Bayesian semiparametric vector autoregressive models

Dr Maria Kalli

(joint work with Professor Jim Griffin)

December 2013

Canterbury Christ Church University

Outline

1. Vector autoregressive models (VARs)
2. Motivation for nonlinear VARs
3. Bayesian nonparametric methods
4. The Bayesian semiparametric VAR(1)
5. Computation
6. Empirical examples
7. Conclusion/Discussion

Outline

1. Vector autoregressive models (VARs)
2. Motivation for nonlinear VARs
3. Bayesian nonparametric methods
4. The Bayesian semiparametric VAR(1)
5. Computation
6. Empirical examples
7. Conclusion/Discussion

Outline

1. Vector autoregressive models (VARs)
2. Motivation for nonlinear VARs
3. Bayesian nonparametric methods
4. The Bayesian semiparametric VAR(1)
5. Computation
6. Empirical examples
7. Conclusion/Discussion

Outline

1. Vector autoregressive models (VARs)
2. Motivation for nonlinear VARs
3. Bayesian nonparametric methods
4. The Bayesian semiparametric VAR(1)
5. Computation
6. Empirical examples
7. Conclusion/Discussion

Outline

1. Vector autoregressive models (VARs)
2. Motivation for nonlinear VARs
3. Bayesian nonparametric methods
4. The Bayesian semiparametric VAR(1)
5. Computation
6. Empirical examples
7. Conclusion/Discussion

Outline

1. Vector autoregressive models (VARs)
2. Motivation for nonlinear VARs
3. Bayesian nonparametric methods
4. The Bayesian semiparametric VAR(1)
5. Computation
6. Empirical examples
7. Conclusion/Discussion

Outline

1. Vector autoregressive models (VARs)
2. Motivation for nonlinear VARs
3. Bayesian nonparametric methods
4. The Bayesian semiparametric VAR(1)
5. Computation
6. Empirical examples
7. Conclusion/Discussion

popularity - 'reduced form' VAR

- Introduced by Sims (1980), VAR is used by macroeconomists
  - to characterise the joint dynamic behaviour of a collection of variables, and
  - to forecast movements of macroeconomic variables based on potential future paths of specified variables.
- We focus on the 'reduced' form VAR, i.e. the stationary VAR model without restrictions,

$$\boldsymbol{y}_t = \boldsymbol{B}\boldsymbol{y}_{t-1} + \epsilon_t \tag{1}$$

for $t = 1, \ldots, T$, where

$\boldsymbol{y}_t = (y_{1,t}, y_{2,t}, \ldots, y_{m,t})'$ is the $m \times 1$ vector of macroeconomic variables at time t,

$\boldsymbol{B}$ is the $m \times m$ matrix of unknown regression coefficients,

$\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t}, \ldots, \epsilon_{m,t})'$ is the $m \times 1$ innovation vector at time t.

popularity - 'reduced form' VAR

- Introduced by Sims (1980), VAR is used by macroeconomists
  - to characterise the joint dynamic behaviour of a collection of variables, and
  - to forecast movements of macroeconomic variables based on potential future paths of specified variables.
- We focus on the 'reduced' form VAR, i.e. the stationary VAR model without restrictions,

$$\mathbf{y}_t = \boldsymbol{B}\boldsymbol{y}_{t-1} + \epsilon_t \tag{1}$$

for $t = 1, \ldots, T$, where

$\boldsymbol{y}_t = (y_{1,t}, y_{2,t}, \ldots, y_{m,t})'$ is the $m \times 1$ vector of macroeconomic variables at time t,

$\boldsymbol{B}$ is the $m \times m$ matrix of unknown regression coefficients,

$\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t}, \ldots, \epsilon_{m,t})'$ is the $m \times 1$ innovation vector at time t.

popularity - 'reduced form' VAR

- Introduced by Sims (1980), VAR is used by macroeconomists
  - to characterise the joint dynamic behaviour of a collection of variables, and
  - to forecast movements of macroeconomic variables based on potential future paths of specified variables.
- We focus on the 'reduced' form VAR, i.e. the stationary VAR model without restrictions,

$$\mathbf{y}_t = \boldsymbol{B}\boldsymbol{y}_{t-1} + \epsilon_t \tag{1}$$

for $t = 1, \ldots, T$, where

$\boldsymbol{y}_t = (y_{1,t}, y_{2,t}, \ldots, y_{m,t})'$ is the $m \times 1$ vector of macroeconomic variables at time t,

$\boldsymbol{B}$ is the $m \times m$ matrix of unknown regression coefficients,

$\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t}, \ldots, \epsilon_{m,t})'$ is the $m \times 1$ innovation vector at time t.

popularity - 'reduced form' VAR

- Introduced by Sims (1980), VAR is used by macroeconomists
  - to characterise the joint dynamic behaviour of a collection of variables, and
  - to forecast movements of macroeconomic variables based on potential future paths of specified variables.
- We focus on the 'reduced' form VAR, i.e. the stationary VAR model without restrictions,

$$\mathbf{y}_t = \boldsymbol{B}\boldsymbol{y}_{t-1} + \epsilon_t \tag{1}$$

for $t = 1, \ldots, T$, where

$\boldsymbol{y}_t = (y_{1,t}, y_{2,t}, \ldots, y_{m,t})'$ is the $m \times 1$ vector of macroeconomic variables at time t,

$\boldsymbol{B}$ is the $m \times m$ matrix of unknown regression coefficients,

$\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t}, \ldots, \epsilon_{m,t})'$ is the $m \times 1$ innovation vector at time t.

- Introduced by Sims (1980), VAR is used by macroeconomists
  - to characterise the joint dynamic behaviour of a collection of variables, and
  - to forecast movements of macroeconomic variables based on potential future paths of specified variables.
- We focus on the 'reduced' form VAR, i.e. the stationary VAR model without restrictions,

$$\mathbf{y}_t = \boldsymbol{B}\boldsymbol{y}_{t-1} + \epsilon_t \tag{1}$$

for $t = 1, \ldots, T$, where

$\boldsymbol{y}_t = (y_{1,t}, y_{2,t}, \ldots, y_{m,t})'$ is the $m \times 1$ vector of macroeconomic variables at time t,

$\boldsymbol{B}$ is the $m \times m$ matrix of unknown regression coefficients,

$\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t}, \ldots, \epsilon_{m,t})'$ is the $m \times 1$ innovation vector at time t.

- Past and current literature finds departures from the linear and Gaussian VAR form. See Koop et al. (1996), Weise(1999), Favero (2012), Wong (2013) to name a few.
- The main argument is that linear models cannot adequately capture 'asymmetries' that may exist in business cycle fluctuations.
- Another part of econometric literature, argues that the use parametric methods to model financial time-series is characterised by parameter instability and poor forecast performance. See Pesaran and Timmermann (1992) and Härdel et al. (1998).

- Past and current literature finds departures from the linear and Gaussian VAR form. See Koop et al. (1996), Weise(1999), Favero (2012), Wong (2013) to name a few.

- The main argument is that linear models cannot adequately capture 'asymmetries' that may exist in business cycle fluctuations.

