Econometrics for Climate Change

Session 3 — Linear Regression & Applications to Climate Change

John Gómez-Mahecha

2026-03-09

Let us remember…

Use statistical methods

Predict

A causal relation between two variables

A predicted value of a variable

… for answering numerically a theoretical question

Let us remember...

Type of data

Time →

Unit ↓

	1
1	X_1
2	X_2
...	
i	X_i
...	
n	X_n

Time →

Unit ↓

	1	2	...	t	...	T
1	X_{11}	X_{12}		X_{t1}		X_{T2}
2	X_{12}	X_{22}		X_{t2}		X_{T2}
...	...	...		...		
i	X_{1i}	X_{2i}	...	X_{ti}	...	X_{Ti}
...	...	...		...		...
n	X_{1n}	X_{2n}	...	X_{tn}	...	X_{Tn}

Let us remember…

Every random variable Y can be represented as

$$Y = E[Y|X] + u$$

where

- Y is known as the **dependent variable**, outcome variable or regressand
- X is usually know as **independent variables** or regressor

The linear regression model assumes a linear form for $E[Y|X]$. That is

$$E[Y|X] = \beta_0 + \beta_1 X$$

Let us remember…

We need an **estimator** for the linear regression model

Concept: Ordinary Least Square Estimators for the Linear Regression

The estimators $(\hat{\beta}_0, \hat{\beta}_1)$ are the OLS estimator if they are the values that solve for

$$Min_{\{\hat{\beta}_j\}} \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2$$

where $\hat{Y}_i \equiv \hat{\beta}_0 + \hat{\beta}_1 X_i$.

Let us remember

- Can OLS be use for any type of variables?
- Does it has to be always linear?
- Does it has to be only one X?

How good a model is will depend on our purpose

Use statistical methods

Predict

A causal relation between two variables

A predicted value of a variable

Which are the properties of the estimator of the β associated to the variable?

… for answering numerically a theoretical question

How good is the fit?

How good a model is will depend on our purpose

- How good is the fit?
  - **R2**: Variance of the variable Y that can be explained by the model's predictions (i.e., $\hat{Y}$)
  - **Adjusted R2** to account for the effect of adding variables to the model
- Properties of our estimator?
  - Which are the properties of an estimator?

Properties of the estimators

Concept: Gauss Markov Assumptions

1. No multicollinearity
2. Homoskedasticity
3. iid sample
4. Linear model
5. Exogeneity (Independence of X from u)

Then, $\hat{\beta}$ will be **unbiased**, **efficient**, **consistent** and we will know its distribution.

The exogeneity assumption

We know that the population model can be written as:

$$Y_i = E[Y_i|X_i] + u_i$$

Under the linear regression model we assume a particular form for $E[Y|X]$. Let us called it $m(X_i) = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots$.

Accounting for that

$$Y_i = m(X_i) + \epsilon_i$$

We do not longer know the properties of ϵ_i, although we can always build $m(X_i)$ so that $E[\epsilon_i] = 0$.

When $\mathrm{E}[\epsilon_i|X_i] \neq E[\epsilon_i]$ we say that the model is **endogenous**.

The exogeneity assumption

- Endogeneity is the same to say:
    - … violation of the exogenity assumption
    - … **confoudedness**
    - … **identification** issues
- The main consequences of exogeneity are:
    - Bias
    - Inconsistency
    - … hypothesis testing
- The three main sources
    1. Omitted variables
    2. Simultaneity
    3. Measurement error

Omitted Variable Bias

- Let us supposed you are interest in finding the effect of X_1 in Y
- You estimate a linear regression of Y on X_1:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + e_i$$

- But the real form for $E[Y|X] = \beta_0 + \beta_1 X_1 + \beta_2 X_2$.
- Notice that $e_i = \beta_2 X_2 + u_i$

If X_1 is correlated with X_2 it will introduce bias:

$$E[\beta_1] = \beta_1 + \beta_2 corr(X_1, X_2)$$

What if X_1 is not correlated with X_2?

Example: Mendelshon, Nordhaus and Shaw (1994)

- One of the first examples for quantifying the economic costs of Weather
- Furthermore, one of the first example of using an **aggregate**
- Data for the US's 1982 and 1978 Census of Agriculture
- Unit of observation: US counties
- The outcome variable is the average farmland values
- The independent variables of interest are temperature and precipitation
- Covariates: Soil characteristics and economic variables

The Impact of Global Warming on Agriculture: A Ricardian Analysis

Author(s): Robert Mendelsohn, William D. Nordhaus and Daigee Shaw

Source: *The American Economic Review* , Sep., 1994, Vol. 84, No. 4 (Sep., 1994), pp. 753-771

