Empirical Modeling Approaches in Marketing and Economics

Professor Peter C. Reiss

Stanford Graduate School of Business
Co-Editor, QME
July, 2013

Columbia-Duke-UCLA Workshop on Quantitative Marketing and Structural Econometrics

Outline

Framework for Appraising Empirical Work

Descriptive Models
    What are They?
    Uses and Abuses
    Data and Regressions

Structural Models
    What are They?
    Identification (briefly)
    Examples
    Advantages and Disadvantages

Summary Observations

Suggested Summer Beach Reading

An Initial Taxonomy

Most empirical studies can usefully be classified along two dimensions:

	Modeling Approach	
	Descriptive	**Structural**
Experimental	A	B
Observational	C	D

Data

These classifications are useful for thinking about the types of inferences a study can draw.

Example: *Does increasing shelf space for a product increase sales?*

Data and Assumptions

The most general object that a researcher can (hope to) describe is the joint density of the observable data.

The joint density characterizes the **Data Generating Process** (DGP).

Data $\quad\quad\quad x : \{x_1, \ldots x_N\} \quad y : \{y_1, \ldots y_N\}$

Joint Density $\quad f(X_1, \ldots X_N, Y_1, \ldots, Y_N)$

Unfortunately, we cannot recover $f(\cdot)$ without assumptions.

Data and Assumptions

The most general object that a researcher can (hope to) describe is the joint density of the observable data.

The joint density characterizes the **Data Generating Process** (DGP).

Data $x : \{x_1, \dots x_N\}$ $y : \{y_1, \dots y_N\}$

Joint Density $f(X_1, \dots X_N, Y_1, \dots, Y_N)$

Unfortunately, we cannot recover $f(\cdot)$ without assumptions.

  Point 1:   <u>ALL</u> EMPIRICAL WORK INVOLVES ASSUMPTIONS!

Data and Assumptions

The most general object that a researcher can (hope to) describe is the joint density of the observable data.

The joint density characterizes the **Data Generating Process** (DGP).

Data $\quad x : \{x_1, \ldots x_N\} \quad y : \{y_1, \ldots y_N\}$

Joint Density $\quad f(X_1, \ldots X_N, Y_1, \ldots, Y_N)$

Unfortunately, we cannot recover $f(\cdot)$ without assumptions.

Point 1: <u>ALL</u> EMPIRICAL WORK INVOLVES ASSUMPTIONS!

Point 2: Most assumptions are maintained and not testable.

Data and Assumptions

The most general object that a researcher can (hope to) describe is the joint density of the observable data.

The joint density characterizes the **Data Generating Process** (DGP).

Data $x : \{x_1, \ldots x_N\}$ $y : \{y_1, \ldots y_N\}$

Joint Density $f(X_1, \ldots X_N, Y_1, \ldots, Y_N)$

Unfortunately, we cannot recover $f(\cdot)$ without assumptions.

Point 1: <u>ALL</u> EMPIRICAL WORK INVOLVES ASSUMPTIONS!

Point 2: Most assumptions are maintained and not testable.

Point 3: Assumptions permit inference, but at the cost of credibility.

Assumptions in Descriptive Studies

In the shelf space example, let $Y =$ **Sales** and $X =$ **Space**. Suppose we know the amount of space given to different brands each week in a store.

To *describe* these data, we would typically assume:

Independent $\qquad\qquad f(X_1, \ldots X_N, Y_1, \ldots, Y_N) = \Pi_{i=1}^{N} f_i(X_i, Y_i).$

Identically Distributed $\qquad \Pi_{i=1}^{N} f_i(X_i, Y_i) = \Pi_{i=1}^{N} f(X_i, Y_i).$

We can now use **nonparametric** or **parametric** methods to estimate $f(X, Y)$, or features of it:

$\mathbf{f(Y|X)}$ - the conditional **density** of Y given X.

$\mathbf{E(Y|X)}$ - the conditional **mean** of Y given X.

$\mathbf{Q_{\tau}(Y|X)}$ - the τth conditional **quantile** of Y given X.

$\mathbf{BLP(Y|X)}$ - the conditional **best linear predictor** of Y given X.

Shelf Space Example

Suppose a regression indicates a positive relation between shelf space and sales. What can we conclude?

Observational Data:

> There is a positive association within a week.
>
> Can't: infer causality; infer behavior; do counterfactuals.
>
> Won't instruments save us?