- Another part of econometric literature, argues that the use parametric methods to model financial time-series is characterised by parameter instability and poor forecast performance. See Pesaran and Timmermann (1992) and Härdel et al. (1998).

Flexible models

- Past and current literature finds departures from the linear and Gaussian VAR form. See Koop et al. (1996), Weise(1999), Favero (2012), Wong (2013) to name a few.
- The main argument is that linear models cannot adequately capture 'asymmetries' that may exist in business cycle fluctuations.
- Another part of econometric literature, argues that the use parametric methods to model financial time-series is characterised by parameter instability and poor forecast performance. See Pesaran and Timmermann (1992) and Härdel et al. (1998).

Flexible models

- Past and current literature finds departures from the linear and Gaussian VAR form. See Koop et al. (1996), Weise(1999), Favero (2012), Wong (2013) to name a few.
- The main argument is that linear models cannot adequately capture 'asymmetries' that may exist in business cycle fluctuations.
- Another part of econometric literature, argues that the use parametric methods to model financial time-series is characterised by parameter instability and poor forecast performance. See Pesaran and Timmermann (1992) and Härdel et al. (1998).

- Past and current literature finds departures from the linear and Gaussian VAR form. See Koop et al. (1996), Weise(1999), Favero (2012), Wong (2013) to name a few.

- The main argument is that linear models cannot adequately capture ‘asymmetries’ that may exist in business cycle fluctuations.

- Another part of econometric literature, argues that the use parametric methods to model financial time-series is characterised by parameter instability and poor forecast performance. See Pesaran and Timmermann (1992) and Härdel et al. (1998).

- We want to merge these two sides of literature, and consider non-linear, semi-parametric VAR models.
- The model we propose uses Bayesian non-parametric methods.
- To be more precise we use the Dirichlet process mixture to construct non-linear first order stationary multivariate VAR processes with non-Gaussian innovations.

Flexible models

- We want to merge these two sides of literature, and consider non-linear, semi-parametric VAR models.
- The model we propose uses Bayesian non-parametric methods.
- To be more precise we use the Dirichlet process mixture to construct non-linear first order stationary multivariate VAR processes with non-Gaussian innovations.

Flexible models

- We want to merge these two sides of literature, and consider non-linear, semi-parametric VAR models.
- The model we propose uses Bayesian non-parametric methods.
- To be more precise we use the Dirichlet process mixture to construct non-linear first order stationary multivariate VAR processes with non-Gaussian innovations.

- A nonparametric model has an infinite number of parameters. For example, in density estimation the parameters can consist of all densities.
- In Bayesian statistics, we need to define a prior for the parameters. This is non trivial since we are working on an infinite parameter space.
- The solution, is to define a stochastic process to be your prior. The Dirichlet process (DP) introduced in Ferguson (1973) is the popular Bayesian nonparametric prior, and the one we will use for our VAR model.

- A nonparametric model has an infinite number of parameters. For example, in density estimation the parameters can consist of all densities.
- In Bayesian statistics, we need to define a prior for the parameters. This is non trivial since we are working on an infinite parameter space.
- The solution, is to define a stochastic process to be your prior. The Dirichlet process (DP) introduced in Ferguson (1973) is the popular Bayesian nonparametric prior, and the one we will use for our VAR model.

- A nonparametric model has an infinite number of parameters. For example, in density estimation the parameters can consist of all densities.
- In Bayesian statistics, we need to define a prior for the parameters. This is non trivial since we are working on an infinite parameter space.
- The solution, is to define a stochastic process to be your prior. The Dirichlet process (DP) introduced in Ferguson (1973) is the popular Bayesian nonparametric prior, and the one we will use for our VAR model.

- A nonparametric model has an infinite number of parameters. For example, in density estimation the parameters can consist of all densities.
- In Bayesian statistics, we need to define a prior for the parameters. This is non trivial since we are working on an infinite parameter space.
- The solution, is to define a stochastic process to be your prior. The Dirichlet process (DP) introduced in Ferguson (1973) is the popular Bayesian nonparametric prior, and the one we will use for our VAR model.

- A nonparametric model has an infinite number of parameters. For example, in density estimation the parameters can consist of all densities.
- In Bayesian statistics, we need to define a prior for the parameters. This is non trivial since we are working on an infinite parameter space.
- The solution, is to define a stochastic process to be your prior. The Dirichlet process (DP) introduced in Ferguson (1973) is the popular Bayesian nonparametric prior, and the one we will use for our VAR model.

- A nonparametric model has an infinite number of parameters. For example, in density estimation the parameters can consist of all densities.
- In Bayesian statistics, we need to define a prior for the parameters. This is non trivial since we are working on an infinite parameter space.
- The solution, is to define a stochastic process to be your prior. The Dirichlet process (DP) introduced in Ferguson (1973) is the popular Bayesian nonparametric prior, and the one we will use for our VAR model.

Dirichlet process mixture model (DPM)

- We the DPM to construct non-linear first order stationary multivariate VAR processes with non-Gaussian innovations. So what is the DPM?
- Introduced by Lo (1984) the DPM model, with Gaussian kernel, is given by

$$f_P(y) = \int \mathrm{N}(y; \mu, \sigma^2)\, \mathrm{d}P(\phi)$$

where $P \sim D(M, P_0)$ - a DP with precision parameter $M > 0$, and base measure, P_0, a distribution on $\mathbb{R} \times \mathbb{R}_+$,

and $\phi = (\mu, \sigma^2)$ with μ to represent the mean and σ^2 the variance of the normal component.

- We the DPM to construct non-linear first order stationary multivariate VAR processes with non-Gaussian innovations. So what is the DPM?
- Introduced by Lo (1984) the DPM model, with Gaussian kernel, is given by

$$f_P(y) = \int \mathrm{N}(y; \mu, \sigma^2)\, \mathrm{d}P(\phi)$$

where $P \sim D(M, P_0)$ - a DP with precision parameter $M > 0$, and base measure, P_0, a distribution on $\mathbb{R} \times \mathbb{R}_+$,

and $\phi = (\mu, \sigma^2)$ with μ to represent the mean and σ^2 the variance of the normal component.

- We the DPM to construct non-linear first order stationary multivariate VAR processes with non-Gaussian innovations. So what is the DPM?
- Introduced by Lo (1984) the DPM model, with Gaussian kernel, is given by

$$f_P(y) = \int \mathrm{N}(y; \mu, \sigma^2)\, \mathrm{d}P(\phi)$$

where $P \sim D(M, P_0)$ - a DP with precision parameter $M > 0$, and base measure, P_0, a distribution on $\mathbb{R} \times \mathbb{R}_+$,

and $\phi = (\mu, \sigma^2)$ with μ to represent the mean and σ^2 the variance of the normal component.

