XXV ENCONTRO BRASILEIRO DE ECONOMETRIA
Porto Seguro - BA, 2003

---

# REVISITING DISTRIBUTED LAG MODELS THROUGH A BAYESIAN PERSPECTIVE

---

Romy R. Ravines, Alexandra M. Schmidt and Helio S. Migon[1]

Instituto de Matemática
Universidade Federal do Rio de Janeiro.

[1] This work is part of the first author's PhD studies, who is being supervised by H.S. Migon and A.M. Schmidt. R.R. Ravines is grateful to CAPES for the the financial support during her PhD studies.

## Outline

- ✓ Review some distributed lag models
- ✓ Write them into a particular class of Bayesian Dynamic Models, the transfer functions models.
- ✓ Perform inference following the Bayesian paradigm. Make use of Markov chain Monte Carlo (MCMC) methods.
- ✓ Computation is made by the use of the software `WinBugs`.
- ✓ An example: the Koyck's consumption function is analyzed using two different approaches, the Koyck's and the Solow's distributed lag models.

# Distributed-lag Models

## Distributed-lag Models

☞ The general form of a linear distributed-lag model is

$$Y_t = \sum_{i=0}^{\infty} \beta_i X_{t-i} + \epsilon_t \tag{1}$$

where any change in $X_t$ will affect $E[Y_t]$ in all the later periods.

☞ The term $\beta_i$ in (1) is the $i$th reaction coefficient, and it is usually assumed that $\lim_{i \to \infty} \beta_i = 0$ and $\sum_{i=0}^{\infty} \beta_i = \beta < \infty$.

## Distributed-lag Models

☞ The general form of a linear distributed-lag model is

$$Y_t = \sum_{i=0}^{\infty} \beta_i X_{t-i} + \epsilon_t \tag{1}$$

where any change in $X_t$ will affect $E[Y_t]$ in all the later periods.

☞ The term $\beta_i$ in (1) is the $i$th reaction coefficient, and it is usually assumed that $\lim_{i\to\infty} \beta_i = 0$ and $\sum_{i=0}^{\infty} \beta_i = \beta < \infty$.

☞ One important aspect to be considered is the number of parameters involved in these distributed lag models.

☞ In order to be parsimonious, it is assumed that the coefficients of lagged variables are functionally related.

## ◇ The Koyck Distributed Lag

$$\beta_i = \alpha\lambda^i, \quad \forall i, \text{ with } 0 < \lambda < 1. \tag{2}$$

Then,

$$Y_t = \alpha X_t + \alpha\lambda^1 X_{t-1} + \alpha\lambda^2 X_{t-2} + \ldots + \epsilon_t. \tag{3}$$

◇ **The Koyck Distributed Lag**

$$\beta_i = \alpha\lambda^i, \quad \forall i, \text{ with } 0 < \lambda < 1. \tag{2}$$

Then,

$$Y_t = \alpha X_t + \alpha\lambda^1 X_{t-1} + \alpha\lambda^2 X_{t-2} + \ldots + \epsilon_t. \tag{3}$$

◇ **The Solow Distributed Lags**

$$\beta_i = \alpha \binom{r+i-1}{i} (1-\lambda)^r \lambda, \quad 0 < \lambda < 1, \ \forall i, r > 0. \tag{4}$$

## ⋄ The Koyck Distributed Lag

$$\beta_i = \alpha\lambda^i, \quad \forall i, \text{ with } 0 < \lambda < 1. \tag{2}$$

Then,

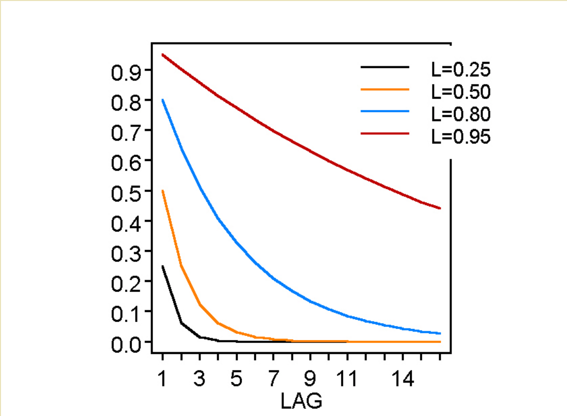
$$Y_t = \alpha X_t + \alpha\lambda^1 X_{t-1} + \alpha\lambda^2 X_{t-2} + \ldots + \epsilon_t. \tag{3}$$

