

厚生労働科学研究費補助金（労働安全衛生総合研究事業）

（分担）研究報告書

経済情勢等が労働災害発生動向に及ぼす影響等に関する研究：
多変量時系列解析による数理モデルの開発と検証（数理モデル班、経済情勢班報告）

研究分担者 榎原 毅 名古屋市立大学大学院医学研究科
池上 徹 （公財）大原記念労働科学研究所
研究協力者 庄司直人 名古屋市立大学大学院医学研究科

研究要旨

マクロ経済学・金融工学等で応用されている多変量時系列解析手法(Kariya, 1993)を用いて、経済情勢が業種別労働災害の発生に及ぼす影響を明らかにすることが最終目的である。平成 28 年度は数理モデル班会合を 1 回、全体会合を 5 回開催し、フォーカスグループ手法により、時系列モデルに投入する主要アウトカム・要因の選定を行った。また、解析に用いるデータ期間、および欠損値補完の処理方針（多重代入法、移動平均、変数変換処理など）について適宜ディスカッションを行い、主要な経済指標に関する変数の定義を行った。更に、時系列モデルへのトライアル解析を行い、指標選定に求められる課題を整理した。

A．研究目的

労働災害(労災)は長期的には減少しているが、小売・飲食業や保健衛生業などの第三次産業では増加傾向にある。第 12 次労働災害防止計画においても、重点業種別の対策が提唱されているが、労働を取り巻く諸環境の要因（経済情勢、産業構造の変化、就業形態、自然・気象条件、産業技術革新等）が及ぼす影響について科学的根拠に基づく解析はほとんど行われておらず、行政政策評価に資する知見が切望されている。

そこで、本研究ではマクロ経済学・金融工学等で応用されている多変量時系列解析手法(Kariya, 1993)を用いて、経済情勢が業種別労働災害の発生に及ぼす影響を明らかにすることが最終目的である。数理モデル班では、経済情勢班と協力し、主に時系列モデルに投入する主要アウトカム・要因の定義の設定と生成、解析プロトコル手順の確立と数理モデルの構築、そしてモデル適合度評価の実施を担っており、平成 28 年度では主要アウトカ

ム・要因の選定と定義の設定を行った。

B．方法

班会合を 1 回（2016 年 8 月）、全体会合を 5 回（2016 年 10 月、12 月、2017 年 1 月、2 月、3 月）開催した。各研究班で調査した各種指標を持ち寄り、各指標の利用可能性についてブレインストーミングを行った。各指標は データ期間、 データ密度（年単位・四半期単位・月単位など）、 データの質（発行元や信頼性）、 データの利用可能性（入手先）、 データ加工の手間、 データ欠損の度合いの 6 側面で検証を行い、最終的に投入する変数の定義方法および優先度について議論を重ねた。

なお本研究では、国が提供・公開している各種データ資源、統計法等、法令の規定に基づく調査データ（連結不可能匿名化後の統計データ）などオープンデータを主に扱うため、文部科学省・厚生労働省「人を対象とする医学系研究に関する倫

理指針（平成 26 年 12 月 22 日）」は適用外である（個々の研究対象者からデータを収集することは行わない）。

C. 研究結果

1. 抽出した変数について

最終的に抽出された変数は計 212 変数であった。

各データについては利用不可能な年または期間が存在する変数もある。年単位データの解析可能期間と想定される 1973～2012 年のデータセットにおいて、完全データの変数（当該期間に欠損がない変数）は全変数の 27%に過ぎなかった。また、同じ統計であっても出処が異なると当該年の統計値が異なるものも存在し、欠損値補完の処理方針（多重代入法、移動平均、変数変換処理など）についても研究組織メンバー内で適宜ディスカッションを行った。時系列データで何年未満の欠損をどのような統計手法で補完するのかについては更なる議論が必要であった。現段階では同一年でも出処により値が異なる変数は、変数 XXXXa, XXXXb のようにコード化しデータシート内では併記しておくこととした（またデータシート内で該当箇所を分かるようにマークしておく）。

