

Quantitative and Analytical Social Science

WMH00001 | Spring 2020

Week 1: Introduction

Week 2: Scientific Inference

Gaku Ito
Hiroshima University



April 15, 2020

Table of Contents

1 Introduction

- Course Description

2 Scientific Inference

- Scientific Inference and Social Science
- Take-Home Assignment 1: Your First Research Proposal
- (Relevant) Research Questions and Literature Review

Introduction



About Me

- ▶ A political scientist
- ▶ PhD in Advanced Social and International Studies (University of Tokyo)
- ▶ Assistant Professor in IDEC (Oct 2018 – present)
 - ▶ Graduate School of Humanities and Social Sciences (Apr 2020 – present)
- ▶ Research interests:
 - ▶ Substantive: Empirical study of IR, with primary focuses on the causes, dynamics, and outcomes of armed conflicts and long-run effects historical institutions.
 - ▶ Methodological: Event data and archival sources; and causal inference, geographic information system, and image analysis
- ▶ My teaching includes
 - ▶ International Relations (conflict/security studies)
 - ▶ Econometrics and statistical programming (not in IDEC)

Key motivations

- ▶ From “I’m interested in X” to social science research
- ▶ Introduction to social science research with a particular focus on the logics of scientific inference
 - ▶ methodological foundations and research designs to explore causal effects
- ▶ Primarily for students with limited methodological training

Key ideas

- ▶ **Social science** research \equiv scientific **inference**
 - ▶ **Inference** \equiv Learning about what we **do not observe** from what we **observe** (individual facts and events)
 - \rightsquigarrow What we do NOT observe...?
 - 1 **Descriptive** inference: **Systematic associations or dependence** between variables (typically correlations)
 - 2 **Causal** inference: **Causal effects** (and/or mechanisms)
- ▶ **Scientific inference**, descriptive or causal, does **NOT** have to be quantitative by definition
 - ▶ You can be **scientific** while being **qualitative**
 - ▶ Being **quantitative or analytical (or computational)** does **NOT necessarily** mean that you are being **scientific**

Key ideas (contd.)

- ▶ Quantitative, qualitative, and analytical approaches share the core logics of scientific inference

⇒ Different styles, but the same logic

- ▶ “Inference, whether descriptive or causal, quantitative or qualitative, is the ultimate goal of all good social science. . . . *the best way to understand a particular event may be by using the methods of scientific inference also to study systematic patterns in similar parallel events.*” (DSI, 34, 43)
- ▶ Concrete understanding of scientific inference is the prerequisite for all the graduate students in social science departments
- ▶ You need to understand the foundations of scientific inference regardless of your methodological approaches

Scientific inference helps your survival in graduate schools

Graduate school is a transition from being a private citizen taking classes to a public figure writing for a big amorphous, ill defined audience that it is your responsibility to define, find, and engage. This is not easy, and it accounts for most of the frustration scholars have with the peer-review process. . . . You'll have to revise more than you want and you thought possible. But try not to get discouraged; they call it **research**, not **search**, for a reason! Be your usual relentless self and get it done (King 2020, 3)

- ↪ No **research** can be done without making scientific inference
- ↪ Summarizing the details of something you are interested in may be called “search,” but not **research**
- ↪ We learn why, as well as how to engage in social science research

This course focuses on the logics of **scientific inference** and relevant **research designs**

- ▶ instead of going into the details of advanced econometric (and analytical) methods
- ▶ while also paying attention to the advances in qualitative studies
- ~> We study the role of counterfactual, identification problem, the roles of confounders and colliders, but not instrument variables, matching, regression discontinuity design, sensitivity analysis, etc.
- ~> We study the “experimental template” for qualitative studies, but not the (problems of) qualitative comparative analysis (QCA)

