Spatial Stochastic Frontier Models

On spatial dependence in stochastic frontier models with an application to European regional economic performance

Thomas de Graaff

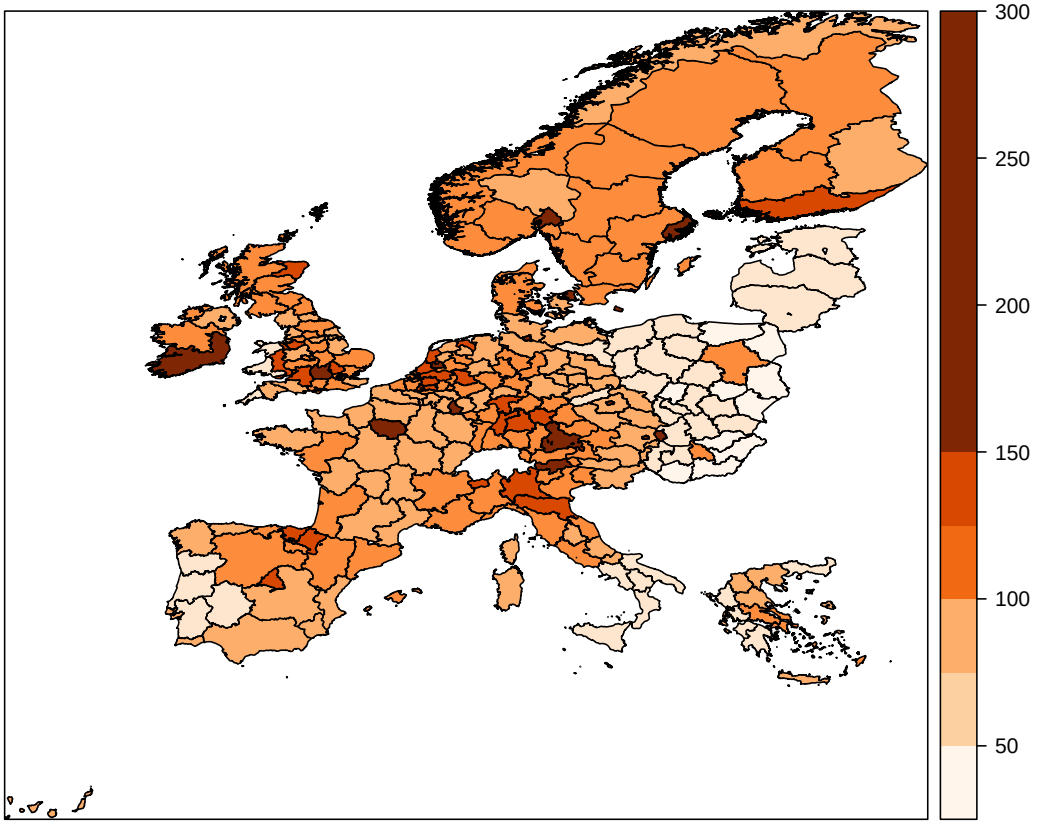
Department of spatial economics, Vrije Universiteit Amsterdam
http://www.thomasdegraaff.nl/

Package development at
https://github.com/Thdegraaff/SpatFrontier/

May 25, 2018

Regional economic performance

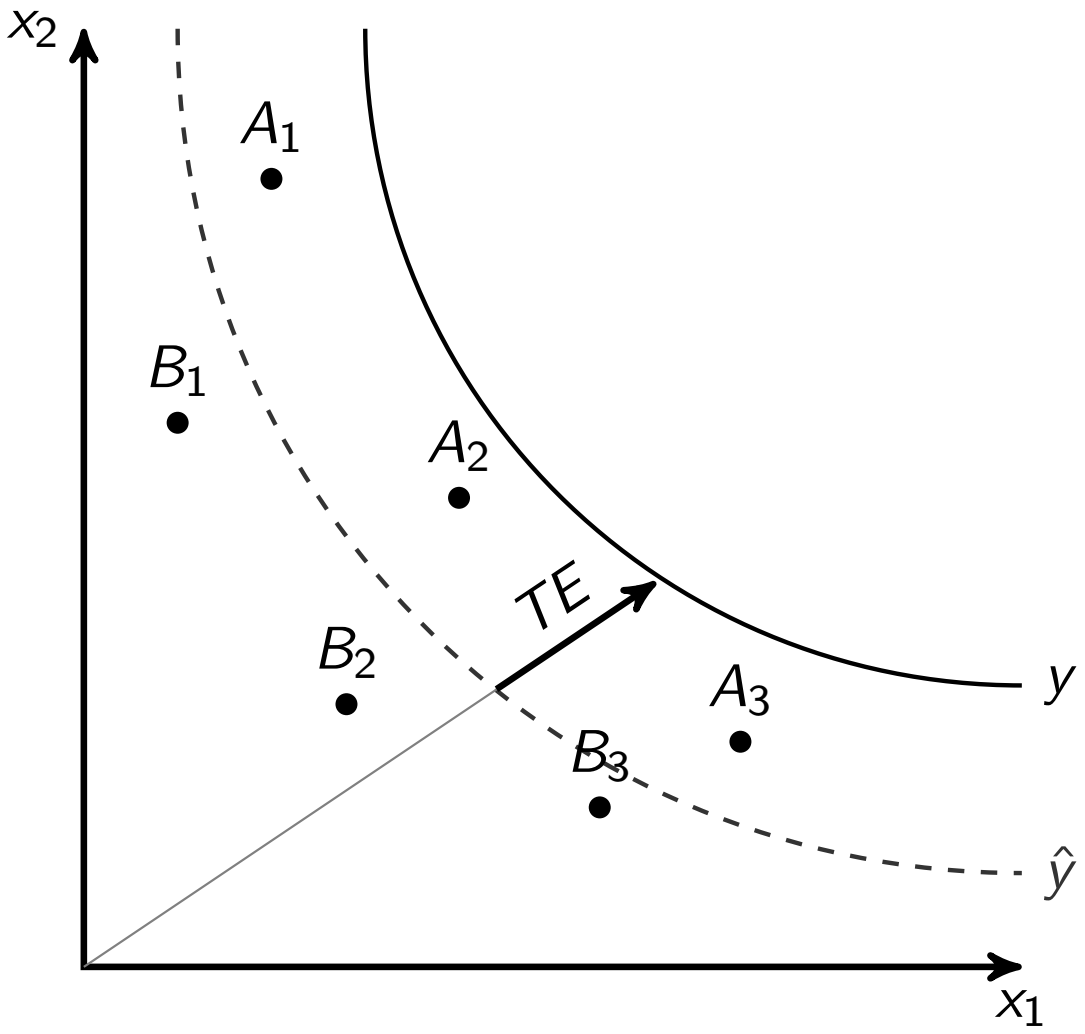
GDP per capita across Europe in 2007—EU27 = 100 (source: cambridge econometrics database



