European Commission

Introduction to Smoothing spline ANOVA models (metamodelling)

M. Ratto

DYNARE Summer School, Paris, June 2015.

Joint Research Centre

www.jrc.ec.europa.eu

Serving society

Stimulating innovation

Supporting legislation

Joint Research Centre

Introduction to ANOVA Models

European Commission

Consider the mapping $Y = f(x)$

where x is a vector of input variables $x = (x_1, x_2, \ldots, x_k)$

and Y is the output.

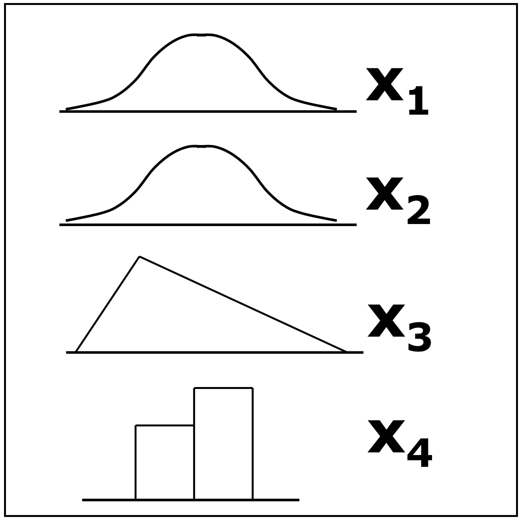
x_i is the i-th element of x varying in $0 \le x_i \le 1$

Specification of the input factors

European Commission

$$X \in \Omega$$

Marginal p.d.f. $p(\vec{x}) = \prod p_i(x_i)$



Marginal p.d.f. + correlation structure

Rank correlation, Dependence-trees, diagonal band, copulas

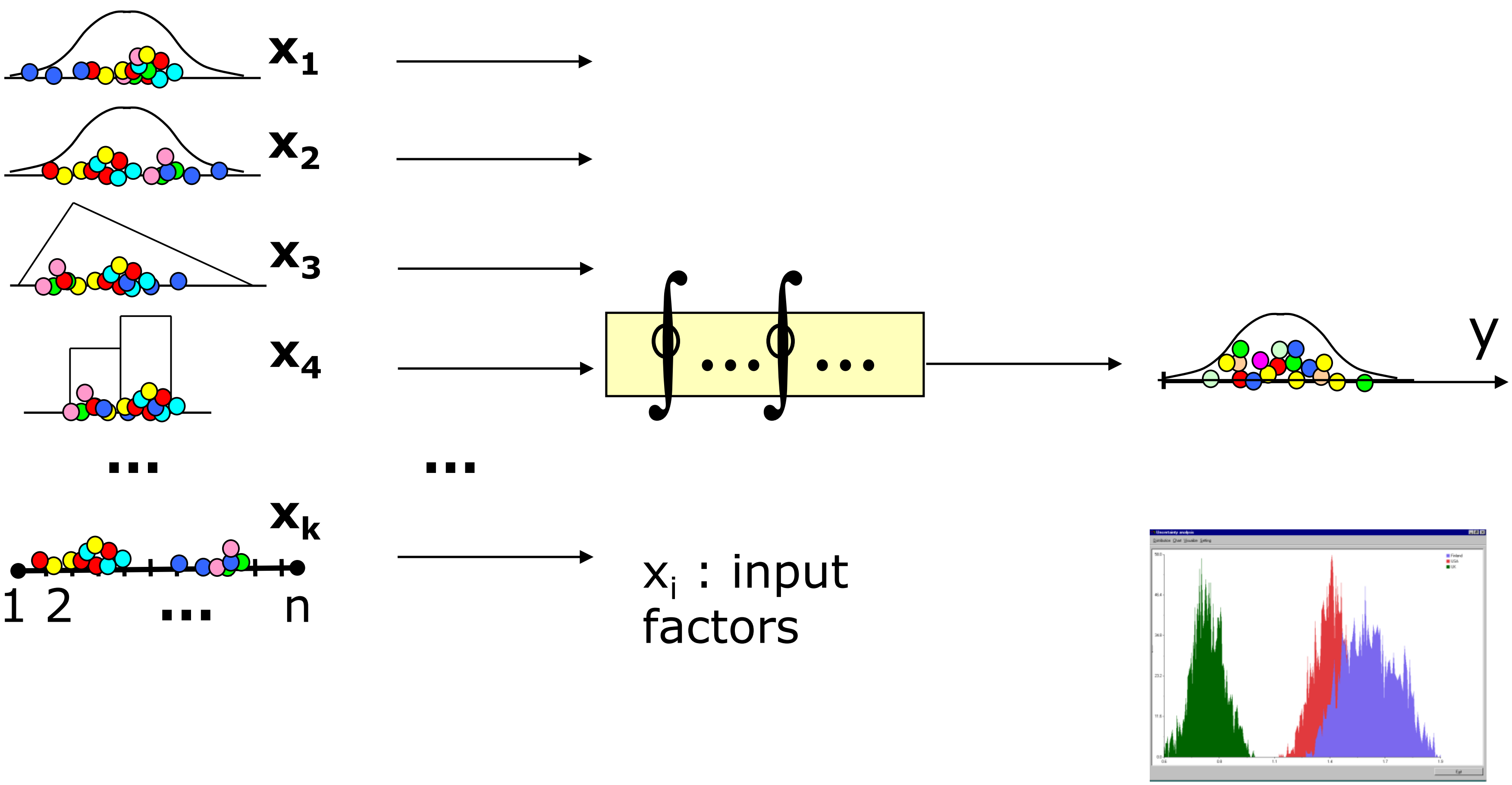
Joint p.d.f. $p(x_1, x_2, \ldots, x_k)$

MCMC (Gibbs, Metropolis, …)

Joint Research Centre

Propagation of uncertainty

Input	*Model*	*Output*
$p(\vec{x})$	$Y = f(\vec{x})$	$p(Y)$



Model approximation (meta-modelling)

$$x \in \Omega \longrightarrow \boxed{f(\mathbf{x})} \longrightarrow Y$$

Realistic Computer models are very costly in terms of run time. We need a way to characterize uncertainty and perform SA based on a limited number of model runs.

