Preliminaries

Fu Zixuan[1]

[1]Last compiled on May 17, 2024

Contents

Chapter 1

Year 2016-2022

1.1 Regression tables

1.1.1 WG, BG, Random effect, Correalted random effect(Mundlak)

Notation

- Dependent variable:

$$\underbrace{y}_{NT\times 1} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \quad \underbrace{y_i}_{T\times 1} = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}$$

- Independent variable:

$$\underbrace{X}_{NT\times K} = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \quad \underbrace{x_i}_{T\times K} = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{iT} \end{pmatrix}$$

- Matrix to calcualte the mean:

$$B_T = d_T(d_T'd_T)^{-1}d_T' \quad \text{where} \quad d_T = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

  Thus

$$B_T y_1 = \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_1 \end{pmatrix}$$

  we set $B = I_N \otimes B_T$

- Matrix to demean the variable:

$$W_T = I_T - B_T$$

  Thus,

$$W_T y_i = \begin{pmatrix} y_{i1} - \bar{y}_1 \\ \vdots \\ y_{iT} - \bar{y}_1 \end{pmatrix}$$

  we set $W = I_N \otimes W_T$

Estimators

- WG:

$$\hat{\beta}_{WG} = (X'WX)^{-1}X'Wy$$

- BG:

$$\hat{\beta}_{BG} = (X'BX)^{-1}X'By$$

- RE: A linear combination of WG and BG
- CRE (Mundlak): equivalent to WG

> In empirical analysis of data consisting of repeated observations on economic units (time series on a cross section) it is often assumed that the coefficients of the quantitiative variables (slopes) are the same, whereas the coefficients of the qualitative variables (intercepts or effects) vary over units or periods. This is the constant-slope variable- intercept framework. In such an analysis an explicit account should be taken of the statistical dependence that exists between the quantitative variables and the effects. It is shown that when this is done, the random effect approach and the fixed effect approach yield the same estimate for the slopes, the "within" estimate. Any matrix combination of the "within" and "between" estimates is generally biased. When the "within" estimate is subject to a relatively large error a minimum mean square error can be applied, as is generally done in regression analysis. Such an estimator is developed here from a somewhat different point of departure.

1.1.2 Specification

Model 1 (Pseudoc Poisson)

$$\begin{aligned} y_{it} &= x_{it1}^{\beta_1} x_{it2}^{\beta_2} x_{it3}^{\beta_3} \theta_i \epsilon_{it} \\ &= f(x_{it};\beta)\theta_i\epsilon_{it} \end{aligned}$$

Tables: Pois

Reference to Koen Jochman's lecture notes on panel data and model with multiplicative effect:

$$y_{it} = f(x_{it};\beta)\theta_i\epsilon_{it} \quad \text{where} \quad \mathbb{E}\left[\epsilon_{it} | x_{i1}, \ldots, x_{iT}, \theta_i\right] = 1$$

Then following the same logic in additive effect, we difference out individual effect to get:

$$\mathbb{E}\left[\frac{y_{it}}{f(x_{it};\beta)} - \frac{y_{i,t-1}}{f(x_{i,t-1};\beta)} \middle| x_{i1}, \ldots, x_{iT}\right] = 0$$

Similarly,

$$\mathbb{E}\left[\frac{y_{it}}{f(x_{it};\beta)} - \frac{\sum_{t=1} y_{it}}{\sum_{t=1} f(x_{it};\beta)} \middle| x_{i1}, \ldots, x_{iT}\right] = 0$$

One of the unconditional moment equation given rise to is

$$\mathbb{E}\left[x_i t \left(\frac{y_{it}}{f(x_{it};\beta)} - \frac{\sum_{t=1} y_{it}}{\sum_{t=1} f(x_{it};\beta)}\right)\right] = 0 \tag{1.1}$$

The moment conditon is often called pseudo-poisson estimator (as if assuming $y_it|x_{i1},\dots,x_{iT},\theta_i \sim Poisson(f(x_{it};\beta)\theta_i)$ and then use maximum likelihood.) When regressors are not strictly exogenous, we can construct a **differencing** based estimator based on sequential moment restrictions. The `fixest` package doesn't provide the differencing based estimator.
TO BE CONSTRUCTED BY HAND LATER.

