BAYESIAN VALIDATION OF GTAP PARAMETERS ON A LARGE SCALE DYNAMIC DATASET

Oleg Lugovoy*
Vladimir Potashnikov**

* Environmental Defense Fund, Washington DC, USA
** The Russian Presidential Academy of National Economy and Public Administration, Moscow, Russia

Outline

- Motivation
- Methodology
- Application
  - Data
  - Model
  - Results
- Concluding remarks

Motivation

- #1 Control parameters predetermine results
  - “..estimation and validation of … substitution elasticities is an important priority for future work.” (Huff, Hanslow, Hertel, Tsigas, 1998)
  - “… results are sensitive to control parameters …” (common conclusion of a sensitivity analysis in a CGE/AGE study)
- #2 Get closer to dynamics
  - Verify GTAP performance on time series data

How to?

- We have:
  - GTAP database (SAM & parameters)
  - Data from various sources:
    - GDP, trade statistics, …
    - Population, labor, migration,…
    - Output/inputs/prices by industries,…
- We want:
  - Estimate/validate the model's exogenous parameters subject to the data, and consistent with the model and our believes

How to? Options?

- Best guess (base on some knowledge or just assumption)
- Estimate econometrically on available data, separately from the model, i.e. on another model
- Estimate econometrically subject to all relevant data, relevant knowledge, and the model

How to? Options?

- Best guess (base on some knowledge or just assumption)

- Estimate econometrically on available data, separately from the model, i.e. on another model

- Estimate econometrically subject to all relevant data, relevant knowledge, and the model

How to? Options?

- Best guess (base on some knowledge or just assumption)

- Estimate econometrically on available data, separately from the model, i.e. on another model

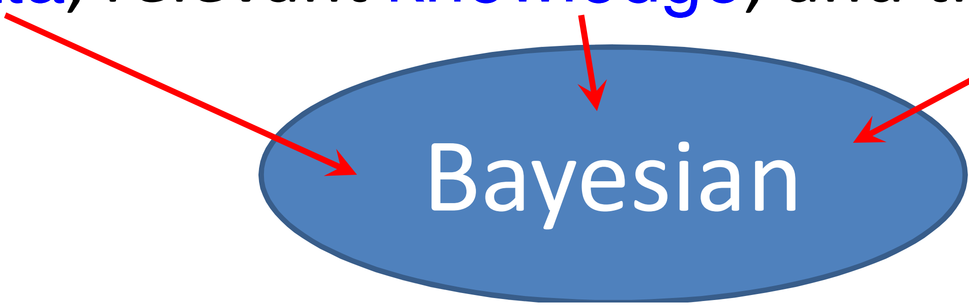
- Estimate econometrically subject to all relevant data, relevant knowledge, and the model

How to? Options?

- Best guess (base on some knowledge or just assumption)

- Estimate econometrically on available data, separately from the model, i.e. on another model

- Estimate econometrically subject to all relevant data, relevant knowledge, and the model



Bayes' rule

$$p(\theta \mid x) \propto p(\theta) \cdot f(x \mid \theta)$$

θ - parameters

x - observed data

$p(\theta)$ - Prior distribution of parameters (our believes/knowledge)

$f(x \mid \theta)$ - Likelihood function

$p(\theta \mid x)$ - Posterior distribution of parameters (combination of information in data and our believes)

Estimation of GTAP Parameters

with

Markov Chains Monte Carlo (MCMC)

Metropolis-Hastings (MH) algorithm

Model modification

Add equations for observed variables:

$$x_i = x_i^* + x_i^{eps}$$

x_i – observed variables (shocks)
x_i^* – endogenous variables
x_i^{eps} – residual
i - industry

New objective function: minimize sum of squares

$$\sum_i w_i \left(\log(x_i) - \log(x_i^*)\right)^2 \xrightarrow[x_i^*]{} \min$$

w_i – weights

Likelihood function & Prior distribution

$$p(\theta \mid x) \propto p(\theta) \cdot f(x \mid \theta)$$

Likelihood function & Prior distribution

$$f(x|\theta) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\sum_{i,t} \frac{\left[\log(x) - \log(x^*)\right]^2}{2\sigma^2}}$$

$$p(\theta | x) \propto p(\theta) \cdot f(x|\theta)$$

Likelihood function & Prior distribution

$$f(x \mid \theta) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\sum_{i,t} \frac{[\log(x) - \log(x^*)]^2}{2\sigma^2}}$$

$$p(\theta \mid x) \propto p(\theta) \cdot f(x \mid \theta)$$

~uniform(-1,10)

METROPOLIS-HASTING ALGORITHM

- *Set initial values* θ*(0).*
- *Set* θ=θ*(t-1)*
- *Generate new candidate values x' from a proposal distribution q(*θ *->* θ*') =q(*θ'|θ*).*
- *Calculate*

$$\alpha = \min\left(1, \frac{f(x \mid \theta')q(\theta' \mid \theta)}{f(x \mid \theta)q(\theta \mid \theta')}\right)$$

- *Update* θ*(t) =*θ' *with probability α and* θ*(t)=*θ*(t-1) with probability (1-α)*
- *Repeat step 2-6, T times.*

Application / results

Model & Data

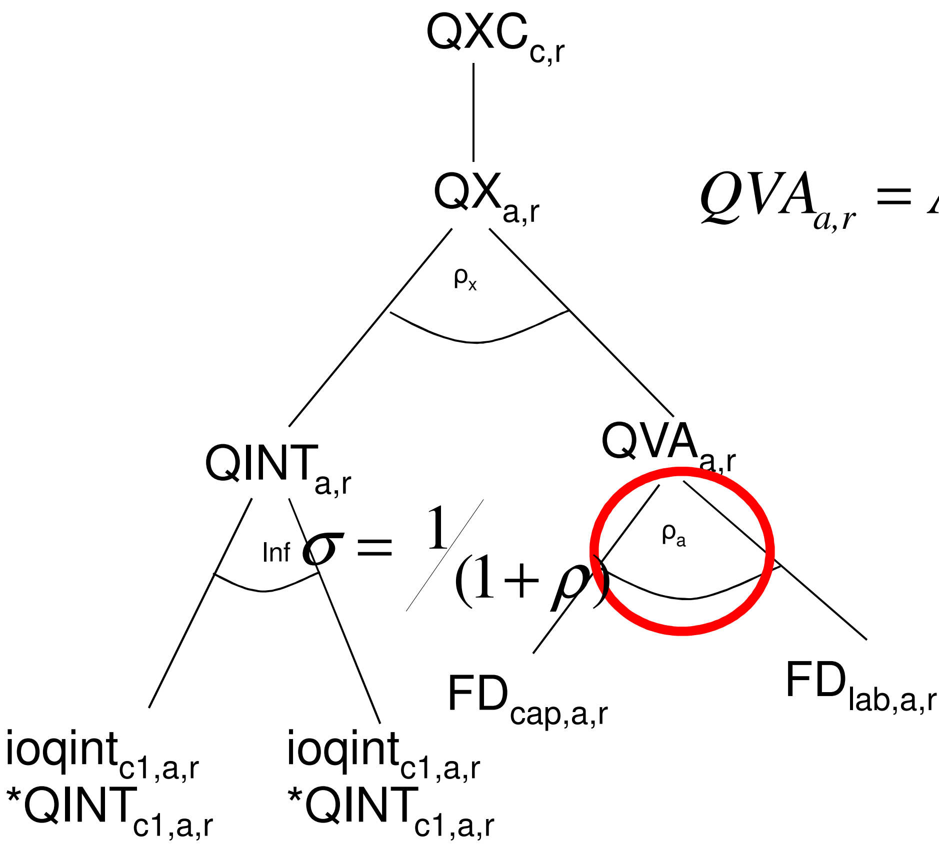
- GTAP model (starting point: GLOBE model in GAMS by S.McDonald, K.Thierfelder, S.Robinson, 2006) :
  - two regions: US & ROW
  - 28 sectors/commodities
- Observed variables:
  - $WF_{f,'us'}$, $FS_{f,'us'}$, $FD_{f,a,'us'}$, $PINT_{a,'us'}$, $QINT_{a,'us'}$, $PVA_{a,'us'}$, $QVA_{a,'us'}$, $PX_{a,'us'}$, $QX_{a,'us'}$, $ADVA_{a,'us'}$
- Estimated parameters
  - elasticity parameter for CES production function for QVA

