Two Recursive Frisch-Waugh-Lovell Algorithms and Applications in Representing Bias with Multiple Omitted Variables

Rae Tiffen

Undergraduate Honors Thesis
University of Vermont, April 2025

Introduction

- The Frisch-Waugh-Lovell theorem shows how each covariate in an OLS-estimated regression can be decomposed into two components: that explained by the other covariates, and the residual component
- The covariate's coefficient can be found by the regression of the outcome variable on the residual component of its decomposition
- What happens when we decompose these decompositions?
- In this thesis, I:
  - Define two recursive Frisch-Waugh-Lovell algorithms
  - Implement both in Python
  - Study their applications for representing bias with multiple omitted variables

Table of Contents

Notation – variables and coefficients

- Need precise notation to identify coefficients and variables in systems of regressions and decompositions
- Centers the projection operator $\cdot$
- $x_{p \cdot abc}$ refers to the residuals of the regression of x_p on covariates x_a, x_b, x_c

Notation – variables and coefficients

- Need precise notation to identify coefficients and variables in systems of regressions and decompositions
- Centers the projection operator $\cdot$
- $x_{p \cdot abc}$ refers to the residuals of the regression of x_p on covariates x_a, x_b, x_c
- $\alpha_{ap \cdot bc}$ refers to the coefficient on x_p in the regression of x_a on x_p, x_b, and x_c

Notation – variables and coefficients

- Need precise notation to identify coefficients and variables in systems of regressions and decompositions
- Centers the projection operator $\cdot$
- $x_{p \cdot abc}$ refers to the residuals of the regression of x_p on covariates x_a, x_b, x_c
- $\alpha_{ap \cdot bc}$ refers to the coefficient on x_p in the regression of x_a on x_p, x_b, and x_c
- Parentheses are sometimes used in coefficient subscripts to clarify which variables (or residuals) are involved in the regression
  - E.g., the coefficient from the regression of $x_{1 \cdot 2}$ on x_3 is $\alpha_{(1 \cdot 2)3}$

Notation – models

- Primary models are denoted with the outcome variable y and with coefficients β:

$$
\begin{aligned}
y &= \beta_{1\cdot 23}x_1 + \beta_{2\cdot 13}x_2 + \beta_{y3\cdot 12}x_3 + y^*_{\cdot 123} \\
y - y^*_{\cdot 123} = \hat{y} &= \beta_{y1\cdot 23}x_1 + \beta_{y2\cdot 13}x_2 + \beta_{y3\cdot 12}x_3
\end{aligned}
$$

Notation – models

- Primary models are denoted with the outcome variable y and with coefficients β:

$$\begin{aligned} y &= \beta_{1\cdot 23} x_1 + \beta_{2\cdot 13} x_2 + \beta_{y3\cdot 12} x_3 + y^*_{\cdot 123} \\ y - y^*_{\cdot 123} = \hat{y} &= \beta_{y1\cdot 23} x_1 + \beta_{y2\cdot 13} x_2 + \beta_{y3\cdot 12} x_3 \end{aligned}$$

- Auxiliary models (decompositions) are denoted with the relevant x variable as the outcome and with coefficients α:

$$\begin{aligned} x_1 &= \alpha_{12\cdot 3} x_2 + \alpha_{13\cdot 2} x_3 + x^*_{1\cdot 23} \\ x_1 - x_{1\cdot 23} = \hat{x}_1 &= \alpha_{12\cdot 3} x_2 + \alpha_{13\cdot 2} x_3 \end{aligned}$$

The Frisch-Waugh-Lovell result

- This system of notation is well-suited for regression algebra. The FWL result can be stated simply as:

  $\alpha_{(a \cdot bc)(p \cdot bc)} = \alpha_{a(p \cdot bc)} = \alpha_{ap \cdot bc}$

