The University of Osaka
Institutional Knowledge Archive

Title	Semiparametric Inference With Nonignorable Nonresponse Data
Author(s)	森川, 耕輔
Citation	大阪大学, 2017, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/69256
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

様式３

論　文　内　容　の　要　旨

氏　　名（　森　川　耕　輔　）	
論文題名	Semiparametric Inference With Nonignorable Nonresponse Data （無視できない無回答データによるセミパラメトリック推測）

論文内容の要旨

欠測値データへの統計的推測法の問題は，経済学，疫学，心理学を含む多くの分野で応用上問題となってきている．無回答データは最も単純な欠測値データの一種であり，標本調査等において，プライバシー・倫理的観点から応答変数に対する回答拒否により起こり得る．この場合，完全データに対する通常の統計的推測法は直接適用できない．そのため，しばしば回答が得られない対象者のデータを削除して，完全データを人工的に作成し，完全データに対する統計手法を用いる“リストワイズ削除”という方法が用いられる．しかし，この方法を用いる場合，推定量に対して，(i)有効性が減少；(ii)バイアスが生じる，という２つの問題が起こる．１つ目の方法はサンプルサイズの減少から起こる．また2つ目の問題は，例えば，年収調査を行う場合年収が低い労働者は回答しにくい場合を考える．この場合，観測されたデータから計算される単純平均は，真の平均年収を過大評価してしまう．本研究では，このような欠測値に適切に対処でき統計的手法の開発を行う．

無視できない欠測値データとは，先ほどの年収の例のように，データが欠測するかどうか（欠測メカニズム）が欠測する応答変数にも依存しているデータである．このようなデータの解析には，欠測メカニズム以外にも応答変数のモデルの特定も要求される．この仮定の強さゆえ，応用上使われることが少ない．そのため，(1) 応答変数に対するモデルの仮定を必要しない柔軟な，また漸近的に最有効推定量となるセミパラメトリック推測法を構築する．さらに，無視できない解析に対する既存の解析手法は，モデルの識別性のため，操作変数の存在といった検証不能な仮定が要求される．本研究では，(2)データから検証可能なモデルの識別性条件を提案する．提案する条件は，観測データから検証可能であり，操作変数も必要としない．本研究では，上記(1),(2)の問題を解決する．

様式７

論文審査の結果の要旨及び担当者

氏　名（　森 川 耕 輔　）		
論文審査担当者	（職）	氏　名
	主 査　教 授	狩野　裕
	副 査　教 授	内田雅之
	副 査　教 授	鈴木　譲

論文審査の結果の要旨

データの一部が想定どおりに取得できないとき，それを欠測または欠損という．欠測値を含むデータの解析は近年極めて重要になっている．欠測値を含むレコード（ユニット，オブザベーション）をデータセットから外し（リストワイズ削除），見かけ上欠測のないデータに整形し統計解析する方法は，推測に大きなバイアスが生じることがあり，危険である．本論文では，統計的推測を実行するに当たり前提とする条件が十分に弱いセミパラメトリック法の枠組みで，一致性を有し且つ漸近有効である推定量を二種類提案した．一つ目は観測されたデータに対するモデルとしてパラメトリックモデルを採用するものであり，他の一つはカーネル法を用いたノンパラメトリックなモデルを採用する．両推定量とも常に一致性と漸近正規性を有し，前者はパラメトリックモデルが正しければ漸近有効（最有効），後者は常に漸近有効になることを数学的に証明した．さらに，これらの推定方法において，反応モデルのパラメータ推定の方法をGeneralized Method of Moments (GMM)によって改良した．また，これらの性質と他の推定量との比較を目的に数値実験が実行され，提案推定量の良さが数値的も確認された．加えて，欠測値データ解析に適用できる統計モデルの識別性条件を導出した．本条件は，観測されたデータのみによってチェック可能という特筆すべき性質を有する．著者は提案した方法をKorea Labor Income Panel Survey Dataの解析に応用し，その有効性を実証した．

欠測値データ解析で重要な概念は，データが観測される確率（i.e. 反応モデル）の欠測値への依存性であり，反応モデルが欠測値に依存する場合は無視不可能と呼ばれ，識別可能な反応モデルを正確に規定しなければならず統計的推測は難しくなる．また，母平均など単純な母数については，欠測のないデータの場合，応答変数のモデルが不必要という広い状況で推定可能であり，欠測がある場合でも応答変数のモデル化をできるだけ避けて推測したいという欲求がある．本論文において，無視不可能な欠測を含むデータに対して，十分に弱い仮定の下で新しい有効な推定方法を構成したことは，先に述べた二つの困難への顕著な学術的貢献であり，博士（理学）の学位論文として価値のあるものと認める．