

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

Scuola Politecnica e delle Scienze di Base

Dipartimento di ingegneria civile, edile ed ambientale



CORSO DI LAUREA TRIENNALE IN
INGEGNERIA PER L'AMBIENTE E IL TERRITORIO

TESI DI LAUREA

**Un nuovo modello di utilità aleatoria capace di cogliere in
maniera efficace gli effetti della eterogeneità dei gusti e delle
correlazioni tra i residui aleatori: il Mixed CoNL**

Relatore

Prof. Ing. Andrea Papola

Correlatore

Ing. Fiore Tinessa

Candidato

Antonio Elia Pascarella

Matr.N49/716

ANNO ACCADEMICO 2017-18

Abstract

Ciascun individuo si trova continuamente di fronte a una scelta, sia durante il proprio cammino di vita sia nella semplice quotidianità. Il tema della scelta è anche tanto caro al filosofo Kierkegaard, che lo avvertiva come fonte di sofferenza: sarà la scelta giusta?

Per quanto sia impossibile prevedere cosa sceglierà il singolo individuo in un dato luogo ed in un dato momento, dal 1600 in poi, grazie a uomini come Pascal, Bernoulli ed altri, si comprese che il caso segue delle regole e “l'uomo medio” diventa un simbolo di questa conquista intellettuale. Questi sono gli anni in cui vengono formalizzate le proprietà della distribuzione normale e si studiano modelli di probabilità. Da tempo (Thurstone, 1927) in letteratura scientifica sono noti i modelli di utilità aleatoria (*random utility models, RUM*), grazie ai quali è possibile calcolare le probabilità che un individuo scelga una certa alternativa all'interno di un insieme discreto di alternative. Tale paradigma ha trovato, storicamente, ampia diffusione di utilizzo per il problema della stima della domanda di mobilità nel campo dell'ingegneria dei trasporti. In particolare, la teoria dell'utilità aleatoria assume che il decisore sia un massimizzatore di beneficio e che, pertanto, scelga l'alternativa di cui percepisce la massima utilità. Tale utilità, essendo ignota all'analista, viene modellata come una variabile aleatoria. Le ipotesi che l'analista fa sulla forma funzionale e sulla distribuzione stocastica di dette utilità determina la bontà del modello risultante in termini di capacità di regressione, nonché di capacità previsionale.

I due problemi su cui storicamente si è maggiormente focalizzata la letteratura sui RUM sono la eterogeneità dei gusti (*taste heterogeneity*) e gli effetti delle correlazioni stocastiche tra le utilità percepite sulle probabilità di scelta di una certa alternativa. Il primo problema è relativo al fatto che ciascun decisore attribuisce un peso differente alle diverse variabili di cui l'utilità è funzione. Il secondo problema è strettamente legato a come la percezione di un insieme di scelta sia influenzata dal grado di similarità tra le alternative. Nel caso in cui tale problema venga trattato in maniera stocastica, le correlazioni tra tali utilità hanno un grande impatto sulle probabilità di scegliere una certa alternativa.

Il mixed logit con specificazione di tipo misto random coefficient ed error component è, teoricamente, una soluzione che risolve entrambi i problemi accennati prima. Purtroppo, le specificazioni di tipo error component sono caratterizzate da complessi problemi di identificabilità dei parametri che spesso le rendono una via poco praticabile.

Recentemente, Papola (2016) ha proposto una nuova classe di modelli di utilità aleatoria, denominata *combination of random utility models (CoRUM)*. Tale classe di modelli consente di ottenere un RUM, sfruttando l'ipotesi che la distribuzione delle utilità percepite sia ottenuta come combinazione lineare di distribuzioni soggiacenti altri RUM. Tale assunzione, in generale, consente di ottenere un modello risultante molto flessibile nel riprodurre diversi effetti, attraverso una mera combinazione di modelli costituenti più semplici. Una particolare specificazione del CoRUM è ottenibile come combinazione lineare di modelli di tipo Nested Logit (NL) ed è denominata CoNL. La specificazione CoNL, come ampiamente dimostrato in Papola (2016), consente di cogliere in maniera molto flessibile gli effetti delle correlazioni, con il vantaggio di disporre di una espressione in forma chiusa per queste ultime. Le proprietà del CoNL sono stata approfondite in lavori successivi, che ne hanno investigato sia la capacità previsionale su scenari futuri (Tinessa et al., 2017), sia la possibilità di applicazione al complesso problema della scelta del percorso (Papola et al., 2018). Le proprietà del CoRUM, e in particolare della specificazione CoNL, possono essere sfruttate per cogliere anche gli effetti della eterogeneità dei gusti sulle probabilità di scelta. Un principio dell'applicazione del CoRUM per questo scopo è contenuto in Tinessa (2019, PhD thesis), attraverso l'applicazione su un caso reale di scelta modale.