抽出した変数については、経済情勢関連については、主に国内総生産、産業別労働生産性、景気動向指数、設備投資・技術革新指数（機械受注指数など）、各種経済指数（企業物価指数、失業率、鉱工業生産指数、消費者物価指数など）など。労働経済関連については、産業構造、就業形態（非正規雇用、外国人労働、高齢労働等）や社会政策、労働力人口（業種別就労人口）などを示す各種指標を、気象・天災関連については、自然災害（台風、豪雨など発生日時および被害数）、気温のほか、大震災など経済活動にも影響を与えられられる事象（震度 5 弱以上、死傷者 100 人以上の災害）については月単位データとして抽出・データ生成を行った。また、従属変数（アウトカム）として利用する労働災害関連の指標については、業種別度数率、強度率、労働損失日数、労災申請件数などの利用を想定しているが、代理アウトカム指標の利用可能性も視野に人口動態統計から死因別死亡率などの情報も収集を行った。

2. 経済動向を示す主要要因変数について

本研究課題として扱う経済動向としては、いわゆる一般的に扱われる事の多い下記の各 Index を主要要因として位置づけ、その他補助的変数の選定をモデル投入時に探索的に行うこととした。

【景気 index】

景気動向指数 DI/CI（内閣府、月次統計）

景気ウォッチャー調査（2000 年～、内閣府）

日銀の短観（全国企業短期経済観測調査）

【家計 index】

家計調査（総務省、月次）：2 人以上の世帯 8000 世帯、単身 1000 世帯のサンプリング調査（個人消費の伸び率データの利用）

家計消費状況調査（総務省、月次）：インターネットを通じた消費行動。家計調査では補足しきれていない部分を補完

毎月勤労統計調査（厚労省、月次）：現金給与総額

消費動向調査（内閣府、月次統計）

ジニ係数の時系列データ：格差を示すローレンツ曲線の AUC 面積データが入手できれば利用

【企業 index】

法人企業会計（財務省、四半期）

労働分配率（景気が良いときは低下、不景気になると上昇する指標なので変動を利用できないか）

【貿易 index】

国際収支統計（財務省、月次）：貿易収支、通関輸出、通関輸入の変数を利用。

為替レート：1973 年以降（1973 までは固定相場制）の円ドルレートの変数

【物価 index】

消費者物価指数（ラスパイラス指数、月次）、消費デフレーター（パーシェ指数、四半期）の 2 指標

国内企業物価指数（日銀、月次）

【雇用 index】

職業安定業務統計（厚労省）：有効求人倍率、有効求職者数、新規求人倍率、新規求人数、充足率

労働力調査（総務省、月次）：完全失業率

【金融 index】

新発国債 10 年物利回り（%）

日経平均株価、TOPIX（東証株価指数）

信用乗数（日銀）：量的緩和政策が物価や景気など実体経済に与える影響を評価する指標
紙幣の所得流通速度(VELOCITY、日銀)

最終的にはモデル投入を行い、変数の選定を行うが、膨大に存在する各種経済・労災・気象指標等について、多様な組み合わせによるモデル投入を探索的に行うにあたり、主要経済動向 Index として上記を用いることとした。

3. 自己回帰モデルによるトライアル解析の実施

時系列データの性質を理解し、必要な事前処理の内容を把握するために、1954 年～2014 年のローデータセットを用い、ARIMA モデルの変数自動選択によりトライアル解析を実施した。死亡災害件数を従属変数として、説明変数には 国内 51 地点での年降水量の基準値との偏差、 全国企業倒産