## ⋄ The Solow Distributed Lags

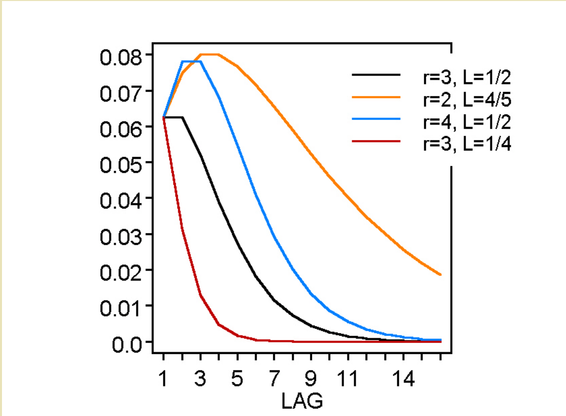
$$\beta_i = \alpha\binom{r+i-1}{i}(1-\lambda)^r\lambda, \quad 0 < \lambda < 1,\ \forall i, r > 0. \tag{4}$$

## ⋄ The Almon Distributed Lags

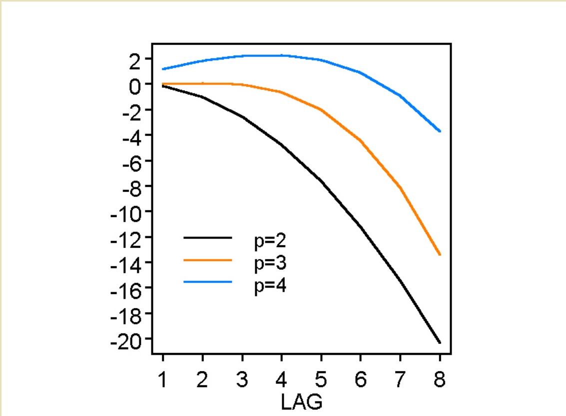
$$\beta_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 + \ldots + \alpha_p i^p = \sum_{k=0}^{p} \alpha_k i^k. \tag{5}$$




(a) **Koyck**




(b) **Solow**




(c) **Almon**

Figure 1: **Examples of different Distributed Lag Models**

## Bayesian Dynamic Models

## Bayesian Dynamic Models

- The general dynamic model is defined by $\{\boldsymbol{F}(.), \boldsymbol{G}(.), \boldsymbol{V}, \boldsymbol{W}\}_t$. $\boldsymbol{F}(.)$ and $\boldsymbol{G}(.)$ are general smooth functions defining the mean of the response variable and the state parameters evolution. $\boldsymbol{V}$ and $\boldsymbol{W}$ represent the variances.
- For each $t$, the univariate Dynamic Linear Model (DLM) is

$$\begin{aligned} \text{Observation equation:} \quad & Y_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \epsilon_t, \quad \epsilon_t \sim N(0, V_t) \\ \text{System equation:} \quad & \boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\boldsymbol{0}, \boldsymbol{W}_t) \\ \text{Initial information:} \quad & (\boldsymbol{\theta}_0 \mid D_0) \sim N[\boldsymbol{m}_0, \boldsymbol{C}_0] \end{aligned} \tag{6}$$

- Extensions for non-linear (unknown $\boldsymbol{G}$) and non-normal Bayesian models are easily introduced.

## Bayesian Dynamic Models

- The general dynamic model is defined by $\{\boldsymbol{F}(.), \boldsymbol{G}(.), \boldsymbol{V}, \boldsymbol{W}\}_t$. $\boldsymbol{F}(.)$ and $\boldsymbol{G}(.)$ are general smooth functions defining the mean of the response variable and the state parameters evolution. $\boldsymbol{V}$ and $\boldsymbol{W}$ represent the variances.
- For each $t$, the univariate Dynamic Linear Model (DLM) is

$$\begin{aligned} \text{Observation equation:} \quad & Y_t = \boldsymbol{F}_t' \boldsymbol{\theta}_t + \epsilon_t, \quad \epsilon_t \sim N(0, V_t) \\ \text{System equation:} \quad & \boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \boldsymbol{W}_t) \\ \text{Initial information:} \quad & (\boldsymbol{\theta}_0 \mid D_0) \sim N[\boldsymbol{m}_0, \boldsymbol{C}_0] \end{aligned} \quad (6)$$

- Extensions for non-linear (unknown $\boldsymbol{G}$) and non-normal Bayesian models are easily introduced.
- The distributed lagged models can be seen as particular cases of DLM. Hence, they can be expressed in the form of (6).