Published by: American Economic Association

Stable URL: https://www.jstor.org/stable/2118029

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Example: Mendelshon, Nordhaus and Shaw (1994)

TABLE 3—REGRESSION MODELS EXPLAINING FARM VALUES

	Cropland weights			Crop-revenue weights	
Independent variables	1982 (i)	1982 (ii)	1978 (iii)	1982 (iv)	1978 (v)
Constant	1,490	1,329	1,173	1,451	1,307
	(71.20)	(60.18)	(57.95)	(46.36)	(52.82)
January temperature	−57.0	−88.6	−103	−160	−138
	(6.22)	(9.94)	(12.55)	(12.97)	(13.83)
January temperature squared	−0.33	−1.34	−2.11	−2.68	−3.00
	(1.43)	(6.39)	(11.03)	(9.86)	(14.11)
April temperature	−137	−18.0	23.6	13.6	31.8
	(10.81)	(1.56)	(2.23)	(1.00)	(2.92)
April temperature squared	−7.32	−4.90	−4.31	−6.69	−6.63
	(9.42)	(7.43)	(7.11)	(9.44)	(11.59)
July temperature	−167	−155	−177	−87.7	−132
	(13.10)	(14.50)	(18.07)	(6.80)	(12.55)
July temperature squared	−3.81	−2.95	−3.87	−0.30	−1.27
	(5.08)	(4.68)	(6.69)	(0.53)	(2.82)
October temperature	351.9	192	175	217	198
	(19.37)	(11.08)	(11.01)	(8.89)	(9.94)
October temperature squared	6.91	6.62	7.65	12.4	12.4
	(6.38)	(7.09)	(8.93)	(12.50)	(15.92)
January rain	75.1	85.0	56.5	280	172
	(3.28)	(3.88)	(2.81)	(9.59)	(7.31)
January rain squared	−5.66	2.73	2.20	−10.8	−4.09
	(1.86)	(0.95)	(0.82)	(3.64)	(1.72)
April rain	110	104	128	82.8	113
	(4.03)	(4.44)	(5.91)	(2.34)	(4.05)
April rain squared	−10.8	−16.5	−10.8	−62.1	−30.6
	(1.17)	(1.96)	(1.41)	(5.52)	(3.35)
July rain	−25.6	−34.5	−11.3	−116	−5.28
	(1.87)	(2.63)	(0.94)	(6.06)	(0.34)
July rain squared	19.5	52.0	37.8	57.0	34.8
	(3.42)	(9.43)	(7.54)	(8.20)	(6.08)
October rain	−2.30	−50.3	−91.6	−124	−135
	(0.09)	(2.25)	(4.45)	(3.80)	(5.15)
October rain squared	−39.9	2.28	0.25	171	106
	(2.65)	(0.17)	(0.02)	(14.17)	(11.25)
Income per capita		71.0	65.3	48.5	47.1
		(15.25)	(15.30)	(6.36)	(7.39)
Density		1.30	1.05	1.53	1.17
		(18.51)	(16.03)	(18.14)	(17.66)
Density squared		-1.72×10^{-4}	-9.33×10^{-5}	-2.04×10^{-4}	-9.38×10^{-5}
		(5.31)	(3.22)	(7.47)	(4.57)
Latitude		−90.5	−94.4	−105	−85.8
		(6.12)	(6.95)	(5.43)	(5.33)
Altitude		−0.167	−0.161	−0.163	−0.149
		(6.09)	(6.41)	(4.72)	(5.20)
Salinity		−684	−416	−582	−153
		(3.34)	(2.20)	(2.59)	(0.81)
Flood-prone		−163	−309	−663	−740
		(3.34)	(6.98)	(8.59)	(11.99)
Wetland		−58.2	−57.5	762	230
		(0.47)	(0.51)	(4.41)	(1.72)
Soil erosion		−1,258	−1,513	−2,690	−2,944
		(6.20)	(8.14)	(8.21)	(11.23)
Slope length		17.3	13.7	54.0	30.9
		(2.91)	(2.49)	(6.24)	(4.54)
Sand		−139	−35.9	−288	−213
		(2.72)	(0.77)	(4.16)	(3.95)
Clay		86.2	67.3	−7.90	−18.0
		(4.08)	(3.47)	(0.22)	(0.63)
Moisture capacity		0.377	0.510	0.206	0.450
		(9.69)	(14.21)	(3.82)	(10.07)
Permeability		−0.002	−0.005	−0.013	−0.017
		(1.06)	(2.53)	(5.58)	(8.61)
Adjusted R^2:	0.671	0.782	0.784	0.836	0.835
Number of observations:	2,938	2,938	2,941	2,941	2,941

Notes: The dependent variable is the value of land and buildings per acre. All regressions are weighted. Values in parenthesis are t statistics.