The Frisch-Waugh-Lovell result

- This system of notation is well-suited for regression algebra. The FWL result can be stated simply as: $\alpha_{(a\cdot bc)(p\cdot bc)} = \alpha_{a(p\cdot bc)} = \alpha_{ap\cdot bc}$
- The coefficient $\alpha_{ap\cdot bc}$ can be obtained from three different regressions:

$$\begin{aligned}
\hat{x}_a &= \alpha_{ap\cdot bc}x_p + \alpha_{ab\cdot pc}x_b + \alpha_{ac\cdot pb}x_c \\
\hat{x}_{a\cdot bc} &= \alpha_{ap\cdot bc}x_{p\cdot bc} \\
\hat{x}_a &= \alpha_{ap\cdot bc}x_{p\cdot bc}
\end{aligned}$$

- The above are: standard, double residual regression, residual regression

Iterated FWL algorithm

- Variant on the Modified Gram-Schmidt algorithm, which is used commonly in LS
- The process:

1. Gather y and the matrix of covariates X $(n \times p)$

Iterated FWL algorithm

- Variant on the Modified Gram-Schmidt algorithm, which is used commonly in LS
- The process:

1. Gather y and the matrix of covariates X $(n \times p)$
2. Starting with the pth column of X, (x_p), regress y and the remaining $p-1$ columns on x_p

Iterated FWL algorithm

- Variant on the Modified Gram-Schmidt algorithm, which is used commonly in LS
- The process:

1. Gather y and the matrix of covariates X $(n \times p)$
2. Starting with the pth column of X, (x_p), regress y and the remaining $p-1$ columns on x_p
    2.1 Replace the values of y and each column of X with the residuals from their respective regression, forming $y_{\cdot p}$ and $X_{\cdot p}$

Iterated FWL algorithm

- Variant on the Modified Gram-Schmidt algorithm, which is used commonly in LS
- The process:

1. Gather y and the matrix of covariates X $(n \times p)$
2. Starting with the pth column of X, (x_p), regress y and the remaining $p-1$ columns on x_p
    2.1 Replace the values of y and each column of X with the residuals from their respective regression, forming $y_{\cdot p}$ and $X_{\cdot p}$
    2.2 Store each regression equation, e.g., $y = \beta_p x_p + y_{\cdot p}$

Iterated FWL algorithm

- Variant on the Modified Gram-Schmidt algorithm, which is used commonly in LS
- The process:

1. Gather y and the matrix of covariates X $(n \times p)$
2. Starting with the pth column of X, (x_p), regress y and the remaining $p-1$ columns on x_p
   2.1 Replace the values of y and each column of X with the residuals from their respective regression, forming $y_{\cdot p}$ and $X_{\cdot p}$
   2.2 Store each regression equation, e.g., $y = \beta_p x_p + y_{\cdot p}$
3. Repeat step 2 for all columns of X, to the final regression of $y_{\cdot 2 \ldots p}$ on $x_{1 \cdot 2 \ldots p}$

Iterated FWL algorithm

- Variant on the Modified Gram-Schmidt algorithm, which is used commonly in LS
- The process:

1. Gather y and the matrix of covariates X $(n \times p)$
2. Starting with the pth column of X, (x_p), regress y and the remaining $p-1$ columns on x_p
   2.1 Replace the values of y and each column of X with the residuals from their respective regression, forming $y_{\cdot p}$ and $X_{\cdot p}$
   2.2 Store each regression equation, e.g., $y = \beta_p x_p + y_{\cdot p}$
3. Repeat step 2 for all columns of X, to the final regression of $y_{\cdot 2 \ldots p}$ on $x_{1 \cdot 2 \ldots p}$
4. Substitute each regression (from step 2) into the last-estimated regression and simplify

Iterated FWL algorithm – example

$$
\begin{aligned}
\text{Original model} \quad & \hat{y} = \beta_{1\cdot 23} x_1 + \beta_{2\cdot 13} x_2 + \beta_{3\cdot 12} x_3 \\
\text{Last-estimated regression} \quad & \hat{y}_{\cdot 23} = \beta_{1\cdot 23} x_{1\cdot 23} \\
\text{Previous regressions} \quad & \hat{y}_{\cdot 23} = \hat{y}_{\cdot 3} - \beta_{2\cdot 3} x_{2\cdot 3}, \;\; x_{1\cdot 23} = x_{1\cdot 3} - \alpha_{12\cdot 3} x_{2\cdot 3} \\
\text{Substitute} \quad & (\hat{y}_{\cdot 3} - \beta_{2\cdot 3} x_{2\cdot 3}) = \beta_{1\cdot 23} (x_{1\cdot 3} - \alpha_{12\cdot 3} x_{2\cdot 3}) \\
& \ldots \\
\text{Result after substitutions} \quad & \hat{y} = \beta_{1\cdot 23} x_1 + (\beta_{2\cdot 3} - \beta_{1\cdot 23} \alpha_{12\cdot 3}) x_2 \\
& \; + (\beta_3 - \beta_{1\cdot 23} \alpha_{13} - \beta_{2\cdot 3} \alpha_{23} - \beta_{1\cdot 23} \alpha_{12\cdot 3} \alpha_{23}) x_3
\end{aligned}
$$

