MAS 420: Frisch-Waugh-Lowell Theorem and Omitted Variable Bias

Adam Rohde
Department of Statistics
University of California, Los Angeles

November 7, 2022

Recommended resources

Useful resources

1. Understanding the Frisch-Waugh-Lovell Theorem - Courthoud (2022) [link]
2. Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis - Lovell (1963) [link]
3. A Simple Proof of the FWL Theorem - Lovell (2008) [link]
4. Making sense of sensitivity: extending omitted variable bias - Cinelli and Hazlett (2020) [link]

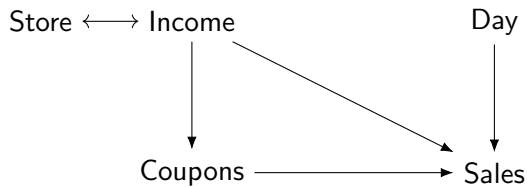
Running example

Suppose we run a chain of 10 grocery stores.
We decide to try to increase sales by providing discounts using coupons.
Coupons are distributed and we observe the share of shoppers that use coupons at each of our stores on each day of the week for a single week.

We want to know whether shoppers using the coupons changed sales.
We suspect that higher income shoppers tend to use the the coupons less but spend more.
So we also record average incomes in the area for each store.

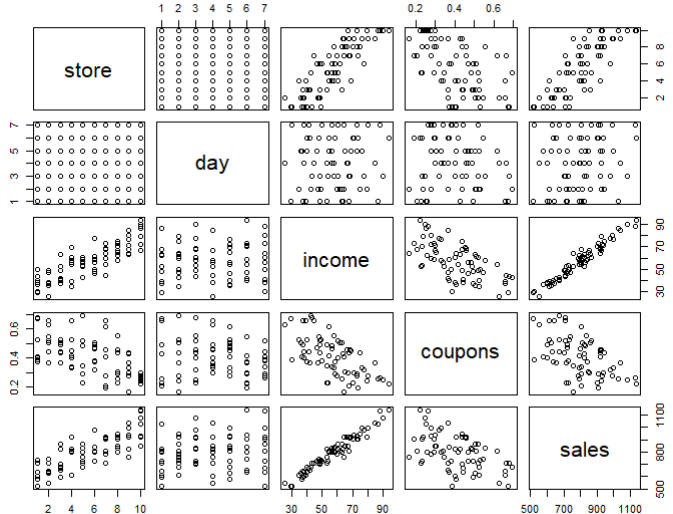
```
n = 70
store = sort(rep(1:10, times=7)) # store ID
day = rep(1:7, times=10) # day ID
income = 5*store + rnorm(n,30,10) # income in 1,000s
coupons = -0.005*income + rnorm(n,0.7,0.1) # pct shoppers using coupon
sales = 200*coupons + 10*income + 10*day + 10*rnorm(n,10,2) # sales
```

Running example



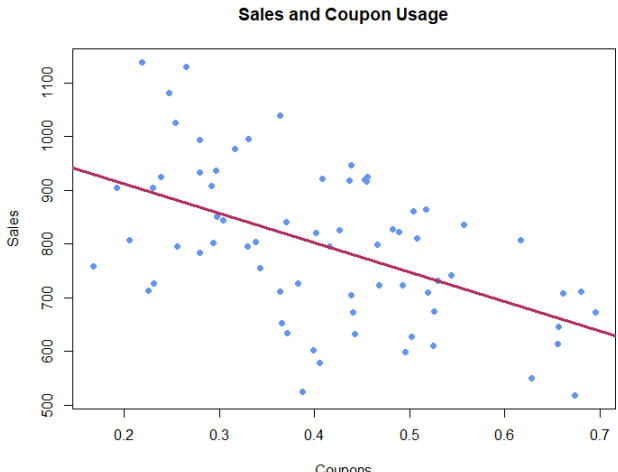

```
n = 70
store = sort(rep(1:10, times=7)) # store ID
day = rep(1:7, times=10) # day ID
income = 5*store + rnorm(n,30,10) # income in 1,000s
coupons = -0.005*income + rnorm(n,0.7,0.1) # pct shoppers using coupon
sales = 200*coupons + 10*income + 10*day + 10*rnorm(n,10,2) # sales
```

Running example



Sales and Coupon Usage

Seems like coupons hurt sales. But we're not adjusting for income.



