AR(4)-B	0.9307	0.8562	0.8095	0.9042	0.8319	0.7599
	(0.0745)	(0.0053)	(0.0650)	(0.0818)	(0.0065)	(0.0114)
			Student-t			
Trend	0.9687	0.9345	0.9174	0.9331	0.9290	0.8977
	(0.6436)	(0.0742)	(0.3665)	(0.3487)	(0.0999)	(0.2410)
Trend-B	0.9276	0.8849	0.9056	0.9338	0.9064	0.8650
	(0.2899)	(0.0306)	(0.3999)	(0.4150)	(0.1929)	(0.2106)
AR(1)	0.8846	0.8422	0.8971	0.9288	0.8657	0.8367
	(0.0729)	(0.0238)	(0.4062)	(0.2997)	(0.1379)	(0.1853)
AR(1)-B	0.8722	0.8108	0.8387	0.8943	0.8056	0.7684
	(0.1122)	(0.0068)	(0.1779)	(0.1587)	(0.0099)	(0.0372)
AR(2)	0.9451	0.8893	0.9293	0.9442	0.8584	0.8482
	(0.1081)	(0.0174)	(0.5327)	(0.2302)	(0.0125)	(0.1081)
AR(2)-B	0.8712	0.7970	0.8079	0.9147	0.7835	0.7468
	(0.0813)	(0.0035)	(0.1154)	(0.1790)	(0.0021)	(0.0238)
AR(4)	0.9435	0.8429	0.8398	0.9369	0.8355	0.7815
	(0.1996)	(0.0104)	(0.1052)	(0.2771)	(0.0322)	(0.0088)
AR(4)-B	0.9413	0.8480	0.8270	0.9239	0.8258	0.7733
	(0.1603)	(0.0092)	(0.0634)	(0.1878)	(0.0173)	(0.0053)

Density forecast: a quick overview

We want to assess how a conditional density forecast approximates the "true" density forecast:

1. The log-score: the density forecast is evaluated at the realization y_{t+h} and it gives higher score to the density forecast with higher prob of y_{t+h}.
2. Density forecasts are ranked according to the log-score and test for the difference between log-scores by Amisano and Giacomini (2007).
3. PIT (Prob Integral Transf): the cdf of candidate density is evaluated at $y_{t+h} \Rightarrow$ we have good approx of the "true" density if the PITs are IIDU(0,1).
4. Diebold (1998): visual inspection of histogram of the PITs to be U(0,1).
5. Berkowitz (2001) computes the inverse normal cdf transf and then test for NID by fitting an AR(1) with intercept and then test LR_3
6. Rossi and Sekhposyan (2013) propose a non-parametric test that considers parameters uncertainty.

Density forecast: a quick overview

We want to assess how a conditional density forecast approximates the "true" density forecast:

1. The log-score: the density forecast is evaluated at the realization y_{t+h} and it gives higher score to the density forecast with higher prob of y_{t+h}.
2. Density forecasts are ranked according to the log-score and test for the difference between log-scores by Amisano and Giacomini (2007).
3. PIT (Prob Integral Transf): the cdf of candidate density is evaluated at $y_{t+h} \Rightarrow$ we have good approx of the "true" density if the PITs are IIDU(0,1).
4. Diebold (1998): visual inspection of histogram of the PITs to be U(0,1).
5. Berkowitz (2001) computes the inverse normal cdf transf and then test for NID by fitting an AR(1) with intercept and then test LR_3
6. Rossi and Sekhposyan (2013) propose a non-parametric test that considers parameters uncertainty.

Density forecast: a quick overview

We want to assess how a conditional density forecast approximates the "true" density forecast:

1. The log-score: the density forecast is evaluated at the realization y_{t+h} and it gives higher score to the density forecast with higher prob of y_{t+h}.
2. Density forecasts are ranked according to the log-score and test for the difference between log-scores by Amisano and Giacomini (2007).
3. PIT (Prob Integral Transf): the cdf of candidate density is evaluated at $y_{t+h} \Rightarrow$ we have good approx of the "true" density if the PITs are IIDU(0,1).
4. Diebold (1998): visual inspection of histogram of the PITs to be U(0,1).
5. Berkowitz (2001) computes the inverse normal cdf transf and then test for NID by fitting an AR(1) with intercept and then test LR_3
6. Rossi and Sekhposyan (2013) propose a non-parametric test that considers parameters uncertainty.