We aim to identify a 'simple' relationship between Xi's and Y that fits well the original model and is less computationally demanding.

Model approximation (meta-modelling)

A Meta-model is a simpler model that mimics the larger computational model.

Evaluations of Meta-Models are much faster.

Model approximation (meta-modelling)

Local approximation methods take f and its derivatives at a base point X_0 and construct a function that matches the properties of *f* in the nearby region (Taylor series).

Model approximation (meta-modelling)

1. Linear Regression
2. Quadratic Response Surface Regression
– Options 1 and 2 work pretty well in some cases. However many realistic models are highly nonlinear and/or periodic and these methods fail.
3. Gaussian Process or Spatial Models
4. Nonparametric Regression Models
– Both 3 and 4 work fairly well for a small number of inputs.
– Variable selection is helpful for a modest number of inputs.

ANOVA Models

European Commission

ANOVA models define a decomposition of Y=*f(x)* into main effects and interactions

$$Y = f(x) = f_0 + \sum_{i=1}^{k} f_i(x_i) + \sum_i \sum_{j>i} f_{ij}(x_i, x_j) + \ldots + f_{1,2,\ldots,k}(x_1, x_2, \ldots, x_k)$$

This is also called the High Dimensional Model Representation (HDMR)

E.g., if k=3:

$$f(x) = f_0 + f_1(x_1) + f_2(x_2) + f_3(x_3) + f_{12}(x_1, x_2) + f_{13}(x_1, x_3) + f_{23}(x_2, x_3) + f_{123}(x_1, x_2, x_3)$$

This decomposition is non-unique for general joint pdf's of **x.**

The total number of summands in the ANOVA is 2^k

Properties of the ANOVA decomposition (orthogonal case)

European Commission

$$Y = f(x) = f_0 + \sum_{i=1}^{k} f_i(x_i) + \sum_i \sum_{j>i} f_{ij}(x_i, x_j) + \ldots + f_{1,2,\ldots,k}(x_1, x_2, \ldots, x_k)$$

If each term is chosen with zero mean ...

$$\int_0^1 f_i(x_i) \quad dp(x_i) = 0, \quad \forall x_i \quad i = 1, 2, \ldots, k$$

$$\int_0^1 \int_0^1 f_{ij}(x_i, x_j) \quad dp(x_i) dp(x_j) = 0, \quad \forall x_i, x_j \quad i < j$$

....

$$\int_\Omega f_{12\ldots k}(x_1, x_2, \ldots, x_k) \quad dp(x_1) dp(x_2) \ldots dp(x_k) = 0.$$

Properties of the ANOVA decomposition (orthogonal case)

European Commission

$$Y = f(x) = f_0 + \sum_{i=1}^{k} f_i(x_i) + \sum_i \sum_{j>i} f_{ij}(x_i, x_j) + \ldots + f_{1,2,\ldots,k}(x_1, x_2, \ldots, x_k)$$

… then the ANOVA decomposition has TWO properties

$$\int_{\Omega} f(x) \; dp(x) = f_0 = E(Y)$$

All the summands are orthogonal:

if $(i_1,\ldots,i_s) \neq (j_1,\ldots,j_l)$ $\quad \int_{\Omega^k} f_{i_1,\ldots,i_s} f_{j_1,\ldots,j_l} dp(x) = 0$

It follows that the ANOVA decomposition is **unique** and each term can be defined as …

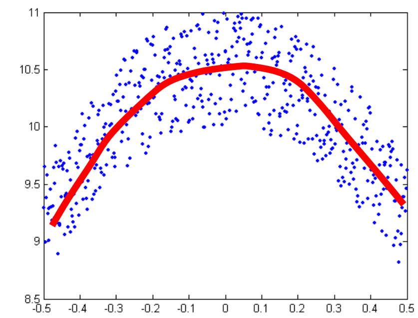
Joint Research Centre

Properties of the ANOVA decomposition

European Commission

$$Y = f(x) = E(Y) + \sum_{i=1}^{k} f_i(x_i) + \sum_i \sum_{j>i} f_{ij}(x_i, x_j) + \ldots + f_{1,2,\ldots,k}(x_1, x_2, \ldots, x_k)$$

$$\int_{\Omega - x_i} \cdot \int f(x)\, dp(x \mid x_i) = f_0 + f_i(x_i) = E(Y \mid x_i)$$



$$\int_{\Omega - \{x_i, x_j\}} \cdot \int f(x)\, dp(x \mid x_i x_j) = f_0 + f_i(x_i) + f_j(x_j) + f_{ij}(x_i, x_j)$$
$$= E(Y \mid x_i, x_j)$$

Joint Research Centre

ANOVA

European Commission

$$Y - E(Y) = \sum_{i=1}^{k} f_i\left(x_i\right) + \sum_i \sum_{j>i} f_{ij}\left(x_i, x_j\right) + \ldots + f_{1,2,\ldots,k}\left(x_1, x_2, \ldots, x_k\right)$$

Being the terms orthogonal, we can square and integrate the eq above over Ω and decompose the variance of *f(x)* into terms of increasing dimensionality (ANOVA)

$$V(Y) = \sum_{i=1}^{k} V_i + \sum_i \sum_j V_{ij} + \sum_i \sum_j \sum_k V_{ijk} \ldots + V_{1,2,\ldots,k}$$

$$V_i = \int f_i^2(x_i)dx_i$$

$$V_{i_1,i_2,\ldots i_s} = \int f^2{}_{i_1,\ldots,i_s} dx_{i_1} dx_{i_2} \ldots dx_{i_s}$$

$$1 = \sum_{i=1}^{k} S_i + \sum_i \sum_j S_{ij} + \sum_i \sum_j \sum_k S_{ijk} \ldots + S_{1,2,\ldots,k}$$

