Philip and Sewall Wright: The Invention of Instrumental Variables Regression

Philip Wright

Sewall Wright
Philip Wright

- Bachelor degree from Tufts—1884
- MA in economics from Harvard—1887
- Professor Lombard College
- Instructor at Harvard under Professor W. Taussig
- US Tariff Commission
- Institute of Economics
March 4, 1926.

Dear Bevall:

It may interest you to see a very simple geometric demonstration which I have worked out for your unit of estimating supply and demand curves without reference to the theory of part coefficients.
Sewall Wright (1889-1988)

- Famous Geneticist
  - Dept. of Zoology at the University of Chicago (1925 to 1955)
- Undergrad: Lombard College
- ScD Harvard, Biology, 1915
- Philip Wright’s oldest son

Sewall Wright
Pictured 1928
Jan. 29, 1926

Dear Father,

Subject: Data different for

Ogive vs. correction to "C + C".

Paragraphs 1 & 2 on your letter I find express the situation very well as does it as well. If you add it to normal output curve, will you

1. I feel that the probable difference is electricity between having the

2. about two supply curves, neither a serious difference, application

3. actual case will come back to the later +

4.
Inventing IV Regression

- Big break through: Instrumental variables can be used to estimate the coefficient on an endogenous variable.

Who authored Appendix B?
• From Stock and Trebbi (2003) stylometric analysis, we know Philip wrote the actual text for appendix B.
Even if Philip wrote appendix B, who was responsible for the main ideas?
Case for Philip

• He authored Appendix B and he was deeply invested in the question as it was vital to his work at the time.
• He also had a deep understanding of the identification problem.
Case for Sewall

• After all, Sewall was the statistical mastermind—he had published “Correlation and Causation” in 1921 he also published a 1925 paper on corn and hog cycles.
  - In this paper the variables are all exogenous and not simultaneous.
  - Sewall had said that he did not know how to manage endogenous variables and simultaneous equations.
Who Invented IV Regression?

The letters that Philip and Sewall exchanged from late 1925 to 1926 shows the development of ideas.

Who invented IV Regression?
Collaborative.
Diagrams using factors

• Below are Phillip’s Jan 26, 1926 rendition of diagrams Sewall introduced to him on Dec 28, 1925
Basics of Path Analysis

• Path Analysis
  • \( X = aA + U \)
  • \( \sigma_x = 1, \sigma_A = 1; a = r_{XA} \)

• In Modern Notation
  \( \hat{\beta} = \hat{a} \frac{\sigma_A}{\sigma_X} \)
  \( \hat{\beta} = \frac{\sum ax}{\sqrt{\sum a^2 \sum x^2}} \times \sqrt{\frac{\sum a^2}{\sum x^2}} \)
  \( \hat{\beta} = \frac{\sum ax}{\sum x^2} \)
Applying Path Analysis to Philip’s work: Inventing the Method of Additional Factors

- From “Correlation and Causation”: a diagram when variables are not independent
The link between 1921 and 1926

• But, Sewall never solved this equation—instead he simplified it

• It was simply an example of what could exist in a system of correlations.

• Sewall finally solves the diagram to the left because Philip prompted the question
Derivation of the IV regressor with Path Analysis

- By definition of elasticity
  - \( e = \frac{O}{P} \times \frac{\sigma_o}{\sigma_p} = \frac{r_{AO1}}{r_{AP1}} \times \frac{\sigma_o}{\sigma_p} \)
- To get \( O/P \), apply path coefficients

Thus
\[
e = \frac{r_{AO1}}{r_{AP1}} \times \frac{\sigma_o}{\sigma_p}
\]
Philip’s Response

“I expect I am stupid but I don’t seem to be able to pick up a new branch of mathematics or quickly, as I could once.”

-Feb. 13, 1926: Philip to Sewall
March 4, 1926.