## ◇ Form-free transfer functions

$$E(Y_t \mid \theta_t) = \sum_{i=0}^{m} \beta_i X_{t-i} = \beta_0 X_t + \beta_1 X_{t-1} + \ldots + \beta_m X_{t-m}, \quad (7)$$

The transfer response function of $X$ is $\begin{cases} \beta_i X & i = 0, 1, \ldots, m; \\ 0 & i > m. \end{cases}$

## ◇ Form-free transfer functions

$$E(Y_t \mid \theta_t) = \sum_{i=0}^{m} \beta_i X_{t-i} = \beta_0 X_t + \beta_1 X_{t-1} + \ldots + \beta_m X_{t-m}, \quad (7)$$

The transfer response function of $X$ is $\begin{cases} \beta_i X & i = 0, 1, \ldots, m; \\ 0 & i > m. \end{cases}$

## ◇ Functional form transfer functions

$$\begin{aligned} Y_t &= \boldsymbol{F}'\boldsymbol{\theta}_t + \epsilon_t & (8a) \\ \boldsymbol{\theta}_t &= \boldsymbol{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\psi}_t X_t + \partial\boldsymbol{\theta}_t & (8b) \\ \boldsymbol{\psi}_t &= \boldsymbol{\psi}_{t-1} + \partial\boldsymbol{\psi}_t & (8c) \end{aligned}$$

The transfer function model (8) can be written as the standard DLM form:

$$\begin{aligned} Y_t &= \tilde{\boldsymbol{F}}'\tilde{\boldsymbol{\theta}}_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \\ \tilde{\boldsymbol{\theta}}_t &= \tilde{\boldsymbol{G}}_t\tilde{\boldsymbol{\theta}}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\boldsymbol{0}, \sigma_\omega^2 \boldsymbol{I}). \end{aligned} \quad (9)$$

# Bayesian Dynamic Models and Distributed Lag Models

# Bayesian Dynamic Models and Distributed Lag Models

☞ The Koyck distributed lag model, can be rewritten as

$$Y_t = E_t + \epsilon_t \tag{10a}$$
$$E_t = \lambda E_{t-1} + \alpha X_t \tag{10b}$$

where $E_t = \alpha X_t + \alpha\lambda^1 X_{t-1} + \alpha\lambda^2 X_{t-2} + \ldots$ and $0 < \lambda < 1$.

- In this model, the transfer response function of X is simply $\alpha\lambda^i X$.
- Here, in the representation (8), we have $n = 1$, $\boldsymbol{\theta}_t = E_t$, the effect variable, $\boldsymbol{\psi}_t = \alpha$, the current effect for all $t$, $\boldsymbol{F} = 1$, $\boldsymbol{G} = \lambda$ and the noise term $\partial\boldsymbol{\theta}_t$ is assumed to be zero.

## Bayesian Dynamic Models and Distributed Lag Models

- The Solow's distributed lag model can be expressed in the form of (9). In this case we have $(1-\lambda L)^r E_t = \alpha(1-\lambda)^r X_t$. An evolution equation can be assigned for $\lambda$ and $r$.

# Bayesian Dynamic Models and Distributed Lag Models

☞ The Solow's distributed lag model can be expressed in the form of (9). In this case we have $(1-\lambda L)^r E_t = \alpha(1-\lambda)^r X_t$. An evolution equation can be assigned for $\lambda$ and $r$.

☞ Also, the Almon's model can be rewritten in a free form transfer function (7) since it is a regression on a fixed and finite number of lagged variables. By using (5) to define the coefficients of the lagged variables, the transfer response function of $X_{t-i}$ in the mean of $Y_t$ will be

$$\begin{cases} \sum_{k=0}^{p} \alpha_k i^k X_{t-i} & i = 0, 1, \ldots, m; \\ 0 & i > m. \end{cases}$$

## Inference Procedure

## Inference Procedure

- From Bayes' theorem: $\pi(\boldsymbol{\theta} \mid x) \propto \ell(x \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})$
- Bayesian inference has experienced a great development since the early 90's due to the introduction of Markov Chain Monte Carlo (MCMC) methods. One of the most popular methods and easy to implement is the Gibbs Sampling.
- For highly structured (complex) models, we usually have to write our own program.