Experimental Data:

> There is a positive association within a week.
>
> Causal story relies on more assumptions and/or a theory.
>
> Can't: infer behavior; do counterfactuals.
>
> Best case is that researcher intervention acts like an instrumental variable; it's unclear what the regression estimate means.

Recap: Descriptive Empirical Work

Descriptive empirical work is primarily about statistical objects.

Useful because it can:

- **Document facts.** e.g., How much space is devoted to a product? Does it vary over time? Is the 2nd shelf better than the 3rd? Facts are useful for empiricists and theorists to know.
    - e.g., Bronnenberg, Dube, and Gentzkow (2012), "The Evolution of Brand Preferences: Evidence from Consumer Migration," *AER*.

- **Identify associations.**

- **Corroborate theory.** Has a theory made useful predictions?

- **Prediction.** What factors best predict behavior?

- **Causal Connections (?)** **Caution:** Relies on experimental control and a theory.

Recap: Descriptive Empirical Work

Further Remarks:

- Descriptive studies have an important role to play in marketing provided they are not over-interpreted.

Recap: Descriptive Empirical Work

Further Remarks:

- Descriptive studies have an important role to play in marketing provided they are not over-interpreted.

- Data description is a lost art; useful figures and tables key; statistical methods should be flexible (as nonparametric as possible).

Recap: Descriptive Empirical Work

Further Remarks:

- Descriptive studies have an important role to play in marketing provided they are not over-interpreted.

- Data description is a lost art; useful figures and tables key; statistical methods should be flexible (as nonparametric as possible).

- Many empiricists believe descriptive work is exclusively about "testing" theories. Descriptive work fundamentally cannot "test" a theory – you need a formal model to do this. If descriptive work produces facts that fit a theory, that does not prove the theory. Similarly, ill-fitting facts do not disprove a theory.

Recap: Descriptive Empirical Work

Further Remarks:

- Descriptive studies have an important role to play in marketing provided they are not over-interpreted.

- Data description is a lost art; useful figures and tables key; statistical methods should be flexible (as nonparametric as possible).

- Many empiricists believe descriptive work is exclusively about "testing" theories. Descriptive work fundamentally cannot "test" a theory – you need a formal model to do this. If descriptive work produces facts that fit a theory, that does not prove the theory. Similarly, ill-fitting facts do not disprove a theory.

- Many structural modelers with interesting data spend too little time describing the data. An all-to-common post-seminar comment: "Wow, what a great dataset. Unfortunately, I didn't learn much about it from the paper!"

A Special Status for Regressions?

The most common descriptive model is a linear regression.

Many researchers believe that because regressions can be expressed mathematically, they have a special status that goes beyond data description.

This special status is reflected in the following comments:

1. β is the "effect" of a one unit change in X on Y.
2. β represents the "partial derivative" of the conditional mean of Y.
3. β is the "reduced form" effect of X on Y.

These comments over-interpret or mis-interpret regression estimates.

What is accurate?

A regression always delivers consistent estimates of the **best linear predictor** (BLP). The BLP is not $E(Y|X)$. Further, both BLP$(Y|X)$ and $E(Y|X)$ are predictive and not causal relationships.

Illustration

Gallileo is famous for his Tower of Pisa dropped ball experiments in which he dropped objects from the tower and recorded the times T it took for the objects to drop distances D.

Suppose Gallileo had regressed observed drop distances $d = (d_1, ..., d_N)$ on times $t = (t_1, ..., t_N)$:

$$d = \alpha_0 + \beta_0 t + \epsilon.$$

How would you describe the meaning of the estimates of α_0 and β_0?

Illustration

Gallileo is famous for his Tower of Pisa dropped ball experiments in which he dropped objects from the tower and recorded the times T it took for the objects to drop distances D.

Suppose Gallileo had regressed observed drop distances $d = (d_1, ..., d_N)$ on times $t = (t_1, ..., t_N)$:

$$d = \alpha_0 + \beta_0 t + \epsilon.$$

How would you describe the meaning of the estimates of α_0 and β_0?

We know that under standard assumptions (i.e., $E(\epsilon) = E(T\epsilon) = 0$):

$$\beta_0 = \frac{\text{Cov}(D, T)}{\text{Var}(T)} \qquad \alpha_0 = E(D) - \beta_0 E(T).$$

Anything else?

Illustration

Gallileo is famous for his Tower of Pisa dropped ball experiments in which he dropped objects from the tower and recorded the times T it took for the objects to drop distances D.