$$
\begin{aligned}
\beta_{2\cdot 13} =& \beta_{2\cdot 3} - \beta_{1\cdot 23} \alpha_{12\cdot 3} \\
\beta_{3\cdot 12} =& \beta_3 - \beta_{1\cdot 23} \alpha_{13} - \beta_{2\cdot 3} \alpha_{23} - \beta_{1\cdot 23} \alpha_{12\cdot 3} \alpha_{23}
\end{aligned}
$$

- Iterated FWL calculates regression coefficients in terms of coefficients from the iterated partitions

Recursive FWL decompositions algorithm

- Extends FWL decompositions by decomposing the decompositions
- Like iterated FWL, can be used to estimate multiple regression coefficients with a series of bivariate regressions

$$\hat{y} = \beta_{1\cdot 2} x_1 + \beta_{2\cdot 1} x_2$$

$$x_1 = \underbrace{\alpha_{12} x_2}_{\text{component explained by other covariate}} + \underbrace{x^*_{1\cdot 2}}_{\text{residual component}} \qquad x_2 = \alpha_{21} x_1 + x^*_{2\cdot 1}$$

Recursive FWL decompositions algorithm – 3-variable case

$$\hat{y} = \beta_{1\cdot 23}x_1 + \beta_{2\cdot 13}x_2 + \beta_{3\cdot 12}x_3$$

$$x_1 = \alpha_{12\cdot 3}x_2 + \alpha_{13\cdot 2}x_3 + x^*_{1\cdot 23} \qquad x_2 = \alpha_{21\cdot 3}x_1 + \alpha_{23\cdot 1}x_3 + x^*_{2\cdot 13} \qquad x_3 = \alpha_{31\cdot 2}x_1 + \alpha_{32\cdot 1}x_2 + x^*_{3\cdot 12}$$

$$x_2 = \alpha_{23}x_3 + x^*_{2\cdot 3} \quad x_3 = \alpha_{32}x_2 + x^*_{3\cdot 2} \quad x_1 = \alpha_{13}x_3 + x^*_{1\cdot 3} \quad x_3 = \alpha_{31}x_1 + x^*_{3\cdot 1} \quad x_1 = \alpha_{12}x_2 + x^*_{1\cdot 2} \quad x_2 = \alpha_{21}x_1 + x^*_{2\cdot 1}$$

- Estimates one "branch" at a time
- To estimate $\beta_{1\cdot 23}$, the algorithm...
- $k!$ last-level, solve and move up

Python demonstration

Standard approach to bias

- X is partitioned into X_1 $(n \times p)$ and X_2 $(n \times k)$
- The coefficient on the pth variable of X_1 in regression of y on X_1 alone can be stated as:
  $\beta_{(1*p)\cdot X_1} = \beta_{(1*p)\cdot X_1 X_2} + \sum_{j=1}^{k} \alpha_{(2*j)(1*p)\cdot X_1} \beta_{(2*j)\cdot X_1 X_2}$

Standard approach to bias

- X is partitioned into X_1 $(n \times p)$ and X_2 $(n \times k)$
- The coefficient on the pth variable of X_1 in regression of y on X_1 alone can be stated as:
  $\beta_{(1*p)\cdot X_1} = \beta_{(1*p)\cdot X_1 X_2} + \sum_{j=1}^{k} \alpha_{(2*j)(1*p)\cdot X_1} \beta_{(2*j)\cdot X_1 X_2}$
- $\beta_{(1*p)\cdot X_1}$ is the coefficient on the pth variable of X_1 in the regression of y on X_1