Data

1. GTAP 7.1 database

2. EU KLEMS dynamic productivity database
   - See O'Mahony, Mary and Marcel P. Timmer (2009), "Output, Input and Productivity Measures at the Industry Level: the EU KLEMS Database", Economic Journal, 119(538), pp. F374-F403, www.euklems.net

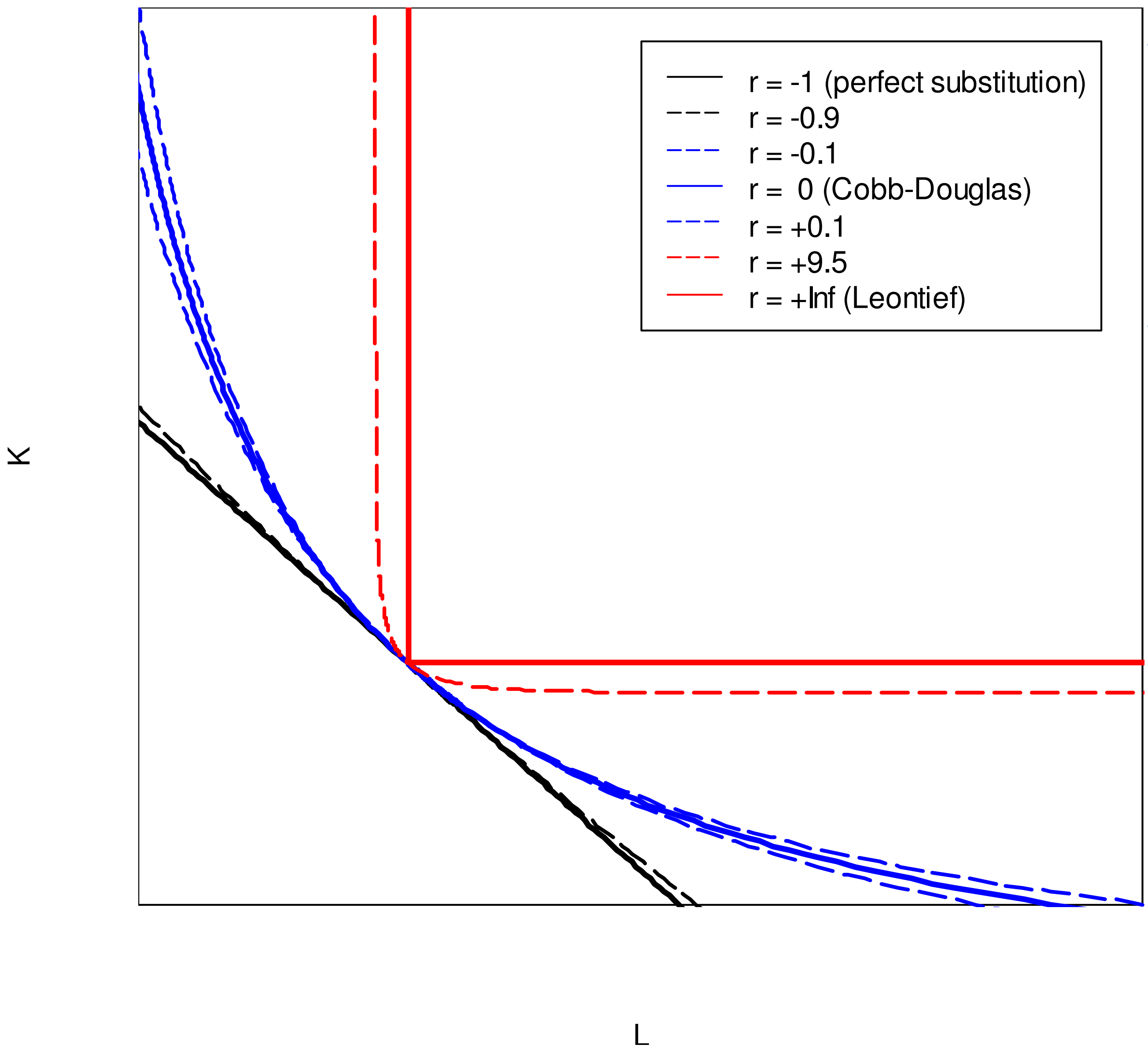
Estimated parameters in CES nest



$$QVA_{a,r} = ADVA_{a,r} * \left(\sum_f d_{f,a,r} * FD_{f,a,r}{}^{-\rho_{a,r}} \right)^{-1/\rho_{a,r}}$$

$$\sigma = 1 \big/ (1+\rho)$$

CES production function & special cases



The switches between:
"linear"
"CES"
"Cobb-Douglas"
and "Leontief"
functions are
added
in the model code

$$\sigma = {}^{1}\!/_{(1+\rho)}$$

Closure

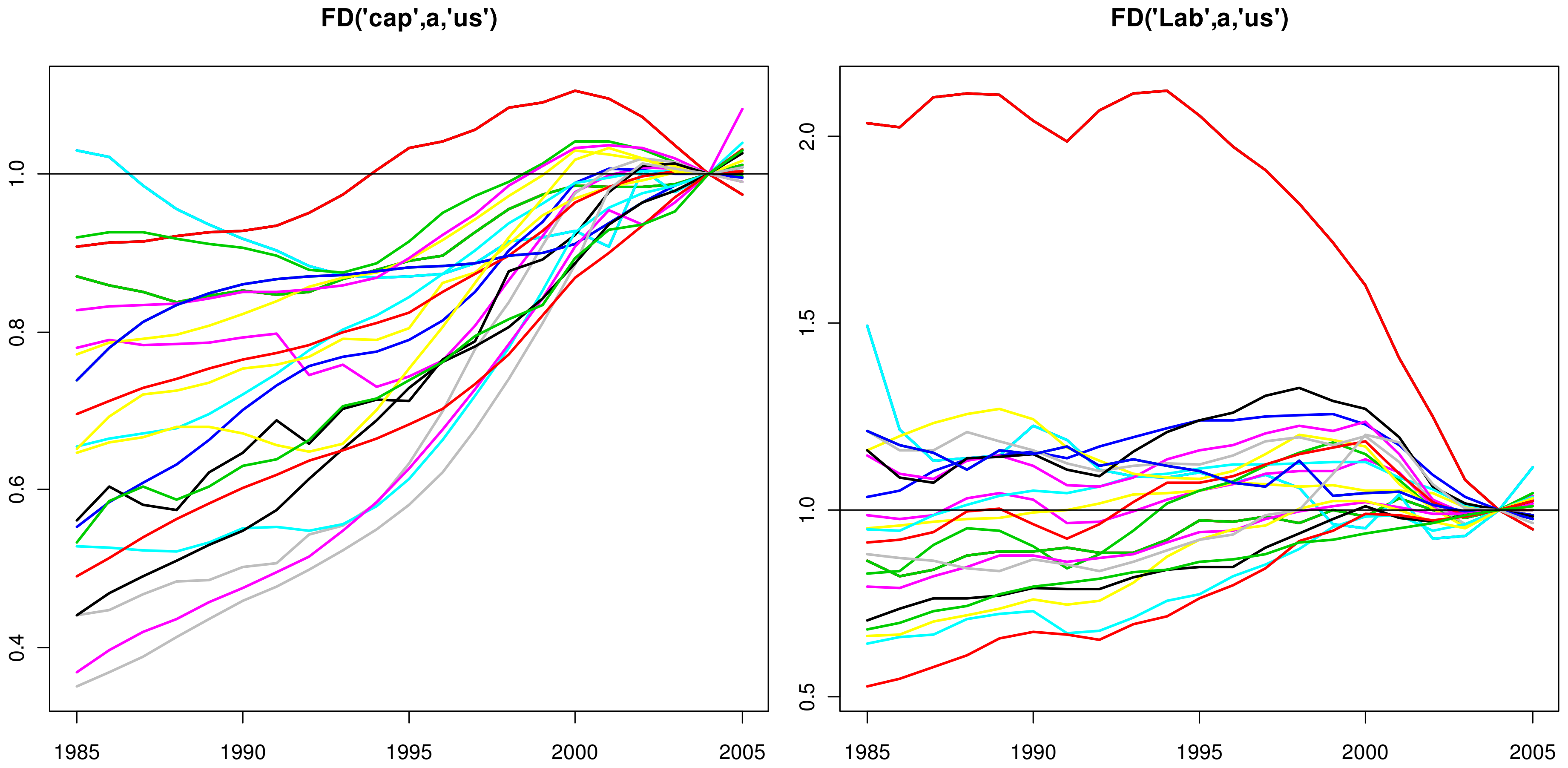
Standard GTAP	For the MCMC estimation
$WF_{f,r}$ – fixed $FS_{f,r}$ – free or $WF_{f,r}$ – free $FS_{f,r}$ – fixed	$WF_{f,r} = WF_{f,r}^{*} + WF^{eps}_{f,r}$ $FS_{f,r} = FS_{f,r}^{*} + FS^{eps}_{f,r}$
in GAMS	
FS.FX(f,'us') = FS0(f,'us') WF.LO(f,'us') = -inf WF.UP(f,'us') = +inf	WF('cap','us') =e= WFcap(t) + eWFcap; FS('cap','us') =e= Fscap(t) + eFScap; WF('Lab','us') =e= WFLab(t) + eWFLab; FS('Lab','us') =e= FSLab(t) + eFSLab;
WF.FX(f,'other') = WF0(f,'other') FS.LO(f,'other') = -inf FS.UP(f,'other') = +inf	