Model 2 (OLS)

$$\log(y_{it}) = \log(x_{it1})\beta_1 + \log(x_{it2})\beta_2 + \log(x_{it3})\beta_3 + \log(\theta_i) + \log(\epsilon_{it})$$

LHS: log(ETP INF);
RHS: log(STAC INPATIENT), log(STAC OUTPATIENT), log(SESSION),CASEMIX
Table: OLS, OLS_lag1
Figure: FixedEffect_OLS

Remark. The Pseudo poisson and Log OLS are equivalent if we assume that ϵ_{it} is independent of x_{it} and θ_i.

Remark. The Mundlak (1978) approach is to include the average of the individual-specific variables $\bar{x_i}$ in the regression. He shows that it is equivalent to within group estimator. (Correlated random effect $\sim$ WG estimator). If the true model is

$$y_{it} = x_{it}\beta + \theta_i + \epsilon_{it}$$

and $E(\theta_i|\bar{x}_i) = \bar{x}_i\gamma + \tilde{\theta}_i$, then

$$y_{it} = (x_{it} - \bar{x}_i)\beta + \bar{x}_i(\gamma + \beta) + \tilde{\theta}_i + \epsilon_{it}(x_{it} - \bar{x}_i)\beta_1 + \bar{x}_i\beta_2 + \tilde{\theta}_i + \epsilon_{it}$$

Only when the θ_i is uncorrelated with x_{it}, the $\beta_1 = \beta_2$.

Pois

Dependent Variable:	ETP_INF	
Model:	(1)	(2)
Variables		
log(SEJHC_MCO)	0.158^a	0.727^a
	(0.028)	(0.026)
log(SEJHP_MCO)	0.044^a	0.107^a
	(0.011)	(0.019)
log(SEANCES_MED)	0.032^a	0.043^a
	(0.008)	(0.011)
CASEMIX	0.007^a	0.014^a
	(0.002)	(0.003)
Fixed-effects		
FI	Yes	
FI_EJ		Yes
Fit statistics		
Observations	3,928	3,928
Squared Correlation	0.995	0.983
Pseudo R^2	0.966	0.956

Signif. Codes: a: 0.01, b: 0.05, c: 0.1

Back

OLS

Dependent Variable:	log(ETP_INF)	
Model:	(1)	(2)
Variables		
log(SEJHC_MCO)	0.182^a	0.673^a
	(0.027)	(0.031)
log(SEJHP_MCO)	0.032^a	0.096^a
	(0.011)	(0.022)
log(SEANCES_MED)	0.021^a	0.037^a
	(0.006)	(0.010)
CASEMIX	0.007^a	0.013^a
	(0.003)	(0.004)
Fixed-effects		
FI	Yes	
FI_EJ		Yes
Fit statistics		
Observations	3,905	3,905
Squared Correlation	0.993	0.980
Pseudo R^2	1.77	1.39

