

A note on statistical discrepancies in the national income accounts of selected Asian economies

Introduction

In 2006, the Philippine economy grew by 5.4%. All of its components contributed positively to this growth, except for “statistical discrepancy,” which deducted 3.9 percentage points from the 9.4 percentage points total contribution of consumption, investment, and net exports. In contrast in 2002, the country posted growth of 4.6%—but this time, the statistical discrepancy made a positive contribution, of 78% of the total. What is this statistical discrepancy? Why does it occur? And how important is it when measuring growth in developing Asia?

Gross domestic product (GDP) can be estimated in three ways: via the expenditure approach (summation of private consumption, government expenditure, investment, and net exports); production approach (total value of final goods and services produced by the economy); and income approach (total income of the factors involved in the production process). In principle, these three approaches should yield identical GDP values because total expenditure on goods and services must by definition equal the value of goods and services produced, which in turn must equal total income paid to the factors that produced those goods and services. Yet as each measurement exercise is independent, there may be differences in the resulting estimates. The difference between these estimates is called the statistical discrepancy (SD).

There has been growing concern about the existence and magnitude of the SD in national accounts statistics. A large and persistent SD is seen as a reflection of the authorities’ inability to generate accurate and reliable national accounts estimates. Considering that these numbers represent a country’s economic activity, data dependability is of utmost importance.

This technical note examines the SDs in the national income accounts of six developing Asian countries: Bangladesh, India, Republic of Korea (Korea), Philippines, Thailand, and Viet Nam. These countries were selected on the basis of availability of sufficient time series national accounts data to allow rigorous analysis. Focusing on the expenditure and production approaches to estimating GDP, this note will assess the data

collection methodologies of the selected countries to identify probable sources of the SD and its allocation across various GDP components.

Framework and methodology

The estimation procedure in this note is based on a framework proposed by Weale (1992), who showed how consistent estimates of the national accounts can be obtained by exploiting statistical information in the data. He applied the approach to allocate the SD between the measurement errors of aggregate expenditure and income estimates of the United States' real gross national product (GNP). These errors averaged 0.6% of GNP over the period 1980–2006. This note applies Weale's approach by examining, at a disaggregated level, the sources of SD using national accounts data of six Asian countries. (A similar exercise for Argentina was undertaken by Bajada [2001]). Weale's method is described in Box 4.1.1. Essentially, it uses information from the variances and covariances of observed national accounts data to draw inferences about their reliability. It then attempts to minimize the "distance" between the revised estimates, which are constrained to be consistent with the national accounts adding-up conditions, and the initial estimates, but allowing larger adjustments for elements that are believed less reliable. Underlying the approach are two assumptions: that initial estimates are subject to measurement errors; and that these measurement errors are independent of the true, but unknown, components. It follows from these assumptions that if there are two estimates of the same variable, the one that exhibits less variance can be considered the more reliable.

To illustrate the procedure described in Box 4.1.1, let \hat{Y}_1 and \hat{Y}_2 be two measures of the true value of GDP, \dot{Y} . Further, assume that \hat{Y}_{1a} and \hat{Y}_{1b} are components of \hat{Y}_1 while \hat{Y}_{2a} and \hat{Y}_{2b} are components of \hat{Y}_2 . Let the covariance matrix of $\hat{Y}_{1a}, \hat{Y}_{1b}, \hat{Y}_{2a}$, and \hat{Y}_{2b} be the positive definite matrix:

$$W = \begin{bmatrix} 2,069 & 57 & 1,549 & 127 \\ 57 & 598 & -86 & 18 \\ 1,549 & -86 & 1,428 & 109 \\ 127 & 18 & 109 & 97 \end{bmatrix}$$

As seen in its diagonal elements, estimates of the four variables have relatively large variances, suggesting that they are unreliable measures. The covariances (i.e., off-diagonal elements) also indicate substantial interrelations among the estimates.

The linear constraint $A = (1 \ 1 \ -1 \ -1)$, thus:

$$WA' = \begin{bmatrix} 450 \\ 723 \\ -74 \\ -61 \end{bmatrix}$$

and $AWA' = 1,308$.