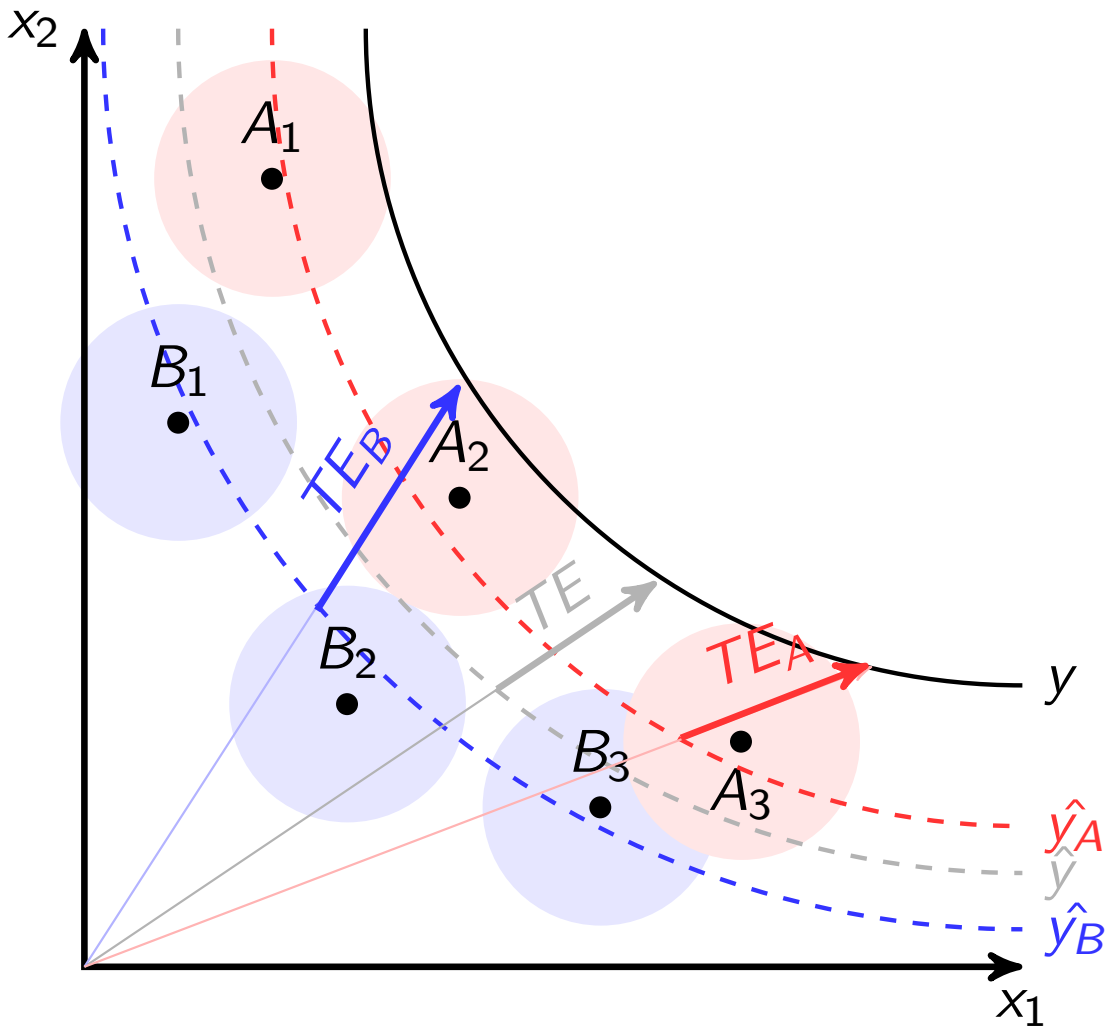
Huge differences between and within countries

- Different use of endowments
- Different relative location within the network

Production function isoquant



Production function isoquant with spatial dependence



Related literature

- Starting point: Weinstein (Technometrics, 1964)
  - Econometric literature: Aigner et al. (1977); Meeusen & van den Broeck (1977).
  - Statistical literature: Azzalini (1985); Azzalini & Capitanio (1999); Dominguez-Molina et al. (2003)

Related literature

- Starting point: Weinstein (Technometrics, 1964)
  - Econometric literature: Aigner et al. (1977); Meeusen & van den Broeck (1977).
  - Statistical literature: Azzalini (1985); Azzalini & Capitanio (1999); Dominguez-Molina et al. (2003)
- Regional stochastic frontiers: Puig-Junoy (2001); Alvarez (2007)

Related literature

- Starting point: Weinstein (Technometrics, 1964)
  - Econometric literature: Aigner et al. (1977); Meeusen & van den Broeck (1977).
  - Statistical literature: Azzalini (1985); Azzalini & Capitanio (1999); Dominguez-Molina et al. (2003)
- Regional stochastic frontiers: Puig-Junoy (2001); Alvarez (2007)
- Recent literature concerning spatial stochastic frontiers:
  - Parametric approach: Pavlyuk (2010); Barrios & Lavado (2010); Fusco & Vidoli (2013); Kinfu & Sawhey (2015); Glass et al. (2016); Jiang et al. (2017)
  - Bayesian approach/sampling: Schmidt et al. (2009); Areal et al. (2012); Tsionas & Michealides (2016)

Related literature

- Starting point: Weinstein (Technometrics, 1964)
  - Econometric literature: Aigner et al. (1977); Meeusen & van den Broeck (1977).
  - Statistical literature: Azzalini (1985); Azzalini & Capitanio (1999); Dominguez-Molina et al. (2003)
- Regional stochastic frontiers: Puig-Junoy (2001); Alvarez (2007)
- Recent literature concerning spatial stochastic frontiers:
  - Parametric approach: Pavlyuk (2010); Barrios & Lavado (2010); Fusco & Vidoli (2013); Kinfu & Sawhey (2015); Glass et al. (2016); Jiang et al. (2017)
  - Bayesian approach/sampling: Schmidt et al. (2009); Areal et al. (2012); Tsionas & Michealides (2016)
- Two `R`-packages:
  - `ssfa`—Fusco & Vidoli (2015)
  - `spfrontier`—Dmitry Pavlyuk (2016)