Signif. Codes: a: 0.01, b: 0.05, c: 0.1

Back

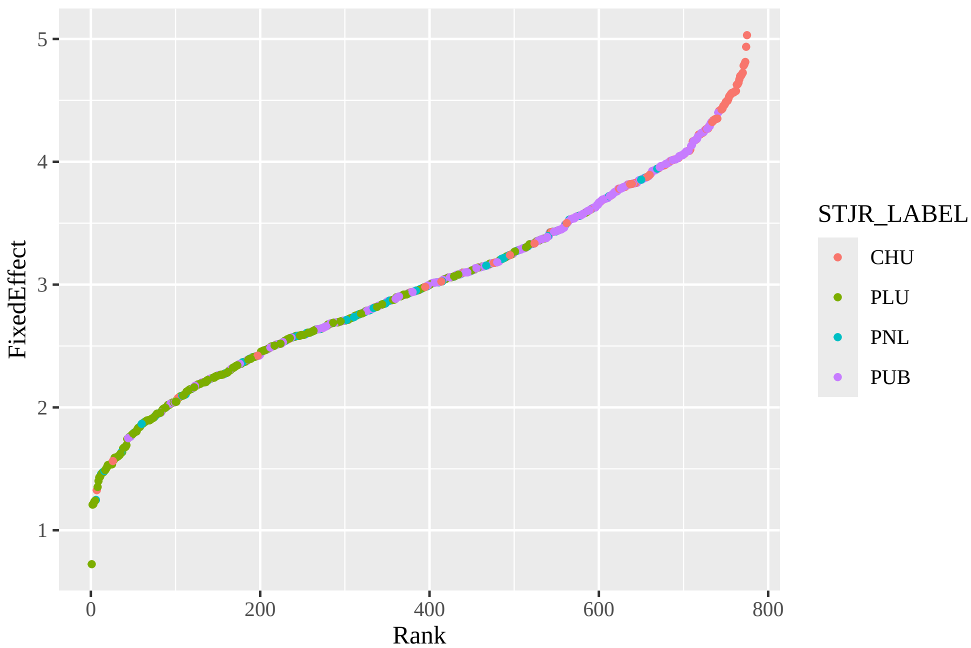
OLS_lag1

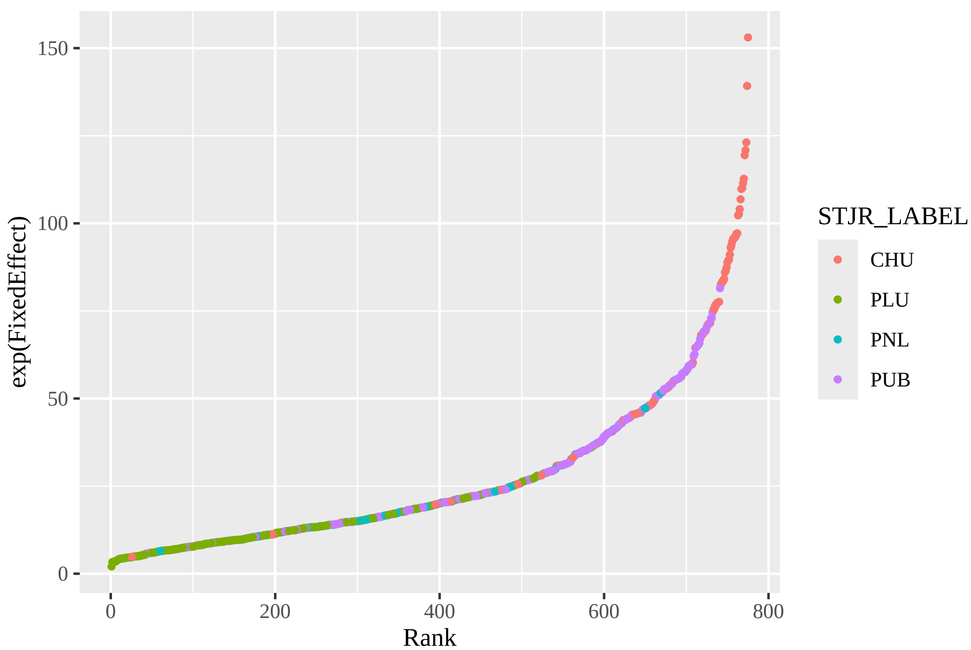
Dependent Variable:	log(ETP_INF)	
Model:	(1)	(2)
Variables		
log(SEJHC_MCO)	0.222^a	0.692^a
	(0.049)	(0.036)
log(SEJHP_MCO)	0.093^a	0.111^a
	(0.025)	(0.032)
log(SEANCES_MED)	0.040^c	0.047^a
	(0.023)	(0.015)
CASEMIX	0.013^a	0.014^a
	(0.003)	(0.005)
Fixed-effects		
FI	Yes	
FI_EJ		Yes
Fit statistics		
Observations	3,151	3,151
Squared Correlation	0.994	0.983
Pseudo R^2	1.84	1.48

Signif. Codes: a: 0.01, b: 0.05, c: 0.1

Back

FixedEffect_OLS Fixed effects extracted from OLS regression with fixed effect on each FI (not FI_EJ). Back





1.2 Mixture model

1.2.1 Location mixture

The model is the following

$$y_{it} = \theta_i + \epsilon_{it} \quad \text{where} \quad \epsilon_{it} \sim N(0,1)$$

Thus

$$\hat{\theta}_i = \frac{1}{T_i} \sum y_{it} \sim N(\theta_i, 1/T_i)$$

The likelihood function $L(F|y)$ (optimizing over distribution function F given the observed y) is

$$L(F|y) = \prod_{i=1}^{N} \int \prod_{1}^{T_i} \phi(y_{it} - \theta_i) dF(\theta_i)$$

Instead of focusing on each observation y_{it} we can also focus on the mean $\hat{\theta}_i = \bar{y}_i$. If we utilize $\hat{\theta}_i - \theta_i \sim N(0, 1/T_i)$, we can write the likelihood function as

$$L(F|y) = \prod_{i=1}^{N} \int \phi((\hat{\theta}_i - \theta_i)\sqrt{T_i})\sqrt{(}T_i) dF$$
$$l(F|y) = \sum_{i=1}^{N} \log \int \phi((\hat{\theta}_i - \theta_i)\sqrt{T_i})\sqrt{(}T_i) dF$$