Substituting the above in equation (12) in Box 4.1.1,

4.1.1 Statistical methodology

In national accounts data, estimates of the true value of gross domestic product (GDP) from the expenditure approach should equal those from the value-added approach, i.e.,

$$GDP_E = GDP_{VA} \quad (1)$$

$$C + I + G + X - M = Ag + In + Se + Ta \quad (2)$$

where

$GDP_E = C + I + G + X - M$ = expenditure measure of GDP

$GDP_{VA} = Ag + In + Se + Ta$ = value-added measure of GDP

C = consumption spending

I = total investment

G = government spending

X = exports of goods and services

M = imports of goods and services

Ag = agriculture

In = industry

Se = services

Ta = indirect taxes less subsidies.

In reported data, however, GDP_E may not equal GDP_{VA} , creating a discrepancy. Differences in estimates arise from varying data collection methods and estimation procedures of GDP components. In general, the statistical discrepancy (SD) is reported on the expenditure side, such that equations (1) and (2) become:

$$GDP_E + SD = GDP_{VA} \quad (3)$$

$$\hat{C} + \hat{I} + \hat{G} + \hat{X} - \hat{M} + SD = \hat{A}g + \hat{I}n + \hat{S}e + \hat{T}a \quad (4)$$

Since SD can be defined as the sum of the measurement errors of each component of GDP, equation 4 can be written as:

$$(\hat{C} + \varepsilon_c) + (\hat{I} + \varepsilon_I) + (\hat{G} + \varepsilon_G) + (\hat{X} + \varepsilon_X) - (\hat{M} + \varepsilon_M) = (\hat{A}g + \varepsilon_{Ag}) + (\hat{I}n + \varepsilon_{In}) + (\hat{S}e + \varepsilon_{Se}) + (\hat{T}a + \varepsilon_{Ta}) \quad (5)$$

where $\varepsilon_c, \varepsilon_I, \varepsilon_G, \varepsilon_X, \varepsilon_M, \varepsilon_{Ag}, \varepsilon_{In}, \varepsilon_{Se}$, and ε_{Ta} are the components of SD.

To show how SD can be allocated to each component, consider \dot{Y} to be a vector of true data, \hat{Y} a vector of

observed values, and ε a vector of measurement errors, i.e.,

$$\dot{Y} = \hat{Y} + \varepsilon \quad (6)$$

With the assumptions that ε and \dot{Y} are uncorrelated and that ε is identically normally distributed with mean 0 and variance V , then the true value can be estimated by maximizing the following log-likelihood function:

$$L(\dot{Y}, V | \hat{Y}) = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(V) - \left[\frac{1}{2} (\dot{Y} - \hat{Y})' V^{-1} (\dot{Y} - \hat{Y}) \right] \quad (7)$$

where N is the number of observations, subject to the condition that:

$$A\dot{Y}_t = 0 \quad (8)$$

where A is a linear constraint. This leads to an estimate of \dot{Y} , i.e.,

$$\tilde{Y} = [I - VA'(AVA')^{-1}A]\hat{Y} \quad (9)$$

where I is an identity matrix and V is the unknown variance-covariance matrix.

Once estimates of the true values of the components are obtained, equation (2) will be satisfied, such that,

$$\tilde{C} + \tilde{I} + \tilde{G} + \tilde{X} - \tilde{M} - \tilde{A}g - \tilde{I}n - \tilde{S}e - \tilde{T}a = 0 \quad (10)$$