Observed variables (EU KLEMS)

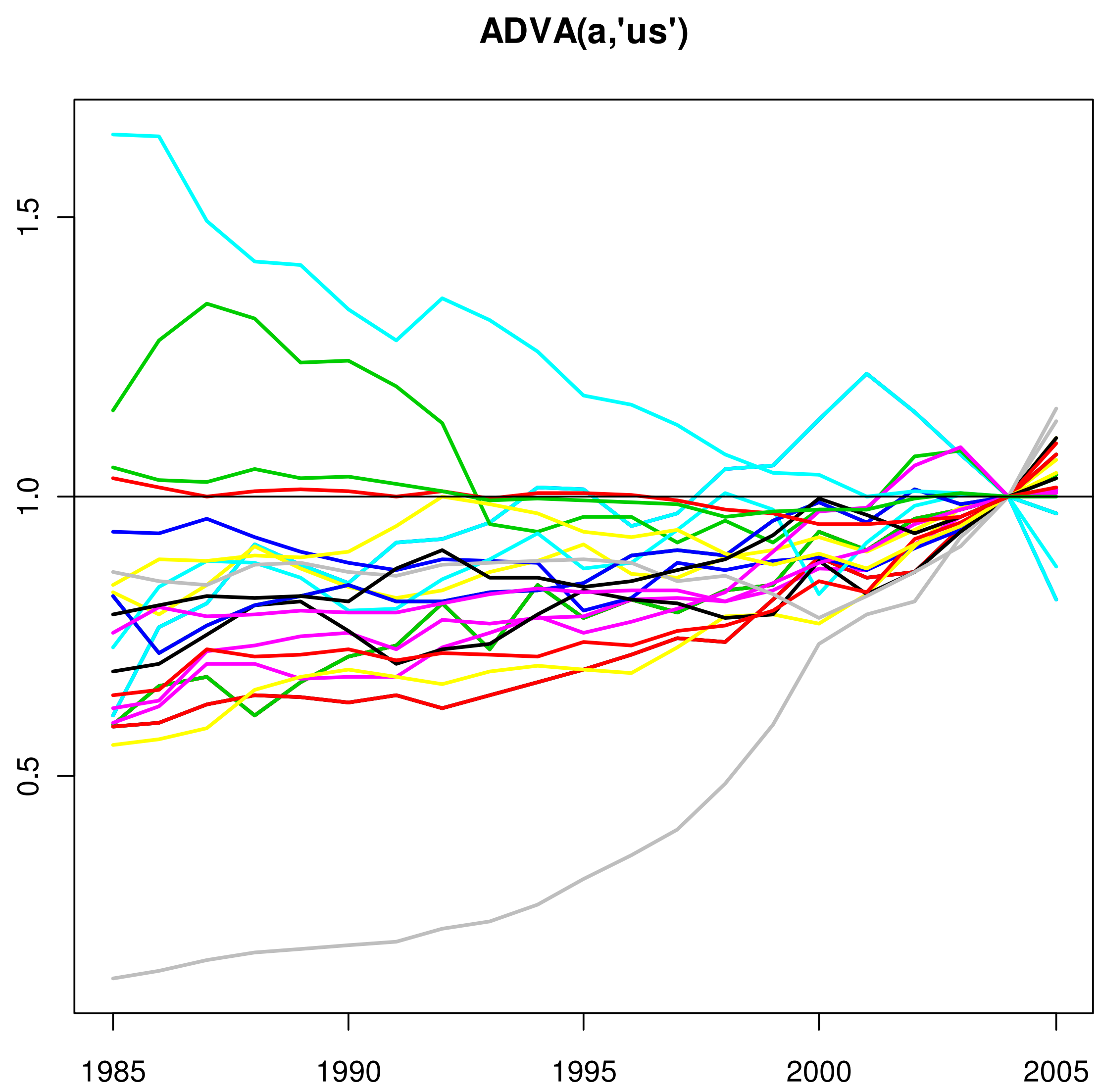
- $WF_{f,\ 'us'}$ *- Price of factor f in USA*
- $FS_{f,\ 'us'}$ *- Supply of factor f in USA*
- $FD_{f,a,\ 'us'}$ *- Demand for factor f by activity a in USA*
- $PINT_{a,\ 'us'}$ *- Price of aggregate intermediate input*
- $QINT_{a,\ 'us'}$ *- Aggregation quantity of intermediates used by activity a in USA*
- $PVA_{a,\ 'us'}$ *- Value added price for activity a in USA*
- $QVA_{a,\ 'us'}$ *- Quantity of aggregate value added for level 1 production*
- $PX_{a,\ 'us'}$*- Composite price of output by activity a*
- $QX_{a,'us'}$*- Domestic production by activity a in USA*
- $ADVA_{a,'us'}$ *- Shift parameter for CES production functions for QVA*

- Summary:
  - 247 observed variables (from 2088 variables in the model)
  - 21 years (1985-2005)
  - 28 estimated parameters