Optimizing over all possible function F neccesitates some kind of discrete approximation. The most common one is the grid approximation. We can also use the EM algorithm to optimize the likelihood function. Let f_j approximate the value of dF on the grid

$$\max_f \left\{ \sum_{i=1}^{N} \log g_i \middle| g = Af, \sum_j f_j \Delta_j = 1, f \geq 0 \right\}$$

Remark. For a reader not as versed in mathematics as she should be. $A_{i*}f = \sum \sqrt{T_i}\phi((\hat{\theta}_i - \theta_j)\sqrt{T_i}) f_j \Delta_j$. We use $\sum f_j \Delta_j$ to approximate the integral $\int dF$ as one can imagine.

This a convex objective function subject to linear equality and inequality constraints. The EM algorithm is a natural choice to optimize this function. The E-step is to calculate the expectation of the log-likelihood function given the observed data and the current estimate of the parameter. The M-step is to maximize the expectation of the log-likelihood function with respect to the parameter. The algorithm iterates between these two steps until convergence. Often, the dual formulation of a convex objective is more efficient than the primal.

$$\max_f \left\{ \sum_{i=1}^{N} \log(v_i) | A^T v = n 1_p, v \geq 0 \right\}$$

Question 1. Derive on your own for practice

Solution. We define the multiplier μ_1', μ_2, λ' for the two equality and one inequality contstraints. The dual objective function is

$$\min_{f,g} \left\{ -\sum_{i=1}^{N} \log g_i + \mu_1'(g - Af) + \mu_2(\sum_j f_j \Delta_j - 1) - \lambda' f \right\}$$

Minimize over each g_i gives the condition

$$\frac{1}{g_i} = \mu_{1i}$$

Minimize over each f_i gives the condition

$$-(\mu_1' A)_j + \mu_2 - \lambda_j = 0$$

Therefore the objective is

$$\begin{array}{ll} & \{\sum \log \mu_{1i} + n - \mu_2\} \\ \text{subject to} & -\mu_1' A + \mu_2 1_p - \lambda = 0 \\ & \lambda \geq 0 \end{array}$$

Thus the dual problem is

$$\begin{array}{ll} \text{maximize}_{\mu_1', \mu_2} & \{\sum \log \mu_{1i} + n - \mu_2\} \\ \text{subject to} & \mu_1' A \leq \mu_2 1_p \\ & \mu_1 \geq 0 \end{array}$$

□

1.2.2 Scale mixture

The model is the following:

$$y_{it} = \sigma_i \epsilon_{it} \quad \text{where} \quad \epsilon_{it} \sim N(0, 1)$$

Similarly

$$s_i = \hat{\sigma}_i^2 = \frac{1}{m_i} \sum_{t=1}^{T_i} y_{it}^2$$

But what's the distribution of $\hat{\sigma}_i$ (which is not so obvious relative to $\hat{\theta}_i$)? Well, $\frac{\sum y_{it}^2}{\sigma_i^2}$ follows a Gamma distribution with shape parameter $r_i = (T_i)/2$ and scale parameter $s_i = \sigma_i^2 / r_i = 2\sigma_i^2 / T_i$.

Question 2. Why is it gamma distribution?

Solution. The sum of k independent standard normal variable X follows a $\chi^2(k)$ distribution. The mean of n independent $\chi^2(k)$ distribution variables K follows a gamma distribution $\gamma(nk/2, 2/n)$. The equivalence lies in here: if $X \sim \gamma(v/2, 2)$ (in the shape-scale parametrization), then X is identical to $\chi^2(v)$, the chi-squared distribution with v degrees of freedom. Conversely, if $Q \sim \chi^2(v)$ and c is a positive constant, then $cQ\ \gamma(v/2, 2c)$. □

Remark. The Gamma distribution $\gamma(k, \theta)$ (shape, scale) has the following distribution function

$$f(x|k,\theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta} \quad F(x|k,\theta) = \frac{1}{\Gamma(k)} \gamma(k, x/\theta)$$

where $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$.
Thus the likelihood function is

$$L(F|y) = \prod_{i=1}^{N} \int \gamma(s_i|r_i, \sigma_i) dF(\sigma_i)$$
$$l(F|y) = \sum_{i=1}^{N} \log \int \gamma(s_i|r_i, \sigma_i) dF(\sigma_i)$$

which we can proceed just as in the location mixture case.