- ▶ Imai, Kosuke. 2017. *Quantitative Social Science: An Introduction*. Princeton: Princeton University Press.
 - ↪ Hereafter **QSS book**
 - ↪ Introduction to statistical causal inference
- ▶ Johnson, Janet Buttolph, H. T. Reynolds, and Jason D. Mycoff. 2020. *Political Science Research Methods*, 9th edition, Thousand Oaks: CQ Press.
 - ↪ Hereafter **PSRM book**
 - ↪ Introduction to empirical research methods (without heavy math)
- ▶ King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
 - ↪ Hereafter **DSI book**
 - ↪ A path-breaking book of social science research methods
- ▶ Other books, book chapters, and journal articles
 - ▶ See syllabus

Prerequisites

- ▶ Again, this is an introductory course
- ▶ Take this course if you **do NOT** understand the following quote

Confounders, colliders, and measurement error in causal inference

The causal structure of **confounding** can be summarized as the presence of **common causes** of treatment and outcome, and the causal structure of **selection bias** can be summarized as conditioning on **common effects** of treatment and outcome (or of their causes). **Measurement bias** arises in the presence of **measurement error**, but there is no single structure to summarize measurement error (Hernán, 2020: 114).

Prerequisites

- ▶ No formal prerequisites
- ▶ We will use four arithmetic operations (and a little bit more)
- ▶ Familiarity with introductory (undergraduate-level) econometrics, experimental methods, and game theory will be helpful, but not required

Requirements and Grade Policy

- ▶ Students are expected to do all the assigned readings, attend to every lecture, and participate in classroom discussion
- ▶ **Grades** will be based on
 - 1 participation in classroom discussion (30%);
 - 2 first research project proposal and in-class presentation (30%); and
 - 3 final research paper and presentation (40%). No exams are assigned
- ▶ Details on the requirements and expectations for the assignments will be announced 2–4 weeks before the deadlines
- ▶ Acts of cheating and plagiarism will be punished according to the Hiroshima University's policy

Requirements and Grade Policy

- ▶ Each week, we have a two- to three-period lecture followed by a one- or two-period classroom discussion.
- ▶ The length of the final research paper will be 3,000–5,000 words in English (or 6,000–10,000 characters in Japanese) and cover a relevant economic, social, or political phenomena
- ▶ Students present an initial research proposal in the next week with core research question(s), literature review, and tentative research design to the class (10–20 minutes), and “revise and resubmit” it as the final research paper with empirical analysis

Scientific Inference

Don'ts

In graduate schools

Q1: What are the details of this case?

Design 1: I go to field and ask some questions!

Q2: What are the challenges in this country?

Design 2: I read and summarize bunches of books and news articles!

Q3: What are the differences between these two cases?

Design 3: I read books and do some interview!

- ▶ These are all “don'ts,” **but why?**
- ▶ These are making **no scientific inference** (thus **NOT** a social science research)
- ▶ **NOT acceptable** at graduate and higher levels

Dos

In graduate schools

Q1: Does resource abundance cause civil war?

Design 1: I investigate the causal effects by comparing otherwise similar countries with and without oil.

Q2: Does economic development foster democratization?

Design 2a (THIS IS **NOT** ACCEPTABLE AS WE LEARN): I compile data and run regressions!

Design 2b: I exploit some “as-if random” shocks to economies to explore the causal effects.

- ▶ These are trying to make **scientific inference**
- ▶ Relevant questions are prerequisite for relevant research projects
- ▶ but relevant projects also require relevant research designs for scientific inference (**as we learn in this course**)

Scientific inference

- ▶ Learning about what we **do not observe** from what we **observe**
- ▶ “**Inference** is the process of using the **facts we know** to learn about **facts we do not know**” (*DSI*, 46)
 - ▶ What we **do** observe: Individual facts and specific events
 - ▶ “the **key distinguishing mark of scientific research** is the goal of **making inferences** that go **beyond the particular observations** collected” (8)
- ▶ Scientific inference = {**descriptive** inference, **causal** inference}
- ▶ Descriptive inference and causal inference are a specific form of scientific inference
 - ▶ They differ in the **target of inference** (what we do not observe)
- ▶ What we observe is individual facts and events, but **what we do NOT observe?**