Weale (1992) shows that the maximum likelihood estimate of WA' converges in probability to VA' , i.e., $\text{plim } WA' = VA'$

$$(11)$$

where W is a maximum likelihood data covariance matrix. Thus, equation (9) can be written as:

$$\tilde{Y} = [I - WA'(AWA')^{-1}A]\hat{Y} \quad (12)$$

From equation (12), the share of each component to SD is estimated from:

$$WA'(AWA')^{-1}A \quad (13)$$

$$\begin{bmatrix} \tilde{Y}_{1a} \\ \tilde{Y}_{1b} \\ \tilde{Y}_{2a} \\ \tilde{Y}_{2b} \end{bmatrix} = \begin{bmatrix} \hat{Y}_{1a} \\ \hat{Y}_{1b} \\ \hat{Y}_{2a} \\ \hat{Y}_{2b} \end{bmatrix} = \begin{bmatrix} 450 \\ 723 \\ -74 \\ -61 \end{bmatrix} [1,308]^{-1} \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \hat{Y}_{1a} \\ \hat{Y}_{1b} \\ \hat{Y}_{2a} \\ \hat{Y}_{2b} \end{bmatrix}$$

hence,

$$\begin{bmatrix} \tilde{Y}_{1a} \\ \tilde{Y}_{1b} \\ \tilde{Y}_{2a} \\ \tilde{Y}_{2b} \end{bmatrix} = \begin{bmatrix} \hat{Y}_{1a} \\ \hat{Y}_{1b} \\ \hat{Y}_{2a} \\ \hat{Y}_{2b} \end{bmatrix} - \begin{bmatrix} 0.34 & 0.34 & -0.34 & -0.34 \\ 0.55 & 0.55 & -0.55 & -0.55 \\ -0.06 & -0.06 & 0.06 & 0.06 \\ -0.05 & -0.05 & 0.05 & 0.05 \end{bmatrix} \begin{bmatrix} \hat{Y}_{1a} \\ \hat{Y}_{1b} \\ \hat{Y}_{2a} \\ \hat{Y}_{2b} \end{bmatrix}$$

The above can be further simplified:

$$\begin{bmatrix} \tilde{Y}_{1a} \\ \tilde{Y}_{1b} \\ \tilde{Y}_{2a} \\ \tilde{Y}_{2b} \end{bmatrix} = \begin{bmatrix} \hat{Y}_{1a} \\ \hat{Y}_{1b} \\ \hat{Y}_{2a} \\ \hat{Y}_{2b} \end{bmatrix} + \begin{bmatrix} 0.34 & (\hat{Y}_{2a} + \hat{Y}_{2b} - \hat{Y}_{1a} - \hat{Y}_{1b}) \\ 0.55 & (\hat{Y}_{2a} + \hat{Y}_{2b} - \hat{Y}_{1a} - \hat{Y}_{1b}) \\ 0.06 & (\hat{Y}_{1a} + \hat{Y}_{1b} - \hat{Y}_{2a} - \hat{Y}_{2b}) \\ 0.05 & (\hat{Y}_{1a} + \hat{Y}_{1b} - \hat{Y}_{2a} - \hat{Y}_{2b}) \end{bmatrix}$$

which leads to the final weights, shown in equation (13), that determine the allocation of SD to each component. Summing up \tilde{Y}_{1a} and \tilde{Y}_{1b} into \tilde{Y}_1 and \tilde{Y}_{2a} and \tilde{Y}_{2b} into \tilde{Y}_2 should lead to balanced values of \tilde{Y}_1 and \tilde{Y}_2 , which are estimates of the true value, \hat{Y} .

$$\begin{bmatrix} \tilde{Y}_{1a} \\ \tilde{Y}_{1b} \\ \tilde{Y}_{2a} \\ \tilde{Y}_{2b} \end{bmatrix} = \begin{bmatrix} \hat{Y}_{1a} \\ \hat{Y}_{1b} \\ \hat{Y}_{2a} \\ \hat{Y}_{2b} \end{bmatrix} + \begin{bmatrix} 0.34 \\ 0.55 \\ -0.06 \\ -0.05 \end{bmatrix} [SD]$$