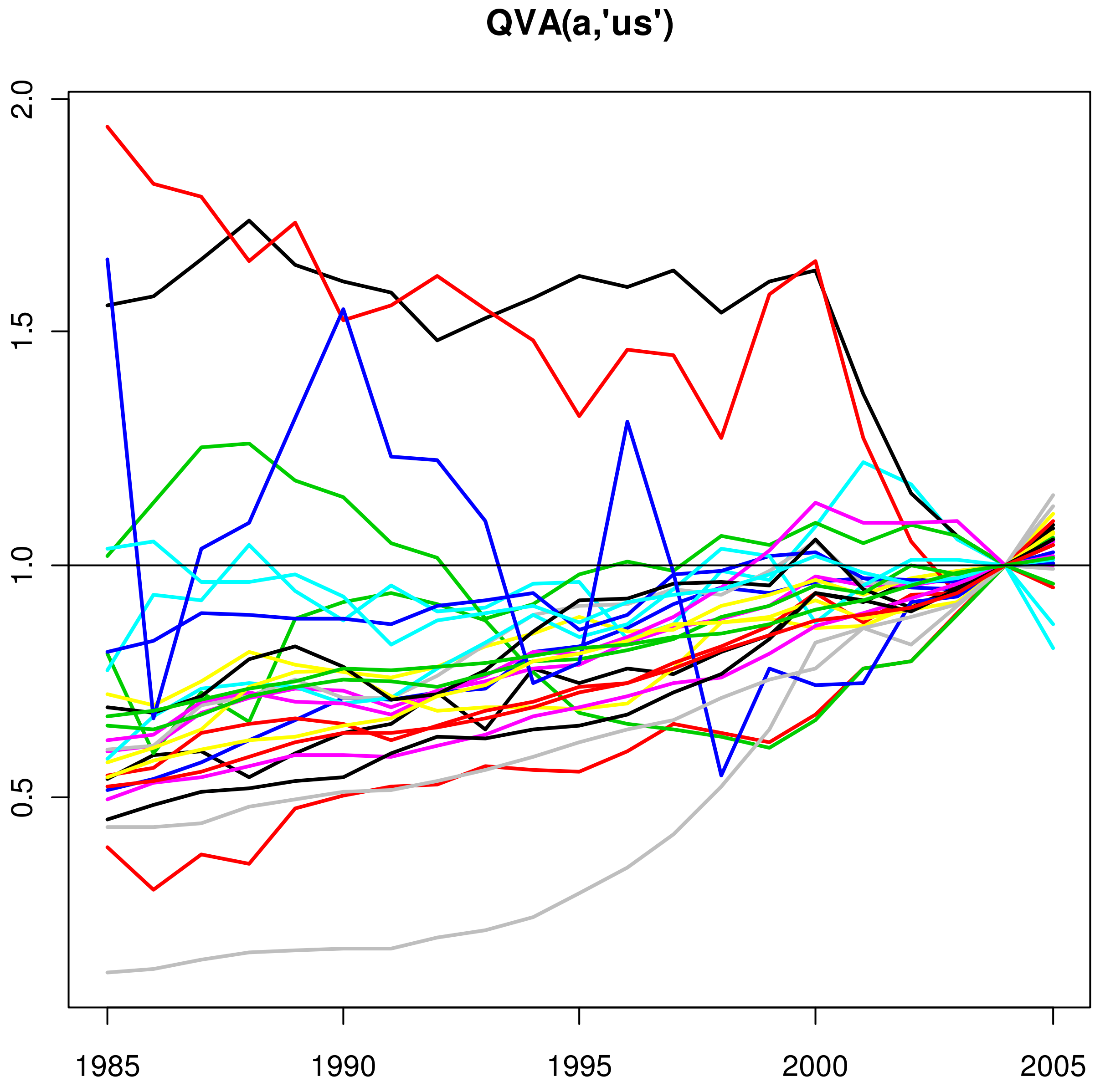
Shocks: factor inputs



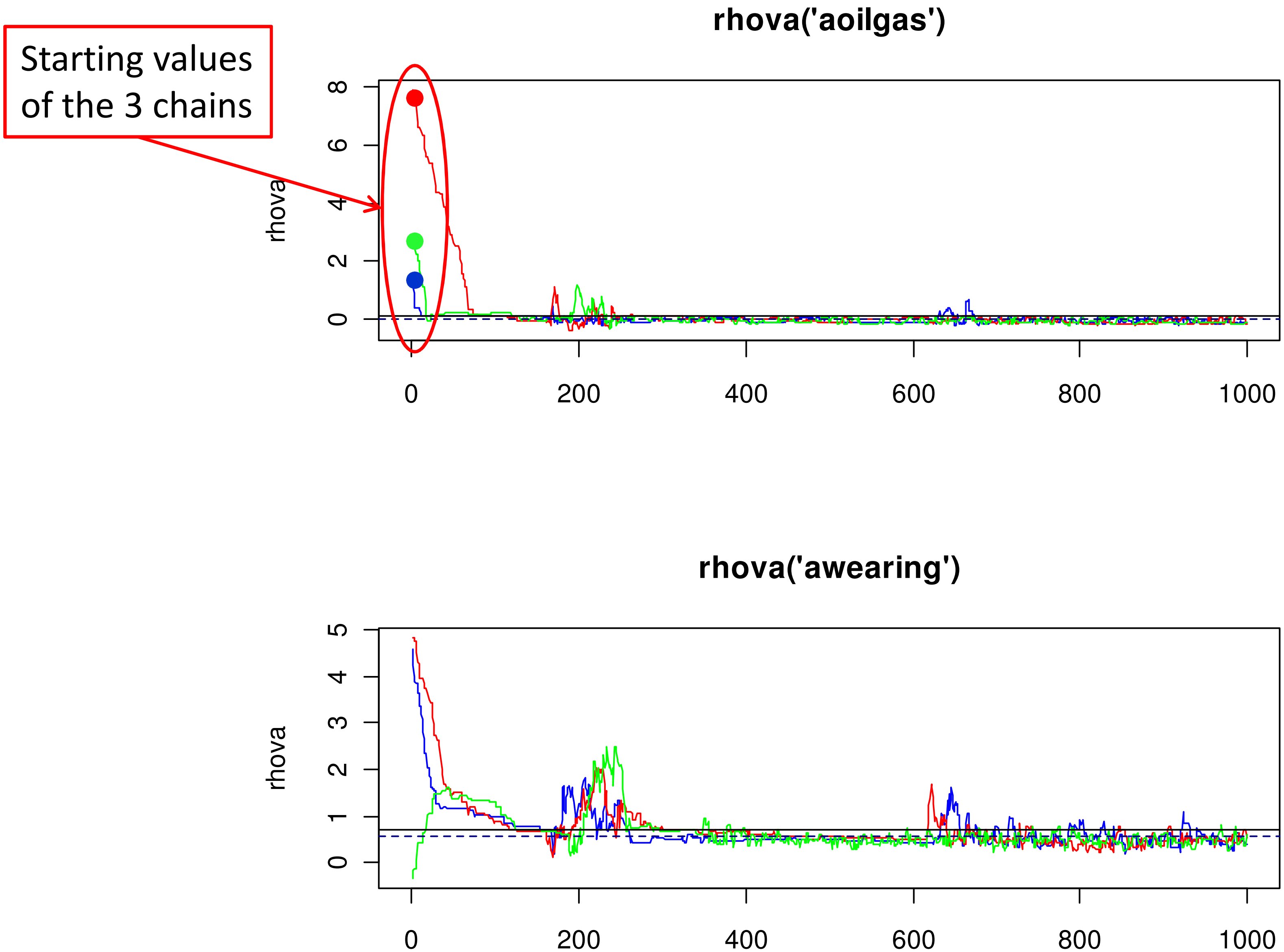
Shocks: productivity



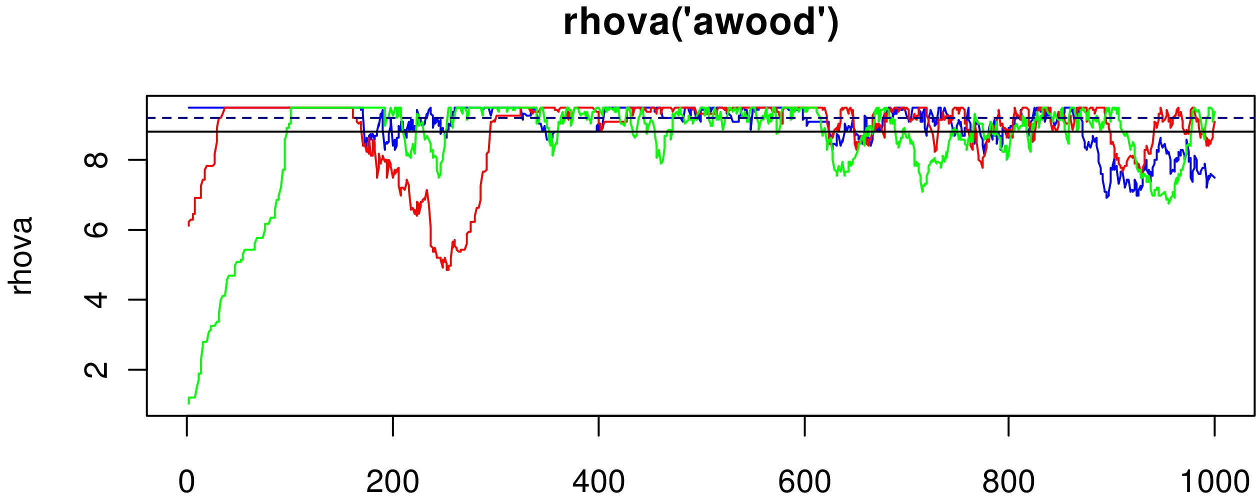