Joint Research Centre

Main effects

If we were to approximate f(x) with a function $g(X_i)$ …

$$L = E[(f(x) - g(x_i))^2]$$

$$g(x_i) = E(Y \mid x_i) \Rightarrow L = L_{\min}$$

$$L_{\min} = E[Var(Y \mid X_i)]$$

… $f_i=E(Y|X_i)$ has the minimum loss L among univariate functions

Joint Research Centre

European Commission

Main effects

$$V_i = Var[E(Y \mid x_i)]$$

$$= Var(Y) - E[Var(Y \mid x_i)]$$

$$L = E[(f(x) - g(x))^2]$$

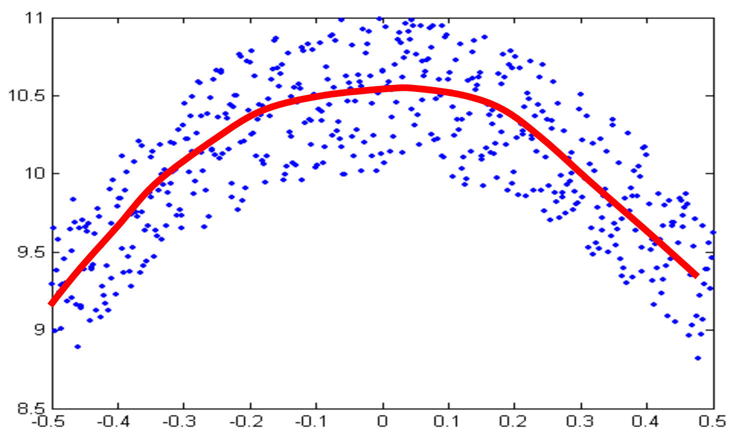
$$g(x) = f_0 \Rightarrow L = V(Y)$$

$$g(x) = f(x) \Rightarrow L = 0$$

$$g(x_i) = E(Y \mid x_i) \Rightarrow L = L_{min}$$

$$L_{min} = E[Var(Y \mid X_i)]$$

when we approximate f(x) with a function $g(x_i)$, $g^* = E(Y|X_i)$ has the minimum loss L



Main effects

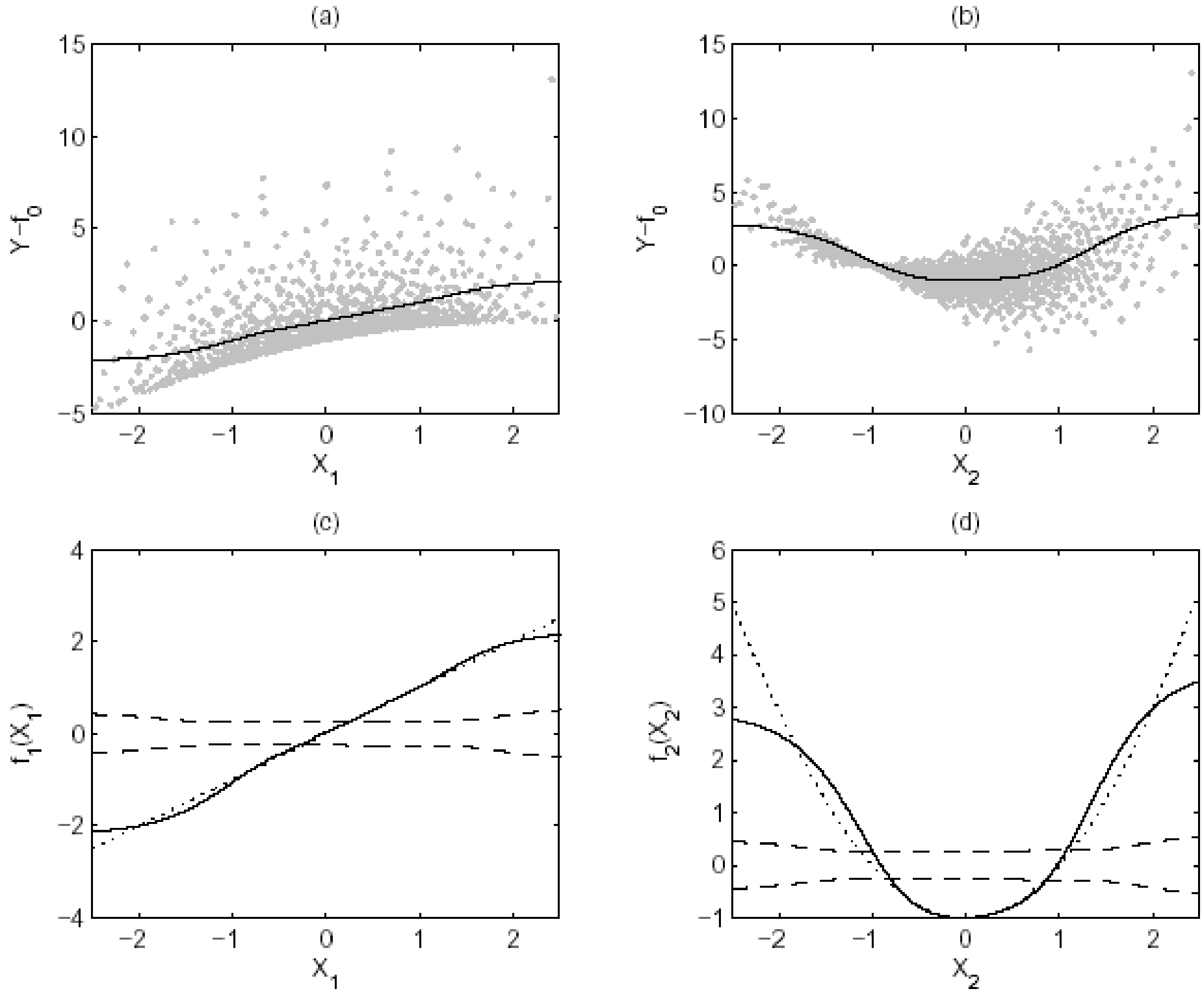
… $f_{ij}=E(Y|X_i,X_j)$ has the minimum loss L among bivariate functions …