1.2.3 Location-scale mixture (independent)

The model is

$$y_{it} = \theta_i + \sigma_i \epsilon_{it} \quad \text{where} \quad \epsilon_{it} \sim N(0,1)$$

The sufficient statistics $\hat{\theta}_i$ and $\hat{\sigma}_i$ are

$$\hat{\theta}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} y_{it} \sim N(\theta_i, \sigma_i^2/T_i)$$
$$\hat{\sigma}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} (y_{it} - \hat{\theta}_i)^2 \sim \gamma(s_i|\alpha = r_i, \beta = \sigma_i^2/r_i)$$

Just as in the previous two cases, we can write the (log) likelihood as

$$l(G_\theta, F_\sigma|y) = \sum_{i=1}^{N} \log \int \int \left[\phi((\hat{\theta}_i - \theta_i)\sqrt{T_i})\sqrt{T_i}\right] [\gamma(s_i|r_i, \sigma_i)]\, dG_\theta(\theta_i) dF_\sigma(\sigma_i)$$

For **estimation**, we can first solve for $\hat{F}_\sigma$ and solve for $\hat{G}_\theta$ given $\hat{F}_\sigma$. There are two compuatation methods.

- Reexpress the Gaussian component as Student's t therefore eliminating the dependence on σ_i.
- Iterate between the Gamma and Gaussian component of the likelihood. (Specific to this independent prior assumption.)

1.2.4 Location-scale mixture (general)

The most general Gaussian location-scale mixture with covariate effects

$$y_{it} = x_{it}\beta + \theta_i + \sigma_i\epsilon_{it} \quad \text{where} \quad \epsilon_{it} \sim N(0,1)$$

Given a true β, it is straightforward that

$$y_{it}|\mu_i, \sigma_i, \beta \sim N(x_{it}\beta + \mu_i, \sigma_i^2)$$

The sufficient statistics for

- μ_i: $\bar{y}_i - \bar{x}_i\beta$

  contains the between information

- σ_i^2: $\frac{1}{T_i-1}\sum_{t=1}^{T_i}(y_{it} - x_{it}\beta - \mu_i)^2$ It is worth mentioning that

  $$S_i|\mu_i, \sigma_i^2, \beta \sim \gamma(r_i, \sigma_i^2/r_i) \quad \text{where} \quad r_i = (m_i - 1)/2$$

  contains the within information (deviations from the individual means)

Remark. The orthogonality between the within and between information no longer holds here. (Why does it hold in the classical Gaussian panel data?)

The likelihood function is

$$\begin{aligned}
l(\beta, h(\theta,\sigma)|y) &= \prod_{i=1}^{N} g_i(\beta, \theta_i, \sigma_i|y_{i1}, \ldots, y_{iT}) \\
&= \prod_{i=1}^{N} \int\int \prod \frac{1}{\sigma}\phi(\frac{y_{it} - x_{it}\beta - \theta_i}{\sigma})h(\theta,\sigma)d\theta d\sigma \\
&= K\prod_{i=1}^{N}\int S_i^{1-r_i}\int\int \frac{1}{\sigma}\phi(\frac{\bar{y}_i - \bar{x}_i\beta - \theta_i}{\sigma})\frac{e^{-R_i}R_i^{r_i}}{S_i\Gamma(r_i)}h(\theta,\sigma)d\theta d\sigma
\end{aligned}$$

where

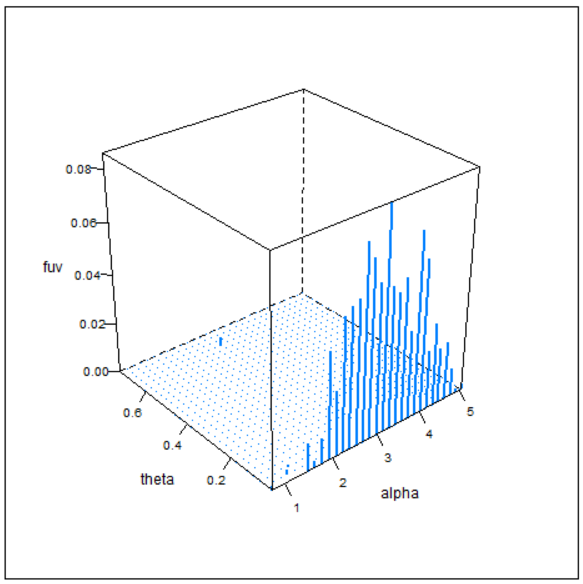
$$R_i = \frac{r_i S_i}{\sigma_i^2} \quad K = \prod_{i=1}^{N}\left(\frac{\Gamma(r_i)}{r_i^{r_i}}(\frac{1}{\sqrt{2\pi}})^{T_i-1}\right)$$