Output

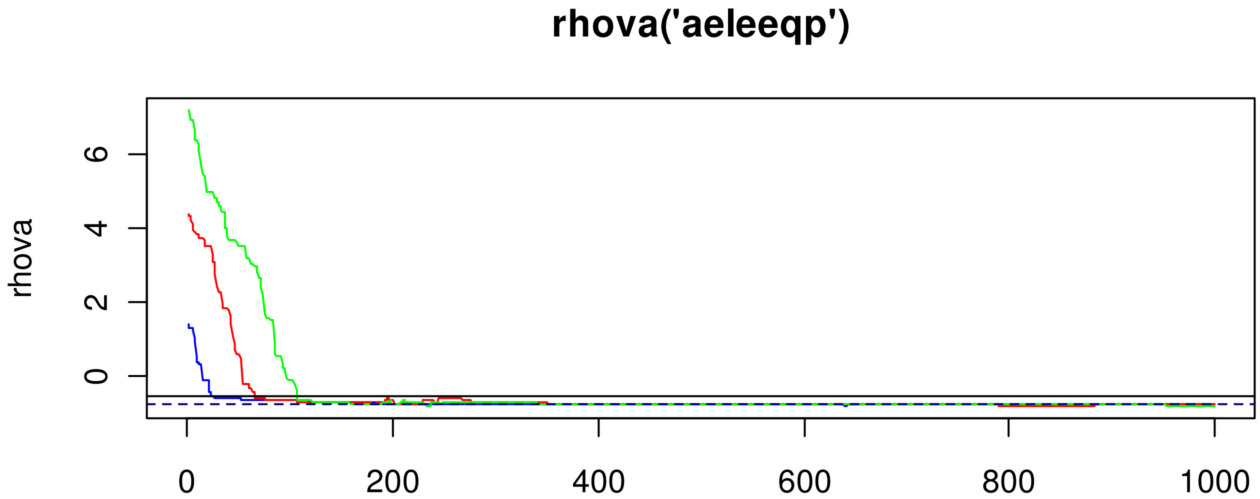


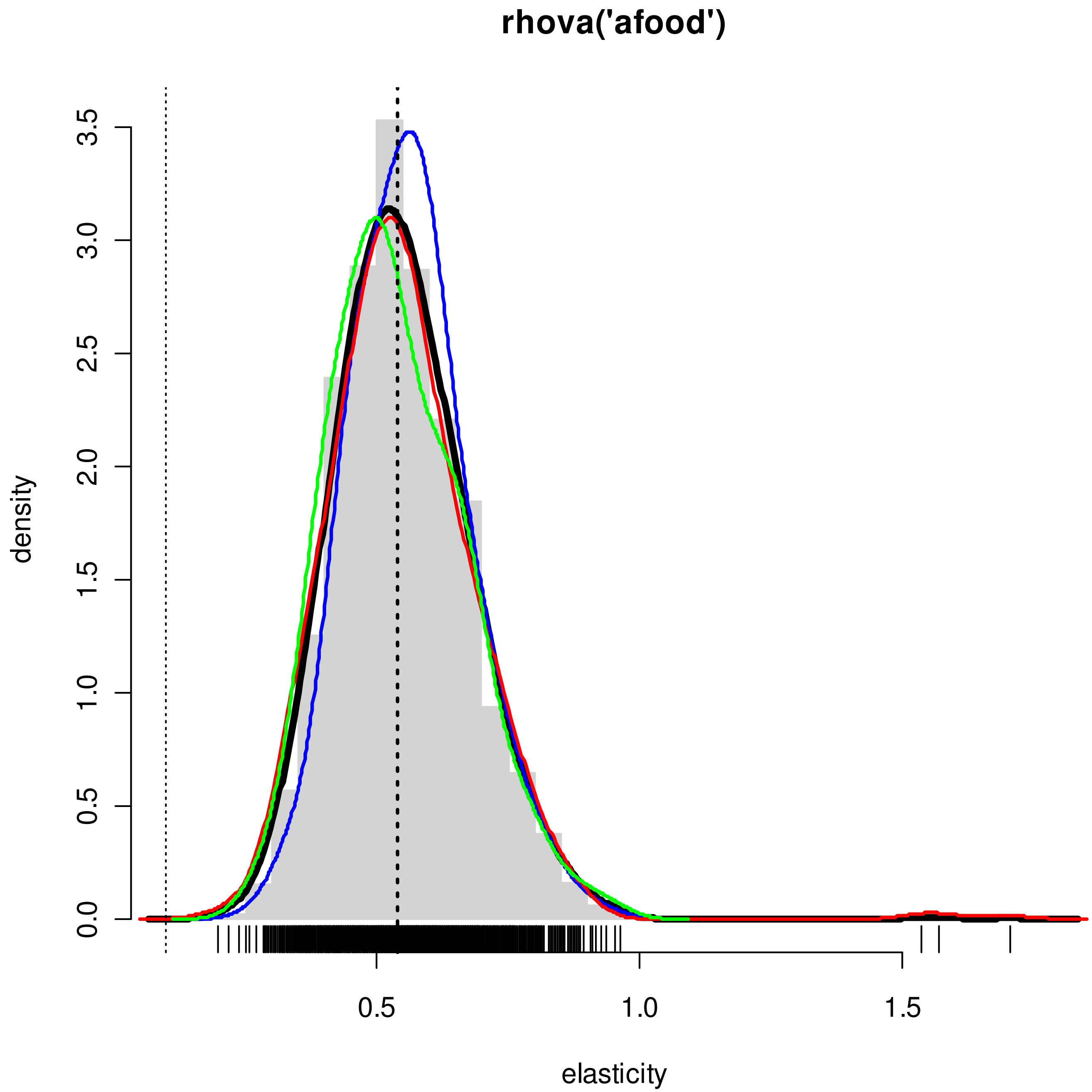
Convergence of MH MCMC algorithm

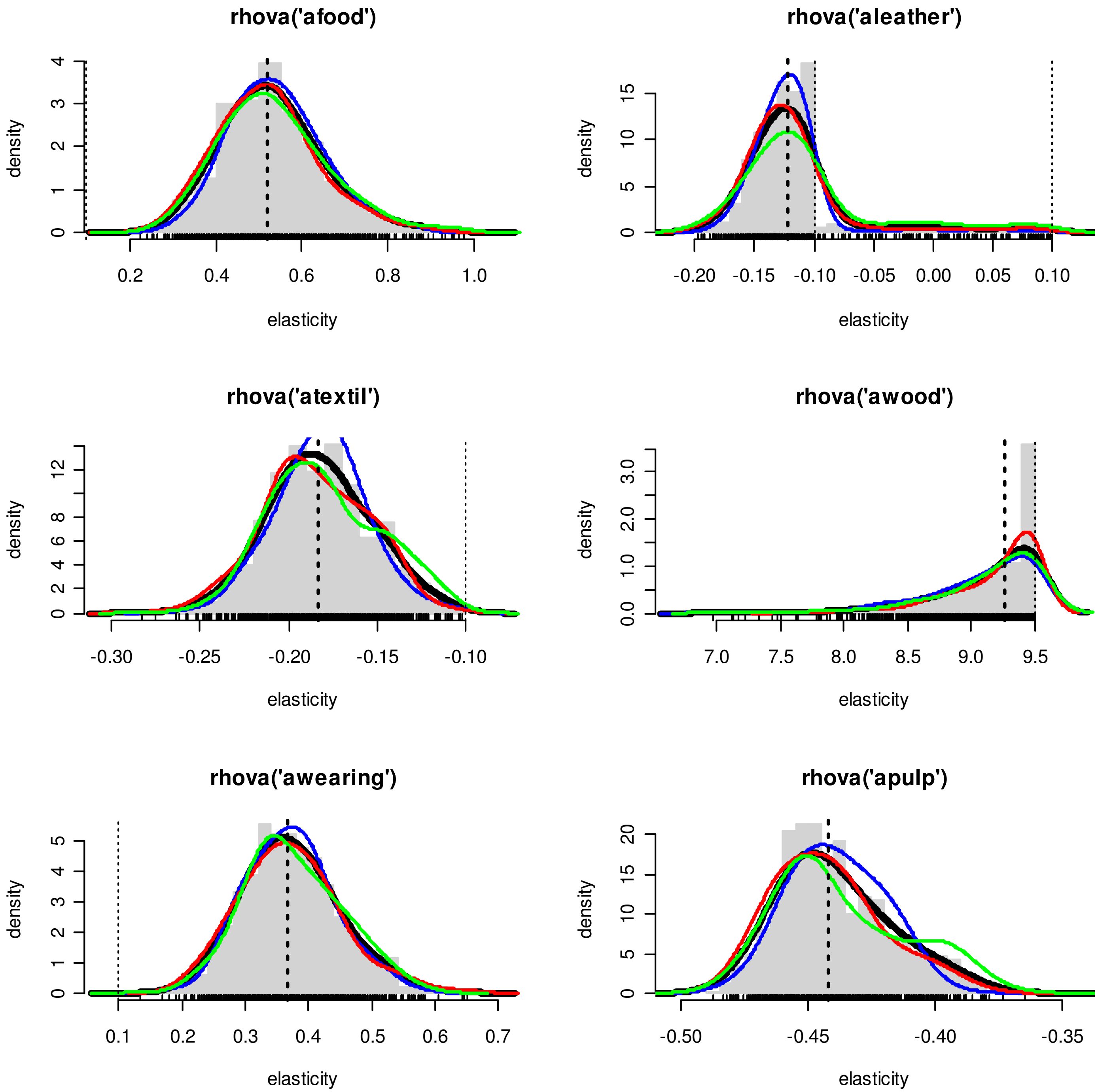