… and so on …

Var(fi)/var(Y): also called non-parameteric R-squared; Pearson’s correlation ratio; …

Joint Research Centre

Main effects



$$f(X_1, X_2) =$$
$$X_1 + X_2^2 + X_1 \cdot X_2$$

$$X_i \sim N(0,1)$$

ANOVA model estimation

European Commission

Tensor product decomposition (reproducing kernel Hilbert space, RKHS)

$$F = \bigotimes_{j=1}^{k} H_j = \{1\} \oplus \left\{ \bigoplus_{j=1}^{k} \bar{H}_j \right\} \oplus \left\{ \bigoplus_{j<i} (\bar{H}_j \otimes \bar{H}_i) \right\} \oplus \cdots$$

Orthogonal functional decomposition

$$F = \{1\} \oplus \left\{ \bigoplus_{j=1}^{q} F_j \right\}$$

ANOVA model estimation

Smoothing methods to estimate ANOVA decompositions, truncated at the 2nd -3rd order terms:

SMOOTHING SPLINES ANOVA MODELS

Smoothing the y vs x mapping (think of an HP-filter), that provides efficient convergence properties to the true ANOVA decomposition.

[this is one possible methods, other are RBF's, kernel regressions, …]

ANOVA model estimation

European Commission

Denote the generic mapping as Y = f(**X**), where $\mathbf{X} \in [0, 1]^k$ and k is the number of parameters.

The simplest example of smoothing spline mapping estimation of z is the additive model:

$$g(\mathbf{X}) = g_0 + \sum_{i=1}^{k} g_i\left(x_i\right)$$

Joint Research Centre

ANOVA model estimation

European Commission

To estimate g we can use a multivariate smoothing spline minimization problem, that is, given λ_j, find the minimizer g(**X**) of:

$$\frac{1}{N}\sum_{n=1}^{N}(y_n - g(\mathbf{X}_n))^2 + \sum_{j=1}^{k}\lambda_j \int_0^1 [g_j''(X_j)]^2 dX_j$$

where a Monte Carlo sample of dimension N is assumed.

This minimization problem requires the estimation of the *k* hyperparameters λ_j (also denoted as smoothing parameters): GCV, GML, etc. (see e.g. Wahba, 1990; Gu, 2002).

Joint Research Centre

ANOVA model estimation

European Commission

We re-formulate the additive model for the general case with interactions as to find the minimizer g(**X**) of:

$$\frac{1}{N}\sum_{n=1}^{N}(y_n - g(\mathbf{X}_n))^2 + \lambda_0 \sum_{j=1}^{q} \frac{1}{\theta_j} \left\| P^j g \right\|_F^2$$

where the q-dimensional vector of θ_j smoothing parameters needs to be optimized ‘somehow’.

Joint Research Centre

ANOVA model estimation

European Commission

Wahba (1990), Gu (2002), Storlie et al. (2007): en-bloc (like HP);

Ratto et al., 2004-2007: based on recursive filtering and smoothing estimation: SDP modelling. (like KF-smoothing version of the HP);

Liu et al. (2002-2006): polynomial basis expansion.

Joint Research Centre

SDP modelling

SDP modeling is one class of non-parametric smoothing approach first suggested by Young (1993).

The estimation is performed with the help of the `classical' recursive (non-numerical) Kalman filter and associated fixed interval smoothing algorithms and has been applied for sensitivity analysis in Ratto et al. (2004-2007).

Mapping/sensitivity strategies

European Commission

OAT (Taylor):
- truncation;
- mapping by knowing all derivatives in a base point
- decomposition with infinite terms

GSA (ANOVA):
- non-parametric regression/smoothing
- mapping on a space-filling MC sample
- decomposition with finite terms

DSGE models

European Commission

Let us consider a generic DSGE model:

$$E_t\{g(y_{t+1}, y_t, y_{t-1}, u_t; X)\} = 0$$

y_t endogenous variables,
u_t exogenous shocks
X structural parameters.

X can be characterised by plausible ranges, expressed
in terms of prior distributions, or by a posterior distribution, as a result of an estimation.

DSGE models

European Commission

The model behaviour is a function of
the values assumed for X within the prior or posterior space of structural parameters.

Let Y be a generic 'output' of the model: a multiplier, a
measure of fit, an IRF.

Y depends on the values of X

$Y=f(X_1, \ldots, X_k)$,

f non-linear
analytic form is unknown

Joint Research Centre

Mapping the reduced form of RE models

Relationship between the reduced form of a rational expectation model and the structural coefficients.

let the reduced form be
$y_t = Ty_{t-1} + Bu_t$,

'outputs' Y of our analysis will be the entries in the transition matrix $T(X_1,\ldots,X_k)$ or in the matrix $B(X_1,\ldots,X_k)$.

Lubik Schorfheide (2005)

European Commission

We analyse the reduced form coefficients describing the relationship between

π_t vs $e_{R,t}$

We sample the structural coefficients from posterior ranges obtained after estimating the model using data for Canada.