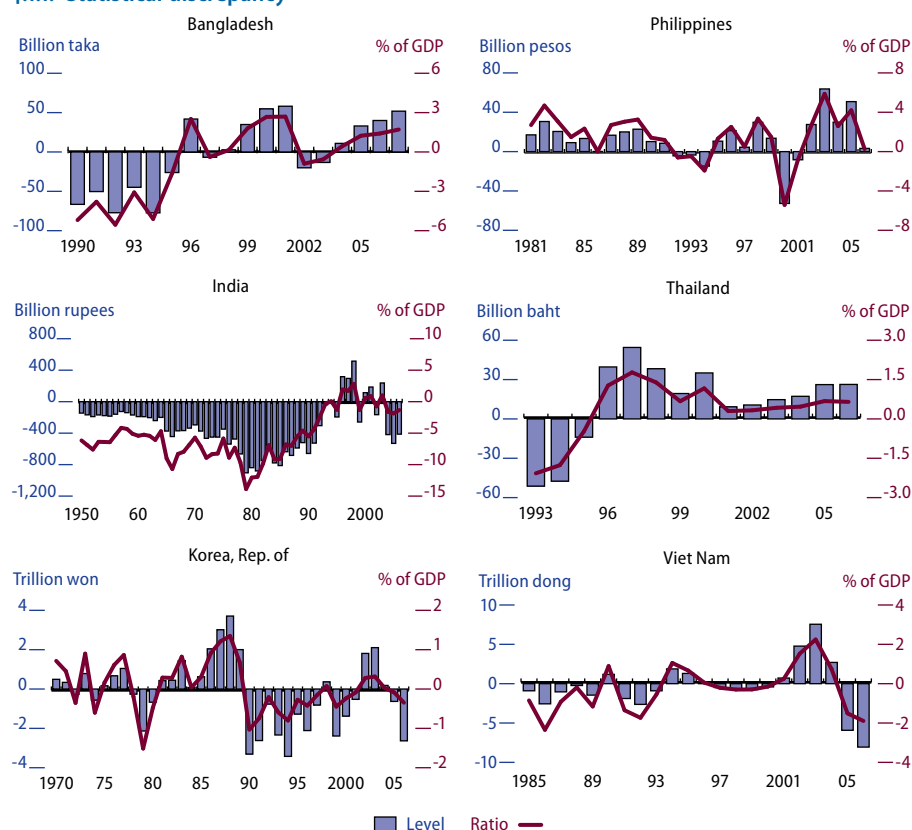
In the illustrative example, the application of the maximum likelihood estimation suggests that 89% of the SD is due to measurement errors in \hat{Y}_1 and 11% is due to \hat{Y}_2 .

Some properties of the statistical discrepancy

Figure 4.1.1 shows the SD level in the six countries. The data confirm that the SD has been volatile and has exhibited no signs of decline. As a ratio to GDP, the SD's volatility has become somewhat less pronounced, but it remains highly visible. In the last 4 or 5 years, the SDs of Bangladesh and the Philippines have generally been positive, meaning that income estimates of GDP exceed expenditure estimates. The opposite is true for India, Korea, and Viet Nam, where SDs have turned negative in recent years. Thailand's SD has been positive since 1996.

Table 4.1.1 presents summary statistics for the observed SDs. Bangladesh, India, and Viet Nam registered negative means ranging from -5.4% to -0.3% of GDP while Korea, Philippines, and Thailand posted positive means ranging from 0.1% to 1.6% of GDP. Since the mean as a measure of central tendency may give misleading results due

4.1.1. Statistical discrepancy



Source: Asian Development Outlook database.

to its sensitivity to the sign of the SD, mean absolute values of the SD are calculated as a proportion of GDP. Korea, Thailand, and Viet Nam have mean absolute rates of at most 1% of GDP. By contrast, India has overestimated its expenditure components by an average of 5.7% of GDP. Bangladesh and the Philippines have statistical discrepancies that average over 2% of GDP.