Convergence (cont.)



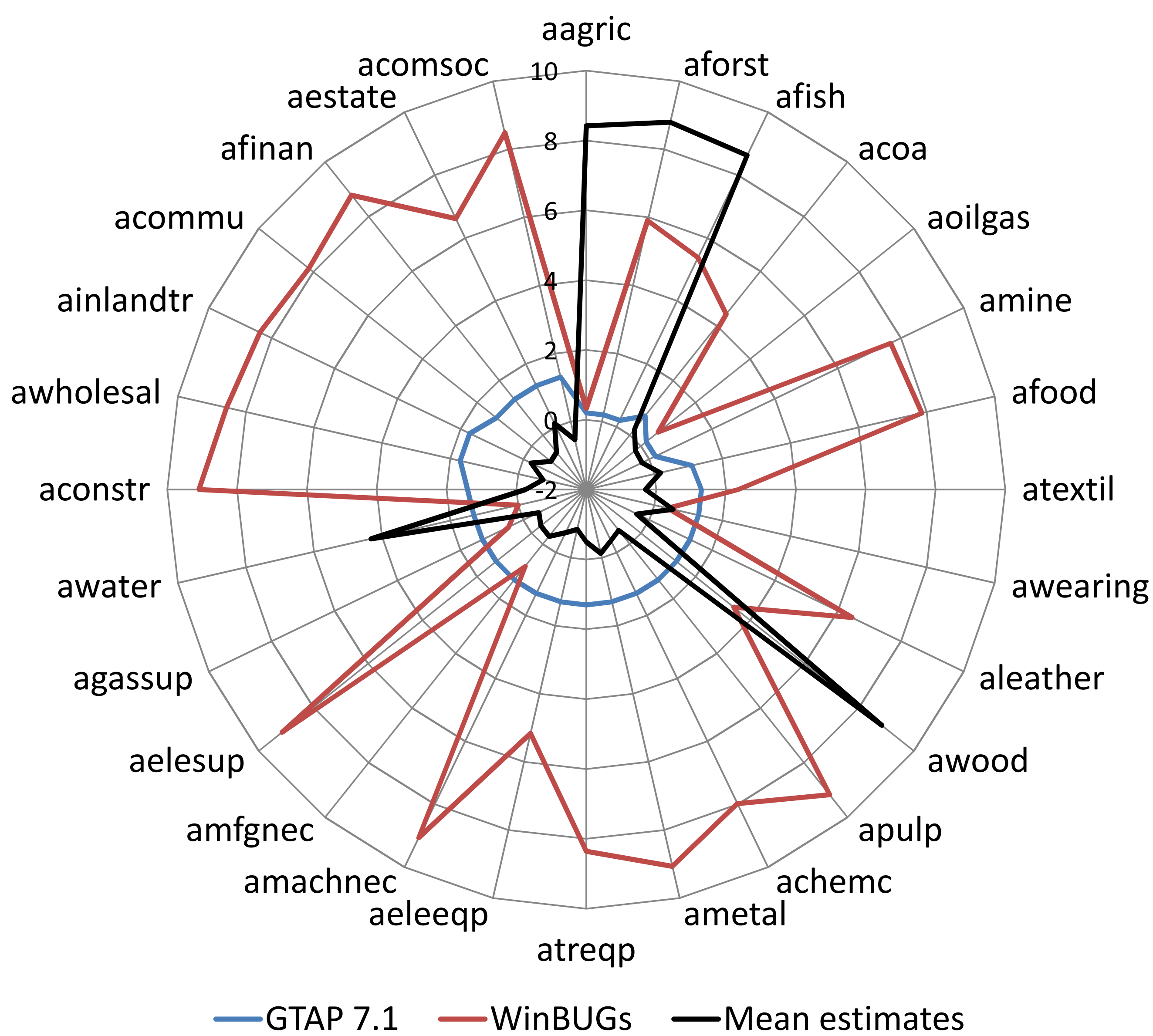



rhova('afood')
density
elasticity
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0.5
1.0
1.5