Joint Research Centre

Lubik Schorfheide (2005)

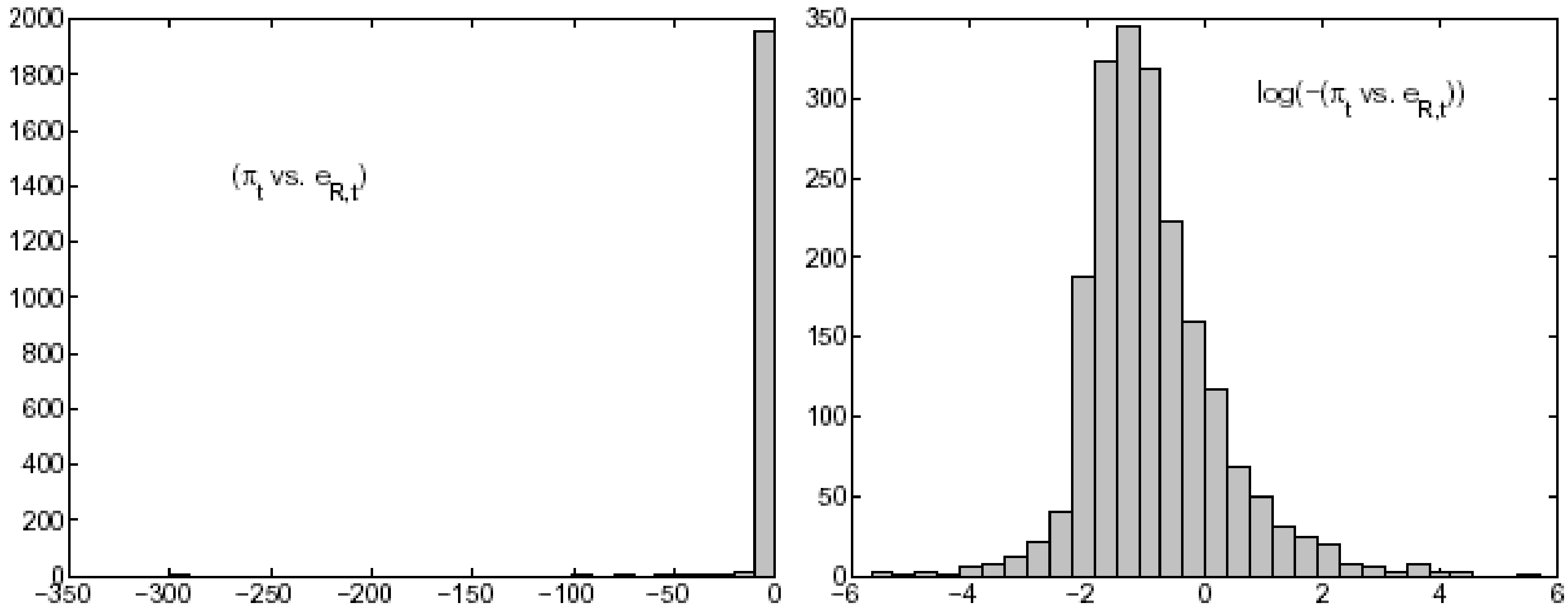


Figure 13: Lubik and Schorfheide model: histograms of the MC sample of the reduced form coefficient $Y = (\pi_t \text{ vs } e_{R,t})$. Left panel: actual values Y; right panel $\log(-Y)$.

Lubik Schorfheide (2005)

European Commission

$\log(-(\pi_t \text{ vs } e_{R,t}))$

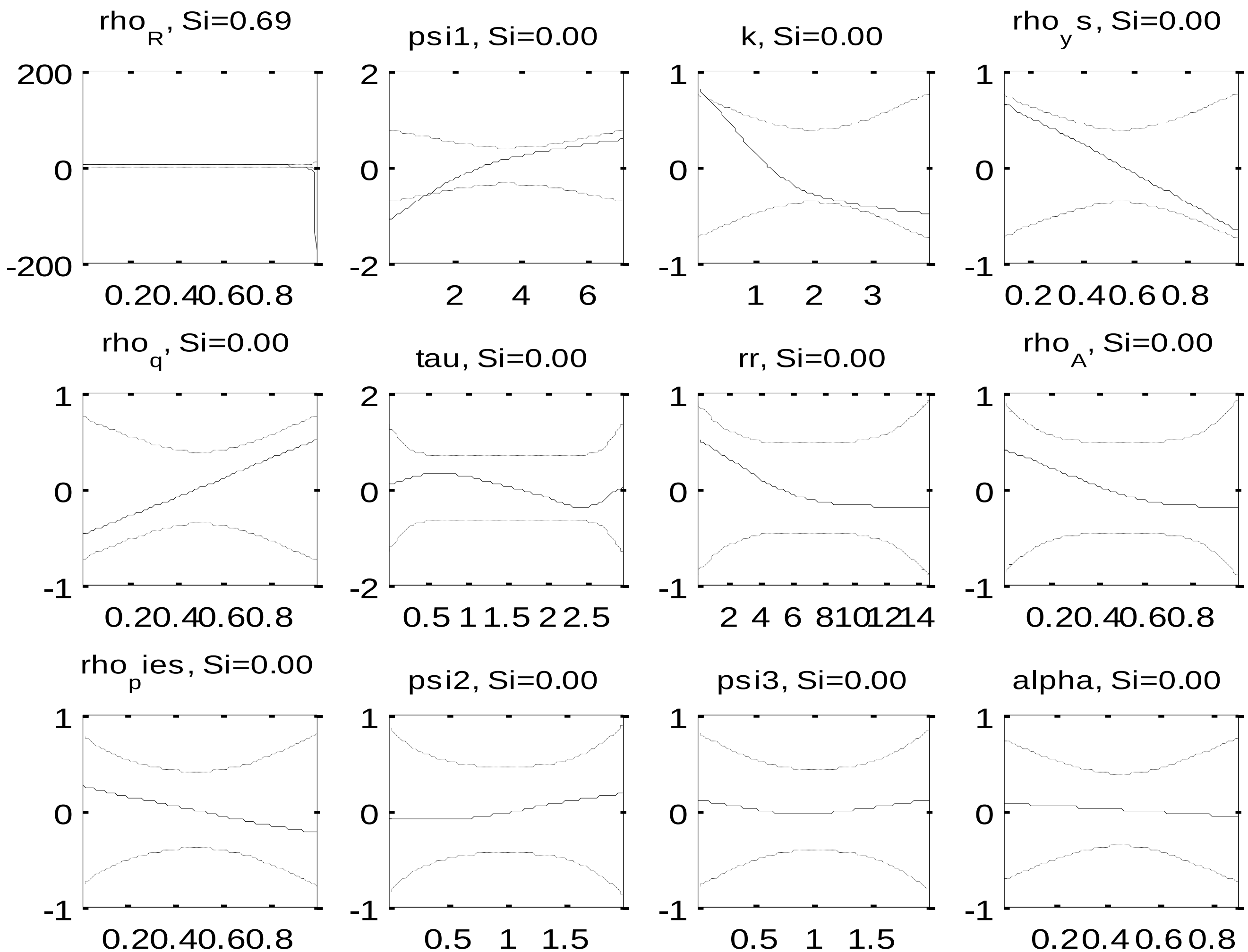
$$Y = -\exp(f_0 + f_1 + \ldots + f_k + e)$$
$$= -\exp(e)\prod_{j=0}^{k}\exp(f_j)$$