Descriptive inference

- ▶ What we **do not** observe: **Systematic pattern(s)** behind realized events
- ▶ Inferring **systematic patterns/associations** from what we observe (individual facts and specific events)
 - ▶ **NOT** mere “description” or the collection of facts
 - ▶ **NOT** simply summarizing facts

e.g., Inferring a systematic association between economic development and democratization

e.g., Inferring a systematic association between oil abundance and civil war

Systematic features and nonsystematic forces (DSI, 55–56)

Naively, we might think that we were directly observing the electoral strength of the Conservatives [in UK] by recording their share of the vote by district and their overall share of seats. But a certain degree of **randomness or unpredictability** is inherent in politics, as in all of social life and all of scientific inquiry. . . . In these examples, the variance in the Conservative vote across districts. . . can be conceptualized as arising from two separate factors: **systematic and nonsystematic differences**. **Systematic differences** in our voter example include **fundamental and predictable** characteristics of the districts, such as differences in ideology, in income, in campaign organization, or in traditional support for each of the parties. In hypothetical weekly replications of the same elections, **systematic differences** would persist, but the **nonsystematic differences** such as turnout variations due to the weather, would vary. . . . With appropriate inferential techniques, we can usually learn about the nature of **systematic differences** even with the ambiguity that occurs in one set of real data due to **nonsystematic, or random, differences**.

Causal inference

- ▶ What we **do not** observe: **Causal effects (and counterfactuals)**
- ▶ Inferring **unobserved causal effects** from what we observe
 - ▶ “the process of learning about a causal effect from existing observations” (*DSI*, 90)

e.g., Inferring the effects of economic development on democratization

e.g., Inferring the effects of peacekeeping troops on civil war violence

e.g., Inferring the effects of aid provision on poverty in recipient countries

But how do we learn something about what we do not actually observe?

~> the key motivating question in this course

Why Inference?

- 1 We cannot observe everything in every detail
 - ▶ “Good archival work or well-done summaries of historical facts may make good descriptive history, but **neither are sufficient to constitute social science**” (*DSI*, 34)
 - ▶ “**Fundamental problem of causal inference**” (Holland 1986): We cannot observe counterfactuals or what did not happen even if we have infinite sample (→ **identification problem**)
- 2 What we do observe (or realized variables/phenomena) is a function of systematic features and nonsystematic factors (random errors)
 - ▶ Inherent randomness and unpredictability of social phenomena
 - ▶ We need to isolate the **systematic patterns** to explain any social phenomena (although nonsystematic components are also important)
- 3 We cannot evaluate the relevance of policy efforts without inferring the (likely) causal effects
 - ▶ Do you risk your savings for some policy prescriptions without evidence?

- ▶ Suppose that you are interested in 14 Japanese national elections during the 2000–2019 period
- ▶ What do we observe? What are the facts we do observe?
- ▶ What do we not observe? What do we need to infer (we do not observe)?
- ▶ What research questions would (not) be appropriate as a social scientific project? Raise some.

Your First Research Proposal

May 13 (ideally, the first off-line day) is YOUR session

- ▶ Prepare your first research proposal and present it to the class in 10–20 minutes with slides
- ▶ Share your research interests with the class — this also helps me preparing course materials
- ▶ Your proposal can be **very, very preliminary**, but is expected to be the initial version of your final research paper
 - ▶ Your presentation does not have to be methodologically sophisticated at this point
 - ▶ Limited knowledge and backgrounds are the main reason you are here
- ▶ But keep in mind the **“don’ts” in previous slides**
 - ▶ You are expected to engage in social science research
- ▶ Also, give some **comments and/or questions** to others