Suppose Gallileo had regressed observed drop distances $d = (d_1, ..., d_N)$ on times $t = (t_1, ..., t_N)$:

$$d = \alpha_0 + \beta_0 t + \epsilon.$$

How would you describe the meaning of the estimates of α_0 and β_0?

We know that under standard assumptions (i.e., $E(\epsilon) = E(T\epsilon) = 0$):

$$\beta_0 = \frac{\text{Cov}(D, T)}{\text{Var}(T)} \qquad \alpha_0 = E(D) - \beta_0 E(T).$$

Anything else?

What about a causal interpretation? (It is after all an experimentally controlled setting!)

Illustration

A law of physics states that (apart from wind resistance)

$$D = \frac{g}{2} T^2 \qquad \text{and} \qquad T = \sqrt{\frac{2}{g}} \sqrt{D} \qquad (1)$$

where g is a gravitational constant.

Suppose we have IID mean zero measurement errors in the experiment and that Gallileo sampled the drop times from a uniform $T \sim U[\underline{T}, \bar{T}]$.

Some algebra reveals:

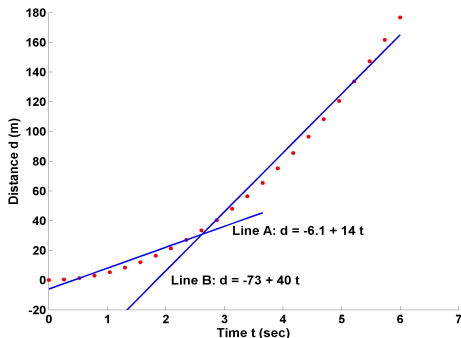$$\beta^* = \frac{Cov(D, T)}{V(T)} = g \left[\frac{\underline{T} + \bar{T}}{2} \right] \qquad (2)$$

and

$$\alpha^* = E(D) - \beta E(T) = -\frac{g}{12} (\underline{T}^2 + \bar{T}^2 + 4 \underline{T} \bar{T}). \qquad (3)$$

These equations illustrate that the BLP coefficients are in general sensitive to the underlying distribution of the data!

Illustration

Specifically, the formulae suggest that changes in the support of the uniform drop times $[\underline{T}, \bar{T}]$ change the BLP coefficients.

This is true even though the (nonlinear) conditional mean function $E(D|T) = m(T)$ remains unchanged!

Structural Models - Definition

In contrast to descriptive statistical models, there are "structural" models.

> **Definition:** *A* **structural model** *is an explicit model of economic behavior that gives rise to the joint density of the data* $f(\mathbf{X}, \mathbf{Y})$, *or properties of this joint density such as* $E(\mathbf{Y}|\mathbf{X})$.

Caution: In marketing, "structural equation modeling" is sometimes used to describe factor models.

A structural model usually has two components:

- **Mathematical equations** derived from an economic or decision-theoretic model (e.g., utility model plus utility maximization; profit function plus profit maximization).
- **A stochastic structure** that maps the theoretical model to a joint density of the data (because theory models rarely fit data perfectly).

Structural Models - Identification

In some structural models, there can be a difference between the joint density of the data $f(X, Y)$ and the density of interest $p(X, Y)$.

   Example: X and Y are censored; p(X,Y) is the uncensored density.

To identify a density (or some feature of it) of interest, we need to show that there exists a mapping from the density of the observed data $f(X, Y)$ to $p(X, Y)$.

A probability model (or its features) are *identified* if the mapping is one-to-one.

A probability model (or its features) are *partially identified* if the mapping places meaningful bounds on the object of interest.

Remember: Identification is a population, not a sample concept.

Identification and Assumptions

RECALL: To identify anything of interest, you need assumptions.

Corollary: The more interesting an economic object, the more assumptions typically required to identify it.

Corollary: The credibility of an identification argument, and therefore an inference, declines in the number of assumptions (Manski).

Example (albeit statistical):

Recall the bivariate regression model:

$$Y = \alpha_0 + \beta_0 X^* + \epsilon. \tag{4}$$

What population conditions identify α_0 and β_0?

Identification and Assumptions

RECALL: To identify anything of interest, you need assumptions.

Corollary: The more interesting an economic object, the more assumptions typically required to identify it.

Corollary: The credibility of an identification argument, and therefore an inference, declines in the number of assumptions (Manski).