- We the DPM to construct non-linear first order stationary multivariate VAR processes with non-Gaussian innovations. So what is the DPM?
- Introduced by Lo (1984) the DPM model, with Gaussian kernel, is given by

$$f_P(y) = \int \mathrm{N}(y; \mu, \sigma^2)\, \mathrm{d}P(\phi)$$

where $P \sim D(M, P_0)$ - a DP with precision parameter $M > 0$, and base measure, P_0, a distribution on $\mathbb{R} \times \mathbb{R}_+$,

and $\phi = (\mu, \sigma^2)$ with μ to represent the mean and σ^2 the variance of the normal component.

- We the DPM to construct non-linear first order stationary multivariate VAR processes with non-Gaussian innovations. So what is the DPM?
- Introduced by Lo (1984) the DPM model, with Gaussian kernel, is given by

$$f_P(y) = \int \mathrm{N}(y; \mu, \sigma^2)\, \mathrm{d}P(\phi)$$

where $P \sim D(M, P_0)$ - a DP with precision parameter $M > 0$, and base measure, P_0, a distribution on $\mathbb{R} \times \mathbb{R}_+$,

and $\phi = (\mu, \sigma^2)$ with μ to represent the mean and σ^2 the variance of the normal component.

- The DPM's stick-breaking representation (see Sethuraman (1992)), is :

$$P = \sum_{j=1}^{\infty} w_j \, \delta_{\phi_j},$$

$\phi_j \overset{\text{iid}}{\sim} P_0$, $w_1 = v_1$, $w_j = v_j \prod_{l<j}(1 - v_l)$ with $v_j \overset{\text{iid}}{\sim} \text{Be}(1, M)$, and we can write

$$f_{v,\phi}(y_i) = \sum_{j=1}^{\infty} w_j \, \text{N}(y_i; \phi_j)$$

To facilitate computation auxiliary allocation variables $(d_1, \ldots, d_n)$ are introduced, so that $p(d_i = j) = w_j$, and $w_1, w_2, \ldots \perp\!\!\!\perp \phi_1, \phi_2 \ldots$

- The DPM's stick-breaking representation (see Sethuraman (1992)), is :

$$P = \sum_{j=1}^{\infty} w_j \, \delta_{\phi_j},$$

$\phi_j \overset{\text{iid}}{\sim} P_0$, $w_1 = v_1$, $w_j = v_j \prod_{l<j}(1 - v_l)$ with $v_j \overset{\text{iid}}{\sim} \text{Be}(1, M)$, and we can write

$$f_{v,\phi}(y_i) = \sum_{j=1}^{\infty} w_j \, \text{N}(y_i; \phi_j)$$

To facilitate computation auxiliary allocation variables $(d_1, \ldots, d_n)$ are introduced, so that $p(d_i = j) = w_j$, and $w_1, w_2, \ldots \perp\!\!\!\perp \phi_1, \phi_2 \ldots$

Model construction

- Villalobos and Walker (2013), construct a Bayesian nonparametric version of the AR(1) model, by expressing both the joint and transition densities as DPM.
- We extend their idea to the multivariate case and model the transition and joint densities using the DPM.
- The joint density will therefore be:

$$f\begin{pmatrix} y_{t-1} \\ y_t \end{pmatrix} = \sum_{j=1}^{\infty} w_j \, \mathrm{N}\left(\begin{pmatrix} y_{t-1} \\ y_t \end{pmatrix} \middle| \begin{pmatrix} \mu_j \\ \mu_j \end{pmatrix} \begin{pmatrix} \Sigma_j & \Omega_j \\ \Omega_j & \Sigma_j \end{pmatrix} \right)$$

  for $j = 1, 2, \ldots, t = 1, \ldots, T$, and $i = 1, \ldots, m$.

- The stationary distribution is then $f(y_t) = \sum_{j=1}^{\infty} \mathrm{N}(y_t | \mu_j, \Sigma_j)$

- Villalobos and Walker (2013), construct a Bayesian nonparametric version of the AR(1) model, by expressing both the joint and transition densities as DPM.
- We extend their idea to the multivariate case and model the transition and joint densities using the DPM.
- The joint density will therefore be:

$$f\begin{pmatrix} y_{t-1} \\ y_t \end{pmatrix} = \sum_{j=1}^{\infty} w_j \, \mathrm{N}\left(\begin{pmatrix} y_{t-1} \\ y_t \end{pmatrix} \middle| \begin{pmatrix} \mu_j \\ \mu_j \end{pmatrix} \begin{pmatrix} \Sigma_j & \Omega_j \\ \Omega_j & \Sigma_j \end{pmatrix}\right)$$

for $j = 1, 2, \ldots,$ $t = 1, \ldots, T$, and $i = 1, \ldots, m$.

- The stationary distribution is then $f(y_t) = \sum_{j=1}^{\infty} \mathrm{N}(y_t|\mu_j, \Sigma_j)$

- Villalobos and Walker (2013), construct a Bayesian nonparametric version of the AR(1) model, by expressing both the joint and transition densities as DPM.
- We extend their idea to the multivariate case and model the transition and joint densities using the DPM.
- The joint density will therefore be:

$$f\begin{pmatrix} y_{t-1} \\ y_t \end{pmatrix} = \sum_{j=1}^{\infty} w_j \, \mathrm{N}\left(\begin{pmatrix} y_{t-1} \\ y_t \end{pmatrix} \middle| \begin{pmatrix} \mu_j \\ \mu_j \end{pmatrix} \begin{pmatrix} \Sigma_j & \Omega_j \\ \Omega_j & \Sigma_j \end{pmatrix}\right)$$

for $j = 1, 2, \ldots, t = 1, \ldots, T$, and $i = 1, \ldots, m$.

- The stationary distribution is then $f(y_t) = \sum_{j=1}^{\infty} \mathrm{N}(y_t|\mu_j, \Sigma_j)$

- Villalobos and Walker (2013), construct a Bayesian nonparametric version of the AR(1) model, by expressing both the joint and transition densities as DPM.
- We extend their idea to the multivariate case and model the transition and joint densities using the DPM.
- The joint density will therefore be:

$$f\begin{pmatrix} y_{t-1} \\ y_t \end{pmatrix} = \sum_{j=1}^{\infty} w_j \, \mathrm{N}\left(\begin{pmatrix} y_{t-1} \\ y_t \end{pmatrix} \middle| \begin{pmatrix} \mu_j \\ \mu_j \end{pmatrix} \begin{pmatrix} \Sigma_j & \Omega_j \\ \Omega_j & \Sigma_j \end{pmatrix} \right)$$

for $j = 1, 2, \ldots, t = 1, \ldots, T$, and $i = 1, \ldots, m$.

- The stationary distribution is then $f(y_t) = \sum_{j=1}^{\infty} \mathrm{N}(y_t | \mu_j, \Sigma_j)$

- Villalobos and Walker (2013), construct a Bayesian nonparametric version of the AR(1) model, by expressing both the joint and transition densities as DPM.
- We extend their idea to the multivariate case and model the transition and joint densities using the DPM.
- The joint density will therefore be:

$$f\begin{pmatrix} y_{t-1} \\ y_t \end{pmatrix} = \sum_{j=1}^{\infty} w_j \, \mathrm{N}\left(\begin{pmatrix} y_{t-1} \\ y_t \end{pmatrix} \middle| \begin{pmatrix} \mu_j \\ \mu_j \end{pmatrix} \begin{pmatrix} \Sigma_j & \Omega_j \\ \Omega_j & \Sigma_j \end{pmatrix} \right)$$

for $j = 1, 2, \ldots, t = 1, \ldots, T$, and $i = 1, \ldots, m$.

