Compatible Trends for ACS Data

Tucker McElroy
U.S. Census Bureau

U S C E N S U S B U R E A U

American Community Survey (ACS)

- ACS replaces Census Long Form
- Sampling error is reduced through rolling samples (pooling over time)
- This induces a time lag in the reported Multi-Year Estimates (MYEs)
- Lag makes diverse ACS data incompatible

ACS

- Data is pooled in 1-year, 3-year, or 5-year windows, according to population size in the region

- Sample is equally weighted over the years in the rolling sample

- Time lag property: 3y MYE is centered at previous year; 5y MYE is centered two years back

ACS Stylized Example

- Say a sample is collected for 3001 through 3005.
- In 3005, the 5y MYE consists of an estimate computed over all past five years
- In 3005, the 3y MYE consists of an estimate computed over years 3003, 3004, 3005
- The 1y MYE is an estimate just based on sample over 3005
- For many types of estimates, the 5y MYE is centered at 3003 value, and the 3y MYE is centered at a 3004 value. Generally true of locally linear MYEs.

ACS Stylized Example

- Suppose we are measuring an increasing quantity (e.g., a population) in two different counties A and B
- A 1y 3005 MYE for county A may be spuriously larger than a 5y 3005 MYE for county B
- We should compare the 1y 3003 MYE for county A to the 5y 3005 MYE for county B
- Drawback: we throw away the 3004 and 3005 1y MYEs for county A. These should be used somehow

Compatible Trends

We propose using trend filters to match up different MYEs such that:

1. The 1y, 3y, and 5y trends are *compatible* (close to identical)
2. Linear dynamics in the time series are treated appropriately
3. The filters are concurrent
4. Filter length is minimal

Compatible Trends

Let the ky MYE be denoted $Y_t^{(k)}$, and use the representation for $k = 3, 5$

$$Y_t^{(k)} = \Theta^{(k)}(B) X_t$$

with B the backshift operator, X_t the 1y MYE, and $\Theta^{(k)}(z)$ is the Simple Moving Average (SMA) polynomial of order k given by

$$\Theta^{(k)}(z) = \frac{1}{k} \left(1 + z + \cdots + z^{k-1} \right).$$

This representation is approximately true for many types of estimates.

Compatible Trends

We seek three concurrent filters $\Psi^{(k)}(B) = \sum_{j \geq 0} \psi_j^{(k)} B^j$ such that when applied to the respective MYEs $Y_t^{(k)}$, they produce the *exact* same series. So this condition is

$$\Psi^{(1)}(B) = \Psi^{(3)}(B)\Theta^{(3)}(B) = \Psi^{(5)}(B)\Theta^{(5)}(B).$$

These conglomerates should pass any linear effects in the X_t series without distortion, so we impose the condition

$$\Psi^{(k)}(B)\Theta^{(k)}(B)\ [at + b] = at + b$$

for any a, b and all times t.

Compatible Trends

Theorem 1. *The minimal length concurrent filters $\Psi^{(k)}$ satisfying these conditions are given by*

$$\Psi^{(5)}(B) = \left(4 + B + B^2 - 3B^3\right)/3$$

$$\Psi^{(3)}(B) = \left(4 + B + B^2 + B^3 + B^4 - 3B^5\right)/5$$

$$\Psi^{(1)}(B) = \left(4 + 5B + 6B^2 + 3B^3 + 3B^4 - B^5 - 2B^6 - 3B^7\right)/15.$$

These will be referred to as the 5y, 3y, and 1y trend filters respectively.

Application of Compatible Trends

- Suppose we have several different counties, and for some the 1y or 3y MYEs are not available

- For each county, apply the 1y, 3y, or 5y trend filters to the 1y, 3y, or 5y MYE as appropriate. Then make comparisons between these trend quantities.

- Advantages: the 5y and 3y MYEs are essentially advanced forward two or one time unit by the trend filters, to be made compatible with the 1y MYEs. This is preferable to lagging the 1y MYE. This forward advance applies principally to the linear aspects of the time series.

- Validity: requires the condition $Y_t^{(k)} = \Theta^{(k)}(B) X_t$ of the rolling sample

Example 1: Mean Travel Time in Bronx, NYC

	MYEs		
Time	1y	3y	5y
99	**40.05**		
00	40.00		
01	41.00	40.00	
02	41.80	41.00	
03	40.80	41.20	40.70
04	40.60	41.00	40.80
05	41.70	41.10	41.20
06	**42.04**	**41.28**	**41.45**

Table 1: MYEs for Mean Travel Time of Bronx, NYC, New York in minutes. Estimates have been backcast and forecast extended to the years 99 and 06, written in bold.

Example 1: Mean Travel Time in Bronx, NYC

- The available data is slightly augmented. It has a slight upward linear trend

- Nearly compatible: sub-diagonals almost constant

- We apply the 1y, 3y, and 5y trend filters to obtain:

$$41.79 \qquad 41.64 \qquad 41.90$$

Very close agreement in this case.

Example 2: Child Poverty in Bronx, NYC

	MYEs		
Time	1y	3y	5y
99	**42.700**		
00	42.200		
01	41.600	42.400	
02	41.300	41.700	
03	40.400	41.200	41.500
04	42.500	41.400	41.100
05	39.600	40.900	40.600
06	**39.080**	**40.525**	**40.150**

Table 2: MYEs for Percent of families below Poverty Rate in last 12 months, with children aged 5 to 17. Estimates have been backcast and forecast extended to the years 99 and 06, written in bold.

Example 2: Child Poverty in Bronx, NYC

- The available data is slightly augmented. Linearity is more questionable

- Compatibility is less clear

- We apply the 1y, 3y, and 5y trend filters to obtain:

$$40.02 \qquad 40.02 \qquad 39.27$$

Close agreement again.

When Compatibility Fails

- Write $Y_t^{(k)} = \Theta^{(k)}(B) X_t + \epsilon_t^{(k)}$; compatibility breaks down when $|\epsilon_t^{(k)}|$ is large

- We propose to measure compatibility for $k = 3, 5$ via

$$C^{(k)} = \max_t \left| \log Y_t^{(k)} - \log \left(\Theta^{(k)}(B) Y_t^{(1)} \right) \right|.$$

- Rule of thumb: $C^{(k)} < 1\%$ required for compatibility.

- Travel: $C^{(3)} = .07\%$, $C^{(5)} = .04\%$; Poverty: $C^{(3)} = 1.06\%$, $C^{(5)} = .52\%$.

Conclusion

- The time-lag effect of using a rolling sample can be countered through proper filter design

- Validity requires compatibility, i.e., rolling sample is well-approximated by a temporal average

- Use caution with comparisons that cross county and MYE-type. We recommend making comparisons only between trend-filtered estimates

- Extensions to turning-point filters and forecast filters are available