Factorisation

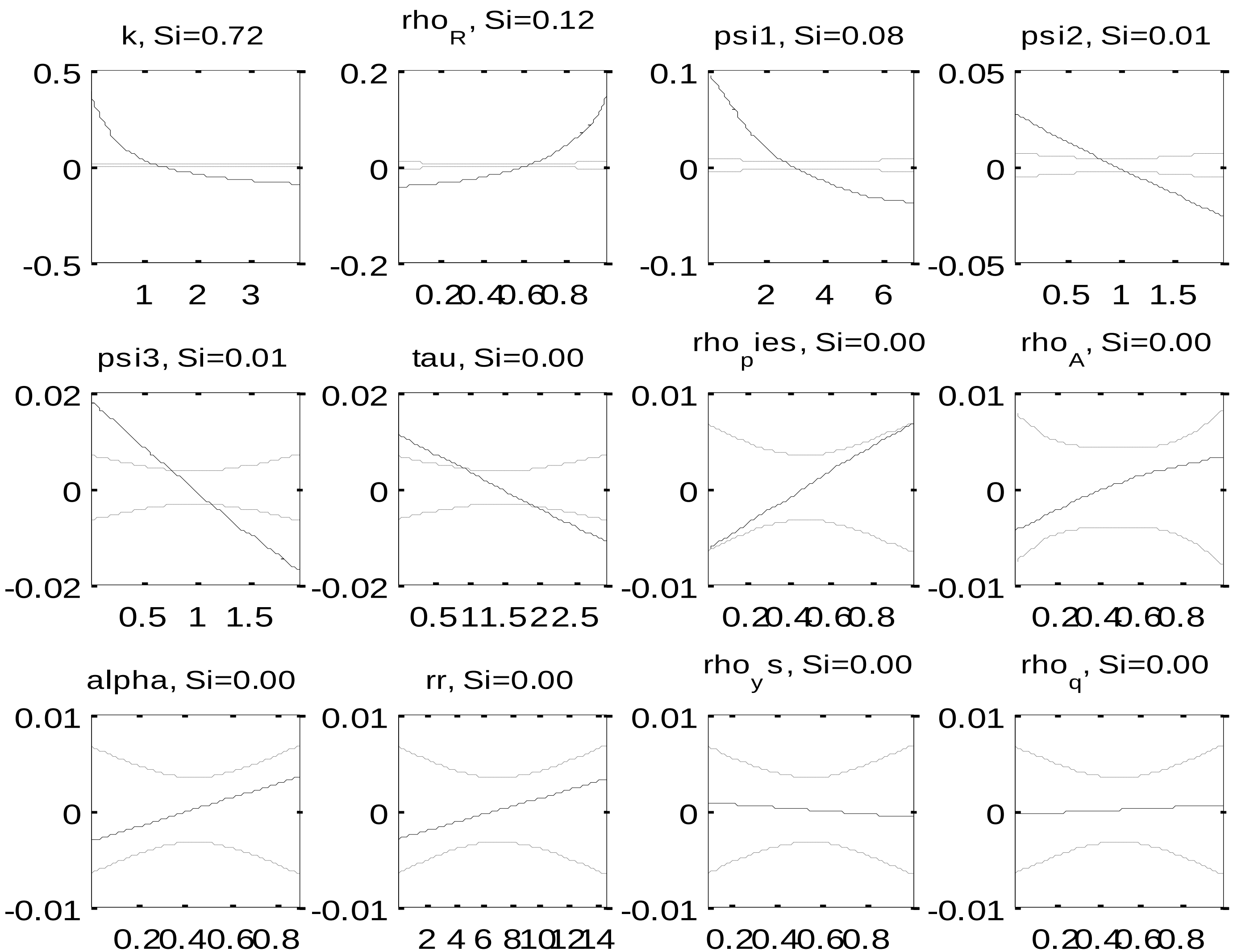
$$Y = -\exp(e)\prod_{j=0}^{k}\exp(f_j) + \sum_{j=0}^{k} g_j + \varepsilon$$

Correction [new in DYNARE 4.5]