Sales and Coupon Usage

Lets try including income and day in our regression.

```
# regress sales on coupons
m1 = lm(sales ~ coupons)
# regress sales on coupons and income
m2 = lm(sales ~ coupons + income)
# regress sales on coupons, income, and day
m3 = lm(sales ~ coupons + income + day)
# create nice latex table
stargazer(m1, m2, m3, align=TRUE)
```

Sales and Coupon Usage

	Dependent variable:		
	sales		
coupons	−547.094***	177.014***	208.353***
	(110.139)	(30.828)	(23.860)
income		9.731***	9.814***
		(0.262)	(0.200)
day			8.711***
			(1.234)
Observations	70	70	70
R^2	0.266	0.966	0.981
Note:			*p<0.1; **p<0.05; ***p<0.01

Frisch-Waugh-Lovell (FWL) Theorem

Suppose we have a model that takes the form

$$Y_i = \beta_1 X_{1,i} + \beta_2 X_{2,i} + \epsilon_i$$

Then the following estimators of β_1 are equivalent:

- OLS estimator from regressing Y on X_1 and X_2
  - $m_0 = \text{lm}(Y \sim X_1 + X_2)$
- OLS estimator from regressing Y on $\tilde{X}_1$, where $\tilde{X}_1$ is the residual from the regression of X_1 on X_2
  - $m_1 = \text{lm}(X_1 \sim X_2)$ and get $\tilde{X}_1 = \text{residuals}(m_1)$
  - $m_2 = \text{lm}(Y \sim \tilde{X}_1)$
- OLS estimator from regressing $\tilde{Y}$ on $\tilde{X}_1$, where $\tilde{Y}$ is the residual from the regression of Y on X_2 and $\tilde{X}_1$ is the residual from the regression of X_1 on X_2
  - $m_3 = \text{lm}(Y \sim X_2)$ and get $\tilde{Y} = \text{residuals}(m_3)$
  - $m_1 = \text{lm}(X_1 \sim X_2)$ and get $\tilde{X}_1 = \text{residuals}(m_1)$
  - $m_4 = \text{lm}(\tilde{Y} \sim \tilde{X}_1)$

Sales and Coupon Usage

Let's test this out with our example.

```
m0 = lm(sales ~ coupons + income)
m1 = lm(coupons ~ income)
m3 = lm(sales ~ income)
coupons_tilde = residuals(m1)
sales_tilde = residuals(m3)
m2 = lm(sales ~ coupons_tilde)
m4 = lm(sales_tilde ~ coupons_tilde)
stargazer(m0, m2, m4, align=TRUE)
```

Sales and Coupon Usage

	Dependent variable:		
	sales	sales	sales_tilde
coupons	177.014***		
	(30.828)		
coupons_tilde		177.014	177.014***
		(164.592)	(30.600)
income	9.731***		
	(0.262)		
Observations	70	70	70
R^2	0.966	0.017	0.330
Note:			*p<0.1; **p<0.05; ***p<0.01

Partialling out

So what is really happening when we regress sales on income and coupons on income before running the regression of the residuals on the residuals?

The process is referred to as "partialling out" or "residualizing" or "orthogonalization". When we residualize, we remove the linear relationship between sales and income (and between coupons and income) and just keep the residuals. This means that the variation in sales (or coupons) explained by income is removed; leaving only some remaining (or residual) variation; the variation in the residuals.

Partialling-out removes the portion of sales that is driven by income and the portion of coupons that is driven by income. Since income is a confounder of the effect of coupons on sales (and everything is linear), doing this leaves us with the causal effect of coupons on sales. This is exactly what regression does when we regress sales on coupons and income (i.e., the coefficient tells you how much the dependent variable increases when the independent variable increases by one, holding other independent variables constant).

Why are “middle” standard errors larger?

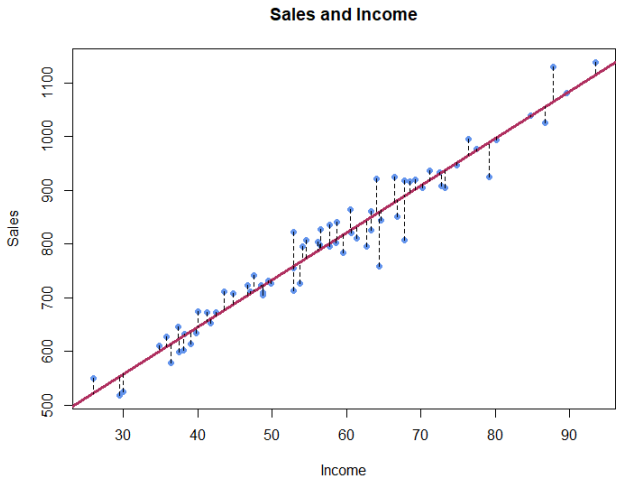
We’re using income to explain coupons but not to explain sales.