- The stationary distribution is then $f(y_t) = \sum_{j=1}^{\infty} \mathrm{N}(y_t | \mu_j, \Sigma_j)$

- The y_t and y_{t-1} are m dimensional vectors,
- the μ_j's are also m dimensional vectors, and we have an infinite number of them,
- the Σ_j's are $m \times m$ positive definite matrices, and we have an infinite number of them,
- and the Ω_j's are $m \times m$ matrices, and we also have an infinite number of them.

- The y_t and y_{t-1} are m dimensional vectors,
- the μ_j's are also m dimensional vectors, and we have an infinite number of them,
- the Σ_j's are $m \times m$ positive definite matrices, and we have an infinite number of them,
- and the Ω_j's are $m \times m$ matrices, and we also have an infinite number of them.

- The y_t and y_{t-1} are m dimensional vectors,
- the μ_j's are also m dimensional vectors, and we have an infinite number of them,
- the Σ_j's are $m \times m$ positive definite matrices, and we have an infinite number of them,
- and the Ω_j's are $m \times m$ matrices, and we also have an infinite number of them.

- The y_t and y_{t-1} are m dimensional vectors,
- the μ_j's are also m dimensional vectors, and we have an infinite number of them,
- the Σ_j's are $m \times m$ positive definite matrices, and we have an infinite number of them,
- and the Ω_j's are $m \times m$ matrices, and we also have an infinite number of them.

- The transition density will then be

$$f(y_t|y_{t-1}) = \frac{\sum_{j=1}^{\infty} w_j \, \mathrm{N}\left(\left(\begin{array}{c} y_{t-1} \\ y_t \end{array} \right) \middle| \left(\begin{array}{c} \mu_j \\ \mu_j \end{array} \right) \left(\begin{array}{cc} \Sigma_j & \Omega_j \\ \Omega_j & \Sigma_j \end{array} \right) \right)}{\sum_{j=1}^{\infty} w_j \, \mathrm{N}(y_{t-1}|\mu_j, \Sigma_j)}$$

- We can then re-write it as locally weighted mixture of VAR(1)'s as follows,

$$f(y_t|y_{t-1}) = \sum_{j=1}^{\infty} p_j(y_{t-1}) \, \mathrm{N}(y_t|\theta_j(y_{t-1}), \Lambda_j(y_{t-1}))$$

where,

- $p_j(y_{t-1}) = \dfrac{w_j \, \mathrm{N}(y_{t-1} | \mu_j, \Sigma_j)}{\sum_{k=1}^{\infty} \mathrm{N}(y_{t-1} | \mu_k, \Sigma_k)}$,
- $\theta_j(y_{t-1}) = \mu_j + \Omega_j' \Sigma_j^{-1} (y_{t-1} - \mu_j)$,

  and
- $\Lambda_j(y_{t-1}) = \Sigma_j - \Omega_j' \Sigma_j^{-1} \Omega_j$

where,

- $p_j(y_{t-1}) = \dfrac{w_j \, \mathrm{N}(y_{t-1}|\mu_j, \Sigma_j)}{\sum_{k=1}^{\infty} \mathrm{N}(y_{t-1}|\mu_k, \Sigma_k)}$,
- $\theta_j(y_{t-1}) = \mu_j + \Omega_j' \Sigma_j^{-1}(y_{t-1} - \mu_j)$,

and

- $\Lambda_j(y_{t-1}) = \Sigma_j - \Omega_j' \Sigma_j^{-1} \Omega_j$

where,

- $p_j(y_{t-1}) = \frac{w_j \, \mathrm{N}(y_{t-1} | \mu_j, \Sigma_j)}{\sum_{k=1}^{\infty} \mathrm{N}(y_{t-1} | \mu_k, \Sigma_k)}$,
- $\theta_j(y_{t-1}) = \mu_j + \Omega_j' \Sigma_j^{-1} (y_{t-1} - \mu_j)$,

and

- $\Lambda_j(y_{t-1}) = \Sigma_j - \Omega_j' \Sigma_j^{-1} \Omega_j$

where,

- $p_j(y_{t-1}) = \dfrac{w_j \, \mathrm{N}(y_{t-1} | \mu_j, \Sigma_j)}{\sum_{k=1}^{\infty} \mathrm{N}(y_{t-1} | \mu_k, \Sigma_k)}$,
- $\theta_j(y_{t-1}) = \mu_j + \Omega_j' \Sigma_j^{-1} (y_{t-1} - \mu_j)$,

and

- $\Lambda_j(y_{t-1}) = \Sigma_j - \Omega_j' \Sigma_j^{-1} \Omega_j$

where,

- $p_j(y_{t-1}) = \dfrac{w_j \, \mathrm{N}(y_{t-1}|\mu_j, \Sigma_j)}{\sum_{k=1}^{\infty} \mathrm{N}(y_{t-1}|\mu_k, \Sigma_k)}$,
- $\theta_j(y_{t-1}) = \mu_j + \Omega_j' \Sigma_j^{-1}(y_{t-1} - \mu_j)$,

and

- $\Lambda_j(y_{t-1}) = \Sigma_j - \Omega_j' \Sigma_j^{-1} \Omega_j$

Priors

- When selecting the priors, we must ensure that the Σ_j matrices are positive definite. For this reason rather than placing priors on the variance covariance matrices, we decompose them in terms of correlation matrices.
- We follow Karlsson (2012), and write Σ_j, Ω_j as follows:

  $\Sigma_j = S_j P_j S_j$ and $\Omega_j = S_j R_j S_j$

  - where, S_j is an $m \times m$ diagonal matrix , with $\sigma_{1,j}, \ldots, \sigma_{m,j}$, in the diagonal.
  - P_j is the $m \times m$ correlation matrix at time t of the VAR(1) variables. Its elements are the correlations between $y_{t,k}$ and $y_{t,l}$ in the j^{th} component.
  - R_j is the cross-correlation matrix at times t and $t-1$. Its elements are the correlations between $y_{t-1,k}$ and $y_{t,l}$ in the j^{th} component.

Priors

- When selecting the priors, we must ensure that the Σ_j matrices are positive definite. For this reason rather than placing priors on the variance covariance matrices, we decompose them in terms of correlation matrices.
- We follow Karlsson (2012), and write Σ_j, Ω_j as follows:

  $\Sigma_j = S_j P_j S_j$ and $\Omega_j = S_j R_j S_j$

  - where, S_j is an $m \times m$ diagonal matrix , with $\sigma_{1,j}, \ldots, \sigma_{m,j}$, in the diagonal.
  - P_j is the $m \times m$ correlation matrix at time t of the VAR(1) variables. Its elements are the correlations between $y_{t,k}$ and $y_{t,l}$ in the j^{th} component.
  - R_j is the cross-correlation matrix at times t and $t-1$. Its elements are the correlations between $y_{t-1,k}$ and $y_{t,l}$ in the j^{th} component.

