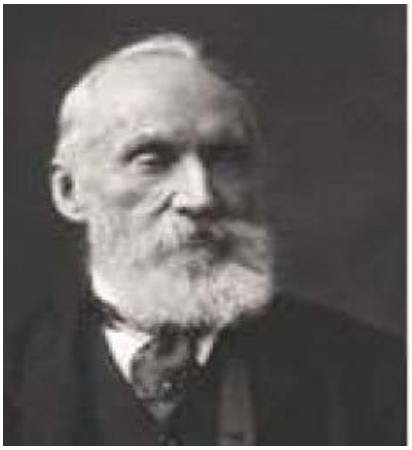
Measuring Innovation in Economic Statistics:

Surveys of Sales of New Products and Scraping New Products

Richard B. Freeman, Harvard University and NBER. May 17, 2016

1. The problem: We lack measures of innovation comparable to GDP based on actual production with which to gauge innovation performance. Indicators of innovation mix per capita measures – R&D/GDP; policies viewed as inducing innovation – rule of law; and aggregates – total S&E graduates – into a potpourri in which small highly advanced countries such as Sweden, Switzerland top global lists.

None of these indicators contain data on "real innovations" – new or significantly improved products (goods or services), or processes available in the market. The result is lots of loose talk about innovation but little substantive analysis.



"When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind."
Lord Kelvin

Two ways to measure innovation

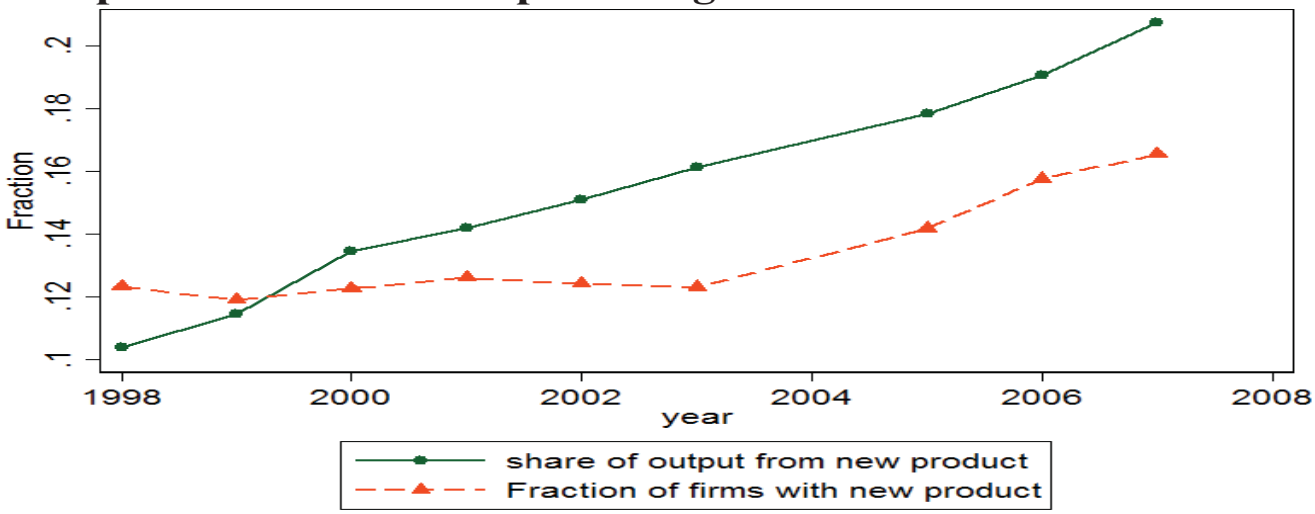
Ask firms whether they brought new products to market and amount sold (NSF BRDIS innovation questions).

Scrape information on new products from announcements/ listings on firm websites, and other places (such as Amazon). To measure value must measure attributes and measure the worth of attributes to purchasers – hedonic pricing.