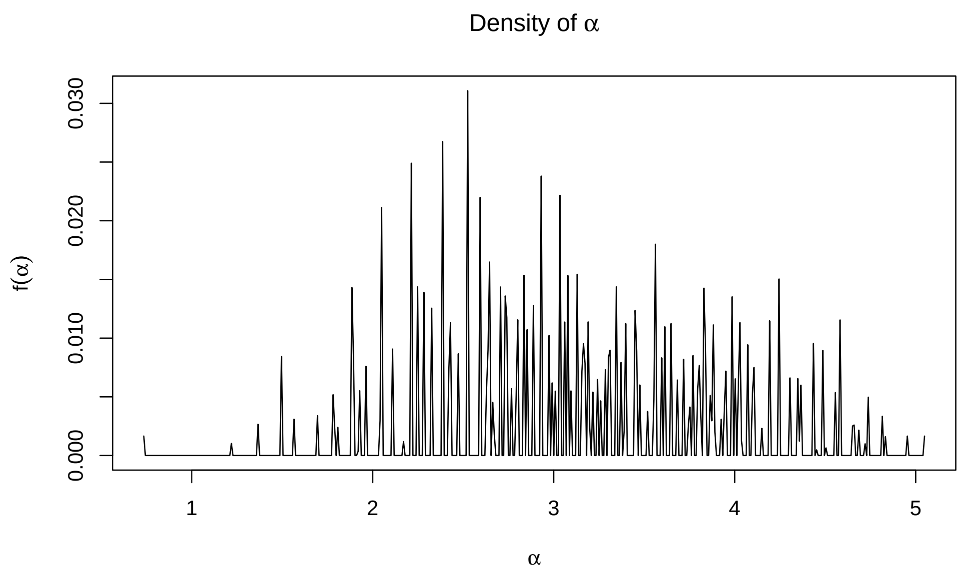
Question 3. The true β is unknown. Therefore, we can not condition on it. How about the so called profile likelihood? How to compare it with the *FORBIDDEN* approach of getting fixed effect estimates from the WG estimation? Since the θ_i is regarded as *NUISANCE* parameters in the WG estimation, how low the status is...! Poor θ_i!

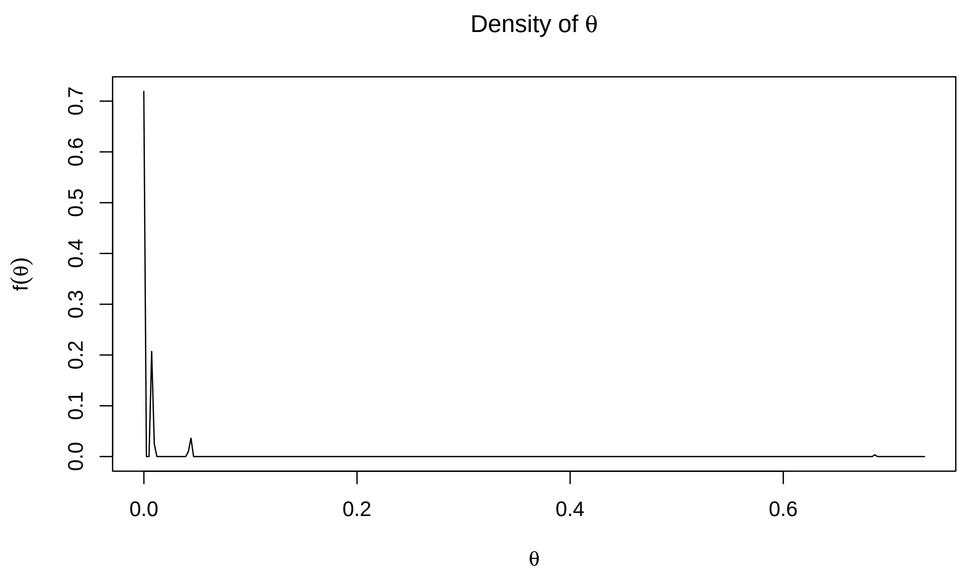
1.2.5 Estimation of mixture density

Dependent of location and scale

Independent location and scale


0.08
0.06
0.04
0.02
0.00
fuv
0.6
0.4
0.2
theta
1
2
3
4
5
alpha


Density of α
f(α)
0.000
0.010
0.020
0.030
1
2
3
4
5
α


Density of θ
f(θ)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.0
0.2
0.4
0.6
θ