Priors

- When selecting the priors, we must ensure that the Σ_j matrices are positive definite. For this reason rather than placing priors on the variance covariance matrices, we decompose them in terms of correlation matrices.
- We follow Karlsson (2012), and write Σ_j, Ω_j as follows:

$$\Sigma_j = S_j P_j S_j \text{ and } \Omega_j = S_j R_j S_j$$

  - where, S_j is an $m \times m$ diagonal matrix , with $\sigma_{1,j}, \ldots, \sigma_{m,j}$, in the diagonal.
  - P_j is the $m \times m$ correlation matrix at time t of the VAR(1) variables. Its elements are the correlations between $y_{t,k}$ and $y_{t,l}$ in the j^{th} component.
  - R_j is the cross-correlation matrix at times t and $t-1$. Its elements are the correlations between $y_{t-1,k}$ and $y_{t,l}$ in the j^{th} component.

- When selecting the priors, we must ensure that the Σ_j matrices are positive definite. For this reason rather than placing priors on the variance covariance matrices, we decompose them in terms of correlation matrices.
- We follow Karlsson (2012), and write Σ_j, Ω_j as follows:
  $\Sigma_j = S_j P_j S_j$ and $\Omega_j = S_j R_j S_j$
  - where, S_j is an $m \times m$ diagonal matrix , with $\sigma_{1,j}, \ldots, \sigma_{m,j}$, in the diagonal.
  - P_j is the $m \times m$ correlation matrix at time t of the VAR(1) variables. Its elements are the correlations between $y_{t,k}$ and $y_{t,l}$ in the j^{th} component.
  - R_j is the cross-correlation matrix at times t and $t-1$. Its elements are the correlations between $y_{t-1,k}$ and $y_{t,l}$ in the j^{th} component.

- When selecting the priors, we must ensure that the Σ_j matrices are positive definite. For this reason rather than placing priors on the variance covariance matrices, we decompose them in terms of correlation matrices.
- We follow Karlsson (2012), and write Σ_j, Ω_j as follows:
  $\Sigma_j = S_j P_j S_j$ and $\Omega_j = S_j R_j S_j$
  - where, S_j is an $m \times m$ diagonal matrix , with $\sigma_{1,j}, \ldots, \sigma_{m,j}$, in the diagonal.
  - P_j is the $m \times m$ correlation matrix at time t of the VAR(1) variables. Its elements are the correlations between $y_{t,k}$ and $y_{t,l}$ in the j^{th} component.
  - R_j is the cross-correlation matrix at times t and $t-1$. Its elements are the correlations between $y_{t-1,k}$ and $y_{t,l}$ in the j^{th} component.

Priors

- When selecting the priors, we must ensure that the Σ_j matrices are positive definite. For this reason rather than placing priors on the variance covariance matrices, we decompose them in terms of correlation matrices.
- We follow Karlsson (2012), and write Σ_j, Ω_j as follows:

  $\Sigma_j = S_j P_j S_j$ and $\Omega_j = S_j R_j S_j$

  - where, S_j is an $m \times m$ diagonal matrix , with $\sigma_{1,j}, \ldots, \sigma_{m,j}$, in the diagonal.
  - P_j is the $m \times m$ correlation matrix at time t of the VAR(1) variables. Its elements are the correlations between $y_{t,k}$ and $y_{t,l}$ in the j^{th} component.
  - R_j is the cross-correlation matrix at times t and $t-1$. Its elements are the correlations between $y_{t-1,k}$ and $y_{t,l}$ in the j^{th} component.

- When selecting the priors, we must ensure that the Σ_j matrices are positive definite. For this reason rather than placing priors on the variance covariance matrices, we decompose them in terms of correlation matrices.
- We follow Karlsson (2012), and write Σ_j, Ω_j as follows:

  $\Sigma_j = S_j P_j S_j$ and $\Omega_j = S_j R_j S_j$

  - where, S_j is an $m \times m$ diagonal matrix , with $\sigma_{1,j}, \dots, \sigma_{m,j}$, in the diagonal.
  - P_j is the $m \times m$ correlation matrix at time t of the VAR(1) variables. Its elements are the correlations between $y_{t,k}$ and $y_{t,l}$ in the j^{th} component.
  - R_j is the cross-correlation matrix at times t and $t-1$. Its elements are the correlations between $y_{t-1,k}$ and $y_{t,l}$ in the j^{th} component.

Priors

- This means that the mean vector and variance matrix for the transition probability can be written as:
  - $\theta_j(y_{t-1}) = \mu_j + S_j R_j' P_j^{-1} S_j^{-1}(y_{t-1} - \mu_j)$
  - $\Lambda_j(y_{t-1}) = S_j(P_j - R_j' P_j^{-1} R_j)S_j$
- We then place priors on R_j, P_j, S_j, and μ_j.)
  - $R_{k,j} \sim \mathrm{U}(-1, 1)$,
  - $P_{k,j} \sim \mathrm{U}(-1, 1)$ for $k \neq j$, with diagonal matrix $P_{j,j} = I$
  - $S_j \sim \mathrm{IG}(S_a, (S_a - 1)S_{\mu_j})$, where $S_{\mu_j} \sim \mathrm{Ga}(1, 5)$, so that $S_{\mu_j} = \mathrm{E}(S_j)$, and $S_a = 4$.
  - $\mu_j \sim \mathrm{N}(\mu_0, \Sigma_0)$. μ_0 is set equal to the sample mean of the data and $\Sigma_{0,j,j} = 1.5^2 Var(y_j)$ and $\Sigma_{0,k,j} = 0$

- This means that the mean vector and variance matrix for the transition probability can be written as:
  - $\theta_j(y_{t-1}) = \mu_j + S_j R_j' P_j^{-1} S_j^{-1}(y_{t-1} - \mu_j)$
  - $\Lambda_j(y_{t-1}) = S_j(P_j - R_j' P_j^{-1} R_j) S_j$
- We then place priors on R_j, P_j, S_j, and μ_j.)
  - $R_{k,j} \sim \mathrm{U}(-1, 1)$,
  - $P_{k,j} \sim \mathrm{U}(-1, 1)$ for $k \neq j$, with diagonal matrix $P_{j,j} = I$
  - $S_j \sim \mathrm{IG}(S_a, (S_a - 1)S_{\mu_j})$, where $S_{\mu_j} \sim \mathrm{Ga}(1, 5)$, so that $S_{\mu_j} = \mathrm{E}(S_j)$, and $S_a = 4$.
  - $\mu_j \sim \mathrm{N}(\mu_0, \Sigma_0)$. μ_0 is set equal to the sample mean of the data and $\Sigma_{0,j,j} = 1.5^2 Var(y_j)$ and $\Sigma_{0,k,j} = 0$