Stochastic production frontiers

In general, regional production, y, can be modeled as:

$$y = f(\mathbf{X}; \beta) TE,$$

Assume Cobb-Douglas and $TE = \exp(-\mu)$, then:

$$\ln y = \ln(\mathbf{X})\beta \underbrace{- \mu + \nu}_{\epsilon = \text{Observed residuals}},$$

Where usually:

- $\mu \sim N^{+}\left(0, \sigma_{\mu}^{2}\right)$ so $\mu \geq 0$
- $\nu \sim N\left(0, \sigma_{\nu}^{2}\right)$

and ν and μ are independent of each other

Tradional (econometric) approach

For likelihood purposes one usually considers the composite stochastic variable $\epsilon = \nu - \mu$ en then find the marginal density of ϵ:

$$\begin{aligned} f(\epsilon) &= \int_0^\infty f(\mu, \epsilon) d\mu \\ &= \frac{2}{\sigma} \phi\left(\frac{\epsilon}{\sigma}\right) \Phi\left(-\frac{\epsilon\lambda}{\sigma}\right) \end{aligned}$$

Technical efficiencies can be retrieved by $f(\mu|\epsilon)$.

Imposing spatial structure

A typology

Note that $\mathbf{A} = (\mathbf{I} - \lambda \mathbf{W})$ and $\mathbf{B} = (\mathbf{I} - \rho \mathbf{W})$

1. Spatial dependence in technical efficiency component μ

$$\ln y = \ln(\mathbf{X})\beta \underbrace{-\mathbf{A}^{-1}\mu + \nu}_{\epsilon}$$

2. Spatial dependence in error term ν:

$$\ln y = \ln(\mathbf{X})\beta \underbrace{-u + \mathbf{A}^{-1}\nu}_{\epsilon}$$

3. Spatial lag separate from technical efficiency component μ

$$\ln y = \mathbf{B}^{-1}\left[\ln(\mathbf{X})\beta\right] \underbrace{+\mathbf{B}^{-1}\nu - \mu}_{\epsilon}$$

4. Spatial lag

$$\ln y = \mathbf{B}^{-1}\left[\ln(\mathbf{X})\beta\right] + \underbrace{\mathbf{B}^{-1}\left[-\mu + \nu\right]}_{\epsilon}$$

Alternative statistical approach

- Alternative specification, skew normal distribution:

$$Z \sim SN(\alpha) = 2\phi(\epsilon)\Phi(\alpha\epsilon)$$

Alternative statistical approach

- Alternative specification, skew normal distribution:

$$Z \sim SN(\alpha) = 2\phi(\epsilon)\Phi(\alpha\epsilon)$$

- Which has two types of genesis:
  1. By convolution: $Z = \delta|U| + \sqrt{1-\delta^2}V$
  2. By conditioning: $Z = (V|U > 0)$

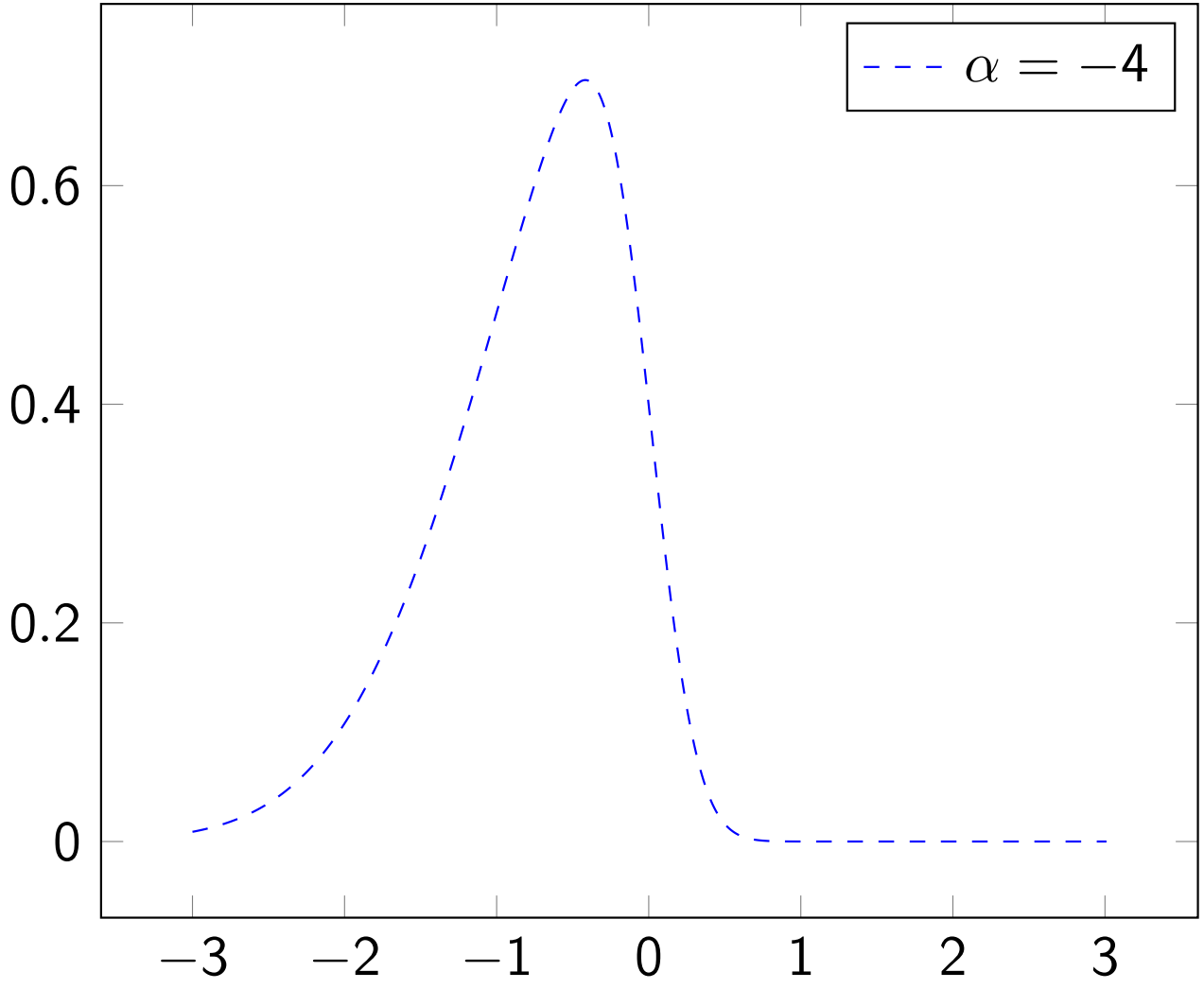
Alternative statistical approach

- Alternative specification, skew normal distribution:

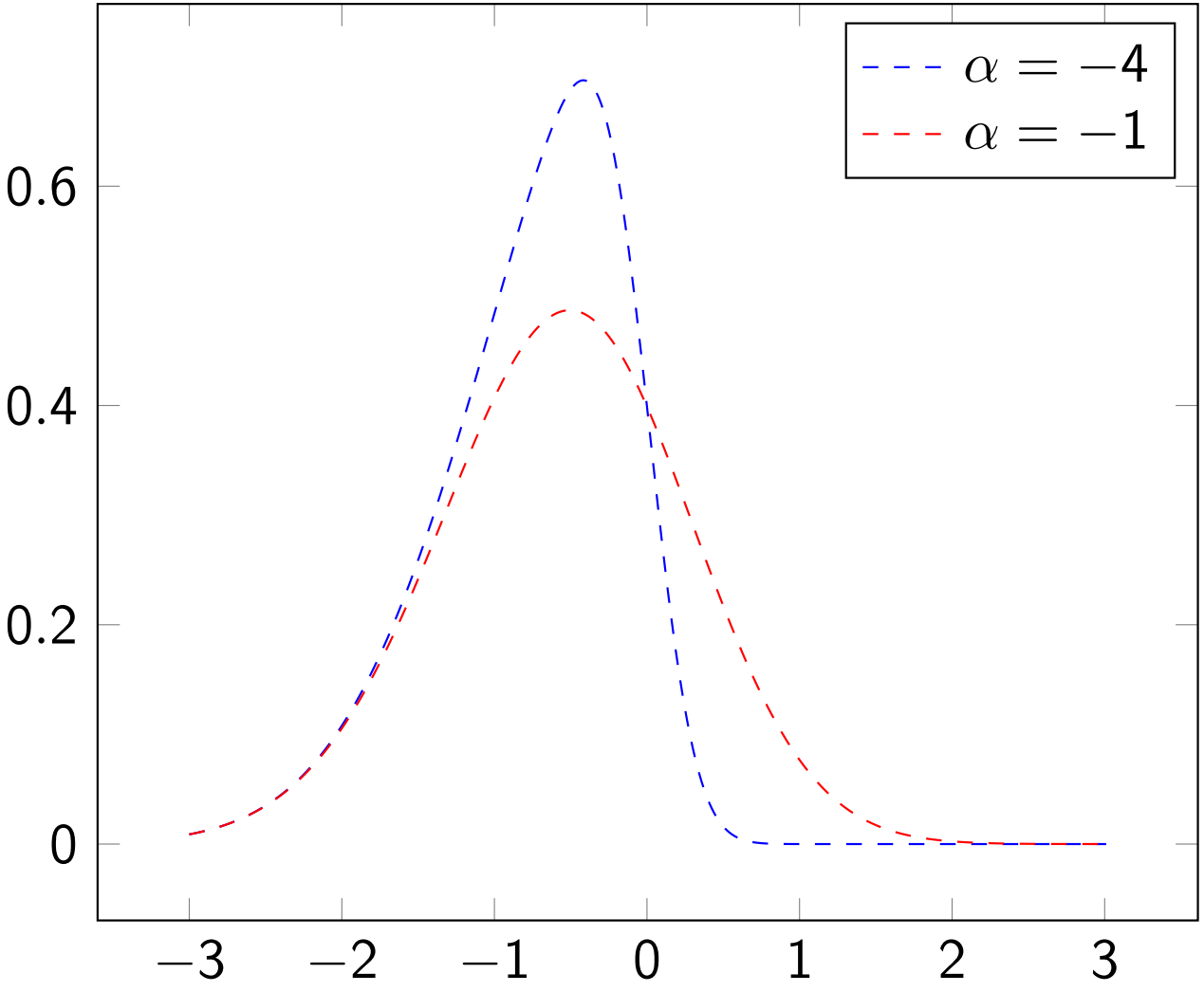
$$Z \sim SN(\alpha) = 2\phi(\epsilon)\Phi(\alpha\epsilon)$$

- Which has two types of genesis:
  1. By convolution: $Z = \delta|U| + \sqrt{1-\delta^2}V$
  2. By conditioning: $Z = (V|U > 0)$
- And is closed under affine transformations:
  - if $Y = \mathbf{X}\beta + \sigma Z$ then $Y \sim SN(\mathbf{X}\beta, \sigma, \alpha)$