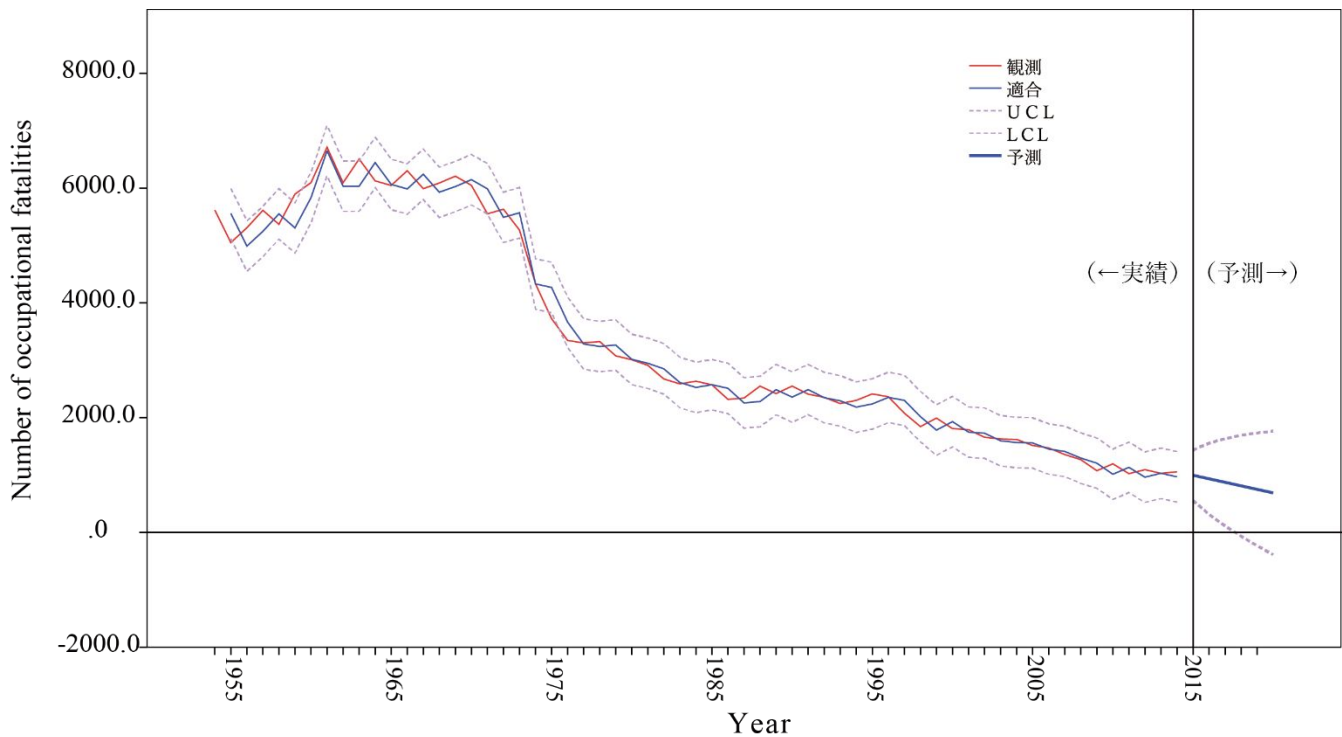


図1 自己回帰モデルを用いた予測モデル例（変数定義の検証のために実施したトライアル解析）

負債総額（百万円） 全国企業倒産件数（件）
 給与総額（百万円） 納税額（百万円） 平均気温の基準値からの偏差、 台風の発生数、
 産業別実労働時間指数（製造業） 国民医療費（億円）の変数を投入した結果、最終的に残った説明変数は死亡災害件数（自己回帰）および全国企業倒産件数（件）の2変数であった。モデル適合統計量である R^2 は0.40であった。また、このモデルにより、2014年までのデータで2015～2018年の死亡災害件数（人）の予測を行ってみた結果（図1）2015年の予測値(95%信頼区間の上限値 UCL～下限値 LCL)：995.6(556.2～1435.0、実統計値：972)、2016年：934.2(312.7～1555.6)、2017年：872.8(111.7～1633.9)、2018年：811.4(-67.5～1890.2)となり、4年後の時点では下限値がマイナスとなった。主に自己回帰モデルに依存した解析では精度が不十分であることが示唆された。

D. 考察

暫定的に年単位データは1973～2012年前後の50年間、月単位データは1992～2012年の約20年間を対象データ期間とし、データセットの整備を進めることで合意した。データについては利用不可能な年または期間が存在する変数もある。年単位データの解析可能期間と想定される1973～2012年のデータセットにおいて、完全データの変数（当該機関に欠損がない変数）は27%に過ぎなかった。また、同じ統計であっても出処が異なると当該年の統計値が異なるものも存在し、欠損値補完の処理方針は今後の解析において重要となっ

てくる。多変量解析を用いた多重代入法などによる欠損値補完を行うか、または状態空間時系列解析手法を採用するのかなど、データセットの欠損パターンなどを包括的視点で勘案した上で、処理方針を今後検討することが必要になるであろう。