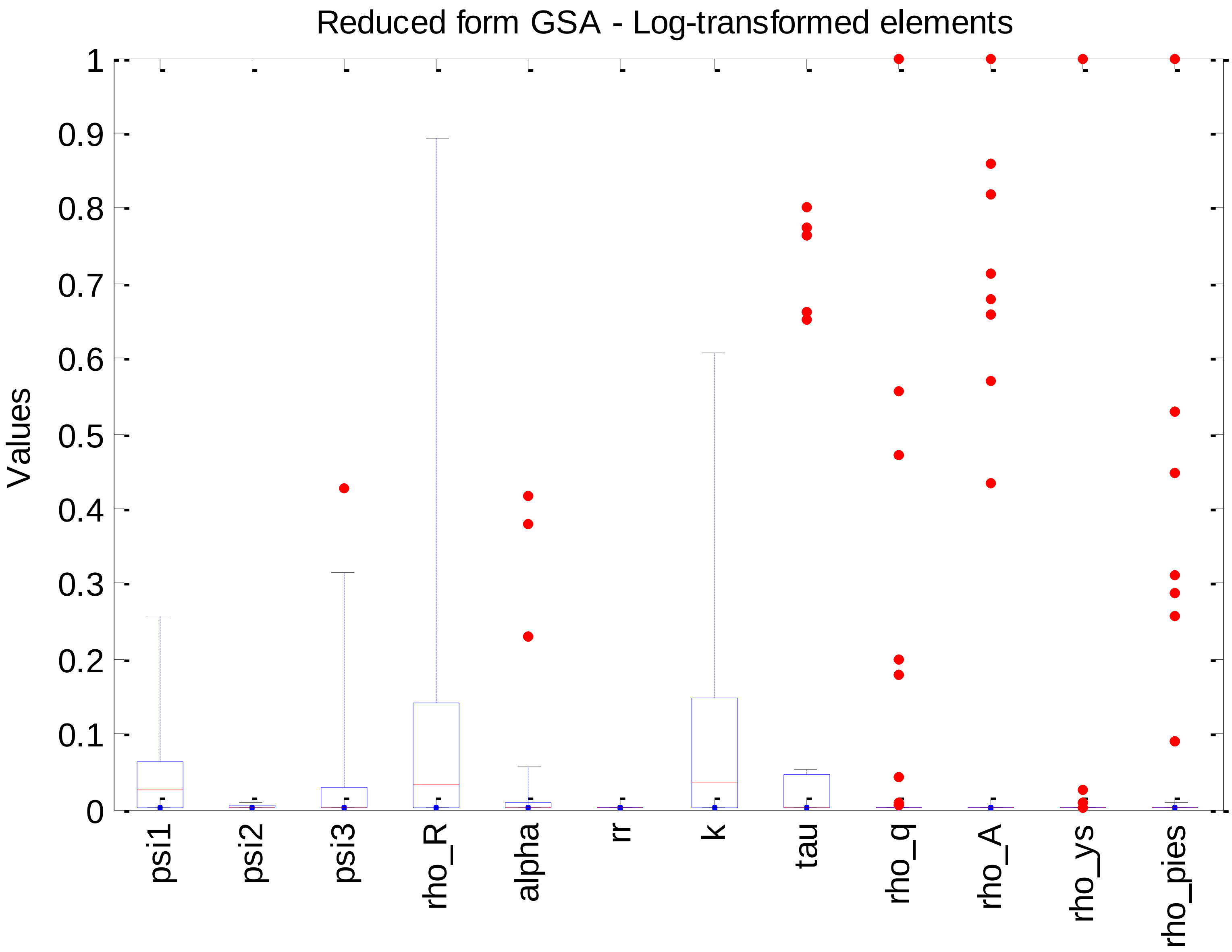
LS2005: pie vs e_R [-log(y)]



LS2005: R vs e_R [-log(y)]



Lubik Schorfheide (2005)



Toolbox documentation

European Commission

The mapping of the reduced form soultion forces the use of samples from prior ranges or prior distributions, i.e.:

pprior=1 (default);

ppost=0 (default);

[unless neighbourhood_width is applied]

Toolbox documentation

option name	default	description
`redform`	0	0 = don't prepare MC sample of reduced form matrices 1 = prepare MC sample of reduced form matrices
`load_redform`	0	0 = estimate the mapping of reduced form model 1 = load previously estimated mapping
`logtrans_redform`	0	0 = use raw entries 1 = use log-transformed entries
`threshold_redform`	[]	[] = don't filter MC entries of reduced form coefficients [max max] = analyse filtered entries within the range [max max]
`ksstat_redform`	0.001	critical p-value for Smirnov statistics d when `threshold_redform` is active plot parameters with p-value<`ksstat_redform`
`alpha2_redform`	0.001	critical p-value for correlation ρ when `threshold_redform` is active plot couples of parameters with p-value<`alpha2_redform`
`namendo`	()	list of endogenous variables
	:	jolly character to indicate ALL endogenous
`namlagendo`	()	list of lagged endogenous variables: analyse entries [`namendo`×`namlagendo`]
	:	jolly character to indicate ALL endogenous
`namexo`	()	list of exogenous variables: analyse entries [`namendo`×`namexo`]
	:	jolly character to indicate ALL exogenous

ss_anova Toolbox

Download here the Recursive SS-ANOVA Toolbox for MATLAB.

http://ipsc.jrc.ec.europa.eu/fileadmin/repository/sfa/finepro/software/ss_anova_recurs.zip