## Inference Procedure

- ☞ From Bayes' theorem: $\pi(\boldsymbol{\theta} \mid x) \propto \ell(x \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})$
- ☞ Bayesian inference has experienced a great development since the early 90's due to the introduction of Markov Chain Monte Carlo (MCMC) methods. One of the most popular methods and easy to implement is the Gibbs Sampling.
- ☞ For highly structured (complex) models, we usually have to write our own program.
- ☞ This task has been simplified after the introduction of the software `BUGS` (Bayesian Analysis using Gibbs Sampling).
- ☞ `BUGS` was developed in the MRC Biostatistics Unit and is available free of charge from `http://www.mrc-bsu.cam.ac.uk/bugs`.

## Application: Consumption Function Estimation

## Application: Consumption Function Estimation

- Zellner and Geisel (1970) proposed this consumption function:

$$Y_t = kX_t^* + \epsilon_t \tag{11}$$

where, for the $t$th period, $t = 1, 2, \ldots, T$, $Y_t$ is measured real consumption, $X_t^*$ is "normal" real income, $k$ is a parameter whose value is unknown, and $\epsilon_t$ is an error term or transitory consumption.

- Assuming that the "normal" income satisfies

$$X_t^* - X_{t-1}^* = (1-\lambda)(X_t - X_{t-1}^*)$$

where the parameter $\lambda$ is such that $0 < \lambda < 1$, the model becomes:

$$Y_t = k(1-\lambda)(X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \ldots) + \epsilon_t \tag{12}$$

- The transfer function model is:

$$
\begin{aligned}
Y_t &= E_t + \epsilon_t \\
E_t &= \lambda E_{t-1} + \psi X_t
\end{aligned}
\tag{13}
$$

where $\psi = k(1-\lambda)$ and $E_t = \psi X_t + \lambda \psi X_{t-1} + \lambda^2 \psi X_{t-2} + \ldots$.

- The transfer function model is:

$$Y_t = E_t + \epsilon_t$$
$$E_t = \lambda E_{t-1} + \psi X_t \tag{13}$$

where $\psi = k(1-\lambda)$ and $E_t = \psi X_t + \lambda\psi X_{t-1} + \lambda^2\psi X_{t-2} + \ldots$.

- Several assumptions can be made about the serial correlation properties of the residual $\epsilon_t$ in (13):
  - Model I: $\epsilon_t \overset{\text{ind}}{\sim} N(0, \sigma^2)$;
  - Model II: $\epsilon_t = \rho\epsilon_{t-1} + \nu_t$. with $\nu_t \overset{\text{ind}}{\sim} N(0, \sigma^2)$

- The transfer function model is:

$$Y_t = E_t + \epsilon_t$$
$$E_t = \lambda E_{t-1} + \psi X_t \tag{13}$$

where $\psi = k(1-\lambda)$ and $E_t = \psi X_t + \lambda\psi X_{t-1} + \lambda^2 \psi X_{t-2} + \ldots$.

- Several assumptions can be made about the serial correlation properties of the residual $\epsilon_t$ in (13):
  - Model I: $\epsilon_t \overset{\text{ind}}{\sim} N(0, \sigma^2)$;
  - Model II: $\epsilon_t = \rho\epsilon_{t-1} + \nu_t$. with $\nu_t \overset{\text{ind}}{\sim} N(0, \sigma^2)$
- A priori we set $\lambda \sim Be(1, 1)$ and $k \sim Be(1, 1)$, $\sigma^2 \sim IG(0.0001, 0.0001)$ and, for Model II, $\rho \sim N(0, 100)$.

- The transfer function model is:

$$\begin{aligned} Y_t &= E_t + \epsilon_t \\ E_t &= \lambda E_{t-1} + \psi X_t \end{aligned} \tag{13}$$

  where $\psi = k(1-\lambda)$ and $E_t = \psi X_t + \lambda \psi X_{t-1} + \lambda^2 \psi X_{t-2} + \ldots$.