So this is sort of like adjusting for a covariate that only explains the treatment, in that it eats up the variation in coupons (the treatment) without eating up any of the variation in sales (the outcome).

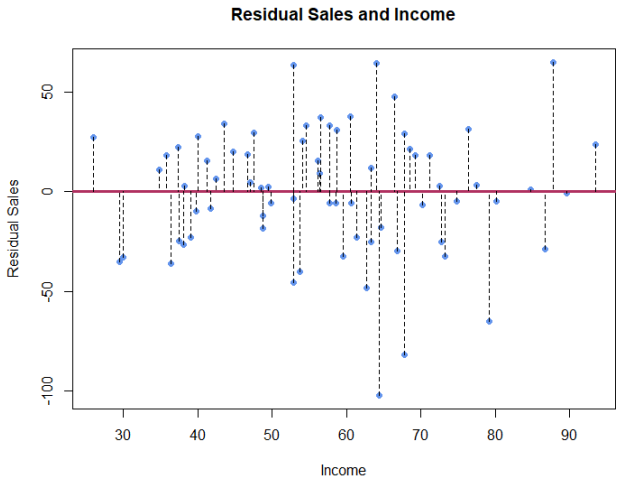
It will break the connection between coupons and income giving us an unbiased estimate, but it will typically hurt precision.

So the better approach is to residualize both the treatment and the outcome.