Your First Research Proposal

Requirements

- ▶ (Broadly) follow the template in van Evera (1997), Chapter 5 (optionally, with additional sections) when preparing your research proposal
- ▶ Chapter 3 of Johnson et al. (2020) is also helpful — if you are just like “I’m interested in X.”
- ↪ Your research design would remain incomplete, but try your best to clarify research question(s) and literature review
- ↪ A little bit more from me on this

Research Question

- ▶ Perhaps first-year students are feeling like “I’m interested in X”
 - ▶ “X” may be some social phenomena, and/or some case(s)
- ▶ Their research questions might look like:
 - ▶ What are the policy challenges in this country?
 - ▶ How many political parties are led by a woman in country Y?
 - ▶ What are the voter turnout for local elections in country Z?
 - ▶ Should Country Z sign the Treaty on the Prohibition of Nuclear Weapons?
- ▶ Unfortunately, these are **NOT acceptable** in graduate schools
 - ▶ Google tells you almost everything on these questions
 - ↪ Factual information, but **not scientific inference**
- ▶ Also note that **normative** questions and conclusions are **inconsistent with (positive) social science research**
 - e.g., Should country X democratize? Is democracy good? Are nuclear weapons bad?
 - ↪ Perhaps these questions are **important while being a private citizen**, but are **inconsistent** with positive social science research

Narrow factual issues or normative conclusions?

A research project will get off on the wrong foot if the question that shapes it fails to address a political phenomenon, is unduly concerned with **discrete facts**, or is focused on reaching **normative conclusions**. . . . Research questions, if they dwell on discrete or narrow factual issues, may limit the significance of a research project. Although important, **facts alone are not enough to yield scientific explanations** (*PSRM*, 50).

Research Question

- ▶ But what is missing in these questions?
- ▶ A **relationship** between variables: **Association, dependence, or covariance** of the values of one variable with the values of another
- ▶ **Exploring a (systematic) relationship** between variables requires **scientific inference**
 - ▶ **Descriptive inference**: Distinguishing systematic associations and nonsystematic forces
 - ▶ **Causal inference**: Distinguishing causal effects and spurious associations
- ▶ **Systematic** relationships are the basis of **generalizable insights**
- ▶ **Better** examples (than previous ones) would look like:
 - ▶ Why is voter turnout for local elections higher in some districts/municipalities than in others?
 - ▶ Do smaller countries sign more multilateral treaties than larger countries, and why?
 - ▶ Why do democracies not fight each other while fighting against non-democracies?

Contributions?

Making contributions? (*DSI*, 16)

Focusing too much on making a contribution to a scholarly literature without some attention to topics that have real-world importance runs the risk of descending to politically insignificant questions. Conversely, attention to the current political agenda without regard to issues of the amenability of a subject to systematic study within the framework of a body of social science knowledge leads to careless work that adds little to our deeper understanding.

- ↪ Students are often interested in real-world phenomena (which is nice). . . .
- ↪ but you need to pay careful attention to scholarly debates and literature

Contributions?

Making contributions? (*DSI*, 16)

[Making contribution] means explicitly locating a research design within the framework of the existing social scientific literature. . . . This ensures that the investigator [YOU!] understand the “state of the art” and minimizes the chance of duplicating what has already been done. It also guarantees that the work done will be important to others, thus improving the success of the community of scholars taken as a whole.

- ↪ How can our research questions and project make significant contributions?
- ↪ How can we make our research questions “better”?

Contributions: How?

Making contributions? (*DSI*, 16–17)

- 1 Choose a **hypothesis seen as important** by scholars in the literature but for which no one has completed a systematic study. If we find evidence in favor of or opposed to the favored hypothesis, we will be making a contribution.

e.g., “Democracy peace” (two democracies do not fight each other). . . . ?

- 2 Choose an **accepted hypothesis** in the literature that we suspect is false (or one we believe has not been adequately confirmed) and investigate whether it is indeed false or whether some other theory is correct.

e.g., Oil prevents democratic transition and consolidation. . . . ?