- Several assumptions can be made about the serial correlation properties of the residual $\epsilon_t$ in (13):
  - Model I: $\epsilon_t \overset{\text{ind}}{\sim} N(0, \sigma^2)$;
  - Model II: $\epsilon_t = \rho\epsilon_{t-1} + \nu_t$. with $\nu_t \overset{\text{ind}}{\sim} N(0, \sigma^2)$

- A priori we set $\lambda \sim Be(1,1)$ and $k \sim Be(1,1)$, $\sigma^2 \sim IG(0.0001, 0.0001)$ and, for Model II, $\rho \sim N(0, 100)$.

- Similarly to Zellner and Geisel (1970), we use the U.S. quarterly price-deflated, seasonally adjusted data on personal disposable income and personal consumption expenditure,1947.I-1960.IV

## Example of a code used in WinBUGS

```
model #(Y=consumption, X=income)
{
    for(t in 2:T)
    {
    mean.Y[t] <- lambda*Y[t-1] + k*(1-lambda)*X[t]
    Y[t] ~ dnorm(mean.Ct[t], tau.v)
    }

# Prior
    lambda ~ dbeta(1,1)
    k ~ dbeta(1,1)
    tau.v ~ dgamma(0.0001,0.0001)
    var.v <- 1/tau.v
}
```