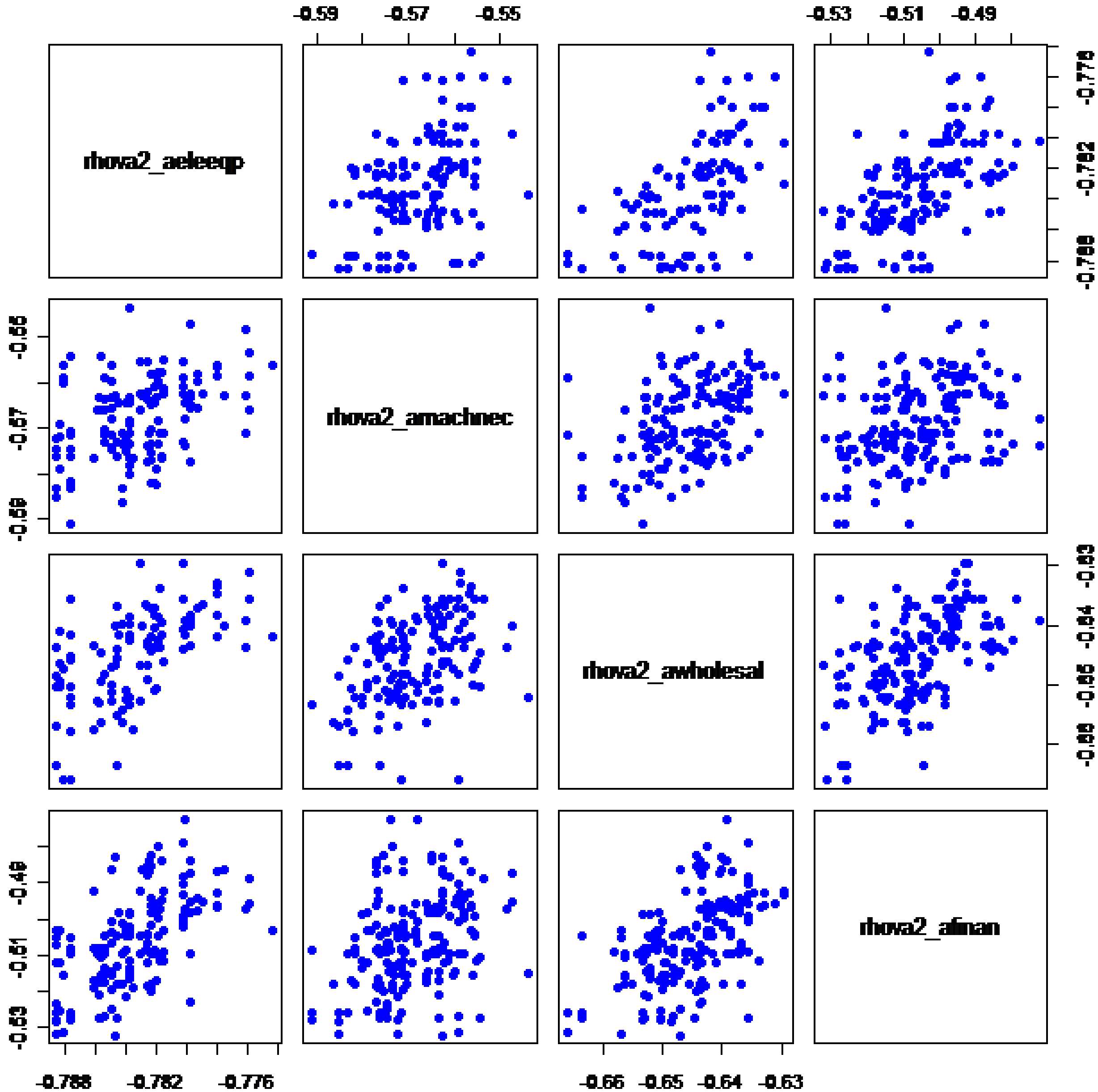

rhova('afood')
density
0
1
2
3
4
0.2
0.4
0.6
0.8
1.0
elasticity
rhova('aleather')
density
0
5
10
15
-0.20
-0.15
-0.10
-0.05
0.00
0.05
0.10
elasticity
rhova('atextil')
density
0
2
4
6
8
12
-0.30
-0.25
-0.20
-0.15
-0.10
elasticity
rhova('awood')
density
0.0
1.0
2.0
3.0
7.0
7.5
8.0
8.5
9.0
9.5
elasticity
rhova('awearing')
density
0
1
2
3
4
5
0.1
0.2
0.3
0.4
0.5
0.6
0.7
elasticity
rhova('apulp')
density
0
5
10
15
20
-0.50
-0.45
-0.40
-0.35
elasticity

	GTAP 7.1	2 equations econometric (WinBUGs)				GTAP+BAYES			
		mean	2.5%	97.5%	compare	mean	2.5%	97.5%	compare
aagric	0.2	0.33	0.28	0.39	<	8.42	5.96	9.50	<
aforst	0.2	5.91	1.03	9.76	<	8.79	6.99	9.50	<
afish	0.2	5.38	0.70	9.76	<	8.63	6.22	9.50	<
acoa	0.7	4.43	1.96	9.00	<	0.21	-0.06	0.54	>
aoilgas	0.2	0.64	0.56	0.72	<	-0.20	-0.29	-0.11	>
amine	0.2	7.67	3.50	9.91	<	-0.23	-0.33	-0.14	>
afood	1.1	7.86	3.85	9.93	<	0.18	-0.09	0.45	>
atextil	1.3	2.35	2.12	2.61	<	-0.31	-0.37	-0.24	>
awearing	1.3	0.40	0.36	0.45	>	0.55	0.28	0.93	>
aleather	1.3	6.44	1.56	9.88	<	-0.40	-0.46	-0.34	>
awood	1.3	3.40	3.12	3.72	<	8.83	7.25	9.50	<
apulp	1.3	9.18	7.22	9.98	<	-0.51	-0.56	-0.47	>
achemc	1.3	7.98	3.75	9.93	<	-0.37	-0.44	-0.31	>
ametal	1.3	9.07	7.02	9.97	<	-0.13	-0.25	0.08	>
atreqp	1.3	8.36	6.05	9.91	<	-0.51	-0.55	-0.47	>
aeleeqp	1.3	5.16	2.48	9.15	<	-0.84	-0.84	-0.83	>
amachnec	1.3	9.06	7.04	9.97	<	-0.63	-0.65	-0.60	>
amfgnec	1.3	0.81	0.76	0.86	>	-0.29	-0.37	-0.21	>
aelesup	1.3	9.14	7.40	9.97	<	-0.34	-0.41	-0.25	>
agassup	1.3	0.47	0.27	0.92	>	-0.49	-0.53	-0.44	>
awater	1.3	-	-	-	-	4.32	-0.50	9.41	OK
aconstr	1.4	9.09	6.96	9.98	<	-0.26	-0.34	-0.16	>
awholesal	1.7	8.56	5.58	9.95	<	-0.73	-0.74	-0.71	>
ainlandtr	1.7	8.38	5.17	9.94	<	-0.24	-0.33	-0.13	>
acommu	1.3	8.16	4.40	9.94	<	-0.71	-0.74	-0.68	>
afinan	1.3	8.79	6.17	9.96	<	-0.64	-0.66	-0.61	>
aestate	1.3	6.62	2.03	9.87	<	0.10	-0.10	0.47	>
acomsoc	1.3	8.49	5.37	9.96	<	-0.53	-0.57	-0.50	>

ρ estimates


aagric
aforst
afish
acoa
aoilgas
amine
afood
atextil
awearing
aleather
awood
apulp
achemc
ametal
atreqp
aeleeqp
amachnec
amfgnec
aelesup
agassup
awater
aconstr
awholesal
ainlandtr
acommu
afinan
aestate
acomsoc
10
8
6
4
2
0
-2
GTAP 7.1
WinBUGs
Mean estimates