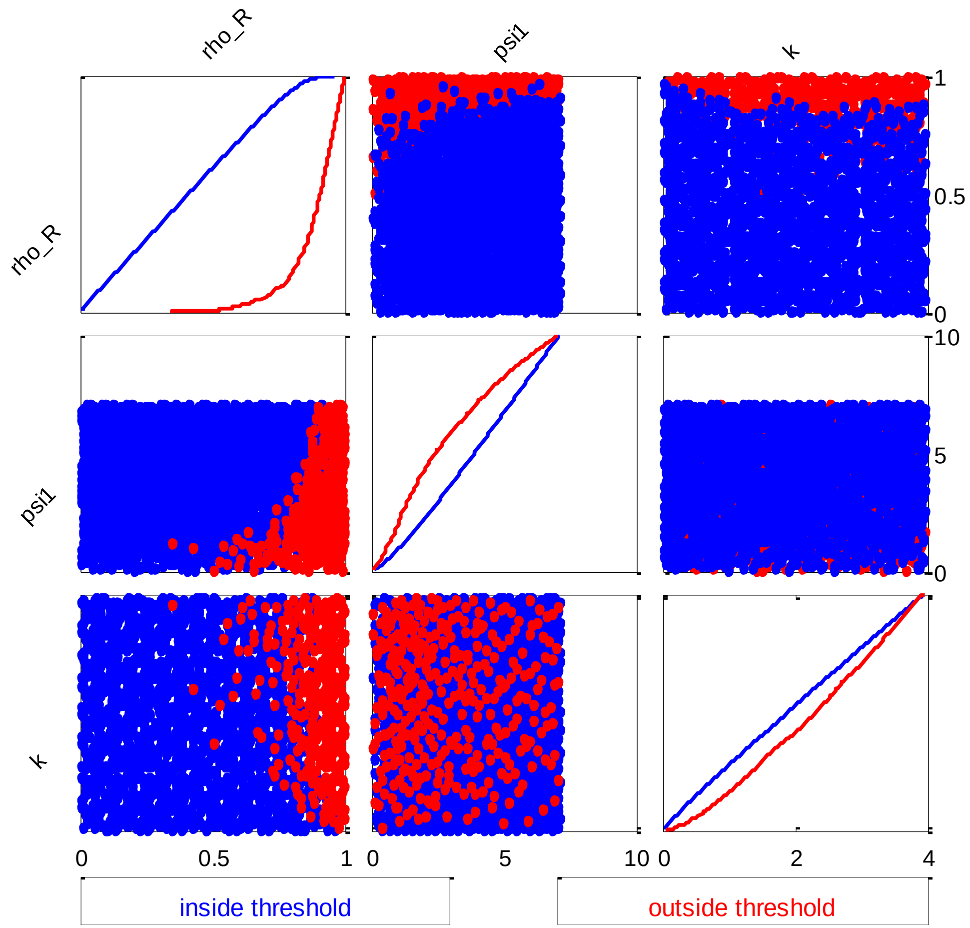
http://ipsc.jrc.ec.europa.eu/fileadmin/repository/sfa/finepro/software/ss_anova_recurs_matlab_ver_less_than_7.5.zip
(MATLAB version < 7.5)

Using Monte Carlo filtering

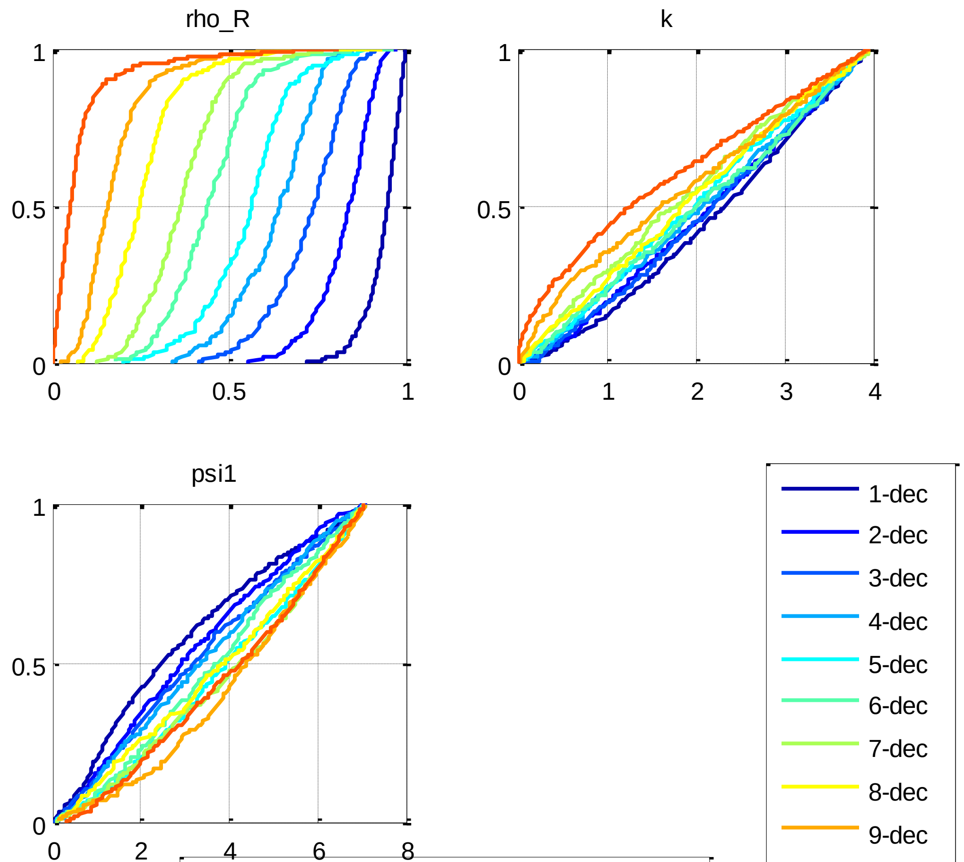
Store the sample of the state space A,B matrices;

For 1-step ahead irf, perform the MCF sensitivity tests for B(i,j) within/outside the specified ranges.

Using Monte Carlo filtering: $y = (\pi_t \text{ vs } R_{t-1}) \in [-1, 0]$



Using Monte Carlo filtering: $y = (\pi_t \text{ vs } R_{t-1}) \in [-\infty, \infty]$



Monte Carlo Filtering for pie vs R

Mapping reduced form solution

option name	default	description
`redform`	0	0 = don't prepare MC sample of reduced form matrices 1 = prepare MC sample of reduced form matrices
`load_redform`	0	0 = estimate the mapping of reduced form model 1 = load previously estimated mapping
`logtrans_redform`	0	0 = use raw entries 1 = use log-transformed entries
`threshold_redform`	[]	[] = don't filter MC entries of reduced form coefficients [`max max`] = analyse filtered entries within the range [`max max`]
`ksstat_redform`	0.001	critical p-value for Smirnov statistics d when `threshold_redform` is active plot parameters with p-value<`ksstat_redform`
`alpha2_redform`	0	critical p-value for correlation ρ when `threshold_redform` is active plot couples of parameters with p-value<`alpha2_redform`
`namendo`	()	list of endogenous variables
	:	jolly character to indicate ALL endogenous
`namlagendo`	()	list of lagged endogenous variables: analyse entries [`namendo`×`namlagendo`]
	:	jolly character to indicate ALL endogenous
`namexo`	()	list of exogenous variables: analyse entries [`namendo`×`namexo`]
	:	jolly character to indicate ALL exogenous