- 3 Attempt to resolve or provide further evidence of one side of a **controversy in the literature**—perhaps demonstrate that the controversy was unfounded from the start.

e.g., Land abundance prevents state-building. . . . ?

Contributions: How?

Making contributions? (*DSI*, 16–17)

- 4 Design research to illuminate or evaluate **unquestioned assumptions** in the literature.
e.g., Is “free-riding” really “free”....?
- 5 Argue that an **important topic has been overlooked** in the literature and then proceed to contribute a systematic study to the area.
e.g., Why are some civil conflict deadlier than others....?
- 6 Show that theories or evidence designed for some purpose in one literature could be **applied in another literature** to solve an existing but apparently unrelated problem.
e.g., Rubinstein’s bargaining game and Nash bargaining solution help us explain why armed conflicts occur

Literature Review: Why?

Why literature review in the first place? This is laborious after all

- ▶ As in previous slides, making contributions requires systematic and appropriate literature review
- ▶ Literature review helps you motivate and sharpen your research questions, research designs, and hypotheses (PSRM, 54)
- ▶ Specifically, it helps you to:
 - 1 see what has (not) been investigated in the literature
 - 2 develop general explanations for observed variations (associations) in a behavior or a phenomenon
 - 3 identify potential relationships between concepts and researchable hypotheses
 - 4 learn how others have defined and measured key concepts
 - 5 identify data sources that other researchers have used
 - 6 develop alternative research designs
 - 7 discover how your research project is related to the work of others (previous studies)

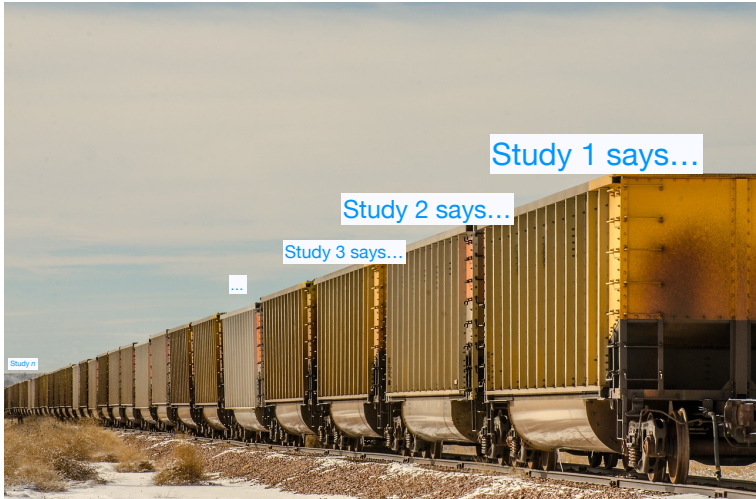
Literature Review: How?

How, then, do we do our literature review?

- ▶ “The key to organizing and writing an effective literature review is to **focus on concepts, ideas, and methods** shared across the literature” (PSRM, 64)
- ▶ Try to organize previous studies **systematically**, **NOT** the “boxcar” **method** (PSRM, 64–65)

Literature Review: How?

Simply (and subsequently) summarizing previous studies you collected (boxcar method) is NOT a relevant literature review



Literature Review: How?

Focusing on concepts, ideas, and methods (PSRM, 64–65)

Suppose that you have collected ten scholarly articles via Google Scholar

- 1 Group the articles according to their **research questions**
 - ▶ This clarifies common research questions in previous studies, and thereby providing an effective start to defining the base of knowledge
- 2 Regroup the articles based on the **data and research designs**
 - ▶ Different methods and samples, different conclusions?
 - ▶ Omitted variables (confounders) might lead to different conclusions
- 3 Sort the articles by the **findings and conclusions**
 - ▶ Conflicting and mixed findings can provide a wonderful motivation for new research
 - ▶ They also establish for the reader the importance and relevance of your research project

Your final representation of literature review may focus on either of 1–3 depending on your own focus and interests