Example: Mendelshon, Nordhaus and Shaw (1994)

- Paradoxically, they find that climate change could improved agriculture production.
  - It will be discredited
- Still some ideas that remain even today:
  - Complex relations
  - Winner and losers

The Impact of Global Warming on Agriculture: A Ricardian Analysis

Author(s): Robert Mendelsohn, William D. Nordhaus and Daigee Shaw

Source: *The American Economic Review* , Sep., 1994, Vol. 84, No. 4 (Sep., 1994), pp. 753-771

Published by: American Economic Association

Stable URL: https://www.jstor.org/stable/2118029

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Example: Schlenker, Hanemann, and Fisher (2005)

- A critic to Mendelsohn and coauthors
- A missing variable: irrigation

In a simplified structure:

$$Land\ value_i = \beta_0 + \beta_1 Temperature_{i,1} + e_i$$

But if the real model is

$$Land\ value_i = \beta_0 + \beta_1 Temperature_{i,1} + \beta_2 Irrigation_i + u_i$$

Then the bias is given by

$$E[\beta_1] - \beta_1 = \beta_2 corr(Temperature, Irrigation)$$

What to do against omitted variables

- Look for the missing variables (?)
- Look for other data structures: Panel model

$$Y_{it} = \beta_0 + \beta_1 X_{it,1} + \cdots + e_{it}$$

- The advantage is that the error can be seprarated in three parts

$$Y_{it} = \beta_0 + \beta_1 X_{it,1} + \cdots + \mu_i + \gamma_t + \tilde{e}_{it}$$

If the missing <u>relevant</u> part is (1) constant in time, (2) or a common temporal shock, the panel data will be able to capture it.

How to estimate a panel model?

- If we are interest on using panel data to prevent omitted variables we usually use two estimations strategies

Option 1 – differentiate

$$Y_{it} = \beta_0 + \beta_1 X_{it,1} + \cdots + \mu_i + \gamma_t + \tilde{e}_{it}$$

(substract $Y_{i,t-1}$)

$$\underbrace{Y_{it} - Y_{i,t-1}} = \delta_0 + \beta_1 \underbrace{\left(X_{it,1} - X_{i,t-1,1}\right)} + \cdots + \varepsilon_{it}$$

$$\Delta Y_{it} = \delta_0 + \beta_1 \Delta X_{it,1} + \cdots + \varepsilon_{it}$$

Estimate through OLS (with corrected errors)

How to estimate a panel model?

- If we are interest on using panel data to prevent omitted variables we usually use two estimations strategies

Option 2 – Fixed effect

$$Y_{it} = \beta_0 + \beta_1 X_{it,1} + \cdots +$$
$$1(Ind = i) +$$
$$1(time = t) +$$

Estimate through OLS

Example: Hsiang (2010)

- How **short-run weather shocks affect economic output** using panel data
- Main conclusion: A **temporary increase of 1 °C in temperature reduces economic output by about 2.5%.**

PNAS

Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America

Solomon M. Hsiang[1]

Columbia University, School of International and Public Affairs, New York, NY 10027

Communicated by Robert M. Solow, Massachusetts Institute of Technology, Cambridge, MA, June 30, 2010 (received for review January 29, 2010)

Example: Hsiang (2010)

- Main conclusion: A temporary increase of 1 °C in temperature reduces economic output by about 2.5%.
- Largest impacts occur outside agriculture

$$
\begin{aligned}
&V_{\{it\}} \\
&= \rho_1 V_{\{i,t-1\}} + \rho_2 V_{\{i,t-2\}} + \beta_T T_{\{it\}} + \beta_C C_{\{it\}} + \beta_R R_{\{it\}} \\
&+ fixed\text{ effects} + \varepsilon_{\{it\}}
\end{aligned}
$$

Table 1. Effect of annual average surface temperature on production (1970–2006)

Industry	%Δ/+1 °C	SE	% output
Total production	–2.5%**	[1.0]	—
Wholesale, retail, restaurants and hotels	–6.1%***	[1.7]	20.4
Other services	–2.2%**	[1.1]	35.0
Transport and communications	–2.2%	[1.7]	10.7
Construction	–0.6%	[3.1]	7.4
Manufacturing	+1.4%	[2.6]	12.0
Agriculture, hunting and fishing	–0.8%	[2.5]	10.5
Mining and utilities	–4.2%*	[2.4]	4.2

***$P < 0.01$, **$P < 0.05$, *$P < 0.1$.

Will panel always work?

- Short answer, NO. But it helps
- There are cases where it would not work
  - For example, what if your variable of interest is constant in time.
- … there are other forms of endogenity anyway.
- Other risks under panel:
  - Eliminate to much of the variance -> Therefore, the specification of the Fixed Effects matter

Why address endogeneity is so important?

It will allow us to claim **causality**!!

