Tornqvist indexing in R

Tqtools package

Mike Jadoo / Economist
Peter B. Meyer / Research Economist
Office of Productivity and Technology,
U.S. Bureau of Labor Statistics

Conference of Statistical Practice
New Orleans, 2/15/2019

Motivation

Provide routines to help individuals to properly aggregate indices using the Tornqvist method.

Vocabulary

Productivity:

-Labor

-Multifactor

NAICS

Nonfarm business sector

Cost of production

Measures

Index :

-Price

-Quantity

Vocabulary

Productivity: a measure of efficiency which compares the amount of goods and services produced (output) with the amount of inputs used.

Vocabulary

$$\textbf{Labor Productivity} = \frac{Output}{Hours\ Worked}$$

Multifactor Productivity

$$= \frac{Output}{Combination\ of\ Labor\ and\ Capital\ Inputs}$$

Vocabulary

NAICS: **N**orth **A**merican **I**ndustry **C**lassification **S**ystem is the standard used by Federal statistical agencies for classifying business establishments

-**Sector:** 2-digit

-**Subsector:** 3-digit

-**Industry group-** 4-digit

Level	NAICS CODE	Description
Sector	44-45	Retail Trade
Subsector	441	Motor Vehicle and Parts dealer
Industry Group	4412	Other Motor Vehicle Dealers

Vocabulary

Nonfarm business sector: includes about 75% of U.S. economy. It excludes general government, farms, and household work.

Vocabulary

Cost of production: cost of producing an item in an industry in current dollars. These can be used as weights in our Index formula.

Measures: Our quantification of a particular concept measured over time

Vocabulary

Index: a statistical representation of a level over time for a particular measure using a base year.

Price index: measurement of price movements in a numerical series

Quantity index: represents a real measure of a particular input or output.

$$\text{Cost} = \text{Price} * \text{Quantity} \quad \text{and} \quad \text{Price} = \frac{Cost}{Quantity}$$

Agenda

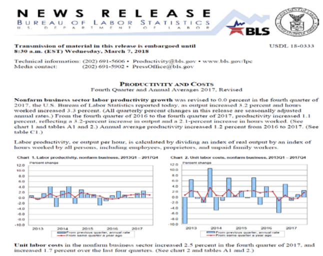
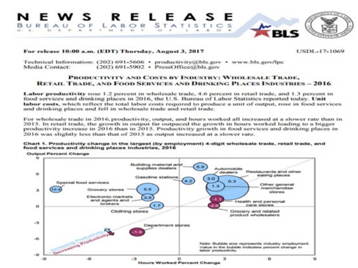
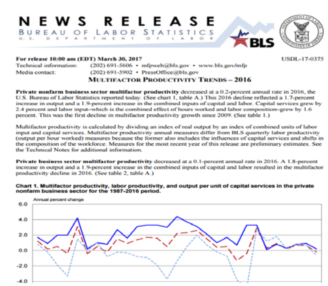
- ❑-Overview of productivity measures
- ❑-Common Index formulas
- ❑-How Tornqvist is used
- ❑-Demonstrations
    1. Information capital measure
    2. Manufacturing sector aggregation

Overview of productivity measures

Why is Productivity Growth Important?

- Enables higher living standards
- Source of potential gains in national income
- Key to competitiveness
- Used in employment projections

Productivity

Measure	NAICS industry coverage	Press release
Quarterly Labor Productivity	• Business and Nonfarm Business Sectors • Manufacturing Sectors • Nonfinancial Corporations	
Annual Labor Productivity	• Manufacturing and Mining Industries • Services Industries • Wholesale and Retail Trade Industries	
Annual Multifactor Productivity	• Private Business • Nonfarm Business • Manufacturing • 60 3-digit NAICS industries	

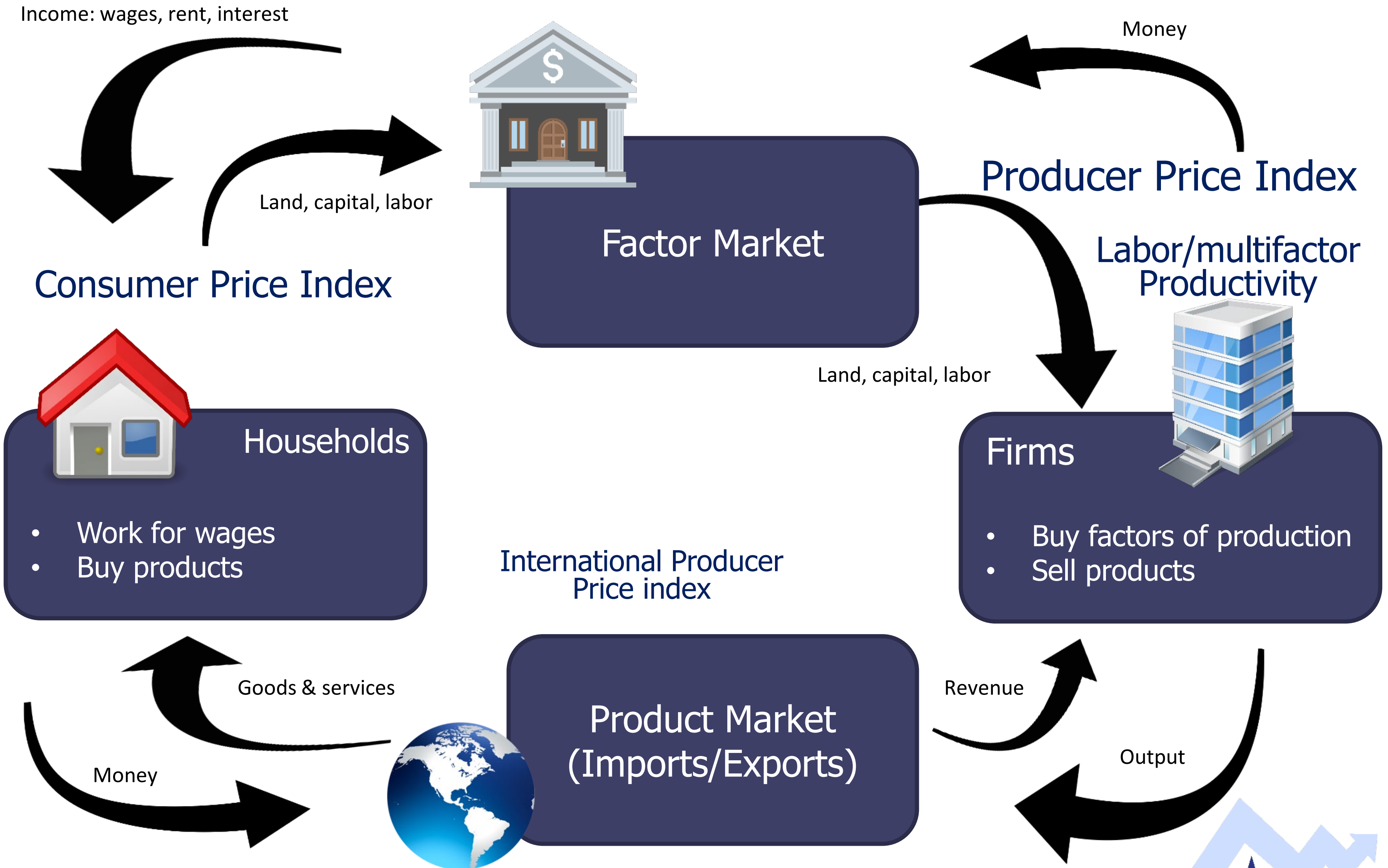
Overview of productivity measures

Common Index formulas

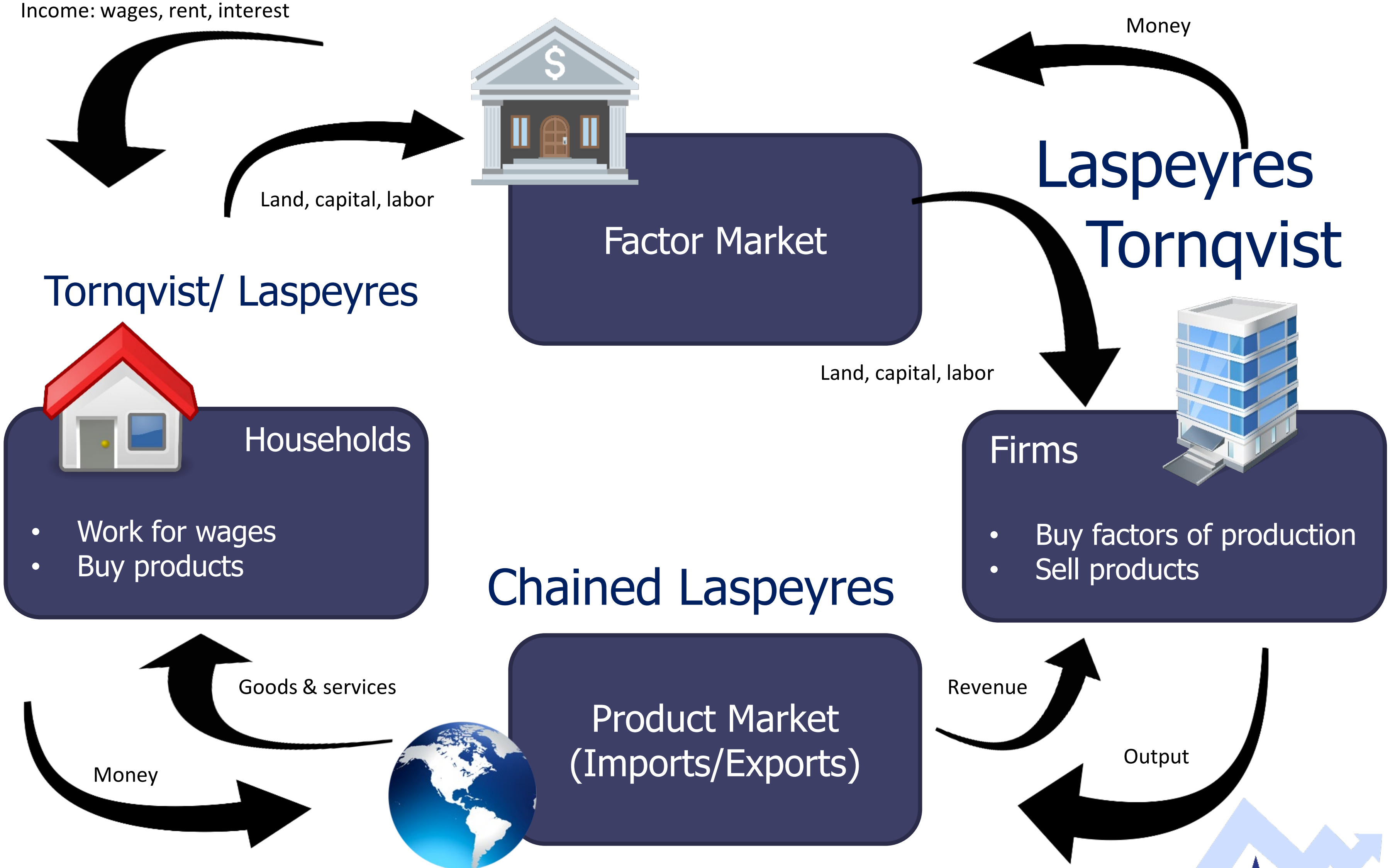
Advantages of Using Indices

- Detailed input data have different units, but the indexes have growth rates, which can be compared.
- The relevant production theory is rearranged to address growth rates only.
- A flexible index function is a good approximation to any smooth production function.
  - Tornqvist index matches the translog production function
- Therefore it is not necessary to specify a production function, and BLS Productivity Office does not pick one.
  - Capital, labor, and intermediate inputs are computed in indexes

Economic Flows



Economic Flows



Common Index formulas

Laspeyres

$$I_L^{0\to1} = \frac{\sum_i q_i^0 * p_i^1}{\sum_i q_i^0 * p_i^0} = \sum_i s_i^0 * \left(\frac{p_i^1}{p_i^0}\right) \; where \; s_i^t = \frac{p_i^t * q_i^t}{\sum_j p_j^t * q_j^t}$$

Paasche

$$I_P^{0\to1} = \frac{\sum_i q_i^1 * p_i^1}{\sum_i q_i^1 * p_i^0} = \left\{\sum_i s_i^1 * \left(\frac{p_i^1}{p_i^0}\right)^{-1}\right\}^{-1}$$

Common Index formulas

Fisher

$$I_F^{0\to1} = \sqrt{I_L^{0\to1} * I_P^{0\to1}}$$

Tornqvist

$$I_T^{0\to1} = \sqrt{\prod_i \left(\frac{p_i^1}{p_i^0}\right)^{s_i^0} * \prod_i \left(\frac{p_i^1}{p_i^0}\right)^{s_i^1}} = \prod_i \left(\frac{p_i^1}{p_i^0}\right)^{(s_i^0+s_i^1)/2}$$

These are "chained" consecutively.

Tornqvist step-by-step

Quantity: Aggregating multiple categories into one

Real dollar Production (based 2009)	2007	2008	2009	2010	2011	2012
Description						
computers equipment	82.087	92.123	100.000	106.682	112.174	116.482
communications equipment	88.651	95.244	100.000	104.228	108.779	113.074
Other equipment	96.484	98.914	100.000	100.449	101.258	102.402

Log change production

Change in natural log Production	2007	2008	2009	2010	2011	2012
Description						
computers equipment		11.5%	8.2%	6.5%	5.0%	3.8%
communications equipment		7.2%	4.9%	4.1%	4.3%	3.9%
Other equipment		2.5%	1.1%	0.4%	0.8%	1.1%

Weights:

Current dollar production cost	2007	2008	2009	2010	2011	2012
Description						
computers equipment	108.094	104.657	100.392	97.303	96.628	94.645
communications equipment	141.665	149.915	134.442	137.370	128.727	138.034
Other equipment	101.840	105.084	100.974	100.363	98.439	102.037
Total production Cost:	351.599	359.656	335.808	335.036	323.794	334.716

Tornqvist step-by-step

Share Weights:

Shares of current dollar cost	2007	2008	2009	2010	2011	2012
Description						
computers equipment	31%	29%	30%	29%	30%	28%
communications equipment	40%	42%	40%	41%	40%	41%
Other equipment	29%	29%	30%	30%	30%	30%
Checking the shares (must sum to 100%)	*100%*	*100%*	*100%*	*100%*	*100%*	*100%*

2 year average shares of cost	2007	2008	2009	2010	2011	2012
Description						
computers equipment		0.299	0.295	0.295	0.294	0.291
communications equipment		0.410	0.409	0.405	0.404	0.405
Other equipment		0.291	0.296	0.300	0.302	0.304

Tornqvist step-by-step

Shares * Log Change

Average shares * Change in natural log	2007	2008	2009	2010	2011	2012
Description						
computers equipment		0.035	0.024	0.019	0.015	0.011
communications equipment		0.029	0.020	0.017	0.017	0.016
Other equipment		0.007	0.003	0.001	0.002	0.003

Sum and chain

Sum data together for aggregate series	2007	2008	2009	2010	2011	2012
Description						
Information Capital		0.071	0.047	0.037	0.034	0.030
(cumlative product cal.)	1	1.074	1.126	1.168	1.209	1.246
Quantity index Information Capital	**0.888**	**0.954**	**1.000**	**1.038**	**1.074**	**1.107**

Common Index formulas

How the Tornqvist Index is used in productivity theory

How it’s used

- Aggregating to create an industry sector
- Combine multiple categories in a particular industry(s)
- Problematic cases for the Tornqvist

Aggregating to create an industry sector

Literature on the theory of index numbers has shown that Tornqvist has desirable properties.

- Exact index for translog structure of production
-Used to combine indices of inputs (Capital and Labor)

*Dean, Harper, Sherwood (1996) page 185

Combine multiple categories in a particular industry(s)

For the same reasons mentioned before the Tornqvist index is used.

Relevant theory only uses growth rates of quantities and prices.

How the Tornqvist Index is used in productivity theory

Demonstrations

1. Information capital measure

2. Manufacturing sector aggregation

R package Dependencies

dplyr

tidyr

R package Dependencies

gather()

```
mutate( )
select( )
arrange()
group_by()
rename()
Left_join()
Inner_join()
```

R package Dependencies

```
mutate( )- create new variables
select( )- keep certain variables
arrange()- sort data frames
group_by()- group data frame for
            following calculations
rename() -  rename variables
left_join()- merge tables toward
             the left most data frame
inner_join()- returns everything
```

```
gather()- transpose
```

R package Dependencies

```
mutate(df, new_var= old_var*9 )

select(df, var1, var2, var3)
arrange(df, var1, var2, var3)

by_cyl<-group_by(df, var1 )
by_cyl2<-summarize(by_cyl, ….)

rename(df, var_newname=var1)
left_join(var1, var2, by=key)
inner_join(var1, var2, by=key)
```

R package Dependencies

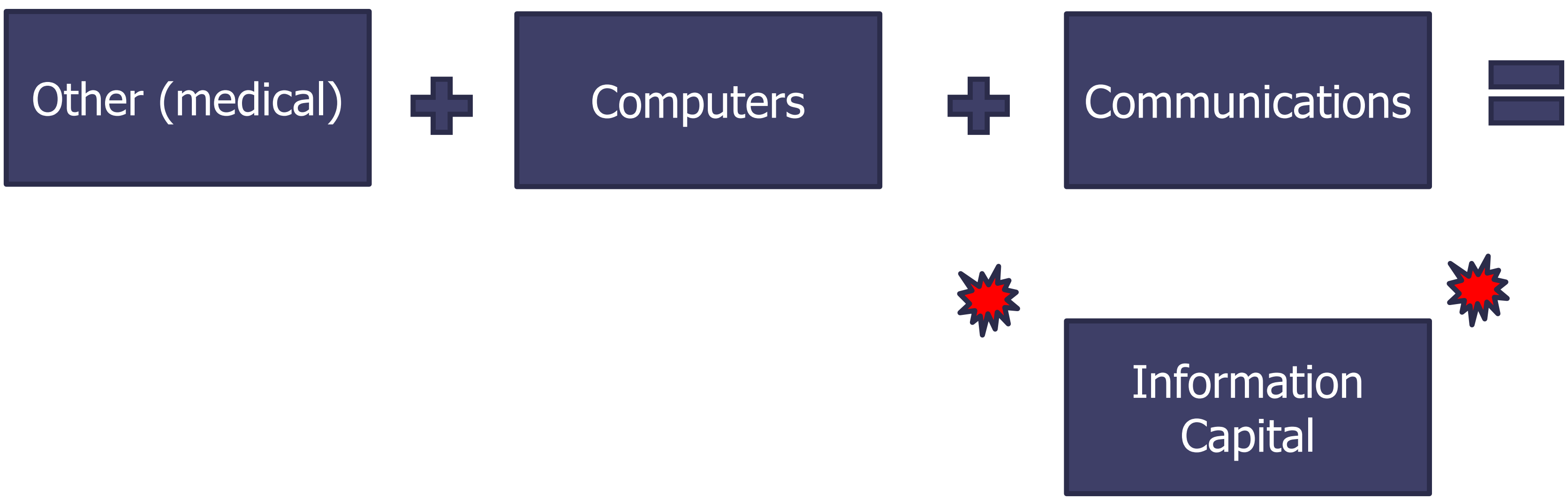
```
gather(df, key="key_var" , value )
```

YEAR	Red	Green
2010	50	80
2011	40	100

VAR	2010	2011
Red	50	40
Green	80	100

Information capital measure

Aggregate multiple measures for each industry to form a new information capital time series.




```
tornqvist_index_m (quantity_var,cost_var,base_year)
```

```
tornqvist_index_m (quantity_var,cost_var,base_year)
```

Data frame

	Year	Computers	Communication	Other	Computers_c	Communication_c	Other_c	sector
1	1987	1.664	19.418	55.062	43.700	44.922	46.803	PG
2	1988	2.039	20.740	56.023	46.765	48.841	48.501	PG
3	1989	2.437	22.068	57.213	56.970	57.428	50.984	PG
4	1990	2.784	23.291	58.508	60.862	61.438	53.932	PG
5	1991	3.043	24.373	59.420	61.666	62.820	54.527	PG
6	1992	3.389	25.417	60.604	63.538	62.939	54.407	PG
7	1993	3.956	26.595	62.713	65.874	66.600	58.315	PG
8	1994	4.785	28.111	65.570	68.374	72.334	64.496	PG
9	1995	6.160	30.107	68.478	72.306	80.288	69.773	PG
10	1996	8.463	32.552	71.459	83.472	84.082	72.919	PG
11	1997	11.976	35.487	74.485	92.559	84.584	74.735	PG
12	1998	17.150	39.007	77.385	104.123	88.189	76.161	PG
13	1999	24.688	43.605	79.517	114.958	93.516	79.685	PG
14	2000	33.839	49.997	80.761	121.164	101.422	81.276	PG

Pre-processing

```
df_cost<-select(df, Computers=Computers_c ,
Communication=Communication_c, Other=Other_c, sector, Year)

df_qty<-select(df, Computers=Computers,
Communication=Communication, Other= Other, sector, Year)
```

```
tornqvist_index_m<-function(quantity_var,cost_var,base_year){
  {
  #count the number of assets to be aggregated in the TQ
  colmn<-ncol(cost_var)
  totvars<-colmn-2

  # total number of variables is stored in object totvars
  # minus 2- remove year and sector

  #block 1
  df_cost_t<-gather(cost_var, ems,cost, 1:totvars)

  df_qty_t<-gather(quantity_var, ems,qty, 1:totvars)

  all<-inner_join(df_cost_t,df_qty_t, by= c("sector"="sector", "ems"="ems","Year"="Year"))

  sort_all<-arrange(all, sector, Year)

  total_cost<-aggregate(cost~sector+Year, sort_all,sum)

  inner_join(sort_all,total_cost, by= c("sector"="sector", "Year"="Year"))%>%mutate(value_share=cost.x/cost.y)->total_cost_2

  tornqvist_data_1<-arrange(total_cost_2, sector,ems, Year)

  tornqvist_data_2<-tornqvist_data_1%>%group_by(sector,ems)%>%mutate(lag_vshare=dplyr::lag(value_share, n=1, default=NA),
                                                                    lag_qty=dplyr::lag(qty, n=1, default=NA))

  tornqvist_data_3<-tornqvist_data_2%>%group_by(sector,ems,Year)%>% mutate(ave_value_share =  ((value_share + lag_vshare )* 0.5),
                                                                          log_quantity = log(qty) - log (lag_qty),
                                                                          Product = log_quantity * ave_value_share )

  tornqvist_data_4<-tornqvist_data_3%>%select(Year, ems, sector, Product,cost.y)

  #block 2

  tornqvist_5<-tornqvist_data_4%>%group_by(sector,Year)%>% summarise(quantity_1=sum(Product), vp=mean(cost.y))

  tornqvist_6<-tornqvist_5%>%group_by(sector)%>% mutate(Q = exp(quantity_1) ,Q=if_else(is.na(Q), 1,Q),q_cumulate = cumprod(Q))

  ty<-tornqvist_6[tornqvist_6$Year == base_year, ]%>%select(sector, Year,q_cumulate)%>%rename(yrbaseyrq_cumul=q_cumulate)

  tornqvist_7<-merge(tornqvist_6,ty, by="sector", all=TRUE)%>%mutate(quantity=q_cumulate/yrbaseyrq_cumul)%>%
    select(-Year.y,-yrbaseyrq_cumul,-Q,-quantity_1,-q_cumulate)%>%rename(Year=Year.x)%>%mutate(price=vp/quantity)
}
  return(tornqvist_7)
}
```

```
tornqvist_index_m<-function(quantity_var,cost_var,base_year){
  {
   [ . . . . . . . CODE . . . . . . .]

  #block 1
       df_cost_t<-gather(cost_var, ems,cost, 1:totvars)

       df_qty_t<-gather(quantity_var, ems,qty, 1:totvars)

       all<-inner_join(df_cost_t,df_qty_t, by= c("sector"="sector", "ems"="ems","Year"="Year"))

       sort_all<-arrange(all, sector, Year)

       total_cost<-aggregate(cost~sector+Year, sort_all,sum)


inner_join(sort_all,total_cost, by= c("sector"="sector", "Year"="Year"))%>%mutate(value_share=cost.x/cost.y)-
>total_cost_2

  tornqvist_data_1<-arrange(total_cost_2, sector,ems, Year)

  tornqvist_data_2<-tornqvist_data_1%>%group_by(sector,ems)%>%mutate(lag_vshare=dplyr::lag(value_share, n=1,
default=NA),
                                                                     lag_qty=dplyr::lag(qty, n=1, default=NA))

  tornqvist_data_3<-tornqvist_data_2%>%group_by(sector,ems,Year)%>% mutate(ave_value_share =  ((value_share +
lag_vshare )* 0.5),
                                                                           log_quantity = log(qty) - log (lag_qty),
                                                                           Product = log_quantity * ave_value_share )

  tornqvist_data_4<-tornqvist_data_3%>%select(Year, ems, sector, Product,cost.y)
```

```
tornqvist_index_m<-function(quantity_var,cost_var,base_year){
  {

 [ . . . . . . . CODE . . . . . . .]

  #block 1
………………… #code from before ……………………………………………………….
                 …………………………………….

inner_join(sort_all,total_cost, by= c("sector"="sector",
"Year"="Year"))%>%mutate(value_share=cost.x/cost.y)->total_cost_2

tornqvist_data_1<-arrange(total_cost_2, sector,ems, Year)

tornqvist_data_2<-tornqvist_data_1%>%group_by(sector,ems)%>%mutate(lag_vshare=dplyr::lag(value_share,
n=1, default=NA), lag_qty=dplyr::lag(qty, n=1, default=NA))

tornqvist_data_3<-tornqvist_data_2%>%group_by(sector,ems,Year)%>% mutate(ave_value_share =
((value_share + lag_vshare )* 0.5), log_quantity = log(qty) - log (lag_qty),

Product = log_quantity * ave_value_share )

tornqvist_data_4<-tornqvist_data_3%>%select(Year, ems, sector, Product,cost.y)
```

```
tornqvist_index_m<-function(quantity_var,cost_var,base_year){
  {

………………… #code from before ……………………………………………………….
                 …………………………………….

#block 2

tornqvist_5<-tornqvist_data_4%>%group_by(sector,Year)%>%
summarise(quantity_1=sum(Product), vp=mean(cost.y))

  tornqvist_6<-tornqvist_5%>%group_by(sector)%>% mutate(Q = exp(quantity_1),
Q=if_else(is.na(Q), 1,Q),q_cumulate = cumprod(Q))

 ty<-tornqvist_6[tornqvist_6$Year == base_year, ]%>%select(sector,
Year,q_cumulate)%>%rename(yrbaseyrq_cumul=q_cumulate)

 tornqvist_7<-merge(tornqvist_6,ty, by="sector", all=TRUE)%>%mutate(quantity=q_cumulate/yrbaseyrq_cumul)%>%
    select(-Year.y,-yrbaseyrq_cumul,-Q,-quantity_1,-q_cumulate)%>%rename(Year=Year.x)%>%mutate(price=vp/quantity)

}
  return(tornqvist_7)
}
```

```
tornqvist_index_m<-function(quantity_var,cost_var,base_year){
  {

………………… #code from before ………………………………………………………….
                ……………………………………….

#block 2

tornqvist_5<-tornqvist_data_4%>%group_by(sector,Year)%>% summarise(quantity_1=sum(Product),
vp=mean(cost.y))

  tornqvist_6<-tornqvist_5%>%group_by(sector)%>% mutate(Q = exp(quantity_1) ,Q=if_else(is.na(Q),
1,Q),q_cumulate = cumprod(Q))

 ty<-tornqvist_6[tornqvist_6$Year == base_year, ]%>%select(sector,
Year,q_cumulate)%>%rename(yrbaseyrq_cumul=q_cumulate)

tornqvist_7<-merge(tornqvist_6,ty, by="sector",
all=TRUE)%>%mutate(quantity=q_cumulate/yrbaseyrq_cumul)%>%

select(-Year.y,-yrbaseyrq_cumul,-Q,-quantity_1,-
q_cumulate)%>%rename(Year=Year.x)%>%mutate(price=vp/quantity)

}
  return(tornqvist_7)
}
```

Data frame

Comp + Comm + Other (PB)

	Year	Computers	Communication	Other	Computers_c	Communication_c	Other_c	sector
1	1987	1.664	19.418	55.062	43.700	44.922	46.803	PG
2	1988	2.039	20.740	56.023	46.765	48.841	48.501	PG
3	1989	2.437	22.068	57.213	56.970	57.428	50.984	PG
4	1990	2.784	23.291	58.508	60.862	61.438	53.932	PG
5	1991	3.043	24.373	59.420	61.666	62.820	54.527	PG
6	1992	3.389	25.417	60.604	63.538	62.939	54.407	PG
7	1993	3.956	26.595	62.713	65.874	66.600	58.315	PG
8	1994	4.785	28.111	65.570	68.374	72.334	64.496	PG
9	1995	6.160	30.107	68.478	72.306	80.288	69.773	PG
10	1996	8.463	32.552	71.459	83.472	84.082	72.919	PG
11	1997	11.976	35.487	74.485	92.559	84.584	74.735	PG
12	1998	17.150	39.007	77.385	104.123	88.189	76.161	PG
13	1999	24.688	43.605	79.517	114.958	93.516	79.685	PG
14	2000	33.839	49.997	80.761	121.164	101.422	81.276	PG

Information capital (PB)

	sector	Year	vp	quantity	price
1	PG	1987	135.425	0.1093872	1238.033
2	PG	1988	144.107	0.1201389	1199.504
3	PG	1989	165.382	0.1311526	1260.989
4	PG	1990	176.232	0.1408814	1250.924
5	PG	1991	179.013	0.1482953	1207.138
6	PG	1992	180.884	0.1571651	1150.917
7	PG	1993	190.789	0.1702614	1120.565
8	PG	1994	205.204	0.1877538	1092.942
9	PG	1995	222.367	0.2119566	1049.116
10	PG	1996	240.473	0.2456828	978.7947
11	PG	1997	251.878	0.2900702	868.3347
12	PG	1998	268.473	0.3466215	774.5423
13	PG	1999	288.159	0.4180275	689.3303
14	PG	2000	303.862	0.4979744	610.196
15	PG	2001	308.538	0.5696786	541.6001
16	PG	2002	300.944	0.620681	484.861
17	PG	2003	308.713	0.663964	464.9544
18	PG	2004	316.516	0.7100554	445.7624
19	PG	2005	349.197	0.7595634	459.7338
20	PG	2006	358.324	0.8186554	437.6982

Manufacturing sector aggregation

Aggregate multiple different industries (NAICS) into one super-sector.

```
tornqvist_index_ind(input,base_year,NAME)
```

Energy

```
tornqvist_index_ind(input,base_year,NAME)
```

Pre-processing

```
input<-select(df, ce , qex, naics, year)
```

```
tornqvist_index_ind<-function(input,base_year,NAME){
  {

    df_2<-data.frame(input)%>%arrange(year)

   #block 1
    #get the total cost of all the industries
    df_3<-aggregate(ce~year, df_2,sum)%>%rename(total_cost=ce)

    df_4<-left_join(input2,df_3, by= c("year"="year"))%>%mutate(value_share=ce/total_cost)

    df_5<-df_4%>%arrange(naics,year)

    df_6<-df_5%>%group_by(naics)%>%mutate(lag_vshare=lag(value_share),
                                      lag_qex=lag(qex), ave_valueshare =  ((value_share + lag_vshare)*0.5) ,
                                      Log_qty_chg = log(qex) - log(lag_qex),
                                      Product = Log_qty_chg * ave_valueshare)

    df_7<-df_6%>%arrange(year,naics)

#block 2
    #get the total contributions of all industries
    df_8<-aggregate(Product~year, df_7,sum)%>%rename(TQ=Product)%>%
      mutate(quant =exp(TQ))

    df_9<-left_join(df_3,df_8, by= c("year"="year"))

    df_10<-select(df_9, year , total_cost, quant)%>%
      mutate(quant= ifelse(is.na(quant),1,quant), q_cumulate = cumprod(quant),NAICS=NAME)

    df_11<-df_10[df_10$year == base_year, ]%>%select( year,q_cumulate)%>%
      rename(yrbaseyrq_cumul=q_cumulate)%>%mutate(NAICS=NAME)

    tq_1<-merge(df_10,df_11,by="NAICS",all=TRUE)%>%
      mutate(quant_index=q_cumulate/yrbaseyrq_cumul)%>%
      select(-year.y,-yrbaseyrq_cumul,-quant,-q_cumulate)%>%rename(year=year.x)%>%mutate(price=total_cost/quant_index)

  }
  return(tq_1)
}
```

```
tornqvist_index_ind<-function(input,base_year,NAME){
  {

    df_2<-data.frame(input)%>%arrange(year)

   #block 1
    #get the total cost of all the industries
 df_3<-aggregate(ce~year, df_2,sum)%>%rename(total_cost=ce)

 df_4<-left_join(input2,df_3, by= c("year"="year"))%>%mutate(value_share=ce/total_cost)

 df_5<-df_4%>%arrange(naics,year)

    df_6<-df_5%>%group_by(naics)%>%mutate(lag_vshare=lag(value_share),
                                 lag_qex=lag(qex), ave_valueshare =  ((value_share + 
lag_vshare)*0.5) ,
                                 Log_qty_chg = log(qex) - log(lag_qex),
                                 Product = Log_qty_chg * ave_valueshare)

    df_7<-df_6%>%arrange(year,naics)
```

```
tornqvist_index_ind<-function(input,base_year,NAME){
  {

    df_2<-data.frame(input)%>%arrange(year)

   #block 1
    #get the total cost of all the industries

    df_3<-aggregate(ce~year, df_2,sum)%>%rename(total_cost=ce)

    df_4<-left_join(input2,df_3, by= c("year"="year"))%>%mutate(value_share=ce/total_cost)

    df_5<-df_4%>%arrange(naics,year)

    df_6<-df_5%>%group_by(naics)%>%mutate(lag_vshare=lag(value_share),
       lag_qex=lag(qex), ave_valueshare =  ((value_share + lag_vshare)*0.5) ,
                                  Log_qty_chg = log(qex) - log(lag_qex),
                                  Product = Log_qty_chg * ave_valueshare)

    df_7<-df_6%>%arrange(year,naics)
```

```
tornqvist_index_ind<-function(input,base_year,NAME){
  {

……………… #code from before ……………………………………………………….
                    …………………………………….

#block 2
    #get the total contributions of all industries
    df_8<-aggregate(Product~year, df_7,sum)%>%rename(TQ=Product)%>%

    mutate(quant =exp(TQ))

    df_9<-left_join(df_3,df_8, by= c("year"="year"))

    df_10<-select(df_9, year , total_cost, quant)%>%

    mutate(quant= ifelse(is.na(quant),1,quant), q_cumulate = cumprod(quant),NAICS=NAME)


    df_11<-df_10[df_10$year == base_year, ]%>%select( year,q_cumulate)%>%
      rename(yrbaseyrq_cumul=q_cumulate)%>%mutate(NAICS=NAME)

    tq_1<-merge(df_10,df_11,by="NAICS",all=TRUE)%>%
      mutate(quant_index=q_cumulate/yrbaseyrq_cumul)%>%
      select(-year.y,-yrbaseyrq_cumul,-quant,-
q_cumulate)%>%rename(year=year.x)%>%mutate(price=total_cost/quant_index)

  }
  return(tq_1)
}
```

```
tornqvist_index_ind<-function(input,base_year,NAME){
  {

……………… #code from before …………………………………………………….
                     …………………………………….

#block 2
    #get the total contributions of all industries
    df_8<-aggregate(Product~year, df_7,sum)%>%rename(TQ=Product)%>%
      mutate(quant =exp(TQ))

    df_9<-left_join(df_3,df_8, by= c("year"="year"))

    df_10<-select(df_9, year , total_cost, quant)%>%
      mutate(quant= ifelse(is.na(quant),1,quant), q_cumulate = cumprod(quant),NAICS=NAME)

 df_11<-df_10[df_10$year == base_year, ]%>%select( year,q_cumulate)%>%

 rename(yrbaseyrq_cumul=q_cumulate)%>%mutate(NAICS=NAME)

 tq_1<-merge(df_10,df_11,by="NAICS",all=TRUE)%>%mutate(quant_index=q_cumulate/yrbaseyrq_cumul)%>%

 select(-year.y,-yrbaseyrq_cumul,-quant,

-q_cumulate)%>%rename(year=year.x)%>%mutate(price=total_cost/quant_index)

  }

  return(tq_1)
}
```

Data frame

NAICS 31-33

	ce	qex	naics	year
1	2064.2662	1.4035696	I361	1987
2	1893.4473	1.3855377	I361	1988
3	1935.1432	1.3681682	I361	1989
4	1920.8109	1.3605698	I361	1990
5	1860.4552	1.3285226	I361	1991
6	1932.6738	1.3881844	I361	1992
7	1963.6331	1.4692636	I361	1993
8	1921.2055	1.4844666	I361	1994
9	1921.0301	1.5463229	I361	1995
10	1982.6745	1.4761900	I361	1996
11	2148.0000	1.5260400	I361	1997
12	2049.0000	1.5114800	I361	1998
13	3134.0000	2.3049200	I361	1999
14	3341.0000	2.2656500	I361	2000
15	4738.0000	2.9203300	I361	2001

Manufacturing sector

NAICS	year	total_cost	quant_index	price
MAN	1997	3062.356	1.373999	2228.791
MAN	1998	3285.677	1.519951	2161.7
MAN	1999	4333.253	2.014493	2151.039
MAN	2000	4404.168	1.897839	2320.623
MAN	2001	6585.495	2.577071	2555.418
MAN	2002	4259.856	1.765701	2412.558
MAN	2003	4005.216	1.480775	2704.811
MAN	2004	3742.877	1.299218	2880.87
MAN	2005	4387.696	1.32016	3323.61
MAN	2006	3989.523	1.14199	3493.482
MAN	2007	4479.381	1.249818	3584.026
MAN	2008	4849.506	1.222089	3968.21
MAN	2009	3472.497	1	3472.497
MAN	2010	3483.168	0.9628227	3617.663
MAN	2011	3113.407	0.829227	3754.59
MAN	2012	3647.99	1.013922	3597.9
MAN	2013	3913.847	1.048001	3734.585
MAN	2014	4449.241	1.128193	3943.687
MAN	2015	3737.494	1.080782	3458.138
MAN	2016	3266.091	0.9977078	3273.595

Creative learning

Why did the statisticians and economists go to the forest?

Limitation of Tornqvist

Equation gives unstable values when

- data has zeros
- missing values
- extreme growth rates from period to period

A quantity near zero, even with a small expenditure share, can greatly affect the index.

General recommendations for hard cases

- Check data tables for zeros or extreme growth rates
- Replace outliers or put cap on growth rates
- Address mainly with *imputation*, *smoothing* and/or *merging* of items
- Avoid imputing hard-coded figures such as 0.01 if possible
- After replacing a value, consider rescaling so totals stay the same

Functions to smooth/merge

- One approach is to merge an item with extreme growth rates into a larger group
- Or, user makes imputation based on knowing the context of the data.
- If neither, these functions now under development, can test an input series to see whether an extreme growth rate (like 20:1) is exceeded and adjust outliers minimally to limit those growth rates.

```
steeptest(vector, .05, 20)
steeplimit(vector, .05, 20)
```

Demonstrations

?

Review

- ✓-Overview of productivity measures
- ✓-Common Index formulas
- ✓-How Tornqvist is used
- ✓-Demonstrations

Contact Information

Mike Jadoo / Economist
Peter B. Meyer / Research Economist

Office of Productivity and Technology,
U.S. Bureau of Labor Statistics
www.bls.gov/mfp
202-691-5601
Jadoo.michael@bls.gov