2. Evidence from Asking firms

What is the output (in monetary value) from new product in your firm this year? where "n*ew product"* is uses technology, new idea from R&D, or is significantly improved in structure, material and technology and performance. (China Survey of Industrial Firms)

Stable incumbent (balanced panel) firms report new product share of output rising from 10% to 20%.

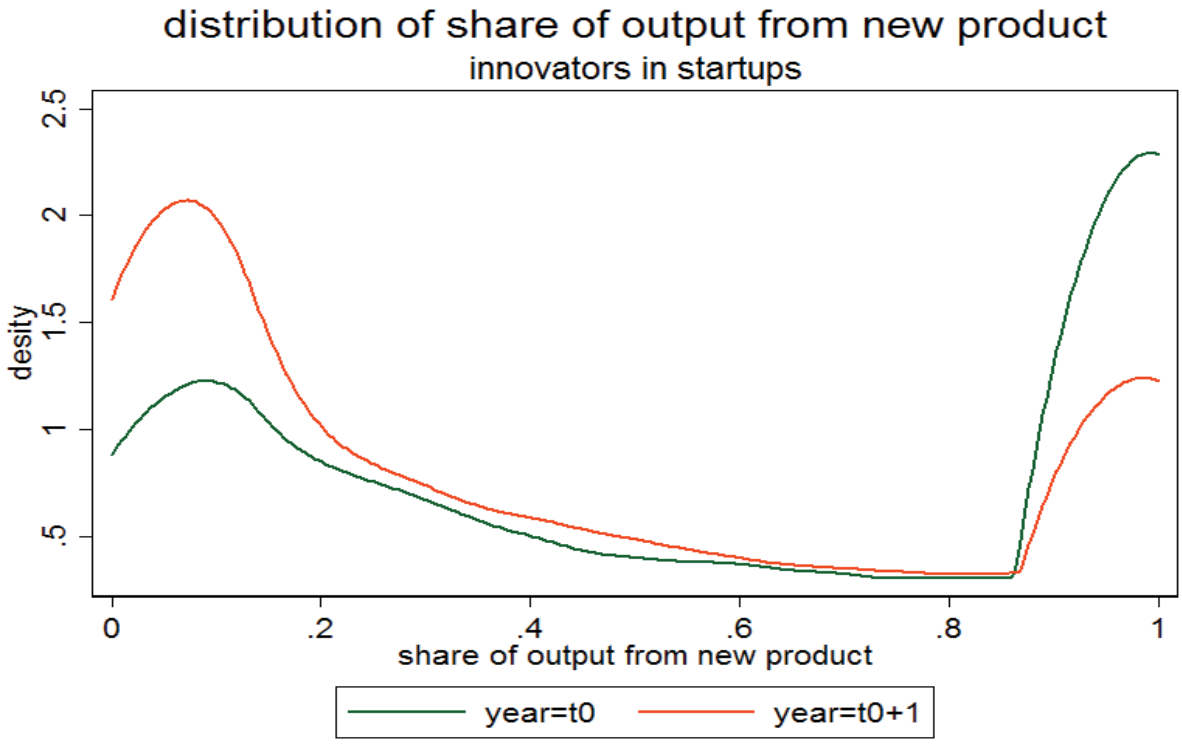


Firms differ in extent/persistence of innovation: some introduce new products continuously and others never doing so

times producing new product	Freq.	Randomly assigned	Freq. (Innovator)	Randomly assigned (Innovator)	share of value from new product output
0	18,899	7,615	0	0	0.00
1	2,855	10,724	2,855	123	2.10
2	1,859	6.809	1,859	760	2.71
3	1,100	2,496	1,100	1,874	4.39
4	620	557	620	2,684	3.80
5	513	92	513	2,236	6.17
6	455	9	455	1,189	3.42
7	387	1	387	426	7.56
8	394	0	394	107	12.29
9	1,221	0	1,221	5	57.57
pool	28,303	28,303	9,404	9,404	100.00

Numbers in column "randomly assigned" are from simulation in binomial distribution with a full sample size on average probability. Numbers in column "randomly assigned (only innovators)" simulate binomial for innovators. Share of sales based on output is deflated by 2 digit industry prices.