Density forecast: a quick overview

We want to assess how a conditional density forecast approximates the "true" density forecast:

1. The log-score: the density forecast is evaluated at the realization y_{t+h} and it gives higher score to the density forecast with higher prob of y_{t+h}.
2. Density forecasts are ranked according to the log-score and test for the difference between log-scores by Amisano and Giacomini (2007).
3. PIT (Prob Integral Transf): the cdf of candidate density is evaluated at $y_{t+h} \Rightarrow$ we have good approx of the "true" density if the PITs are IIDU(0,1).
4. Diebold (1998): visual inspection of histogram of the PITs to be U(0,1).
5. Berkowitz (2001) computes the inverse normal cdf transf and then test for NID by fitting an AR(1) with intercept and then test LR_3
6. Rossi and Sekhposyan (2013) propose a non-parametric test that considers parameters uncertainty.

Density forecast: a quick overview

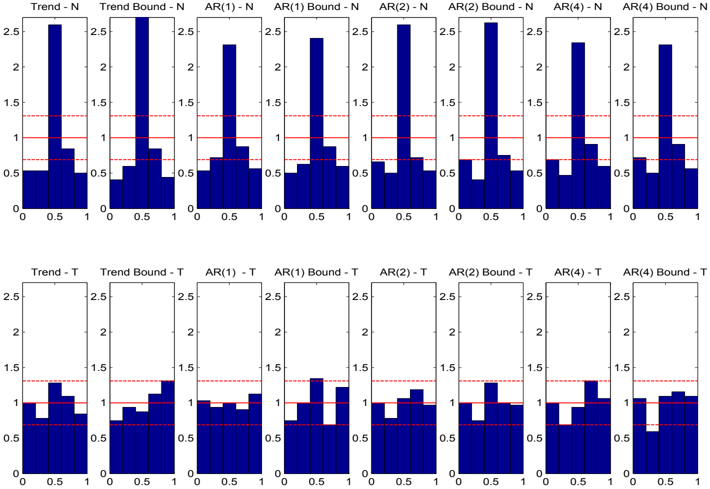
We want to assess how a conditional density forecast approximates the "true" density forecast:

1. The log-score: the density forecast is evaluated at the realization y_{t+h} and it gives higher score to the density forecast with higher prob of y_{t+h}.
2. Density forecasts are ranked according to the log-score and test for the difference between log-scores by Amisano and Giacomini (2007).
3. PIT (Prob Integral Transf): the cdf of candidate density is evaluated at $y_{t+h} \Rightarrow$ we have good approx of the "true" density if the PITs are IIDU(0,1).
4. Diebold (1998): visual inspection of histogram of the PITs to be U(0,1).
5. Berkowitz (2001) computes the inverse normal cdf transf and then test for NID by fitting an AR(1) with intercept and then test LR_3
6. Rossi and Sekhposyan (2013) propose a non-parametric test that considers parameters uncertainty.

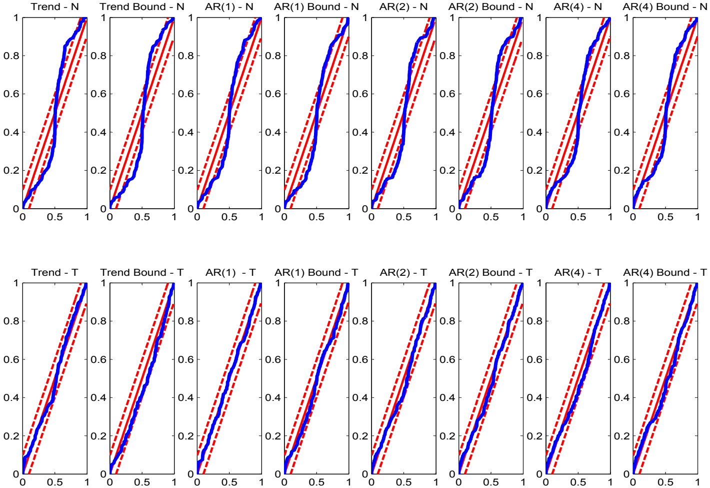
Density forecast

	Normal			Student-t		
	Av Log Score	LR	$\kappa^{CS}_{\alpha,P}$	Av Log Score	LR	$\kappa^{CS}_{\alpha,P}$
Trend	-2.8237	0.0001	5.7760	-1.5999	0.5694	0.9923
Trend-B	-3.0188	0.0001	6.6422	-1.6353	0.0124	1.5210
AR(1)	-2.7127	0.0055	4.0960	-1.6065	0.6715	0.1322
AR(1)-B	-2.6537	0.3831	4.7610	-1.6223	0.5172	0.5760
AR(2)	-2.7784	0.0129	4.7610	-1.6145	0.1988	1.1560
AR(2)-B	-2.6932	0.0121	4.7610	-1.6146	0.3501	0.4623
AR(2)	-2.9495	0.1794	4.4223	-1.6313	0.2424	0.9923
AR(4)-B	-2.7859	0.0822	4.0960	-1.6603	0.1826	0.9923