- This means that the mean vector and variance matrix for the transition probability can be written as:
    - $\theta_j(y_{t-1}) = \mu_j + S_j R_j' P_j^{-1} S_j^{-1}(y_{t-1} - \mu_j)$
    - $\Lambda_j(y_{t-1}) = S_j(P_j - R_j' P_j^{-1} R_j) S_j$
- We then place priors on R_j, P_j, S_j, and μ_j.)
    - $R_{k,j} \sim \mathrm{U}(-1, 1)$,
    - $P_{k,j} \sim \mathrm{U}(-1, 1)$ for $k \neq j$, with diagonal matrix $P_{j,j} = I$
    - $S_j \sim \mathrm{IG}(S_a, (S_a - 1)S_{\mu_j})$, where $S_{\mu_j} \sim \mathrm{Ga}(1, 5)$, so that $S_{\mu_j} = \mathrm{E}(S_j)$, and $S_a = 4$.
    - $\mu_j \sim \mathrm{N}(\mu_0, \Sigma_0)$. μ_0 is set equal to the sample mean of the data and $\Sigma_{0,j,j} = 1.5^2 Var(y_j)$ and $\Sigma_{0,k,j} = 0$

- This means that the mean vector and variance matrix for the transition probability can be written as:
  - $\theta_j(y_{t-1}) = \mu_j + S_j R_j' P_j^{-1} S_j^{-1}(y_{t-1} - \mu_j)$
  - $\Lambda_j(y_{t-1}) = S_j(P_j - R_j' P_j^{-1} R_j) S_j$
- We then place priors on R_j, P_j, S_j, and μ_j.)
  - $R_{k,j} \sim \mathrm{U}(-1, 1)$,
  - $P_{k,j} \sim \mathrm{U}(-1, 1)$ for $k \neq j$, with diagonal matrix $P_{j,j} = I$
  - $S_j \sim \mathrm{IG}(S_a, (S_a - 1)S_{\mu_j})$, where $S_{\mu_j} \sim \mathrm{Ga}(1, 5)$, so that $S_{\mu_j} = \mathrm{E}(S_j)$, and $S_a = 4$.
  - $\mu_j \sim \mathrm{N}(\mu_0, \Sigma_0)$. μ_0 is set equal to the sample mean of the data and $\Sigma_{0,j,j} = 1.5^2 Var(y_j)$ and $\Sigma_{0,k,j} = 0$

- This means that the mean vector and variance matrix for the transition probability can be written as:
    - $\theta_j(y_{t-1}) = \mu_j + S_j R_j' P_j^{-1} S_j^{-1}(y_{t-1} - \mu_j)$
    - $\Lambda_j(y_{t-1}) = S_j(P_j - R_j' P_j^{-1} R_j) S_j$
- We then place priors on R_j, P_j, S_j, and μ_j.)
    - $R_{k,j} \sim \mathrm{U}(-1, 1)$,
    - $P_{k,j} \sim \mathrm{U}(-1, 1)$ for $k \neq j$, with diagonal matrix $P_{j,j} = I$
    - $S_j \sim \mathrm{IG}(S_a, (S_a - 1)S_{\mu_j})$, where $S_{\mu_j} \sim \mathrm{Ga}(1, 5)$, so that $S_{\mu_j} = \mathrm{E}(S_j)$, and $S_a = 4$.
    - $\mu_j \sim \mathrm{N}(\mu_0, \Sigma_0)$. μ_0 is set equal to the sample mean of the data and $\Sigma_{0,j,j} = 1.5^2 Var(y_j)$ and $\Sigma_{0,k,j} = 0$

- This means that the mean vector and variance matrix for the transition probability can be written as:
  - $\theta_j(y_{t-1}) = \mu_j + S_j R_j' P_j^{-1} S_j^{-1}(y_{t-1} - \mu_j)$
  - $\Lambda_j(y_{t-1}) = S_j(P_j - R_j' P_j^{-1} R_j)S_j$
- We then place priors on R_j, P_j, S_j, and μ_j.)
  - $R_{k,j} \sim \mathrm{U}(-1, 1)$,
  - $P_{k,j} \sim \mathrm{U}(-1, 1)$ for $k \neq j$, with diagonal matrix $P_{j,j} = I$
  - $S_j \sim \mathrm{IG}(S_a, (S_a - 1)S_{\mu_j})$, where $S_{\mu_j} \sim \mathrm{Ga}(1, 5)$, so that $S_{\mu_j} = \mathrm{E}(S_j)$, and $S_a = 4$.
  - $\mu_j \sim \mathrm{N}(\mu_0, \Sigma_0)$. μ_0 is set equal to the sample mean of the data and $\Sigma_{0,j,j} = 1.5^2 Var(y_j)$ and $\Sigma_{0,k,j} = 0$

Priors

- This means that the mean vector and variance matrix for the transition probability can be written as:
  - $\theta_j(y_{t-1}) = \mu_j + S_j R_j' P_j^{-1} S_j^{-1}(y_{t-1} - \mu_j)$
  - $\Lambda_j(y_{t-1}) = S_j(P_j - R_j' P_j^{-1} R_j) S_j$
- We then place priors on R_j, P_j, S_j, and μ_j.)
  - $R_{k,j} \sim \mathrm{U}(-1, 1)$,
  - $P_{k,j} \sim \mathrm{U}(-1, 1)$ for $k \neq j$, with diagonal matrix $P_{j,j} = I$
  - $S_j \sim \mathrm{IG}(S_a, (S_a - 1)S_{\mu_j})$, where $S_{\mu_j} \sim \mathrm{Ga}(1, 5)$, so that $S_{\mu_j} = \mathrm{E}(S_j)$, and $S_a = 4$.
  - $\mu_j \sim \mathrm{N}(\mu_0, \Sigma_0)$. μ_0 is set equal to the sample mean of the data and $\Sigma_{0,j,j} = 1.5^2 Var(y_j)$ and $\Sigma_{0,k,j} = 0$

- This means that the mean vector and variance matrix for the transition probability can be written as:
  - $\theta_j(y_{t-1}) = \mu_j + S_j R_j' P_j^{-1} S_j^{-1}(y_{t-1} - \mu_j)$
  - $\Lambda_j(y_{t-1}) = S_j(P_j - R_j' P_j^{-1} R_j) S_j$
- We then place priors on R_j, P_j, S_j, and μ_j.)
  - $R_{k,j} \sim \mathrm{U}(-1, 1)$,
  - $P_{k,j} \sim \mathrm{U}(-1, 1)$ for $k \neq j$, with diagonal matrix $P_{j,j} = I$
  - $S_j \sim \mathrm{IG}(S_a, (S_a - 1)S_{\mu_j})$, where $S_{\mu_j} \sim \mathrm{Ga}(1, 5)$, so that $S_{\mu_j} = \mathrm{E}(S_j)$, and $S_a = 4$.
  - $\mu_j \sim \mathrm{N}(\mu_0, \Sigma_0)$. μ_0 is set equal to the sample mean of the data and $\Sigma_{0,j,j} = 1.5^2 Var(y_j)$ and $\Sigma_{0,k,j} = 0$

key points

- Standard MCMC methods for infinite mixture models cannot be used here.
- We use the adaptive truncation method of Griffin(2013).
- We sample a sequence of posteriors for truncated versions of the model, with different levels of truncation.
- The algorithm provides a method for choosing when to stop sampling this sequence, in such a way so that large truncation errors are avoided. The final posterior provides an approximation to the posterior of the infinite dimensional model.