- What is causality?
  - Moving X will move Y, just because the movement of X.
  - It requires
    1. The ability to manipulate X
    2. Temporality: I move X then Y moves
    3. Correlation
  - Different operator

$$E[Y|X] \neq E[Y|do(X)]$$

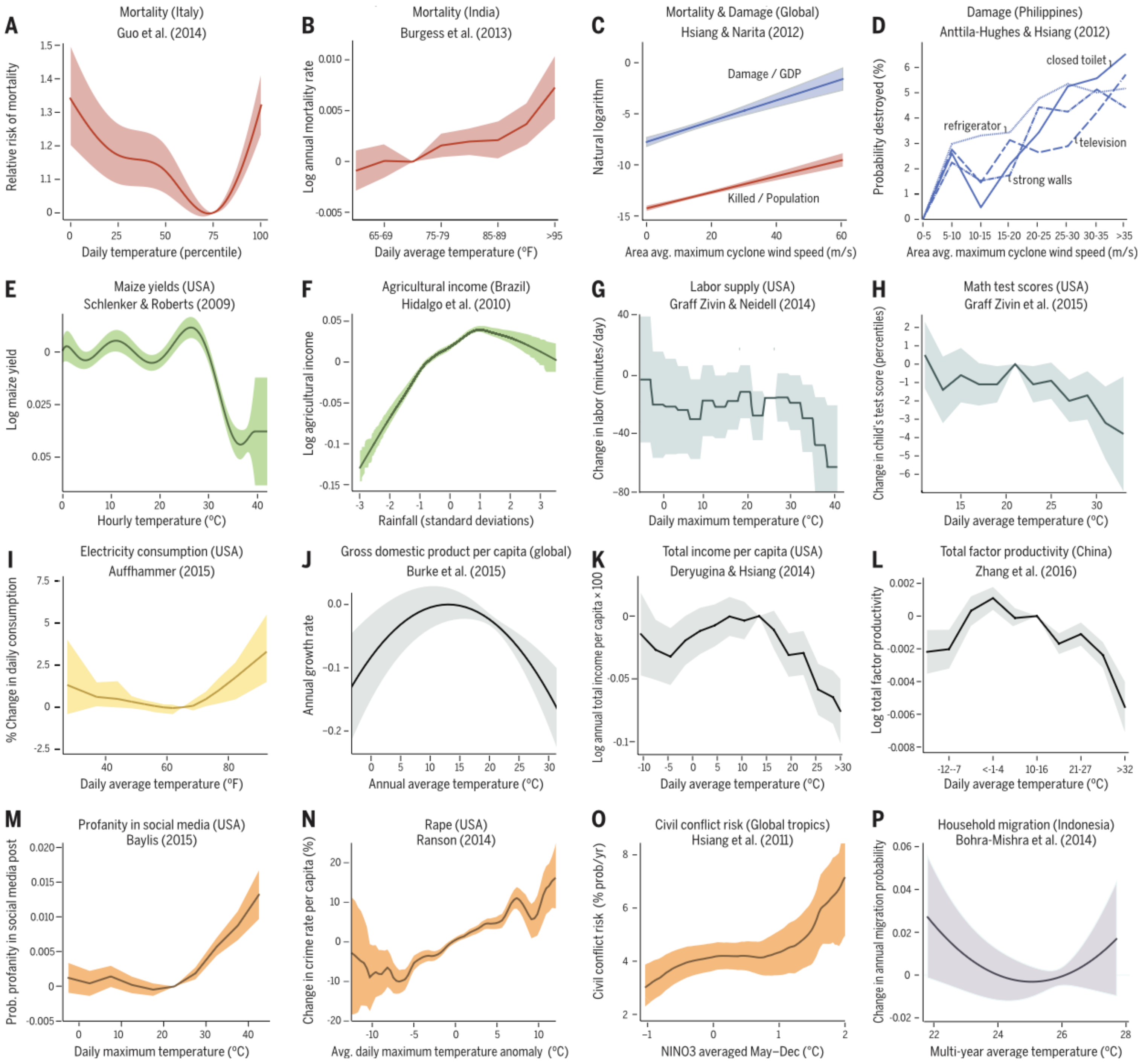
Carleton and Hsiang (2016)

REVIEW SUMMARY

SOCIAL SCIENCES

Social and economic impacts of climate

Tamma A. Carleton* and Solomon M. Hsiang*†



Last message

For your **projects**, you could consider statistical models to estimate an uncertain number that you need for the technology evaluation.

My suggestions:

- Can this uncertain number be express as the correlation between two variables? Are them observable?
- Can you estimate this with cross-sectional or a panel data?
- Can you write their relation in the form of al linear regression?
- What will be the ideal unit of observation?
- Will this regression violate the exogeneity assumption?