Standard approach to bias

- X is partitioned into X_1 $(n \times p)$ and X_2 $(n \times k)$
- The coefficient on the pth variable of X_1 in regression of y on X_1 alone can be stated as:
  $\beta_{(1*p)\cdot X_1} = \beta_{(1*p)\cdot X_1 X_2} + \sum_{j=1}^{k} \alpha_{(2*j)(1*p)\cdot X_1} \beta_{(2*j)\cdot X_1 X_2}$
- $\beta_{(1*p)\cdot X_1}$ is the coefficient on the pth variable of X_1 in the regression of y on X_1
- $\beta_{(1*p)\cdot X_1 X_2}$ is the coefficient on the same variable when X_2 is also included

Standard approach to bias

- X is partitioned into X_1 $(n \times p)$ and X_2 $(n \times k)$
- The coefficient on the pth variable of X_1 in regression of y on X_1 alone can be stated as:
  $\beta_{(1*p)\cdot X_1} = \beta_{(1*p)\cdot X_1 X_2} + \sum_{j=1}^{k} \alpha_{(2*j)(1*p)\cdot X_1} \beta_{(2*j)\cdot X_1 X_2}$
- $\beta_{(1*p)\cdot X_1}$ is the coefficient on the pth variable of X_1 in the regression of y on X_1
- $\beta_{(1*p)\cdot X_1 X_2}$ is the coefficient on the same variable when X_2 is also included
- The final summation is the bias term:
  - $\alpha_{(2*j)(1*p)\cdot X_1}$ is the coefficient on the pth variable of X_1 in the regression of the jth omitted variable on the included variables

Standard approach to bias

- X is partitioned into X_1 $(n \times p)$ and X_2 $(n \times k)$
- The coefficient on the pth variable of X_1 in regression of y on X_1 alone can be stated as:
  $\beta_{(1*p)\cdot X_1} = \beta_{(1*p)\cdot X_1 X_2} + \sum_{j=1}^{k} \alpha_{(2*j)(1*p)\cdot X_1} \beta_{(2*j)\cdot X_1 X_2}$
- $\beta_{(1*p)\cdot X_1}$ is the coefficient on the pth variable of X_1 in the regression of y on X_1
- $\beta_{(1*p)\cdot X_1 X_2}$ is the coefficient on the same variable when X_2 is also included
- The final summation is the bias term:
  - $\alpha_{(2*j)(1*p)\cdot X_1}$ is the coefficient on the pth variable of X_1 in the regression of the jth omitted variable on the included variables
  - $\beta_{(2*j)\cdot X_1 X_2}$ is the coefficient on the jth omitted variable in the regression of y on X_1 and X_2

Recursive decompositions approach to bias

- Goldberger (1991, p.186), after introducing the standard approach to bias, notes:
  "The situation here is quite reminiscent of the distinction between partial and total derivatives in calculus. Indeed $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{1\cdot 2} + F\boldsymbol{\beta}_{2\cdot 1}$ has the same pattern as $\frac{dy}{dx_1} = \frac{\partial y}{\partial x_1} + (\frac{dx_2}{dx_1})(\frac{\partial y}{\partial x_2})$."
- Recursive decompositions approach shows that Goldberger's observation is strictly correct
- Uses the chain rule, treating each RHS variable as a function of its covariates

Recursive decompositions approach – General process

1. Construct the regression's decomposition tree

Recursive decompositions approach – General process

1. Construct the regression's decomposition tree
2. Substitute all decompositions of omitted variables into their parent regressions, such that all are substituted into the primary regression

Recursive decompositions approach – General process

1. Construct the regression's decomposition tree
2. Substitute all decompositions of omitted variables into their parent regressions, such that all are substituted into the primary regression
3. Differentiate with respect to the covariate of interest

Recursive decompositions approach – General process

1. Construct the regression's decomposition tree
2. Substitute all decompositions of omitted variables into their parent regressions, such that all are substituted into the primary regression
3. Differentiate with respect to the covariate of interest

- With P included variables, K omited variables, and covariate of interest x_1, this process obtains the relation of $\beta_{1 \cdot P}$ to $\beta_{1 \cdot PK}$
- Algebraically equivalent to the standard approach

Recursive decompositions approach – Example

- Regression of y on x_1, x_2, x_3, and x_4
  $\hat{y} = \beta_{1 \cdot 234} x_1 + \beta_{2 \cdot 134} x_2 + \beta_{3 \cdot 124} x_3 + \beta_{4 \cdot 123} x_4$
- The bias on the coefficient for x_1 from omitting x_3 and x_4 can be found by...