Dear [Name]:

It may interest you to see a very simple geometric demonstration which I have worked out for your use of estimating supply and demand curves without reference to the theory of price coefficients.

\[
\begin{align*}
\eta &= \frac{a-b}{p} \quad \text{(linear \( \eta \) is negative)} \\
B &= \text{factor uncorrelated with } D
\end{align*}
\]

P = price, D in demand, S is supply under mean price, and B is demand under mean price, all expressed as percentage deviations from mean.

then \( e = \frac{1}{1-p} \)

A is factor uncorrelated with S

\[
\begin{align*}
e_P &= o_s - s, \\
e_A, P &= A_o - A_s, \\
e_A, P &= A_o - A_s \\
e_A, P &= A_o - A_s \\
\sum e_A P &= \sum o_s - \sum A_s \\
e &= \frac{\sum o_s}{\sum A_P}
\end{align*}
\]
Philip’s Derivation

“I have worked out for your math of estimating supply and demand curves without reference to the theory of path coefficients.”
A is factor uncorrelated with $S$

\[ e_{P_1} = 0, -S, \]
\[ e_{A_1 P_1} = A, 0, -A, S, \]
\[ e_{A_1 P_2} = A_1, 0, -A_1, S_2 \]
\[ e_{A_1 P_3} = A_3, 0, -A_3, S_3 \]

\[ e_{\Sigma AP} = \Sigma AO - \Sigma AS \]

\[ \therefore e = \frac{\Sigma AO}{\Sigma AP} \]
“I discovered it by a singularly round about process.”
Introducing Error

“I should suppose ... with a finite number of observations [ΣAS] would seldom be precisely 0. If an expression could be developed showing the value of the most probable deviation from 0 if this product it would be of the nature of a ‘probable error.’”

– March 6, 1926: Philip to Sewall

• \( e = (ΣAO-\text{probable error})/(ΣAP) \)
Testing his method

- March 4: a simulation for known elasticities to test his formula—basically a one draw Monte Carlo Simulation.
- March 15: Estimate elasticity of demand for spring wheat
  - Finds elasticity of demand of -0.88
  - Uses building permits as the instrument
  - Concludes this “a result obviously absurd”
Philip’s Data

• 22 observations: years 1903 to 1924
• Real Price, Wheat Output (millions of bushels)
• Variables:
  – Acreage, yield per acre, rainfall (inches), ratio of harvest of flaxseed to spring wheat, building permits
• Uses deviations of data from straight line trends.
Philip’s Test

- Output_Dev = -0.72 price_dev (0.96)
- Instrument: building permits
- First stage F-statistic: 1.31
- A weak instrument by modern standards
“In the case of any specific commodity is it possible to find factors which have such distinct caused relations with output or demand conditions?”

“Such factors, I fear, especially in the case of demand conditions, are not easy to find.”

-Philip: March 15, 1926
Using one of Philip’s different variables

- Output_Dev = -0.48 price_dev
  \( (0.74) \)
- Instrument: rainfall
- First stage F-statistic: 12.66
- Not a weak instrument by modern standards
Additional Diagrams
2SLS

- Ideological groundwork for 2SLS

First Stage of 2SLS

Second Stage of 2SLS
Summary: Philip and Sewall’s Contribution

- Introduced the idea of additional factors
- 2 separate derivations that lead to the IV estimator
- Ideological framework for 2SLS
- A first draw Monte Carlo simulation
- Empirical Applications
- Further discussion of weak variables
What Next?
What next?

“When I was in Cambridge, I spoke to Dr. Taussig about my study of supply and demand curves and he suggested that I prepare an article for the Quarterly. I hope I can handle it so as to do justice to your analysis—I think it will prove the most valuable part of the article.”

– Philip to Sewall, February 9, 1926
Philip’s submission for the Aug. 1926 edition was rejected.

Taussig was the editor at the time.
Conclusions

• Philip used the IV estimator in his work for his 1928 book and included the findings on IV regression in Appendix B.

• He included both Sewall’s and his own derivations.

• They never solved a multiple factor of variables diagram to get the 2SLS equation, though the ideological framework was present and understood in the 1926 correspondence.
Philip to Sewall: March 15, 1926

“With all your anxieties and problems connected with your new position I am afraid I have been asking too much.”
“My only excuse is that it is a mathematical problem which interests me very much and which I have thought you also might find not without interest.”
“I hope all this work you have put into it is not going to prove obsolete. I hope we shall be able to see some results in practical application.”