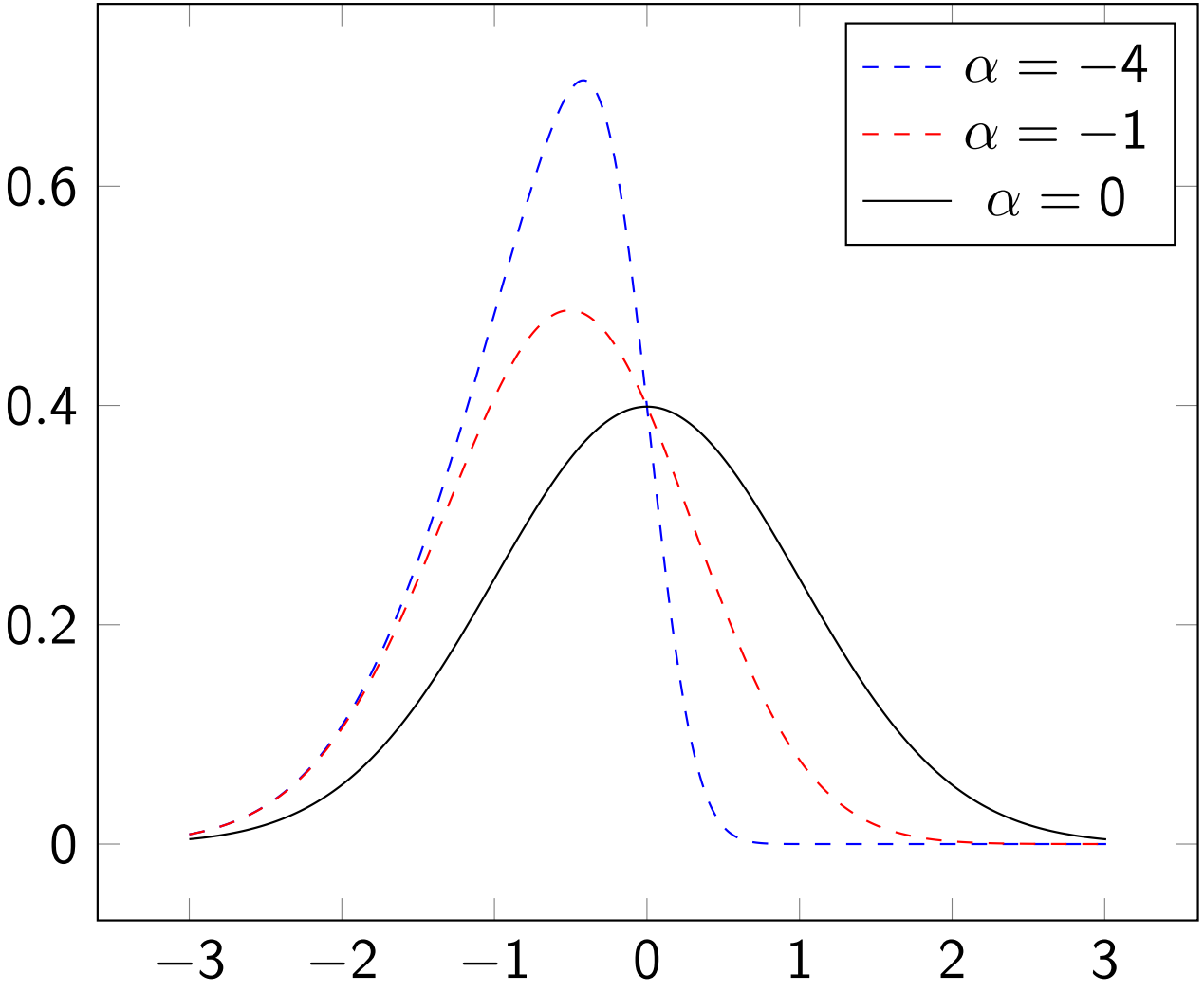
Varying the shape parameter α



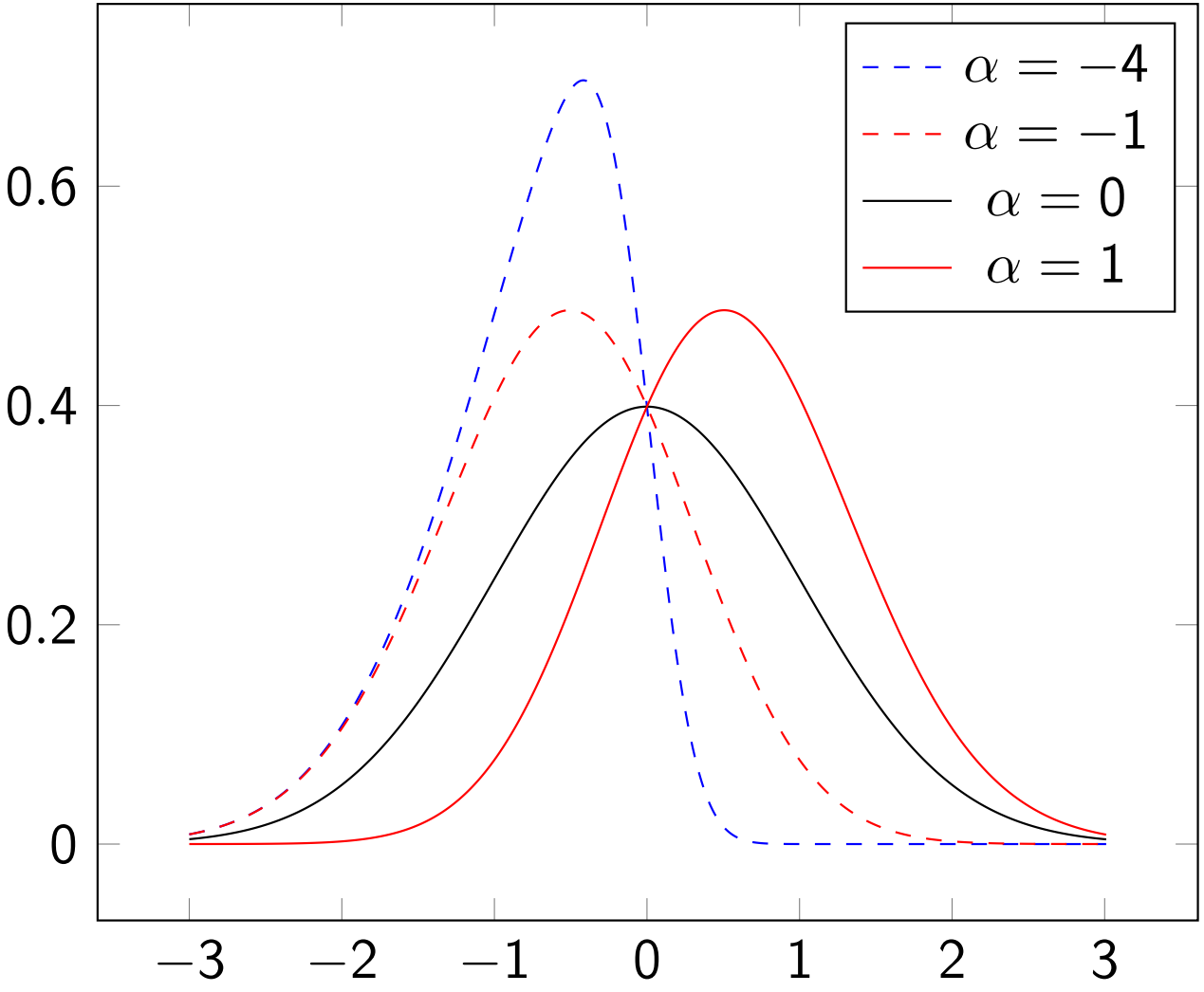
Varying the shape parameter α



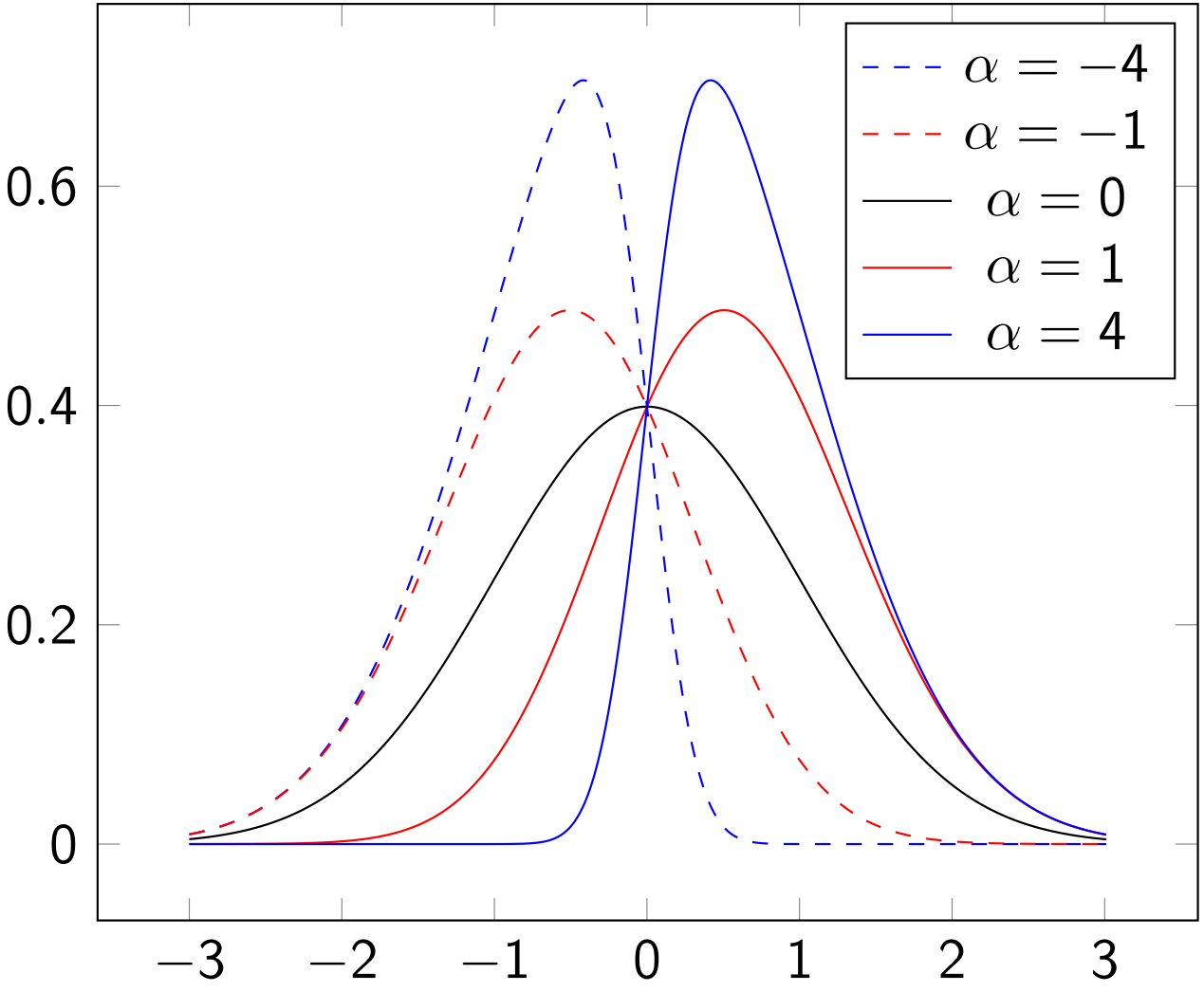
Varying the shape parameter α



Varying the shape parameter α



Varying the shape parameter α



Multivariate skew-normal distributions

Via conditioning

If we partition *Psi* in component U and V of size 1 and n respectively, such that:

$$\Psi = \begin{bmatrix} U \\ V \end{bmatrix} \sim N_{1+n}(0, \Sigma^*), \qquad \Sigma^* = \begin{bmatrix} 1 & \delta^t \\ \delta & \Sigma \end{bmatrix}$$

then:

$$Z = (V|U > 0) \sim SN_n(0, \Sigma, \alpha), \qquad \alpha = (1 - \delta^t \Sigma \delta)^{\frac{1}{2}} \sigma^{-1} \delta$$

Combining spatial structure with stochastic frontiers

For the spatial error model

If the model is: $\ln y = \ln(\mathbf{X})\beta \underbrace{-\mu + \mathbf{A}^{-1}\nu}_{\epsilon}$ then

$$\epsilon \sim SN(\ln y - \ln(\mathbf{X})\beta, \sigma^2(\mathbf{A}^t\mathbf{A})^{-1}, \alpha)$$

For the spatial "lag" model

If the model is: $\ln y = \mathbf{B}^{-1}\left[\ln(\mathbf{X})\beta\right] \underbrace{-\mu + \mathbf{B}^{-1}\nu}_{\epsilon}$ then:

$$\epsilon \sim SN(\mathbf{B}^{-1}(\ln y - \mathbf{X}\beta), \sigma^2(\mathbf{B}^t\mathbf{B})^{-1}, \alpha)$$

Simulation set-up

Simulation results

Regional production function data

- Cambridge Econometrics Dataset for Employment
- Regional value added and capital stock from regional supply and use framework (Thissen et al. 20013a, 2013b and 2013c)
  - Regional trade database between 256 European Nuts2 regions over 15 sectors (2000–2010)
  - No Perpetual Inventory Method but value added for capital $\rightarrow$ Using country specific interest rates we can then find capital stock ($VA^{capital} = rK$)
- Sample period: mean over 2000–2010
- Spatial weights matrix $\mathbf{W}$: 4 nearest neighbours