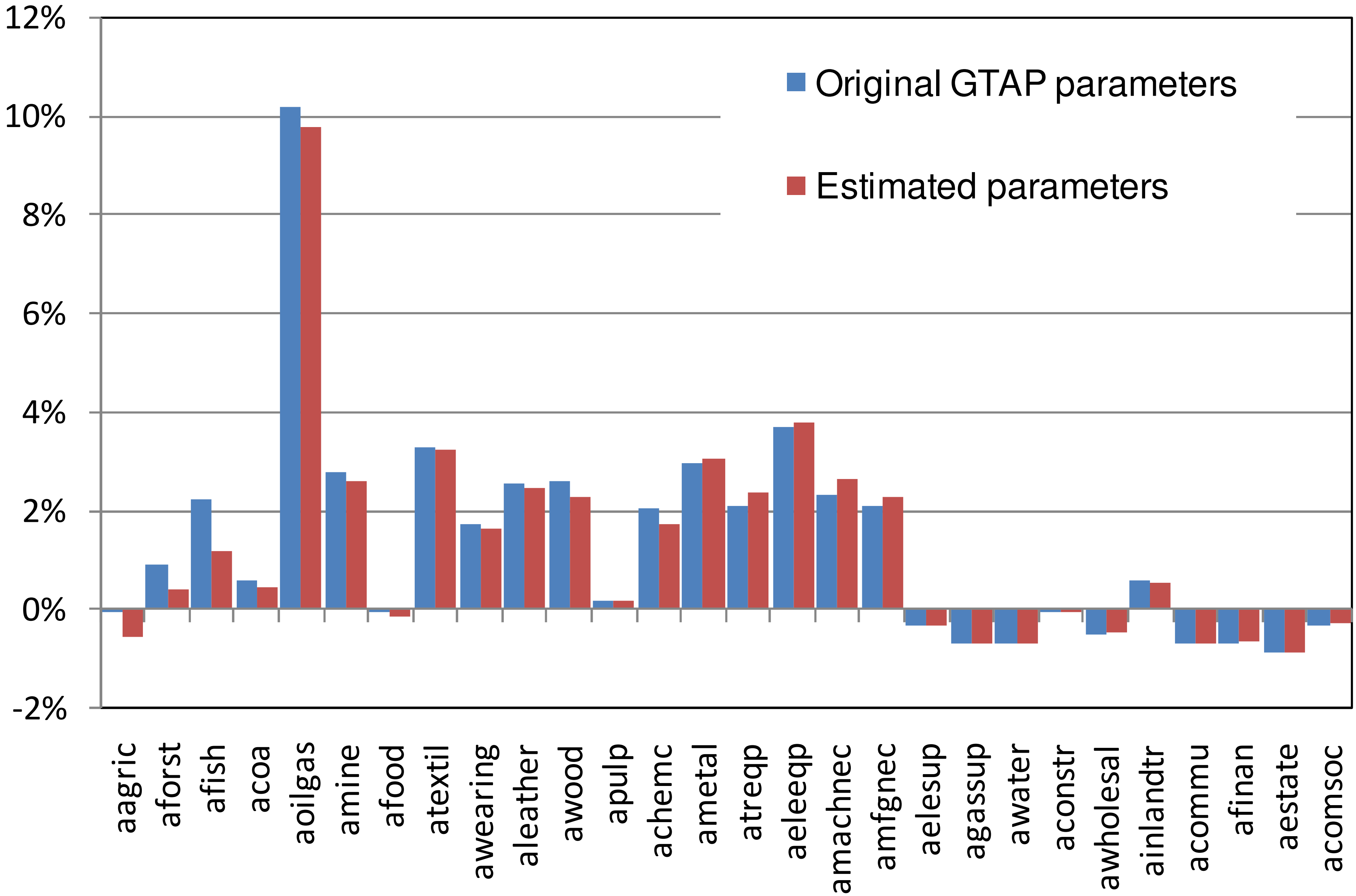
Correlations of the estimated parameters



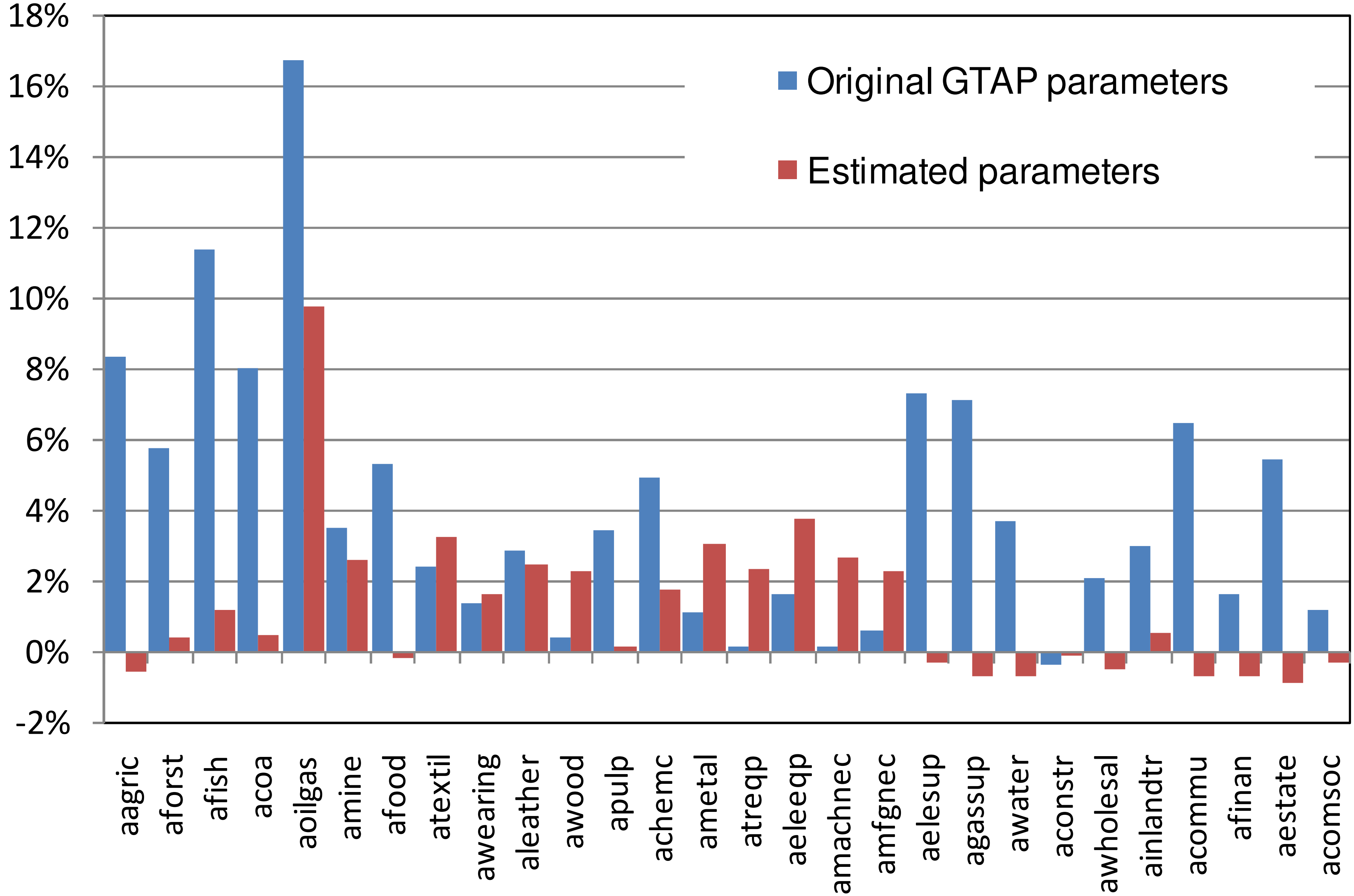
Does it makes any difference?

Comparison of experiments with original GTAP parameters and estimated on EU-KLEMS data

Experiment 1: 10% $tx_{a,r}$ increase



Experiment 2: 10% $FS_{cap,r}$ increase



Concluding remarks (1/3)

- Elasticity parameters are crucial and their validation is always on demand
- The proposed Bayesian approach has several features:
  - It combines information from observed data, our believes, and the model in consistent way
  - Very flexible in data requirements: almost any data or information can be naturally taking into account with the approach
  - The samples from joint distribution of parameters might be used for sensitivity analysis
  - It works (!), i.e. produces “reasonable” results

Concluding remarks (2/3)

- Caveats:
  - Estimated parameters depend on:
    - Data (f.i. number of observable variables)
    - Number and set of estimated parameters
    - Benchmark SAM (2004 in our case is fixed for all years)
    - Prior information for the estimation
  - Estimated parameters are historical
    - Future parameters are not obligatory equal to historical parameters

Concluding remarks (2/3)

- Caveats:
  - Estimated parameters depend on:
    - Data (f.i. number of observable variables)
    - Number and set of estimated parameters
    - Benchmark SAM (2004 in our case is fixed for all years)
    - Prior information for the estimation
  - Estimated parameters are historical
    - Future parameters are not obligatory equal to historical parameters

Concluding remarks (3/3)

- Further steps:
  - Extend number of observed variables and estimated parameters
    - Cover more countries
    - Add trade data
    - Estimate CET/Armington parameters
    - Fuzzy SAM/IO coefficients
  - Try other MCMC algorithms
    - Gibbs sampler
    - Adaptive MCMC algorithms
  - Try open source solvers

Thank you for your attention!

Comments, Notes, Critiques are welcome

olugovoy[at]hotmail[dot]com
potashnikov.vu[at]gmail[dot]com