Example (albeit statistical):

Recall the bivariate regression model:

$$Y = \alpha_0 + \beta_0 X^* + \epsilon. \qquad (4)$$

What population conditions identify α_0 and β_0?

$$E(\epsilon) = E(X^* \epsilon) = 0$$

Identification and Assumptions

Formally

$$\begin{aligned}
\text{Cov}(X^*, Y) &= \text{Cov}(X^*, \alpha_0 + \beta_0 X^* + \epsilon) \\
&= \beta_0 \text{Var}(X^*) + \text{Cov}(X^*, \epsilon)
\end{aligned}$$

and

$$E(Y) = \alpha_0 + \beta_0 E(X^*) + E(\epsilon).$$

Under the maintained assumptions,

$$\begin{bmatrix} \beta_0 \\ \alpha_0 \end{bmatrix} = \begin{bmatrix} \dfrac{\text{Cov}(X^*, Y)}{\text{Var}(X^*)} \\ E(Y) - \beta_0 E(X^*) \end{bmatrix}. \tag{5}$$

Suppose we now are forced to back off on the assumption that we observe X^*. Suppose instead we observe a noisy version of X^*, $X = X^* + \eta$.

In this case, we cannot obtain α_0 and β_0 from the first and second population moments. We lack identification!

Identification and Assumptions

But we can "purchase" identification with more assumptions. These assumptions may, however, reduce the credibility of our inferences/estimates.

Specifically, consider the classical measurement error assumptions: $E(\eta) = \text{Cov}(X^*, \eta) = \text{Cov}(\eta, \epsilon) = 0$. Now

$$\text{Cov}(X, Y) = \beta_0 \text{Var}(X^*)$$

and

$$E(Y) = \alpha_0 + \beta_0 E(X^*)$$

but we do not observe $\text{Var}(X^*)$. The coefficients still are not identified!

Identification and Assumptions

However, they are partially identified. In particular, for positive values of β_0 we have the bounds (Frisch (1934))

$$\plim_{N \to \infty} \frac{1}{\hat{d}} \geq \beta_0 \geq \plim_{N \to \infty} \hat{b}$$

where $\hat{d}$ is the slope coefficient from regressing X on Y.

Thus, more assumptions has gotten us partial identification.

Does this mean the cost of the assumptions was worth it?

Structural Models

Back to structural models ...

Recall that the structure in structural models is reflected in the joint density of the observed data.

Two sources of structure:

1. Economic equations of the form $g(X, Y, \epsilon) = 0$.

2. Stochastic structure of the form $f_{X,\epsilon}(X, \epsilon)$.

Example:

Demand	q_t^D	$= \beta_{10} + \beta_{11}x_{1t} + \gamma_{12}p_t + \epsilon_{1t}$
Supply	p_t	$= \beta_{20} + \beta_{22}x_{2t} + \gamma_{22}q_t^S + \epsilon_{2t}$
Equilibrium	q_t^D	$= q_t^S$

Structural Models

Why is this a structural model?

Structural Models

Why is this a structural model?

Answer: The supply function describes the *behavior of firms* and the demand function describes the *behavior of consumers*. These are the objects of economic interest.

Notice how the structural model induces a distribution on the (conditional) distribution of **observed** prices and quantities - $\mathbf{f}(\mathbf{P}, \mathbf{Q}|\mathbf{X})$.

e.g., the conditional mean is the "reduced form" $y_t' = x_t'\Pi + v_t'$.

This structural model has **content** to the extent that we can uniquely recover the behavioral parameters β and γ from the parameters of the conditional distribution (the Π and $\text{Var}(v_t)$). (This is what identification is about.)

Structural Models - Example

The reduced form parameters of this structural model permit us to make causal ("behavioral") statements about the impact of a change in x_t on y_t.

CAUTION. Many researchers mistakenly label regressions of y_t on exogenous variables x_t as a "reduced form"

$$y_t' = x_t'\Pi + v_t'.$$

They go on to interpret the Π coefficients as the causal effect of x_t on y_t.

DO NOT DO THIS !

A causal reduced form only exists when it has been derived from stuctural model. Consider how a linear "reduced form" would represent the following demand and supply system:

Demand	$\ln q_t^D$	$=$	$\beta_{10} + \beta_{11}x_{1t} + \gamma_{12}\exp(p_t) + \epsilon_{1t}$
Supply	$\ln p_t$	$=$	$\beta_{20} + \beta_{22}x_{2t} + \gamma_{22}\exp(q_t^S) + \epsilon_{2t}$
Equilibrium	q_t^D	$=$	q_t^S

Structural Models - Example

The reduced form for this model is not available in closed form.

Thus what meaning can we attach to our using the regression

$$p_t = \pi_0 + \pi_1 x_{1t} + \pi_2 x_{2t} + v_{1t} \ ?$$

Structural Models - Example

The reduced form for this model is not available in closed form.