key points

- Standard MCMC methods for infinite mixture models cannot be used here.
- We use the adaptive truncation method of Griffin(2013).
- We sample a sequence of posteriors for truncated versions of the model, with different levels of truncation.
- The algorithm provides a method for choosing when to stop sampling this sequence, in such a way so that large truncation errors are avoided. The final posterior provides an approximation to the posterior of the infinite dimensional model.

key points

- Standard MCMC methods for infinite mixture models cannot be used here.
- We use the adaptive truncation method of Griffin(2013).
- We sample a sequence of posteriors for truncated versions of the model, with different levels of truncation.
- The algorithm provides a method for choosing when to stop sampling this sequence, in such a way so that large truncation errors are avoided. The final posterior provides an approximation to the posterior of the infinite dimensional model.

key points

- Standard MCMC methods for infinite mixture models cannot be used here.
- We use the adaptive truncation method of Griffin(2013).
- We sample a sequence of posteriors for truncated versions of the model, with different levels of truncation.
- The algorithm provides a method for choosing when to stop sampling this sequence, in such a way so that large truncation errors are avoided. The final posterior provides an approximation to the posterior of the infinite dimensional model.

key points

- Standard MCMC methods for infinite mixture models cannot be used here.
- We use the adaptive truncation method of Griffin(2013).
- We sample a sequence of posteriors for truncated versions of the model, with different levels of truncation.
- The algorithm provides a method for choosing when to stop sampling this sequence, in such a way so that large truncation errors are avoided. The final posterior provides an approximation to the posterior of the infinite dimensional model.

- The preliminary results are based on a Var(1) with three variables. They include:
  - heat plots of $f(y_{t,k}|y_{t-1,j})$, and
  - plots of the median $E(y_{t,k}|y_{t-1,j})$ together with the 95% credible interval.

    Recall that at each iteration of the sampler we have different $E(y_{t,k}|y_{t-1,j})$ and that's why we choose the median as our point estimate.
- The purpose of this analysis is to gain insight on the co-movements of macroeconomic variables, and how changes in previous lags of the same variables, as well as other variables affect their expected value.

- The preliminary results are based on a Var(1) with three variables. They include:
  - heat plots of $f(y_{t,k}|y_{t-1,j})$, and
  - plots of the median $E(y_{t,k}|y_{t-1,j})$ together with the 95% credible interval.
    Recall that at each iteration of the sampler we have different $E(y_{t,k}|y_{t-1,j})$ and that's why we choose the median as our point estimate.
- The purpose of this analysis is to gain insight on the co-movements of macroeconomic variables, and how changes in previous lags of the same variables, as well as other variables affect their expected value.

- The preliminary results are based on a Var(1) with three variables. They include:
  - heat plots of $f(y_{t,k}|y_{t-1,j})$, and
  - plots of the median $E(y_{t,k}|y_{t-1,j})$ together with the 95% credible interval.

    Recall that at each iteration of the sampler we have different $E(y_{t,k}|y_{t-1,j})$ and that's why we choose the median as our point estimate.
- The purpose of this analysis is to gain insight on the co-movements of macroeconomic variables, and how changes in previous lags of the same variables, as well as other variables affect their expected value.

- The preliminary results are based on a Var(1) with three variables. They include:
  - heat plots of $f(y_{t,k}|y_{t-1,j})$, and
  - plots of the median $E(y_{t,k}|y_{t-1,j})$ together with the 95% credible interval.

    Recall that at each iteration of the sampler we have different $E(y_{t,k}|y_{t-1,j})$ and that's why we choose the median as our point estimate.
- The purpose of this analysis is to gain insight on the co-movements of macroeconomic variables, and how changes in previous lags of the same variables, as well as other variables affect their expected value.

- The preliminary results are based on a Var(1) with three variables. They include:
  - heat plots of $f(y_{t,k}|y_{t-1,j})$, and
  - plots of the median $E(y_{t,k}|y_{t-1,j})$ together with the 95% credible interval.
    Recall that at each iteration of the sampler we have different $E(y_{t,k}|y_{t-1,j})$ and that's why we choose the median as our point estimate.
- The purpose of this analysis is to gain insight on the co-movements of macroeconomic variables, and how changes in previous lags of the same variables, as well as other variables affect their expected value.

data

- We constracted our data set using data series obtained from FRED, the economic database of the Federal Reserve Bank of St Louis.
- The sample period is from the second quarter of 1965 to the first quarter of 2011.
- The three variables are:
  - GDP deflator,
  - GDP growth (difference in logs of real GDP), and
  - Employment growth (differences in logs of non farm payroll).

- We constracted our data set using data series obtained from FRED, the economic database of the Federal Reserve Bank of St Louis.
- The sample period is from the second quarter of 1965 to the first quarter of 2011.
- The three variables are:
  - GDP deflator,
  - GDP growth (difference in logs of real GDP), and
  - Employment growth (differences in logs of non farm payroll).

data

- We constracted our data set using data series obtained from FRED, the economic database of the Federal Reserve Bank of St Louis.
- The sample period is from the second quarter of 1965 to the first quarter of 2011.
- The three variables are:
    - GDP deflator,
    - GDP growth (difference in logs of real GDP), and
    - Employment growth (differences in logs of non farm payroll).

data

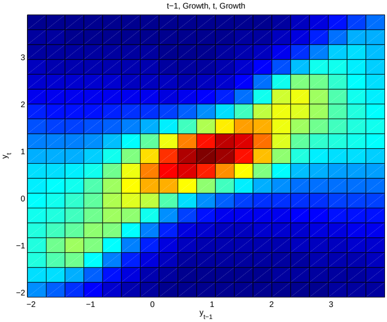
- We constracted our data set using data series obtained from FRED, the economic database of the Federal Reserve Bank of St Louis.
- The sample period is from the second quarter of 1965 to the first quarter of 2011.
- The three variables are:
    - GDP deflator,
    - GDP growth (difference in logs of real GDP), and
    - Employment growth (differences in logs of non farm payroll).

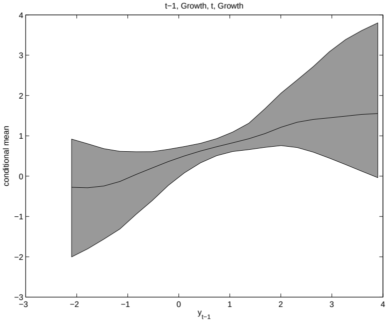
data

- We constracted our data set using data series obtained from FRED, the economic database of the Federal Reserve Bank of St Louis.
- The sample period is from the second quarter of 1965 to the first quarter of 2011.
- The three variables are:
  - GDP deflator,
  - GDP growth (difference in logs of real GDP), and
  - Employment growth (differences in logs of non farm payroll).