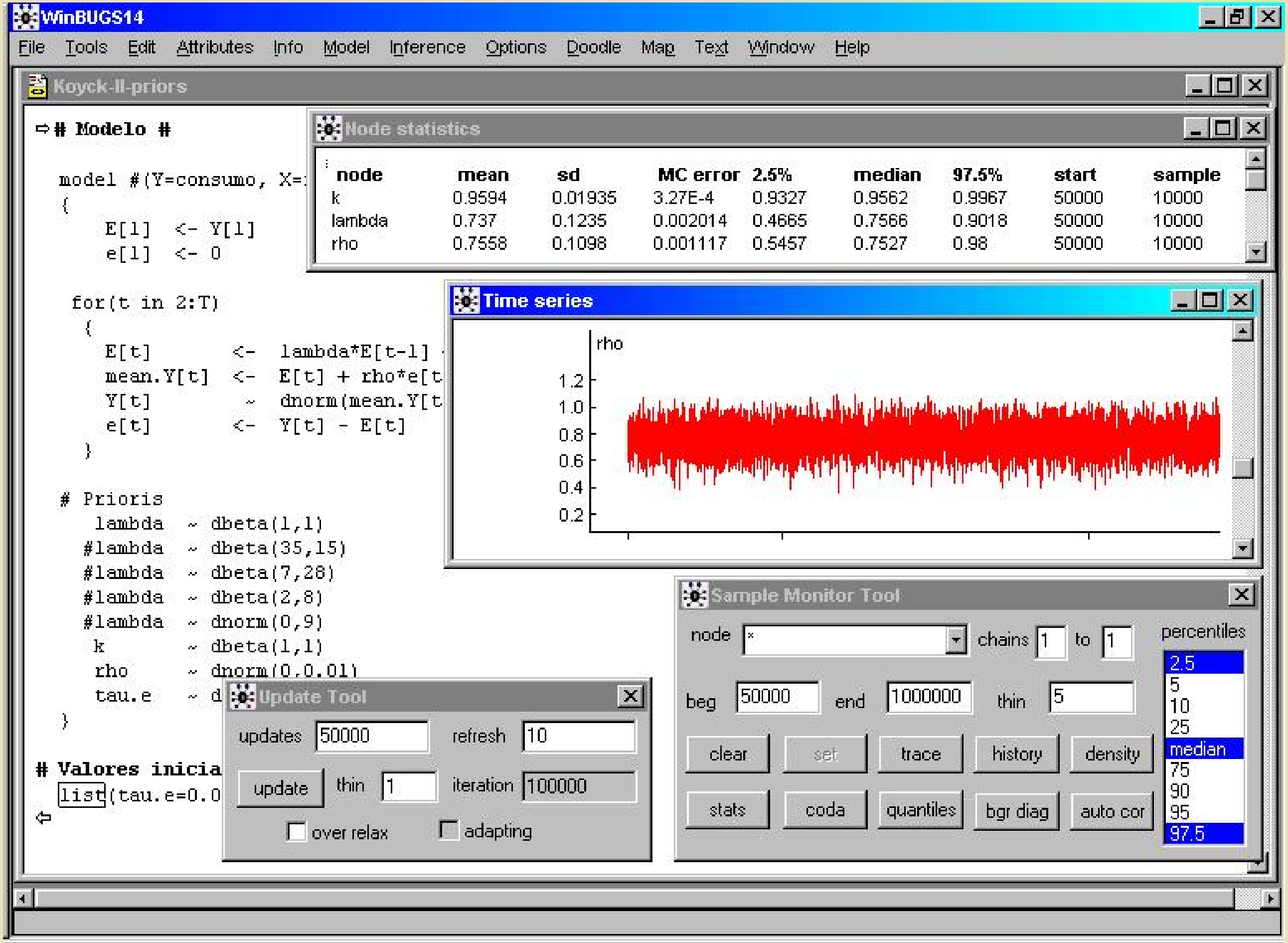


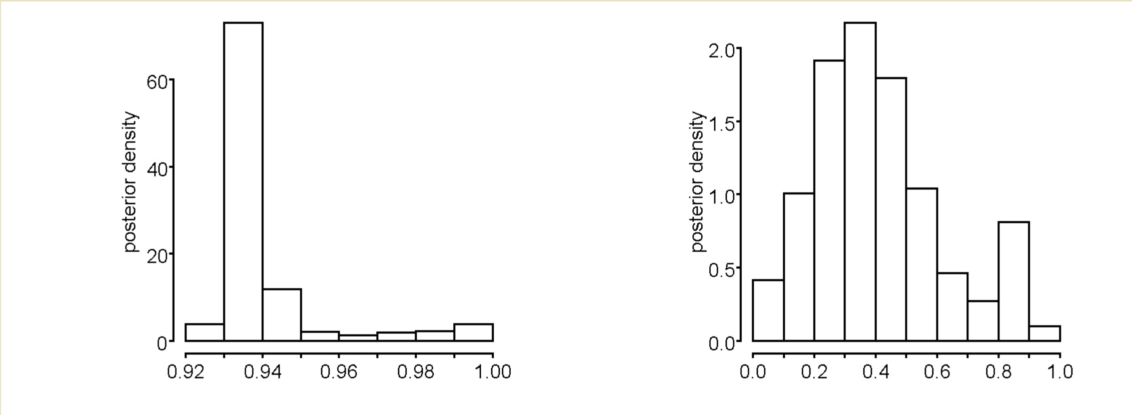

Figure 2: **Interface do `WinBUGS`**

First Prev Next Last Go Back Full Screen Close Quit

Table 1: **Posterior summaries associated with parameters in Model I**

| | $\boldsymbol{k}$ | | | $\boldsymbol{\lambda}$ | | |
|---|---|---|---|---|---|---|
| | mean | modes | sd | mean | modes | sd |
| $ZG$[1] | 0.948 | 0.940 - 1.000 | 0.020 | 0.508 | 0.380 - 0.900 | 0.254 |
| $KT$[2] | 0.941 | 0.934 - 0.995 | 0.015 | 0.411 | 0.351 - 0.934 | 0.210 |
| $TF$[3] | 0.941 | 0.935 - 0.995 | 0.015 | 0.409 | 0.347 - 0.878 | 0.213 |

[1] ZG = Zellner & Geisel, [2] KT = Koyck's transformation, [3] TF = Transfer function