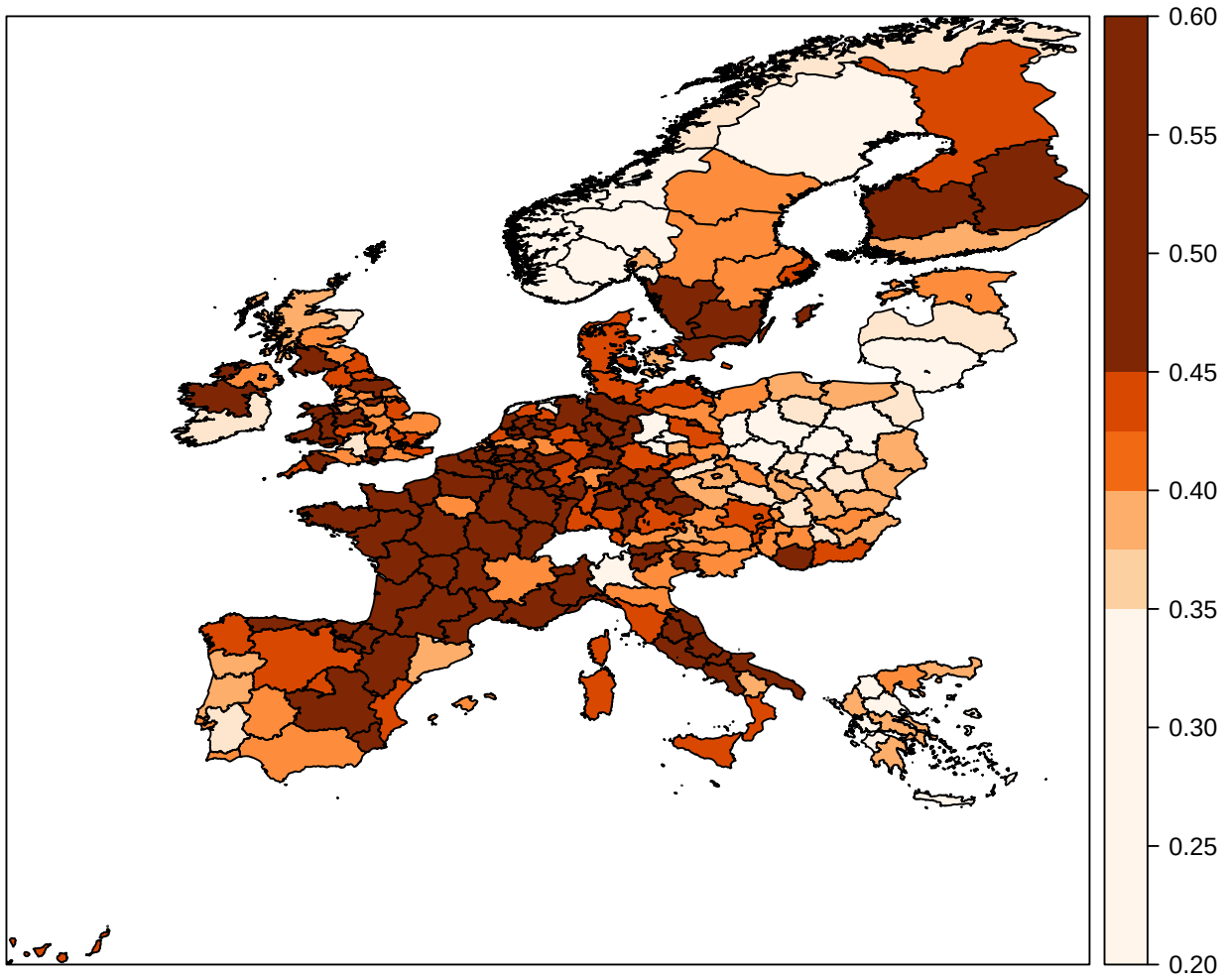
Results for energy and manufacturing sector

	OLS	SEM	SAR	Frontier	SEM fr.	SAR fr.
Constant	2.40^{***}	3.16^{***}	0.46	2.50^{***}	4.03^{***}	0.49
	(0.16)	(0.14)	(0.26)	(0.18)	(0.25)	(1.22)
ln(Capital)	0.77^{***}	0.58^{***}	0.70^{***}	0.76^{***}	0.58^{***}	0.70^{***}
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
ln(Employment)	0.30^{***}	0.44^{***}	0.32^{***}	0.30^{***}	0.44^{***}	0.32^{***}
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
σ	0.33^{***}	0.22^{***}	0.29^{***}	0.33^{***}	0.29^{***}	0.29^{*}
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.12)
λ		0.79^{***}			0.78^{***}	
		(0.09)			(0.09)	
ρ			0.25^{***}			0.25^{***}
			(0.03)			(0.03)
α				-1.75^{***}	-1.41^{***}	-0.13
				(0.53)	(0.37)	(5.09)
Log Lik.	-80.10	92.28	41.45	11.03	94.86	41.45

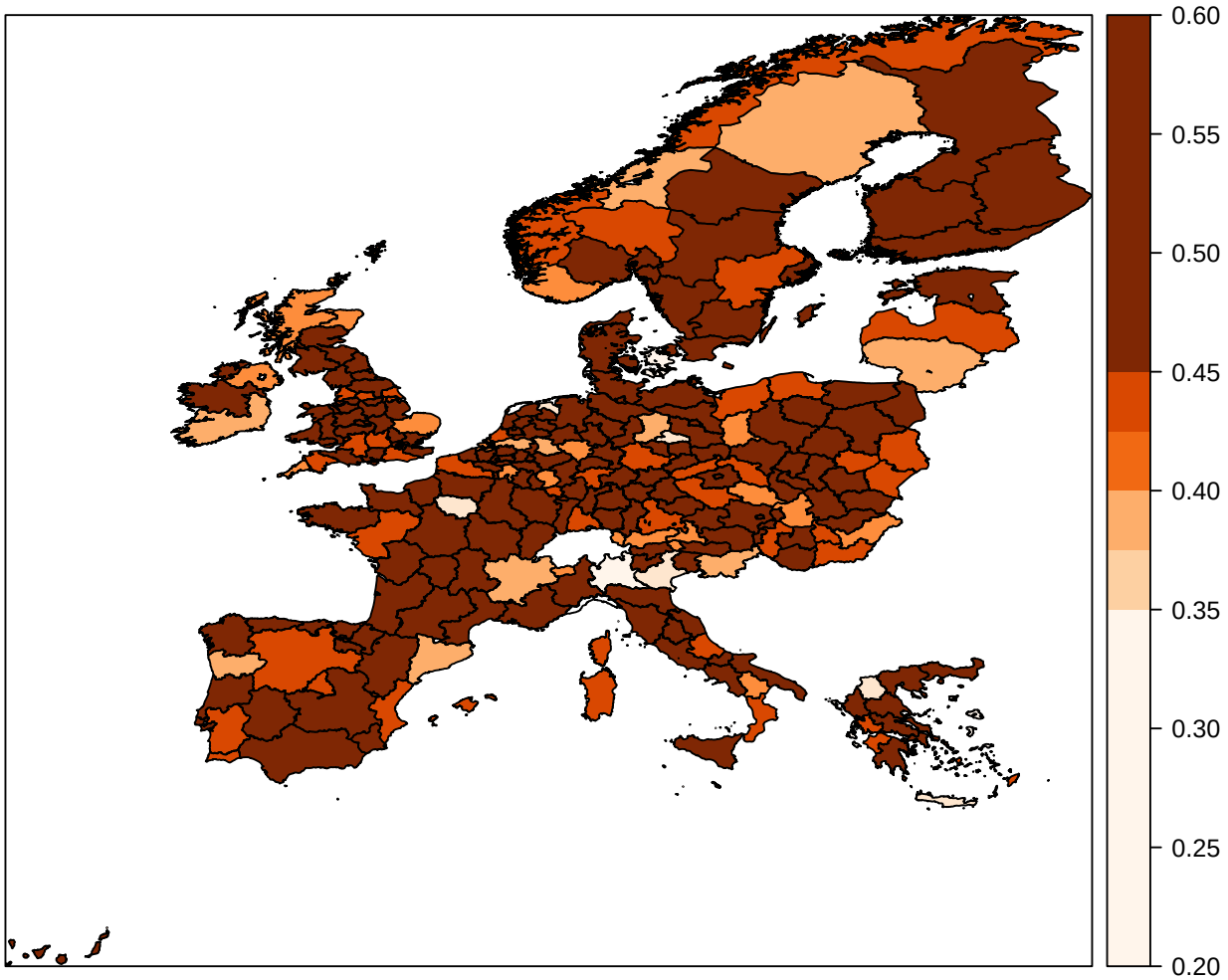
Results all sectors for frontier model with spatial error