Startups should have all products as new to market, so we look at year of entry and next year



Our analysis of sales of new products in China's survey of industrial firms from 1998 to 2007 further shows that:

1. Innovative firms have greater sales growth, high and increasing productivity, and higher survival over time than other firms.
2. Innovative firms have greater R&D and design and utility patents.
3. Innovators gain some sales on new products at the expense of their own products and some at the expense of competitors but also gain sales by growing the market.

Needed: questions about life cycle of new products, details on most important new product in given year; inputs in developing it; patents associated with it.

3. Web scraping Micro-unit Innovation

Original plan: web scrape by firms, many of whom list new products on web site, but firm web sites are idiosyncratic so must crowd-source, which makes it hard to keep up.

2nd Plan: web scrape lists of innovations – R&D magazine, and many others, but again variation so better done with human research assistants. Virtue of getting "important innovations" but no accompanying Ps or Qs

3rd Plan: Marketplace – Amazon data
Cell phones in China for attributes
Boehnke & Mendel "Amazon's Price and Sales-Rank Data"

Cell Phone Database and Analysis

The database contains the prices and attributes of 1,536 cell phone models produced by 18 major brands, from 2011-2015, with estimates of quantities sold based on rank of sales

Source of data is Amazon China and ZOL websites via WebArchive
Amazon has prices for top-selling cell-phone models. We obtained prices for others from ZOL . The prices are reported at different times.
To estimates quantities, we used sales ranks on Amazon, where higher rank means bigger sales. To transform ranks into quantities we fit power law from quantities sold by *brand* and assumed the same power law linking quantities to ranks holds for models as for brands.

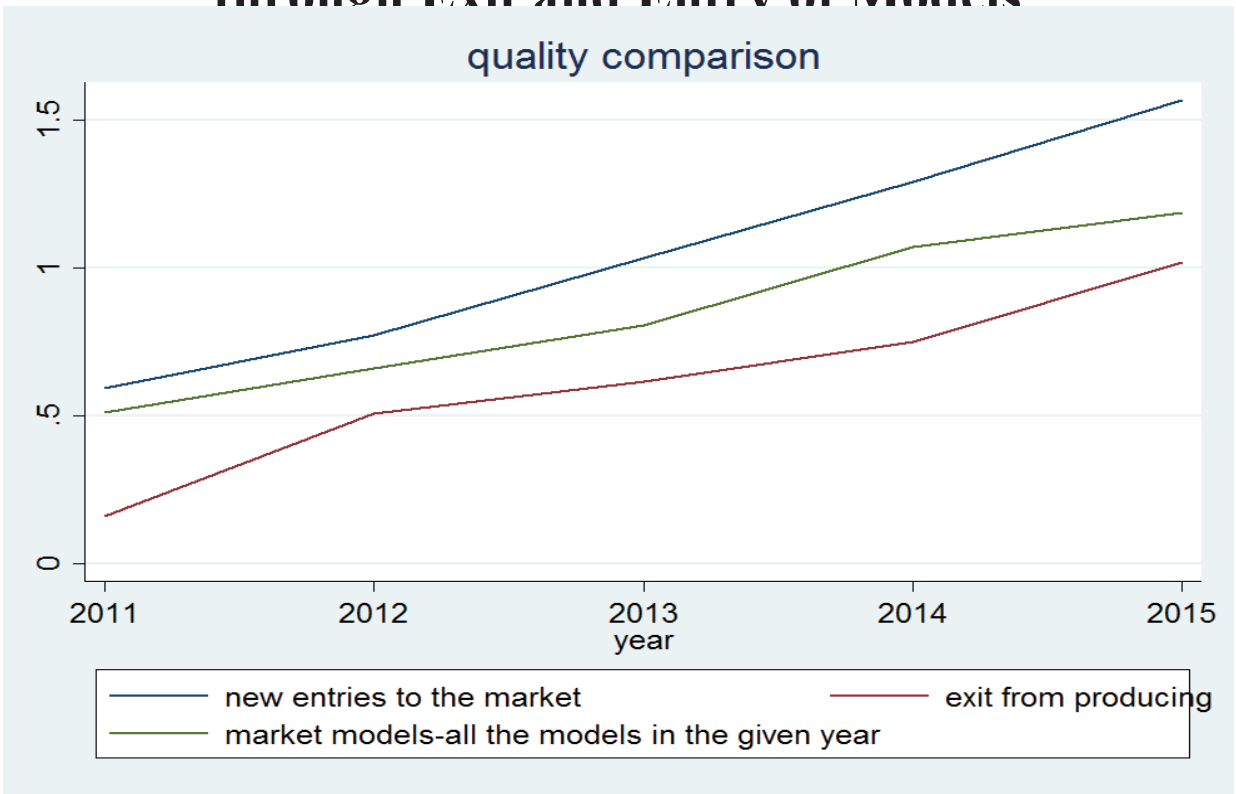
We created measures of average quality of a model using hedonic price regressions, where we weighted observations by estimated quantities sold. Since we did not have ranks for the smallest models, they are deleted in calculations that weighted the models by their sales.