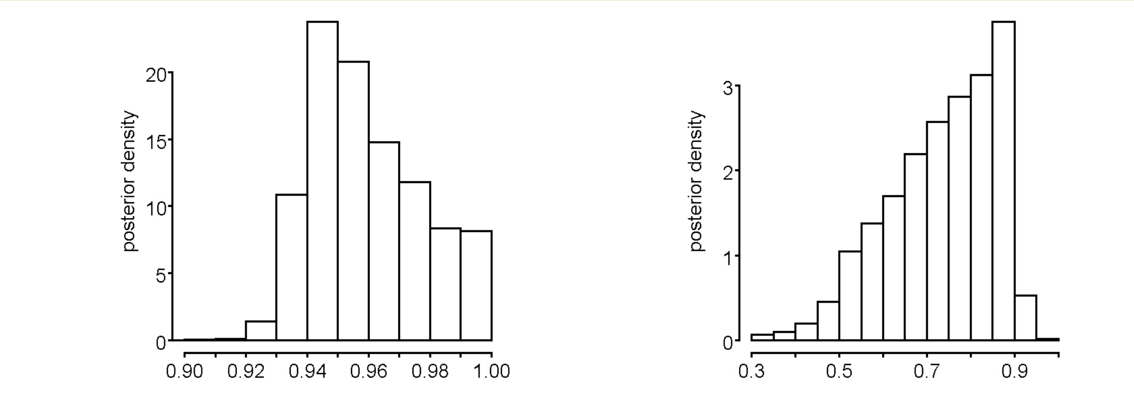

(a) $\boldsymbol{k}$ (b) $\boldsymbol{\lambda}$

Figure 3: **Posterior densities of $k$ and $\lambda$ in Model I, using TF**

Table 2: **Posterior summaries associated with parameters in Model II**

| | $\boldsymbol{k}$ | | | $\boldsymbol{\lambda}$ | | | $\boldsymbol{\rho}$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | mean | median | sd | mean | median | sd | mean | median | sd |
| ZG[1] | 0.878 | 0.940 | 0.201 | 0.597 | 0.610 | 0.184 | | | |
| KT[2] | 0.960 | 0.957 | 0.015 | 0.767 | 0.776 | 0.086 | 0.703 | 0.698 | 0.120 |
| TF[3] | 0.960 | 0.956 | 0.021 | 0.734 | 0.753 | 0.127 | 0.756 | 0.752 | 0.108 |

[1] ZG = Zellner & Geisel, [2] KT = Koyck's transformation, [3] TF = Transfer function




(a) $\boldsymbol{k}$ (b) $\boldsymbol{\lambda}$

Figure 4: **Posterior densities of $k$ and $\lambda$ in Model II, using TF**

☞ Comparing Koyck's models

- Comparing Koyck's models
- The Deviance Information Criterion (DIC) can be used to assess model complexity and compare different models. DIC is given by

$$\text{DIC} = \bar{D} + p_D = D(\bar{\theta}) + 2p_D \tag{14}$$

- The `WinBUGS` version 1.4 computes the DIC automatically. The model with the smallest DIC is considered to be the best one.

- Comparing Koyck's models
- The Deviance Information Criterion (DIC) can be used to assess model complexity and compare different models. DIC is given by

$$\text{DIC} = \bar{D} + p_D \; = \; D(\bar{\theta}) + 2p_D \tag{14}$$

- The WinBUGS version 1.4 computes the DIC automatically. The model with the smallest DIC is considered to be the best one.

Table 3: **Model comparison through DIC**

| | Deviance | DIC |
|---|---|---|
| Model I | | |
| Koyck's transformation | 312.01 | 312.49 |
| Transfer function | 313.13 | 311.40 |
| Model II | | |
| Koyck's transformation | 273.81 | 280.37 |
| Transfer function | 277.33 | 279.20 |

☞ We use the same dataset to estimate some particular Solow's models

- We use the same dataset to estimate some particular Solow's models
- First, we adjusted three Solow's models for fixed $r = 1, 2, 3$ in (4), assuming that $\epsilon_t \overset{\text{ind}}{\sim} N(0, \sigma^2)$ in (11)
- Then, we allowed $r$ to vary in the discrete set $\{1, 2, 3\}$. We used a Categorical priori distribution for $r$, that is, $r \sim Categorial(\boldsymbol{p})$ with $\boldsymbol{p} = (0.2, 0.3, 0.5)$, and started the chain with $r_0 = 3$.

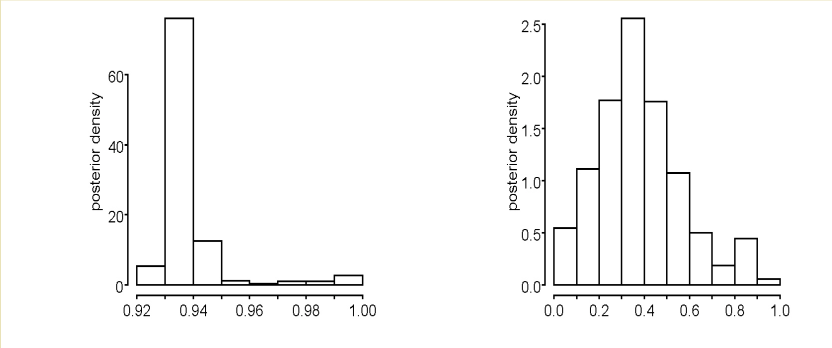
☞ We use the same dataset to estimate some particular Solow's models