	Agric.	E & M	Constr.	Distri.	Serv.	NM. Serv
Constant	2.43^{***}	4.03^{***}	2.11^{***}	1.24^{***}	2.05^{***}	1.74^{***}
	(0.71)	(0.25)	(0.22)	(0.17)	(0.29)	(0.19)
ln(Capital)	0.68^{***}	0.58^{***}	0.61^{***}	0.69^{***}	0.71^{***}	0.61^{***}
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
ln(Employment)	0.21^{***}	0.44^{***}	0.39^{***}	0.30^{***}	0.29^{***}	0.39^{***}
	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)
σ	0.26^{***}	0.29^{***}	0.28^{***}	0.24^{***}	0.17^{***}	0.25^{***}
	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
α	-0.03	-1.41^{***}	-1.19^{***}	-1.54^{***}	-0.02	-1.68^{***}
	(1.44)	(0.37)	(0.29)	(0.34)	(0.75)	(0.32)
λ	0.57^{***}	0.78^{***}	0.76^{***}	0.78^{***}	0.59^{***}	0.81^{***}
	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)
Log Lik.	62.92	94.86	94.19	149.78	164.88	143.52

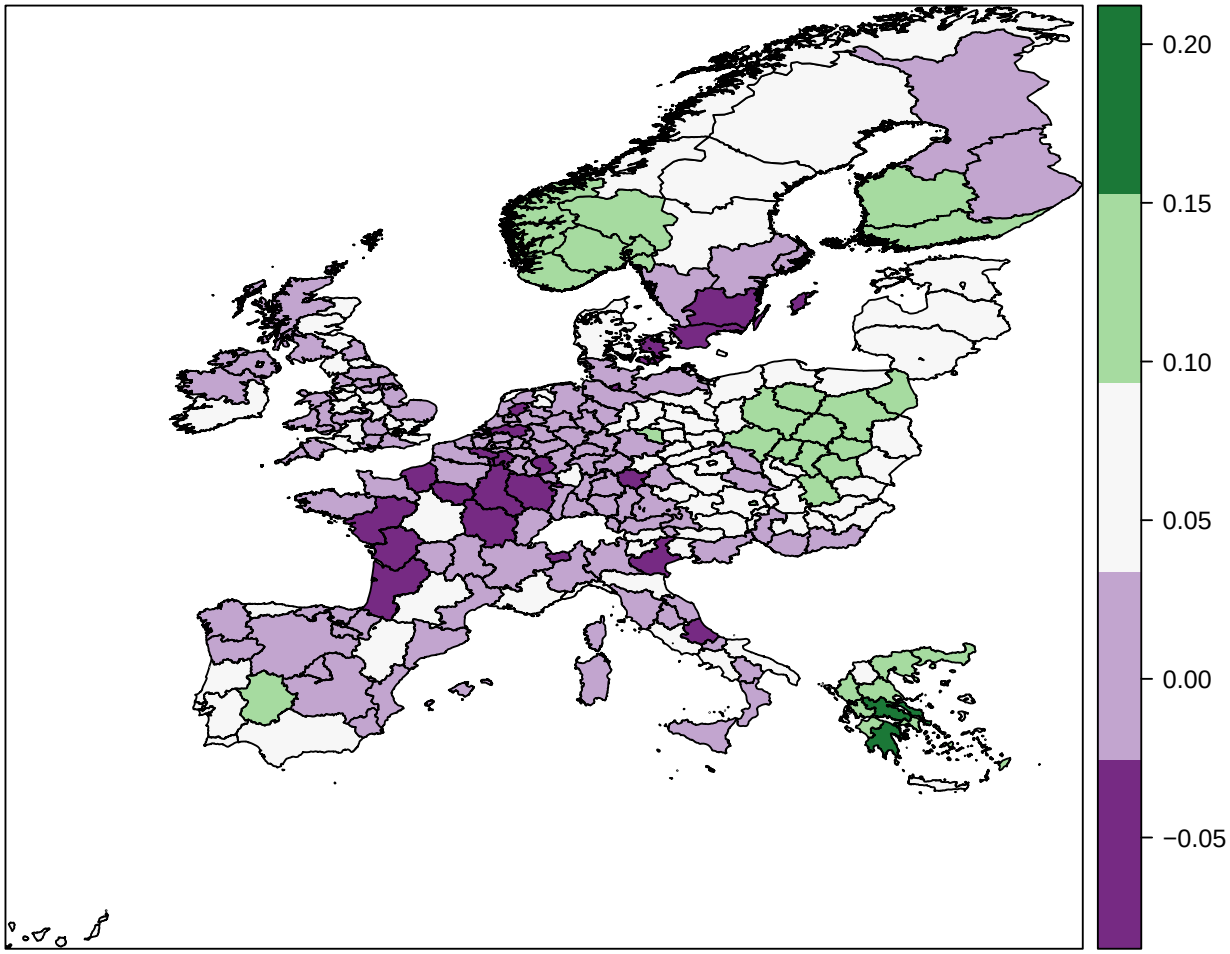
Regional efficiency manufacturing



Regional efficiency manufacturing controlling for distance



Difference frontier vs. spatial error frontier for E & M



In conclusion & further research

- Introducing spatial structure in stochastic frontier models is possible and for some models relatively straightforward, but
- Multivariate distribution of error term should be explicitly taken into account
- Extension:

$$\Psi \sim \begin{pmatrix} U \\ V \end{pmatrix} \sim N_{d+n}(0, \Omega^*), \qquad \Omega^* = \begin{pmatrix} \Phi & \Delta' \\ \Delta & \Omega \end{pmatrix}.$$

But then $Z = (V|U > 0)$ becomes more complex and estimation requires Bayesian estimation in combination with simulation techniques