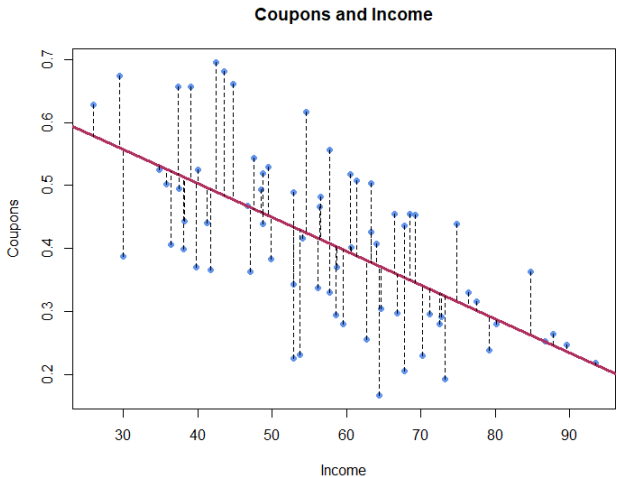
Partialling out



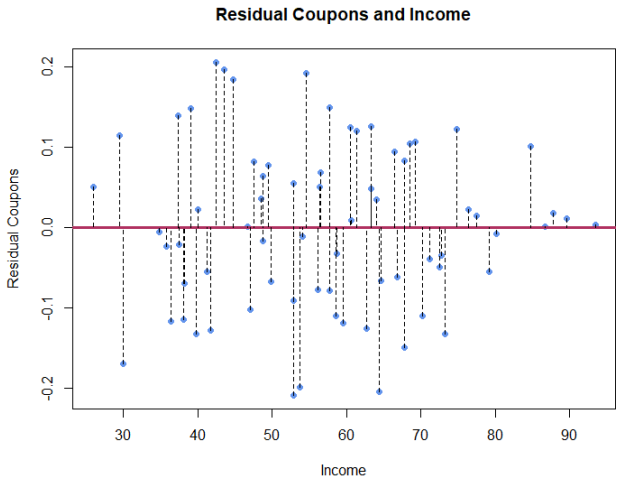
Partialling out



Partialling out

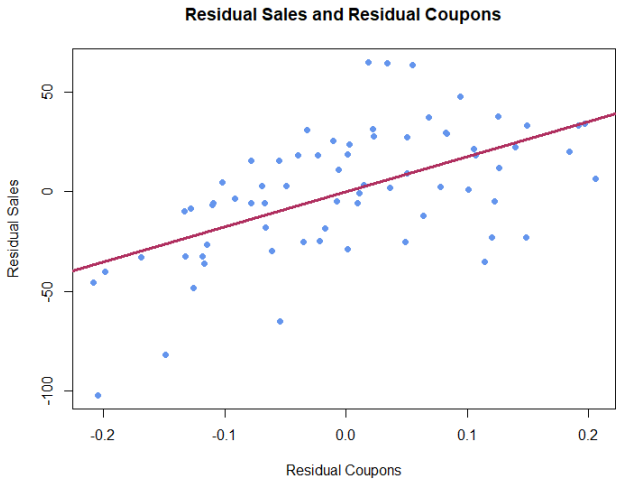


Partialling out



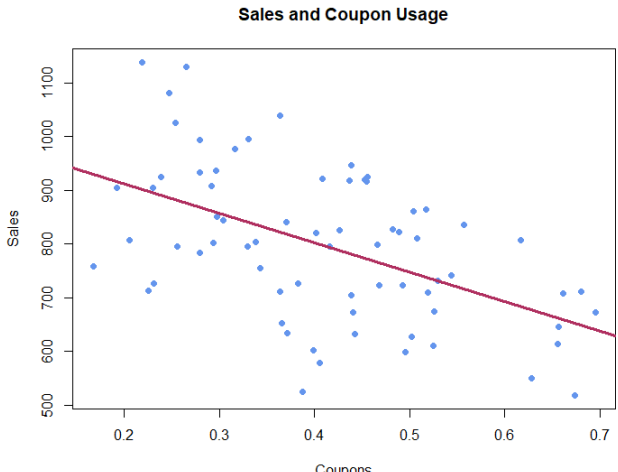
Causal relationship

So this gives us the causal relationship.



Biased relationship

Recall this is the biased relationship, where we don't partial out the effect of income.



Proof sketch of FWL theorem

Consider the regression including both coupons and income:

$$\text{Sales} = \hat{\alpha}\mathbf{1} + \hat{\beta}\text{Coupons} + \hat{\gamma}\text{Income} + \hat{\epsilon} \tag{1}$$

Also consider the partialling out regressions, where $\hat{\epsilon}_{\text{Sales}}$ and $\hat{\epsilon}_{\text{Coupons}}$ are the residualized Sales and Coupons:

$$\begin{aligned} \text{Sales} &= \hat{a_1}\mathbf{1} + \hat{a_2}\text{Income} + \hat{\epsilon}_{\text{Sales}} \\ \text{Coupons} &= \hat{b_1}\mathbf{1} + \hat{b_2}\text{Income} + \hat{\epsilon}_{\text{Coupons}} \end{aligned} \tag{2}$$

We want to show that regressing $\hat{\epsilon}_{\text{Sales}}$ on $\hat{\epsilon}_{\text{Coupons}}$ recovers $\hat{\beta}$.
Substituting (2) into (1), we get

$$\begin{aligned} [\hat{a_1}\mathbf{1} + \hat{a_2}\text{Income} + \hat{\epsilon}_{\text{Sales}}] &= \hat{\alpha}\mathbf{1} + \hat{\beta}\left[\hat{b_1}\mathbf{1} + \hat{b_2}\text{Income} + \hat{\epsilon}_{\text{Coupons}}\right] + \hat{\gamma}\text{Income} + \hat{\epsilon} \\ \implies \hat{\epsilon}_{\text{Sales}} &= \hat{\beta}\hat{\epsilon}_{\text{Coupons}} + \left(\hat{\alpha} + \hat{\beta}\hat{b_1} - \hat{a_1}\right)\mathbf{1} + \left(\hat{\gamma} + \hat{\beta}\hat{b_2} - \hat{a_2}\right)\text{Income} + \hat{\epsilon} \end{aligned} \tag{3}$$

Proof sketch of FWL theorem

$$\hat{\epsilon}_{\text{Sales}} = \hat{\beta}\hat{\epsilon}_{\text{Coupons}} + \left(\hat{\alpha} + \hat{\beta}\hat{b_1} - \hat{a_1}\right)\mathbf{1} + \left(\hat{\gamma} + \hat{\beta}\hat{b_2} - \hat{a_2}\right)\text{Income} + \hat{\epsilon} \tag{3}$$

- Recall that residuals from a least squares regression are uncorrelated with the explanatory variables. So $\hat{\epsilon}_{\text{Sales}}$ and $\hat{\epsilon}_{\text{Coupons}}$ are uncorrelated with Income and $\mathbf{1}$.
- Also recall that the coefficients on explanatory variables that are uncorrelated with both the dependant and explanatory variables are zero. So the coefficients on Income and $\mathbf{1}$ in (3) must be zero.
- Therefore, Income and $\mathbf{1}$ drop out of (3) and we get the residual on residual regression in (4), where we see that the coefficient on $\hat{\epsilon}_{\text{Coupons}}$ is the same as on Coupons in (1), namely $\hat{\beta}$.