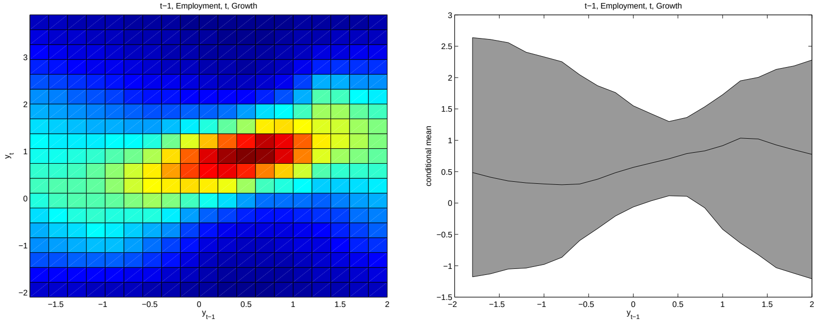
- We constracted our data set using data series obtained from FRED, the economic database of the Federal Reserve Bank of St Louis.
- The sample period is from the second quarter of 1965 to the first quarter of 2011.
- The three variables are:
  - GDP deflator,
  - GDP growth (difference in logs of real GDP), and
  - Employment growth (differences in logs of non farm payroll).

GDP growth at t and $t-1$

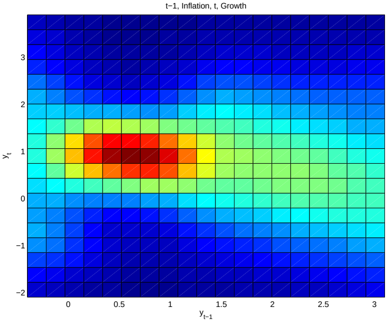


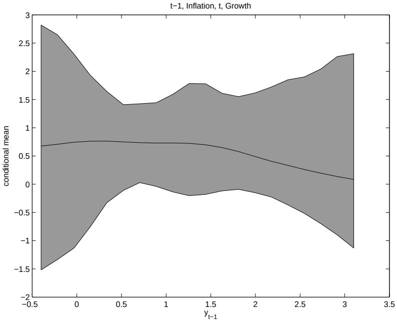


GDP growth at t and Employment growth at $t-1$

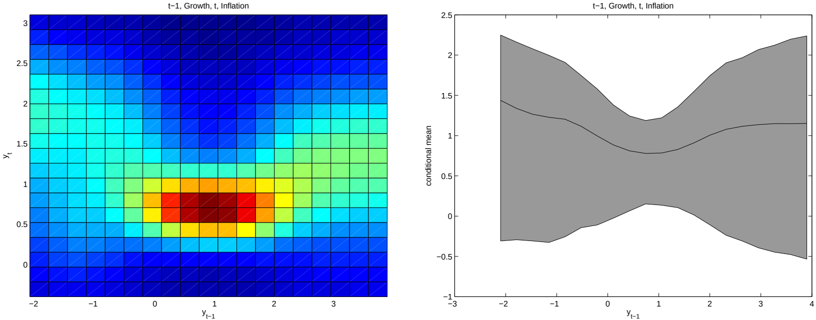


GDP growth at t and GDP deflator at $t-1$

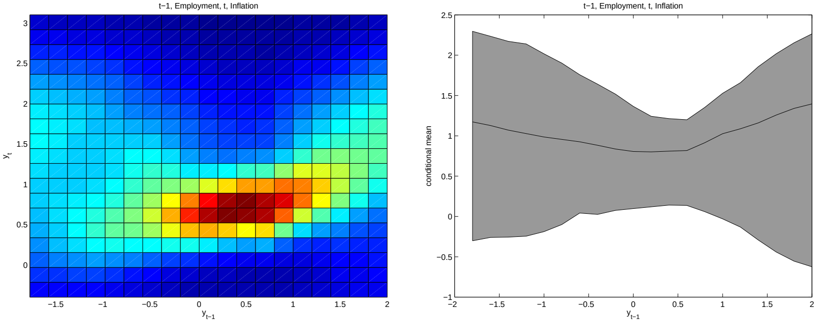




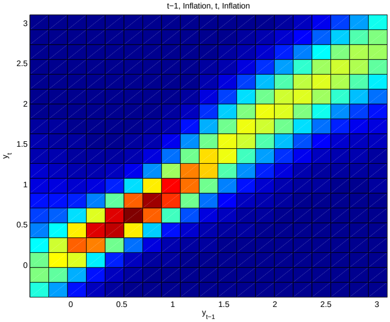
GDP deflator at t and GDP growth at $t-1$

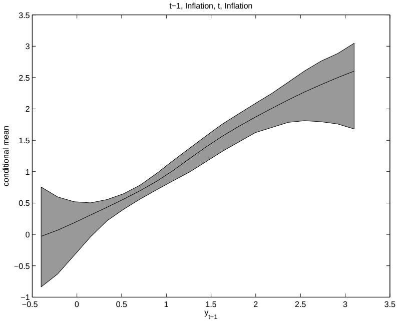


GDP deflator at t and Employment growth at $t-1$

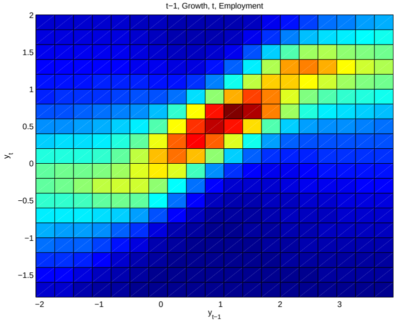


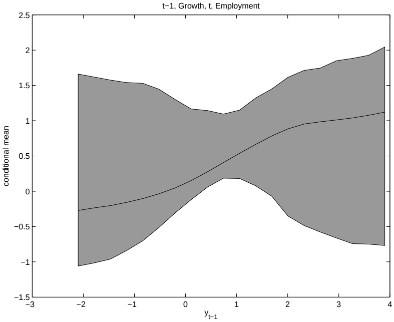
GDP deflator at t and $t-1$



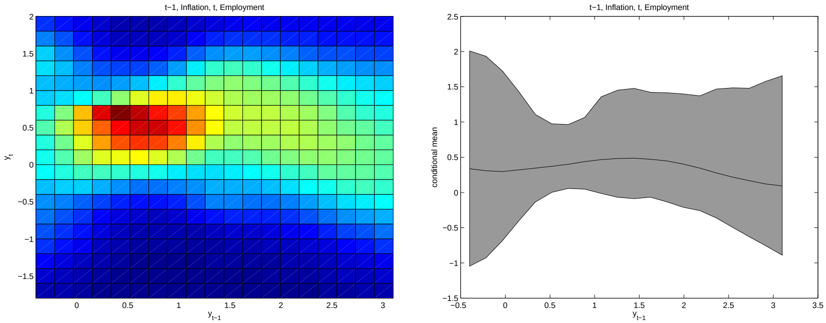


Employment growth at t and GDP growth at $t-1$





Employment growth at t and GDP deflator at $t-1$



Employment growth at t and $t-1$