

# History Carried Forward: Mixing Time Series of Differing Frequencies (Version 0.9)

Mark Keintz

PhilaSUG

Oct 30, 2018

# Mixing Time Series Frequencies

## The “Take Home” Points

- SET statement variables are retained over data step iterations - until a SET reading the same variables is executed. (MERGE too).
- So conditional SETs (if xxx then set yyy) can preserve variables over many iterations, providing a means of carrying forward “historic” data.

# Mixing Time Series Frequencies

## The “Take Home” Points

- For sorted data sets, this is much faster than commonly used PROC SQL approach.
- Using strategically placed “sentinel” variables provides a way to set “stale” data to missing values.

# The Problem: How to “carry forward” Low Frequency Data Records

- Start with data set YEAR (one obs per year) and QTR (four per year)
  - Both sorted by ID and DATE
  - Otherwise with mutually exclusive variables
- Task: propagate the yearly data through all subsequent quarterly records preceding next yearly DATE

# Sample YEAR Data (2 ID's, 3 Years)

ID	DATE	AstA (\$mm)	LbtA (\$mm)
XX	30JUN2014	821	281
XX	30JUN2015	799	303
XX	30JUN2016	804	322
YY	31DEC2014	1,401	904
YY	31DEC2015	1,427	950
YY	31DEC2016	1,550	962

# Sample QTR Data (showing 1 ID)

ID	DATE	SalQ (\$mm)	EmpQ(1,000's )
XX	30JUN2014	132.3	13.5
XX	30SEP2014	128.9	12.6
XX	31DEC2014	138.3	13.5
XX	31MAR2015	112.0	11.3
XX	30JUN2015	115.5	11.7
XX	31SEP2015	140.2	14.2
XX	...	...	...
XX	31MAR2017	98.9	9.9

# Sample Yearly Data Carried Forward Into Quarterly Series

ID	DATE	YRDATE	AstA	LbtA	SalQ	EmpQ
XX	30JUN2014	30JUN2014	821	281	132.3	13.5
XX	30SEP2014	30JUN2014	821	281	128.9	12.6
XX	31DEC2014	30JUN2014	821	281	138.3	13.5
XX	31MAR2015	30JUN2014	821	281	112.0	11.7
XX	30JUN2015	30JUN2015	799	303	115.5	14.2
XX	31SEP2015	30JUN2015	799	322	140.2	10.6
XX	...	...	...	...	...	...
XX	31MAR2017	31JUN2016	804	322	98.9	9.9

# Yearly Data Carry Forward via PROC SQL

## An Expensive Solution

```
proc sql noprint;
  create table yrqtr as
  select *
    from qtr
  left join
    year (rename=(date=yrcode))

  on year.id=qtr.id
  and yrcode<=date<intnx('year',yrcode,1,'s')
  order by id,date;
quit;
```



# Much Faster with Selective SET Statements

```
data want;
  set YEAR (in=inyear keep=id date)
      QTR (in=inq) ;
  by id date;

  /*Read ALL the year variables */
  if inyear then set YEAR (rename=(date=yrdate));

  if inq then output;

  /* Don't carry data across ID boundaries*/
  if last.id then call missing(of _all_);
run;
```

# More than Two Series with Selective SET Statements

```
data YQM ;
  set YEAR  (in=inY keep=id date)
      QTR   (in=inQ keep=id date)
      MONTH (in=inM); /*No KEEP for highest freq*/
  by id date;

  if inY then set YEAR (rename=(date=YR_date));
  if inQ then set QTR  (rename=(date=QTR_date));

  if inM then output;
  if last.id then call missing(of _all_);
run;
```

# Using Sentinel Variables to Prevent Stale Historic Data

```
data YQTR_holes (drop=_sentinel:);
  set YEAR  (in=inyear keep=id date)
      QTR    (in=inq);
  by id date;
  retain _sentinel1 .;
  if inyear then set YEAR (rename=(date=YR_date));
  retain _sentinel2 .;
  if yr_date^=. and intck('qtr',yr_date,date)>3
then call missing(of _sentinel1--_sentinel2);

  if inq then output;
  if last.id then call missing(of _all_);
run;
```

# Sample Yearly Data With Missing Year Data for 30JUN2015

ID	DATE	YRDATE	AstA	LbtA	SalQ	EmpQ
XX	30JUN2014	30JUN2014	821	281	132.3	13.5
XX	30SEP2014	30JUN2014	821	281	128.9	12.6
XX	31DEC2014	30JUN2014	821	281	138.3	13.5
XX	31MAR2015	30JUN2014	821	281	112.0	11.7
XX	30JUN2015	.	.	.	115.5	14.2
XX	31SEP2015	.	.	.	140.2	10.6
XX	...	...	...	...	...	...
XX	31MAR2017	31DEC2016	804	322	98.9	9.9

# Irregular Series Data Set

## Admit/Discharge

ID	DATE	ACTION	ACTIONMMD
101	15JAN2015	A	X2
101	20JAN2015	D	C5
102	15FEB2014	A	X1
102	25FEB2014	D	C9

# Irregular Series Data Set Services

ID	DATE	SVCID	CHGUNITS
101	18JAN2015	323	3.2
101	19JAN2015	488	1.2
102	15FEB2014	101	3.0
102	16FEB2014	229	1.7

# Irregular Series Data Set TESTS

ID	DATE	TSTLAB	TSTTYPE
101	21JAN2015	788	VIS
102	15JAN2015	823	HRT

# Mixing Irregular Series

- **data** want;
- merge admdis (in=inA keep=id date)
- srvcs (in=inS keep=id date)
- tests (in=inT keep=id date);
- by id date;
- if inA then set admdis (rename=(date=date\_A));
- if inS then set srvcs (rename=(date=date\_S));
- if inT then set tests (rename=(date=date\_T));
- output;
- if last.id then call missing (of \_all\_);
- **run;**



# Mixing Irregular Series: Results

ID	DATE	A/D	AID	SID	CHG	LAB	LTYPE
101	15JAN2015	A	X2				
101	18JAN2015	A	X2	323	3.2		
101	19JAN2015	A	X2	488	1.2		
101	20JAN2015	D	C5	488	1.2		
101	21JAN2015	D	C5	488	1.2	788	VIS
102	15FEB2014	A	X1	101	3.0	823	HRT
102	15FEB2014	A	X1	229	1.7	823	HRT
102	25FEB2015	D	C9	229	1.7	823	HRT

# Questions?

Mark Keintz

[mkeintz@wharton.upenn.edu](mailto:mkeintz@wharton.upenn.edu)