Il presente lavoro di tesi si propone di integrare il lavoro di Tinessa (2019) attraverso lo studio del CoRUM su una popolazione sintetica. Tale studio offre il vantaggio di disporre di un contesto nel quale, a differenza di un caso reale, tutte le componenti non osservabili dell'utilità siano note. Ciò consente di lavorare in un ambiente controllato, nel quale tutti i valori veri delle variabili siano noti a priori. In particolare, nel presente lavoro, si intende investigare l'applicazione del CoNL con una specificazione del tipo random coefficient. Nel seguito, ci si riferirà a tale specificazione con il nome di Mixed CoNL.

Il lavoro di tesi si articola come segue: il capitolo 2 riprende i passi salienti dello stato dell'arte sui RUM e sul paradigma del CoRUM; il capitolo 3 descrive la formulazione del Mixed CoNL, il setup della popolazione sintetica e i risultati del confronto, in termini di bontà di regressione, del Mixed CoNL rispetto ad diversi RUM; il capitolo 4 riporta una sintesi sulle principali conclusioni del lavoro e offre qualche suggerimento su sviluppi futuri.

Il dataset sintetico, costruito in ambiente Matlab, è stato definito mediante l'assunzione di attributi e utilità marginali distribuite secondo una distribuzione Normale monovariata. In aggiunta, ulteriori componenti non osservabili sono state introdotte all'interno delle utilità individuali, assumendo per esse una distribuzione Normale Multivariata, con una certa matrice di correlazione target.

I risultati confermano che il mixed CoNL esibisce migliori misure di bontà di accostamento, quali la verosimiglianza, il ρ^2 e il ρ^2 corretto. In fase di stima dei parametri il mixed CoNL, oltre che una migliore performance in termini di probabilità, ha dato origine a parametri quasi sempre statisticamente significativi.

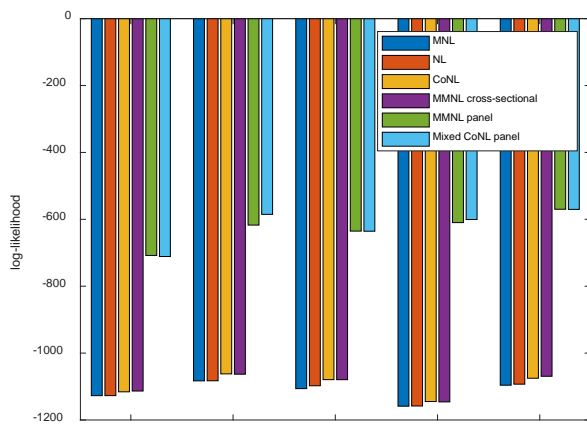


Grafico1:log-likelihood con numero individui=100

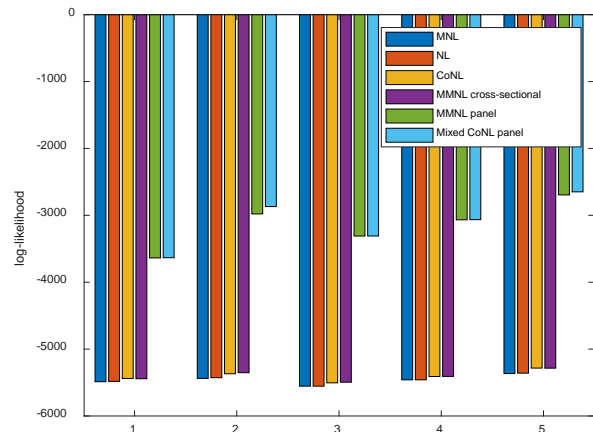


Grafico2:log-likelihood con numero individui=500

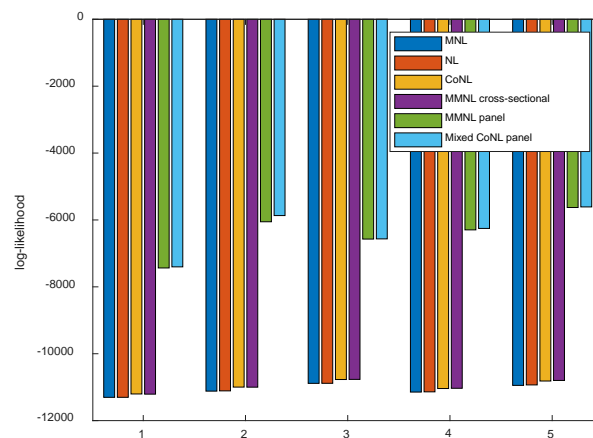


Grafico3:log-likelihood con numero individui=1000

La sperimentazione mostra, in conclusione, che l'idea di costruire un modello di utilità aleatoria capace di cogliere gli effetti della eterogeneità dei gusti e delle correlazioni fra utilità trova una soddisfacente espressione nel Mixed CoRUM. In particolare, il Mixed CoNL consente di ottenere risultati migliori del Mixed Logit random coefficient e di evitare onerose specificazioni di tipo Error Component