Thus what meaning can we attach to our using the regression

$$p_t = \pi_0 + \pi_1 x_{1t} + \pi_2 x_{2t} + v_{1t} \ ?$$

Answer: This is <u>not</u> a reduced form. We are instead estimating a descriptive model and getting the best linear predictor (BLP) of price given x_t.

Without being clear on the structural model, this BLP does not reveal anything about demand and supply behavior. The BLP coefficients π do have the following descriptive interpretation:

If we draw two observations on $\{p, x_1, x_2\}$ from the population (for which the demand and supply model is relevant), and these two observations have the same x_1 value and their x_2 values differ by one, then the best (in a mean squared error sense) prediction for the difference in prices is π_2.

Advantages of Structural Models

- Contain economic or behavioral parameters that are of marketing interest (E.g., the price elasticity of demand, marginal utility of income, or marginal cost).

- Once estimated, a structural model can perform counterfactuals. For example, Would a vertically integrating firm increase price or change advertising?

- Can test the "fit" or predictive performance of two or more competing theories. Here it should be noted that the test depends on the maintained structure of the competing models. For example, a test of collusive versus Bertrand pricing presumes a given functional form for demand and costs.

- Structural models (hopefully) make clear what assumptions are used to produce a given set of behavioral parameter estimates.

Cautions on Structural Models

- Structural models should refect the institutional realities of data. (e.g., "We estimate a model for small and large cars.")

- Theory rarely delivers complete structural models. Researchers must add functional form assumptions, parameters and variables. For example, consider the indirect utility function in discrete choice models:

$$V_i(p, y) = \beta_{0i} + \beta_{1i}p_i + \beta_{2i}y_i + \beta_{3i}(\text{Advertising, Attributes,...}).$$

- Structural modelers should take care to verify that (i) their functional form assumptions do not "deliver the result" (e.g., Logit cross-elasticities); and (ii) their results are not sensitive to model elements not tied to theory.

- Structural models are usually based on highly stylized theories. This is because it is difficult to generate flexible, yet estimable models (e.g., dynamic game models).

Recap - Descriptive Empirical Work

Statistical models are about describing data. They have an important role to play in marketing provided they are not over-interpreted.

At a general level, data description is about (flexibly) characterizing the joint density of the data - $\mathbf{f}(\text{Data}) = \mathbf{f}(\mathbf{X}, \mathbf{Y})$.

Descriptive methods in economics and marketing usually seek to characterize conditional densities (or their properties - means, medians).

The most common conditional model is the linear regression. It always delivers consistent estimates of the BLP. The BLP is not $E(Y|X)$. Further, both $\text{BLP}(Y|X)$ and $E(Y|X)$ are predictive and not causal relationships.

Recap - Structural Empirical Work

A **structural model** is an explicit model of economic behavior that characterizes the joint density of data $f(\mathbf{X}, \mathbf{y})$, or properties of this joint density in terms of the behavioral parameters.

Structural models facilitate: parameter estimation; counterfactuals; comparisons of theories; and make clear assumptions needed to estimate a quantity.

Theory rarely delivers complete structural models. Researchers must add functional form assumptions, parameters, variables and errors.

Structural modelers should verify that their functional form assumptions do not "deliver the result" and that their results are not sensitive to model elements not tied to theory.

Parting Wisdoms

Structural models are not about high-tech statistics or fancy techniques. (If you believe this, you are not alone, but you have lost sight of what good empirical work is about.)

Descriptive work is just as important as structural work. Indeed, any sensible structural paper does a good job documenting the main features of the data. Additionally, a good paper documents whether the structural model has done a reasonable job explaining the "facts".

Structural modeling is difficult because it requires: (i) knowledge of theory; (ii) knowledge of econometrics; (iii) knowledge of the real world; and (iv) an ability to put all these pieces together. Make sure you work on each of these skills.

Suggested Summer Beach Reading:

1. *Marketing Science* Nov-Dec issue, 2011; summary of workshop papers.

2. P. Reiss and F. Wolak, "Structural Econometric Modeling: Rationales and Examples from Industrial Organization." *Handbook of Econometrics, Vol 6a*, 2007.

3. P. Reiss. "Descriptive, Structural, and Experimental Empirical Methods in Marketing Research", *Marketing Science*, 2011, pp. 950-964.

4. P. Reiss. Economic Data and Economic Inference, Book in Progress.

Note. These slides soon available at: