

**MEMO TO:** Case study teams  
**FROM:** Steve O'Connell  
**DATE:** December 3, 2002  
**SUBJECT:** Differences between half-decadal *dyn* and *gyso*

### *1. Introduction*

In the O'Connell/Ndulu (1999) paper we used two different series for growth in real GDP per capita: *dyn* for the pooled conditional model, and *gyso* for the Hoeffler augmented Solow model. The purpose of this note is to explain the differences between half-decadal values of these two variables. As a general caveat, we urge country teams to take the following approach:

Do not allow differences in the two series to get in the way of the presentation and analysis of stylized facts in your macro/growth section. Particularly in your exposition, choose one series and stick with it. Do not present different growth rates in different tables except where necessary, and if you do so add a footnote to each table noting that the two series are different and indicating whether the differences are critical to your interpretation or not.

In most cases, the differences between the two series are not large enough to alter the overall story or the key puzzles of growth in your country. In the few cases in which differences are really large, you must decide whether anything can be learned by examining the differences. While the guidelines below may be helpful in this respect, be prepared to come to terms with the fact that both series are subject to substantial measurement error. To the degree possible, emphasize themes that are supported by both series. Where differences are large, use what you know about your country's growth record to choose which of the two series to emphasize.

Remember that a key reason for differences between the Hoeffler and Ndulu/O'Connell decompositions is that the *empirical models* are different (see the O'Connell and Ndulu framework paper, "African Growth Performance: A Focus on Sources of Growth," April 1999). The Hoeffler model is a structural model in which investment appears on the right-hand side and the parameters are estimated by instrumental variables; the Ndulu/O'Connell model is a reduced-form in which investment is excluded and the parameters are estimated by OLS. Adjust your use of these models accordingly. For example, a hallmark of the Hoeffler approach is that the use of instrumental variables resulted in a very large parameter on investment. Moreover, as you know, African investment rates are often very low at international prices. When very low investment is multiplied by a very large parameter, actual growth may be dramatically under-predicted because the estimated 'contribution' of investment is so small. In such cases, do not let the large average residual throw you off; check whether changes in the Hoeffler residuals from episode to episode largely agree with those coming from the Ndulu/O'Connell model. The watch-word is: be creative, look for the common stylized themes, and then focus on the rest of your paper.

## 2. The two growth variables

The two variables are defined as follows, using the half-decade between 1970 and 1975 as an example:

$$dyn = \frac{1}{5} \sum_{t=1970}^{1974} 100 \cdot \left( \frac{y_t - y_{t-1}}{y_{t-1}} \right)$$

$$gyso = \frac{1}{5} \sum_{t=1971}^{1975} 100 \cdot (\ln \tilde{y}_t - \ln \tilde{y}_{t-1}) = \frac{1}{5} \cdot 100 \cdot (\ln \tilde{y}_{1975} - \ln \tilde{y}_{1970}),$$

where  $y$  is real GDP per capita in constant local currency units and  $\tilde{y}$  is real GDP per capita in constant US dollars at 1985 international prices (sometimes referred to as PPP-adjusted real GDP).

While these two variables should tell roughly the same story, they are not identical. The correlation between the two series, for countries in SSA and excluding the 1990s (for reasons we explain in 3(a) below), is only 0.68. It is apparent from the definitions above that there are three explanations for the lack of a perfect correlation between the two series: the underlying annual growth rates are calculated differently, the half-decades include different sets of years, and the underlying real GDP variables are defined differently. I briefly review these below, from the empirically *least* important to the *most* important.

Difference 1: The method for calculating growth is different because *dyn* uses the discrete-time formula while *gyso* uses the difference in logs. This is not an important difference, however, when time averages are used.

- (a) The two formulas are often used interchangeably because the approximation  $\ln(1+x) \cong x$  is very good for small values of  $x$ . If we define  $x$  as

$$\frac{y_t - y_{t-1}}{y_{t-1}}, \text{ then } \ln(1+x) \text{ is exactly equal to } \ln y_t - \ln y_{t-1}, \text{ and the}$$

approximation therefore tells us that  $\frac{y_t - y_{t-1}}{y_{t-1}} \cong \ln y_t - \ln y_{t-1}$ . Figure 1 shows

the approximation, which is clearly better the closer the growth rate is to zero.

- (b) The figure also shows that  $x > \ln(1+x)$  for any nonzero value of  $x$ . The discrete-time formula will therefore systematically produce a slightly higher average value for growth if the same underlying series and set of years is used (in the annual data, for example, applying the discrete-time formula rather than the log formula to  $\tilde{y}$  increases average growth in SSA by about 0.28 percentage points). This difference will tend to be larger when the variance in growth is higher, because as the figure shows, the *gyso* variable dampens positive outliers and exaggerates negative outliers by comparison with *dyn*.