$$\hat{\epsilon}_{\text{Sales}} = \hat{\beta}\hat{\epsilon}_{\text{Coupons}} + \hat{\epsilon} \tag{4}$$

Residuals uncorrelated with explanatory variables

Let the residuals from an OLS regression be $\hat{\epsilon}_i = Y_i - \hat{Y}_i = Y_i - X_i\hat{\beta}$.
The covariance between the $\hat{\epsilon}_i$s and X_is is (sample covariance formula)

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i X_i - \left[\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i\right]\left[\frac{1}{n}\sum_{i=1}^{n}X_i\right]$$

If the model includes an intercept then $\sum_{i=1}^{n}\hat{\epsilon}_i = 0$ and covariance becomes $\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i X_i$.

But the way we estimate OLS is by solving the normal equations
$0 = \mathbb{X}^\top(\mathbf{Y} - \mathbb{X}\hat{\beta}) = \mathbb{X}^\top(\mathbf{Y} - \hat{\mathbf{Y}}) = \mathbb{X}^\top\hat{\epsilon} = \sum_{i=1}^{n}\hat{\epsilon}_i X_i$.

So the covariance $\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i X_i = 0$ and so correlation is also zero.

Omitted variable bias

We can cast the confounding problem as omitted variable bias.
We initially ran the regression: $\text{Sales} = \hat{\alpha}_s + \hat{\beta}_s \text{Coupons} + \hat{\epsilon}_s$.
But we should have run the regression: $\text{Sales} = \hat{\alpha}_\ell + \hat{\beta}_\ell \text{Coupons} + \hat{\gamma}_\ell \text{Income} + \hat{\epsilon}_\ell$.
The income variable was omitted from the regression, which caused a biased estimate.

We can now try to understand the bias $\hat{\beta}_s - \hat{\beta}_\ell$.

$$\hat{\beta}_s = \frac{\widehat{\text{Cov}}(\text{Coupons}, \text{Sales})}{\widehat{\text{Var}}(\text{Coupons})} = \frac{\widehat{\text{Cov}}(\text{Coupons}, \hat{\alpha}_\ell + \hat{\beta}_\ell \text{Coupons} + \hat{\gamma}_\ell \text{Income} + \hat{\epsilon}_\ell)}{\widehat{\text{Var}}(\text{Coupons})}$$

$$= \frac{\hat{\beta}_\ell \widehat{\text{Cov}}(\text{Coupons}, \text{Coupons}) + \hat{\gamma}_\ell \widehat{\text{Cov}}(\text{Coupons}, \text{Income})}{\widehat{\text{Var}}(\text{Coupons})} = \hat{\beta}_\ell + \hat{\gamma}_\ell \frac{\widehat{\text{Cov}}(\text{Coupons}, \text{Income})}{\widehat{\text{Var}}(\text{Coupons})}$$

$$= \hat{\beta}_\ell + \hat{\gamma}_\ell \hat{\delta}, \text{ where } \hat{\delta} \text{ is the reg. coef. from } \text{lm}(\text{Income} \sim \text{Coupons})$$

Omitted variable bias

So we see that the bias from omitting the income variable from our regression can be written as

$$\hat{\beta}_{\mathbf{s}} - \hat{\beta}_{\ell} = \hat{\gamma}_{\ell}\hat{\delta}$$

which captures the relationship between sales and income ($\hat{\gamma}_{\ell}$) and between coupons and income ($\hat{\delta}$).
Both multiple regression (i.e., regressing sales on coupons and income) and partialling out are equivalent ways to remove the effect of income on coupons and of income on sales to estimate the effect of coupons on sales.

Omitted variable bias

	Dependent variable:		
	sales	sales	income
coupons	−547.094***	177.014***	−74.414***
	(110.139)	(30.828)	(11.053)
income		9.731***	
		(0.262)	
Observations	70	70	70
R^2	0.266	0.966	0.400
Note:			*p<0.1; **p<0.05; ***p<0.01

$\hat{\beta}_\ell = 177.014$ and $\hat{\beta}_s - \hat{\gamma}_\ell \hat{\delta} = 547.094 - (9.731)(74.414) = 177.014$

Any remaining time

questions / break