China Hedonic Price Regressions based on attributes

$$(LnP = \beta_0 + \beta_1 Z1 + \beta_2 Z2 + \ldots + \delta t + \varepsilon)$$

VARIABLES	(1) control time	(2) control brand	(3) share weighted
display_size	-0.0993***	-0.00153	0.135**
	(0.0154)	(0.0157)	(0.0536)
display_resolution	0.0899***	0.0216*	0.260***
	(0.0146)	(0.0119)	(0.0517)
network	0.106***	-0.0503***	0.0689*
	(0.0165)	(0.0118)	(0.0396)
processor_core	-0.0462***	-0.00312	-0.00817
	(0.00489)	(0.00399)	(0.0104)
clock_speed	0.274***	0.264***	0.386***
	(0.0243)	(0.0212)	(0.0895)
ram	0.142***	0.213***	0.231***
	(0.0182)	(0.0150)	(0.0502)
rom	0.0222***	0.0117***	0.00935***
	(0.000803)	(0.000774)	(0.00180)
battery	0.0140	0.0864***	-0.136*
	(0.0179)	(0.0163)	(0.0711)
has_front_cam(dum)	0.402***	0.214***	0.123***
	(0.0156)	(0.0137)	(0.0379)
front_camp	0.0210***	0.00588**	-0.0390***
	(0.00344)	(0.00265)	(0.00902)
back_camp	0.0200***	0.0143***	0.0312***
	(0.00163)	(0.00127)	(0.00647)
Constant	7.085***	8.250***	7.744***
	(0.216)	(0.154)	(0.198)
Observations	7,507	7,507	679
R-squared	0.591	0.753	0.911

Data show Attributes Change in raising quality through Exit and Entry of Models



Boehnke & Mendel (2016)

Use 3rd party price tracking sites to obtain prices and quantities to obtain a panel data set representing daily prices and sales ranks for 135,000 products on sale at Amazon from 9 countries, 2008 to present.

Detailed products but we need attributes of products over time or ideally of models with fixed attributes, as with the cell phones. Computers would be another good product. Some data on 3rd party used.

Goal is an innovation index for products, with sales so can mimic the % of sales from new products/processes but must obtain qualities, form product groups, to allow for hedonic prices to value changes in attributes. But know firms, dates when introduced and no longer sold (save as used).