- (c) Overall, the use of a different formula for annual growth does not explain much of the difference between half-decadal values of the two series. If we use the discrete time formula for both series, the correlation between half-decadal values (again excluding the 1990s) rises only very slightly, from 0.68 to 0.69. At the individual country level, the difference should matter only in ‘outlier’ half-decades, i.e. half-decades containing annual observations with extreme growth or decline.

Difference 2: The time aggregation is different because *dyn* is an average of annual growth rates for 1960-64, 1965-69, etc, while *gyso* is an average of annual growth for 1961-65, 1966-70, etc. This is a fairly important difference, given the variance of growth rates in SSA.

- (a) For the *dyn* variable we used the more common convention for aggregating by half-decades or decades. Neither approach is obviously superior, however.
- (b) This particular difference between the two series is moderately important empirically. The table below shows correlations between different versions of the variables, using the discrete-time formula to calculate all underlying growth rates. The correlations on the diagonal are below 1 solely because of the time shift. Those off the diagonal reflect both the time shift and the difference in definition of real GDP ( $y$  versus  $\tilde{y}$ ). The definitional difference is clearly the most important empirically, but the timing does matter.

Correlations using alternative timing

		<i>Half-decades calculated using (e.g.) 1971-75 growth rates</i>	
		<i>dyn</i>	<i>gyso</i> *
<i>Half-decades calculated using (e.g.) 1970-74 growth rates</i>	<i>dyn</i>	0.86	0.69
	<i>gyso</i> *	0.61	0.83

\*For this table, *gyso* is calculated using the discrete-time formula rather than the difference in logs.

Difference 3: The definition of real GDP per capita is different because *dyn* is calculated at constant local prices and *gyso* is calculated at constant 1985 international prices. This is the most important difference between the two variables. If we eliminate other differences, the correlation between the two series rises from its current value of 0.68 only to 0.75 (again excluding the 1990s).

- (a) While the PPP-adjusted series is superior for international comparisons of *levels* of real GDP, neither variable is clearly preferable for comparisons of growth rates. The reason we used the *dyn* variable in the pooled full specification is that it was available in the 1990s, unlike the PPP-adjusted data from the Penn World Tables which was available only until 1992. (Anke Hoeffler’s model was run using half-decadal observations only through 1990. In the data spreadsheets we distributed, any observations for *gyso* after 1992 are based on a World Bank proxy for real GDP at international prices. This

proxy was constructed by using historical international-price GDP shares to aggregate constant-local currency growth rates for the components of GDP. Differences between *gyso* and *dyn* are therefore, by construction, smaller in the 1990s than in other half-decades.)

- (b) If we eliminate other differences by using the same years for the half-decadal averages and the same method for calculating growth (i.e. either percentage changes or log differences), the correlation between half-decadal growth rates of  $y$  and  $\tilde{y}$  rises from 0.68 to about 0.73 for SSA (excluding the 1990s).
- (c) What then explains the substantial remaining differences between the two series? Both series use constant base-year prices to aggregate real growth rates of underlying GDP components. Differences can arise from differing relative prices, differing base years, or differing real series for the underlying GDP components. Unfortunately, it is difficult to say anything systematic about the likely net effect of these influences. Below we identify two systematic influences ((d) and (e)) arising from differences between international and domestic prices. But there is little we can say about differences arising from differences in the base year or in the disaggregated real growth series.
- (d) Suppose therefore that the same base year and disaggregated real growth rates were used. Then the domestic price data would be expected to give lower relative weight to nontraded components than traded components by comparison with the PPP-adjusted data, consistent with the well-known tendency for nontraded goods and services to be relatively cheaper at domestic prices than at international prices in low-income countries. This means *dyn* would tend to grow faster (slower) than *gyso*, other things equal, during periods of relatively rapid growth in the traded (nontraded) sectors – e.g. periods of substantial real depreciation (appreciation).
- (e) Under the same assumption about base year and underlying real growth rates, an additional difference arises within the traded goods sector. The domestic currency data would tend to give higher relative weight to tariff- or quota-protected traded goods sectors than the international price data. This means that under a very aggressive import substitution policy, for example, rapid growth of import-protected sectors could produce a situation in which the traded component of local currency real GDP is growing faster than the traded component of international-price real GDP. Brecher and Diaz-Alejandro demonstrated years ago that in this situation the traded component of international-price real GDP could even be falling (they actually referred to total real GDP, because all goods were traded in their model). This outcome can emerge if growth is driven by capital accumulation and the import-protected sector is relatively capital intensive. The traded component of international-price real GDP can fall in this case because the capital accumulation goes mainly into the import-competing sector, which is already ‘too large’ at international prices. The traded component of real GDP rises at distorted domestic prices, but may actually fall at international prices because of the worsening distortion.

**Figure 1:** Alternative growth rate formulas

*dyn, gys0*