Density forecast: inspection of the PITs



Density forecast: inspection of the PITs–RS test



Pairwise comparison. Forecasting sample 1973q1–2012q4

	Normal								Student-t						
	Trend	Trend-B	AR(1)	AR(1)-B	AR(2)	AR(2)-B	AR(4)	AR(4)-B	Trend	Trend-B	AR(1)	AR(1)-B	AR(2)	AR(2)-B	AR(4)
Normal															
Trend-B	0.003														
AR(1)	0.018	0.000													
AR(1)-B	0.033	0.000	0.167												
AR(2)	0.158	0.000	0.183	0.101											
AR(2)-B	0.024	0.000	0.741	0.605	0.025										
AR(4)	0.077	0.460	0.008	0.010	0.021	0.008									
AR(4)-B	0.548	0.003	0.298	0.148	0.893	0.141	0.003								
Student-t															
Trend	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000							
Trend-B	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.526						
AR(1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.882	0.569					
AR(1)-B	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.666	0.826	0.670				
AR(2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.721	0.703	0.832	0.879			
AR(2)-B	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.750	0.697	0.804	0.858	0.998		
AR(4)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.490	0.950	0.639	0.883	0.670	0.771	
AR(4)-B	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.243	0.717	0.351	0.557	0.331	0.452	0.190

Summary and future research

1. We propose a score-driven approach for TVPs models and we concentrate on the AR model
2. We show how the implied algorithms are related to the learning algorithms
3. We extend existing adaptive algorithms to the case of changes in volatility and heavy-tails based on the score-driven criterion
4. We show how to impose restrictions on the time-varying parameters
5. Application to the US inflation leads to interesting results in fitting and forecasting $\Rightarrow$ the effects of Student-t distribution for density forecast
6. Future research: multivariate VARs and skew-t distribution

Summary and future research

1. We propose a score-driven approach for TVPs models and we concentrate on the AR model
2. We show how the implied algorithms are related to the learning algorithms
3. We extend existing adaptive algorithms to the case of changes in volatility and heavy-tails based on the score-driven criterion
4. We show how to impose restrictions on the time-varying parameters
5. Application to the US inflation leads to interesting results in fitting and forecasting ⇒ the effects of Student-t distribution for density forecast
6. Future research: multivariate VARs and skew-t distribution

Summary and future research

1. We propose a score-driven approach for TVPs models and we concentrate on the AR model
2. We show how the implied algorithms are related to the learning algorithms
3. We extend existing adaptive algorithms to the case of changes in volatility and heavy-tails based on the score-driven criterion
4. We show how to impose restrictions on the time-varying parameters
5. Application to the US inflation leads to interesting results in fitting and forecasting $\Rightarrow$ the effects of Student-t distribution for density forecast
6. Future research: multivariate VARs and skew-t distribution

Summary and future research

1. We propose a score-driven approach for TVPs models and we concentrate on the AR model
2. We show how the implied algorithms are related to the learning algorithms
3. We extend existing adaptive algorithms to the case of changes in volatility and heavy-tails based on the score-driven criterion
4. We show how to impose restrictions on the time-varying parameters
5. Application to the US inflation leads to interesting results in fitting and forecasting ⇒ the effects of Student-t distribution for density forecast
6. Future research: multivariate VARs and skew-t distribution

Summary and future research

1. We propose a score-driven approach for TVPs models and we concentrate on the AR model
2. We show how the implied algorithms are related to the learning algorithms
3. We extend existing adaptive algorithms to the case of changes in volatility and heavy-tails based on the score-driven criterion
4. We show how to impose restrictions on the time-varying parameters
5. Application to the US inflation leads to interesting results in fitting and forecasting $\Rightarrow$ the effects of Student-t distribution for density forecast
6. Future research: multivariate VARs and skew-t distribution