

Tesi di Dottorato in
Economia e Politiche dei Mercati e delle Imprese – curriculum Metodi Statistici

ciclo XXX
A.A 2016/2017

High-dimensional Statistics for complex data
Massimo Pacella

ABSTRACT (ITA)

L'analisi dei dati ad alta dimensione è diventata negli ultimi anni un argomento di ricerca popolare, a causa dell'emergere di svariate applicazioni in diversi campi scientifici in cui si rende necessario lavorare con enormi set di dati. Una delle sfide principali nell'analisi dei dati ad alta dimensione riguarda l'interpretabilità dei modelli stimati e l'efficienza computazionale delle procedure adottate. Tale scopo può essere raggiunto attraverso l'identificazione di variabili rilevanti che influenzano realmente il fenomeno di interesse, in tal modo è possibile costruire modelli di facile interpretabilità ed utilità nel risolvere problemi pratici.

I primi tre capitoli della tesi sono dedicati allo studio della selezione di variabili in contesti di dati ad alta dimensionalità. Nella prima parte è riportata una breve ed esauriente rassegna sulle principali tecniche di selezione per modelli non parametrici. Infine, nel capitolo 3 viene presentata una proposta di ricerca riguardante un approccio di *screening* per modelli non additivi sviluppato utilizzando informazione condizionata nella procedura di selezione.

Diversamente, la seconda parte della tesi tratta dei modelli spazio-temporali sempre in contesto di dati ad alta dimensione. Nell'ultimo decennio, una particolare classe di tali modelli ha subito un rapido sviluppo, essi sono definiti *Spatial Dynamic Panel Data models* (SDPD). Sono state proposte diverse versioni dei modelli SDPD, basate su diverse ipotesi dei parametri del modello e sulle diverse proprietà degli stimatori. La versione standard ha i parametri di regressione costanti rispetto alle unità spaziali, mentre di recente, ne è stata proposta una generalizzazione in cui i parametri risultano adattativi rispetto alle unità spaziali. Tale assunzione è motivata da evidenze pratiche ed empiriche, da cui risulta chiaro che, considerare una parametrizzazione costante rispetto alle singole unità può risultare limitante. Il capitolo 4 è dedicato alla rassegna sulle caratteristiche principali di tale classe di modelli, mentre nel capitolo 5 viene proposta una strategia di *testing* per identificare la struttura del modello SDPD. Tale procedura di *multiple testing* consente di scegliere tra la versione del modello con parametri adattativi e altre caratterizzazioni derivate imponendo particolari vincoli sui parametri. Il test multiplo viene eseguito in alta dimensionalità mediante tecnica alla Bonferroni e la distribuzione della statistica test è derivata attraverso uno schema di ricampionamento bootstrap.