Recursive decompositions approach – Example

- Regression of y on x_1, x_2, x_3, and x_4
  $\hat{y} = \beta_{1\cdot 234}x_1 + \beta_{2\cdot 134}x_2 + \beta_{3\cdot 124}x_3 + \beta_{4\cdot 123}x_4$
- The bias on the coefficient for x_1 from omitting x_3 and x_4 can be found by...

$$
\begin{aligned}
\hat{y} =& \beta_{1\cdot 234}x_1 + \beta_{2\cdot 134}x_2 \\
&+ \beta_{3\cdot 124}(\alpha_{31\cdot 24}x_1 + \alpha_{32\cdot 14}x_2 + \alpha_{34\cdot 12}x_4) \\
&+ \beta_{4\cdot 123}(\alpha_{41\cdot 23}x_1 + \alpha_{42\cdot 13}x_2 + \alpha_{43\cdot 12}x_3) \\
\hat{y} =& \beta_{1\cdot 234}x_1 + \beta_{2\cdot 134}x_2 \\
&+ \beta_{3\cdot 124}(\alpha_{31\cdot 24}x_1 + \alpha_{32\cdot 14}x_2 + \alpha_{34\cdot 12}(\alpha_{41\cdot 2}x_1 + \alpha_{42\cdot 1}x_2)) \\
&+ \beta_{4\cdot 123}(\alpha_{41\cdot 23}x_1 + \alpha_{42\cdot 13}x_2 + \alpha_{43\cdot 12}(\alpha_{31\cdot 2}x_1 + \alpha_{32\cdot 1}x_2))
\end{aligned}
$$

Recursive decompositions approach – Example

$$
\begin{aligned}
\frac{dy}{dx_{1\cdot 2}} = \beta_{1\cdot 2} = &\beta_{1\cdot 234} \\
&+ \beta_{3\cdot 124}(\alpha_{31\cdot 24} + \alpha_{34\cdot 12}\alpha_{41\cdot 2}) \\
&+ \beta_{4\cdot 123}(\alpha_{41\cdot 23} + \alpha_{43\cdot 12}\alpha_{31\cdot 2}) \\
\beta_{1\cdot 2} = &\beta_{1\cdot 234} + \beta_{3\cdot 124}\alpha_{31\cdot 2} + \beta_{4\cdot 123}\alpha_{41\cdot 2}
\end{aligned}
$$

- Simplifies to the result of the standard approach
- Extends derivative of a regression function to the total derivative

Iterated FWL approach to bias

- Computes coefficients in terms of previously-estimated coefficients
- First-partitioned variable's coefficient is always calculated with its bivariate coefficient minus a bias term
  $\beta_{3\cdot 12} = \beta_3 - \beta_{1\cdot 23}\alpha_{13} - \beta_{2\cdot 3}\alpha_{23} - \beta_{1\cdot 23}\alpha_{12\cdot 3}\alpha_{23}$
- Same computed result as standard approach, but not easy to simplify to it

Conclusion

- While not straightforward, provides more granular perspective on how the standard result is formed

Conclusion

- While not straightforward, provides more granular perspective on how the standard result is formed
- Contribute to the discussion around the algebra of misspecification in OLS by:
  - Extending the derivative of a regression equation to the total derivative
  - Showing that the result is equivalent to the standard approach to representing bias

Conclusion

- While not straightforward, provides more granular perspective on how the standard result is formed
- Contribute to the discussion around the algebra of misspecification in OLS by:
  - Extending the derivative of a regression equation to the total derivative
  - Showing that the result is equivalent to the standard approach to representing bias
- Recursive view of FWL decompositions holds true – the FWL theorem extends recursively