また、5回の全体会合を重ねる中で、個々の指標について質的側面についても適切に理解しておく必要性が提案され、議論を重ねてきた。例えば、労働災害統計自体、1次・2次産業の減少と3次産業の増大など産業構造の変化に伴い70年代以降は働き方が変革していること、また各種法制度の整備の状況や労働災害の質的变化（メンタル疾患の増加など）などもあり、単に量的側面だけで捉えることが困難になりつつある。各指標について、それぞれ想定される背景因子が存在すると考えられるため、モデル投入を行う前に適切な調整法（モデル人口調整、死亡率調整、基準年調整などによる各種補正值）を検証することも必要となるため、平成29年度では各指標について下記プロフィール情報を一覧整理することとした。

<各指標のプロファイル情報>

- ・ 出処情報
- ・ サンプルング方法：データ収集方法、分母・分子の定義ほか。
- ・ 指標の求め方・算出方法
- ・ 基準年の有無
- ・ 変数の特徴：トレンド、定常・非定常、ランダムウォーク、周期性に関する傾向
- ・ 変数の調整方法（既存研究での事例検索）

- ・要配慮点・特筆すべき事項：
- ・時系列変動データ（図示）

これら各指標の特性を把握し、モデル投入する変数の優先度を付け、基礎解析として時系列解析へ変数投入すべきかの検証を行うこととした（ホワイトノイズ解析、交差相関、自己相関分析）。また変数の特徴としては、定常・非定常時系列、トレンド・ランダムウォーク性の解析、スペクトル解析による時系列推定モデルの検討（Holtトレンド、Brownトレンド、減衰トレンドなど）を行い、各変数の持つ時系列情報の特性を事前に明らかにしておくこととした。

また、時系列解析の特性を把握し、事前データクリーニングの方針を検討するにあたり、ARIMAモデルの変数自動選択によるトライアル解析を実施した。死亡災害件数を従属変数としてモデルを生成した結果、最終的に残った説明変数は死亡災害件数（自己回帰）および全国企業倒産件数（件）の2変数であったこと、また、モデル適合統計量である R^2 は約40%の説明力であり不十分であった。予測モデルによる2015年以降の推定値の信頼区間は広く、自己回帰モデルに依存した解析では実用性は低いと考えられた。一方で、自己回帰モデルによる低い説明力は、適切に補正された各種指標を組み合わせることでモデル適合度の向上が期待される可能性が十分残っていることも示唆される。現在、200以上の指標（変数）を収集しており、これらの指標を多様に組み合わせることで、モデルの精度を上げることを目指しているが、その前段階として、現状のローデータをそのまま投入するのではなく、各指標のプロファイルを整理し、適切な補正を行うデータの下処理が重要といえよう。なお、現在収集済みの指数を見直していく中で、解析の枠組みもアプローチを分ける必要性が明らかになってきた。すなわち、本研究課題で扱うデータはマクロデータによる解析のため、利用可能な指標が一定期間存在しない変数については扱うことができない。そこで、解析の粒度設計として長期的なトレンドで解析するアプローチと、即時的反応が期待される変動指数を用いた短期的アプローチの2つに分けて、経済動向が労働災害に与える影響を調べる必要があると示唆された。長年にわたって継続的に蓄積された指標については年単位、時代の要請により指標の重要性が提唱され、近年、取得され始めた指標

で月単位で利用可能なものもあり、これら2つの異なる粒度でアプローチすることで、説明力を高めて行くことを目指している。

E. 結論

経済情勢が業種別労働災害の発生に及ぼす影響を明らかにするために、主に時系列モデルに投入する主要アウトカム・要因の抽出と定義の設定を行った。現段階で抽出された変数は計200以上におよび、それらオープンデータの利用可能期間は1947年（昭和22年）～2017年（平成29年）の約70年の期間にまたがっていた。各変数により収集期間が異なるため、最大公約数的な期間を検討した結果、暫定的に年単位データは1973～2012年前後の50年間、月単位データは1992～2012年の約20年間を対象データ期間とするのが合理的と判断された。ただし、完全データの変数（当該機関に欠損がない変数）は27%に過ぎないこと、また自己回帰モデルによるトライアル解析からも死亡災害件数の予測には自己回帰モデルは不十分なことから、200以上の指標（変数）について各プロファイルを明らかにして、モデル投入を行う前に適切な調整法を検証することが重要となることが示唆された。

F. 健康危険情報

該当なし

G. 研究発表

1. 論文発表

平成28年度はなし

2. 学会発表

平成28年度はなし

H. 知的財産権の出願・登録状況

1. 特許取得

該当なし

2. 実用新案登録

該当なし

3. その他

該当なし