☞ First, we adjusted three Solow's models for fixed $r = 1, 2, 3$ in (4), assuming that $\epsilon_t \overset{\text{ind}}{\sim} N(0, \sigma^2)$ in (11)

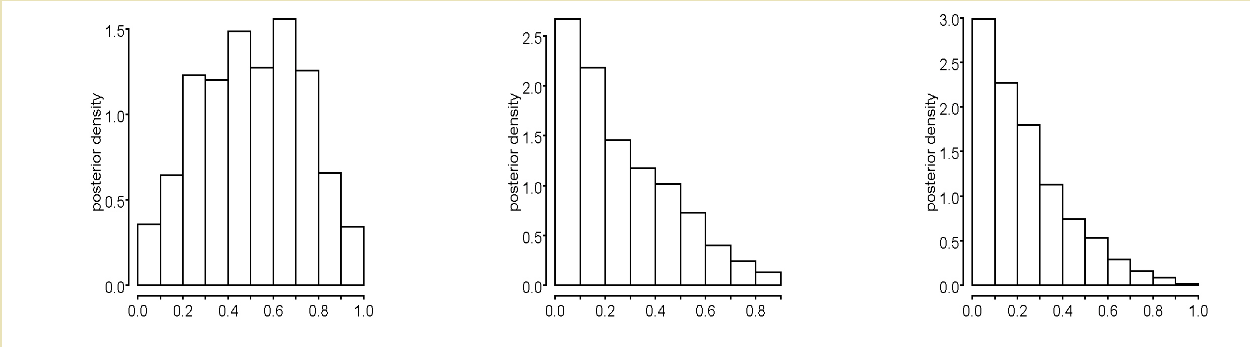
☞ Then, we allowed $r$ to vary in the discrete set $\{1, 2, 3\}$. We used a Categorical priori distribution for $r$, that is, $r \sim Categorial(\boldsymbol{p})$ with $\boldsymbol{p} = (0.2, 0.3, 0.5)$, and started the chain with $r_0 = 3$.

Table 4: **Posterior summaries and DIC associated with the Solow's models**

| | $k$ | | | $\lambda$ | | | DIC |
|---|---|---|---|---|---|---|---|
| | mean | modes | sd | mean | modes | sd | |
| r=1 | 0.939 | 0.935 - 0.995 | 0.013 | 0.387 | 0.369 - 0.871 | 0.200 | 313.602 |
| r=2 | 0.939 | 0.937 | 0.008 | 0.016 | 0.012 | 0.015 | 319.816 |
| r=3 | 0.938 | 0.937 | 0.007 | 0.008 | 0.005 | 0.007 | 319.778 |
| r=1,2,3 | 0.939 | 0.936 - 0.996 | 0.012 | 0.380 | 0.348 - 0.884 | 0.191 | |




(a) $\boldsymbol{k}$ (b) $\boldsymbol{\lambda}$




(c) $\mathbf{Prob}(\boldsymbol{r} = 1 \mid \boldsymbol{y})$ (d) $\mathbf{Prob}(\boldsymbol{r} = 2 \mid \boldsymbol{y})$ (e) $\mathbf{Prob}(\boldsymbol{r} = 3 \mid \boldsymbol{y})$

Figure 5: **Posterior distribution of the parameters in Solow's Model with a categorical prior for** $r$

# Discussion and Future work

## Discussion and Future work

☞ We discussed here the inference of distributed lag models under a Bayesian framework.

## Discussion and Future work

- ☞ We discussed here the inference of distributed lag models under a Bayesian framework.
- ☞ Inference of distributed lag models can be performed using MCMC algorithms. Use of the software `WinBUGS`

## Discussion and Future work

- We discussed here the inference of distributed lag models under a Bayesian framework.
- Inference of distributed lag models can be performed using MCMC algorithms. Use of the software `WinBUGS`
- More complex models can be analyzed under the Bayesian framework: Migon (1985), Alves, Gamerman and Ferreira (2003).

## Discussion and Future work

- ☞ We discussed here the inference of distributed lag models under a Bayesian framework.
- ☞ Inference of distributed lag models can be performed using MCMC algorithms. Use of the software `WinBUGS`
- ☞ More complex models can be analyzed under the Bayesian framework: Migon (1985), Alves, Gamerman and Ferreira (2003).
- ☞ An important theoretical issue, which is currently under research, is how to choose the form of the transfer function.