In Table 4.1.2, the impact of the SD on annual growth estimates is shown. In India for instance, the contribution of SD to GDP growth peaked at a remarkable 4,700% (in absolute terms) of GDP. This happened in 1966 when the Indian economy was estimated to have contracted by 0.03%, but in a context where the statistical discrepancy deducted 1.68 percentage points from growth. Indeed, there have been several instances in which the contribution to growth of SD has been larger than the

4.1.1 Descriptive statistics of the statistical discrepancy (% of GDP)

	Period covered	Mean	Mean absolute	Minimum	Maximum	Standard deviation	% of sample with positive SD
Bangladesh	1990–2007	-0.6	2.3	-5.6	2.7	2.8	50.0
India	1950–2006	-5.4	5.7	-13.9	2.8	3.8	12.3
Korea, Rep. of	1970–2006	0.1	0.5	-1.5	1.4	0.6	51.4
Philippines	1981–2006	1.6	2.3	-5.4	5.9	0.0	80.8
Thailand	1993–2006	0.3	1.0	-2.1	1.8	1.1	79.0
Viet Nam	1985–2006	-0.3	1.0	-2.4	2.3	1.2	36.4

SD = statistical discrepancy.

Source: Authors' calculations.

4.1.2 Statistical discrepancy and GDP growth

	Bangladesh	India	Korea, Rep. of	Philippines	Thailand	Viet Nam
Real GDP growth (%)						
Mean	5.2	4.8	6.9	2.9	4.0	6.8
Minimum	3.3	-5.2	-6.9	-7.3	-10.5	2.8
Maximum	6.6	9.6	12.0	6.8	9.2	9.5
Median	5.2	5.2	7.1	4.3	5.0	7.1
Statistical discrepancy contribution to GDP growth (percentage points)						
Mean absolute value	1.7	1.3	0.6	2.3	0.5	1.0
Minimum	-3.6	-4.3	-1.8	-7.2	-0.9	-2.4
Maximum	4.3	3.3	1.4	4.6	1.8	2.1
Median	0.5	0.0	0.2	0.1	0.1	-0.1
Statistical discrepancy contribution to GDP growth (% share)						
Mean absolute value	35.0	123.5	9.0	94.5	12.4	18.0
Minimum	-82.2	-265.2	-69.9	-529.1	-39.7	-55.0
Maximum	92.5	4,700.0	13.0	260.0	30.8	41.2
Median	10.2	5.0	0.8	5.5	1.7	-1.8

Source: Authors' calculations.

contribution of the relevant expenditure components (among others, as mentioned in the first paragraph of this technical note, the Philippines).

A simple correlation analysis does not suggest that any particular GDP component is strongly related to the SD (Table 4.1.3). Indeed, except for Bangladesh and Thailand, the SDs of all the countries show no strong correlation with any single component of GDP, whether from the expenditure side or from the economic activity side.

Given the inconclusiveness of the correlation results, a more detailed analysis to identify sources of the SD is implemented.

4.1.3 Correlation of statistical discrepancy with GDP components

	Private consumption	Government consumption	Investment	Exports	Imports	Net exports	Agriculture	Industry	Services
Bangladesh	0.67*	0.59*	0.71*	0.70*	0.77*		0.72*	0.71*	0.69*
India	0.21	0.22	0.17	0.23	0.30		0.22	0.23	0.25
Korea, Rep. of	-0.27	-0.28	-0.35	-0.17	-0.25		-0.15	-0.25	-0.26
Philippines	0.15	0.00	0.03	0.10	0.15		0.19	0.20	0.18
Thailand	0.41	0.54*	-0.27	0.40	0.27		0.42	0.42	0.48
Viet Nam	-0.05	-0.01	-0.02			-0.27	0.03	-0.05	-0.01

* Significantly higher than +/- 0.5 at 5% level.

Source: Authors' calculations.

Reconciling the estimates

Before applying the methods of Box 4.1.1, two statistical issues need to be dealt with. The estimate of the covariance matrix that is needed to apportion the SD is potentially vitiated by heteroskedasticity and by non-stationarity in the data. To circumvent these difficulties, the covariance matrix is estimated using the level differences in the estimates, normalized on the lagged level of GDP. This transformation does not affect the accounting identity, but results in homoskedastic and stationary data from which the covariance can then be calculated. Results are presented in Table 4.1.4. The numbers report the shares of the observed

SD that the estimation procedure attributes to the expenditure and income components of the national accounts.

The yellow-shaded cells show the component with the highest contribution to SD. The green-colored cells show the component next in line. Either the trade or investment variables account for the highest proportion of SD. Notable exceptions are India and Viet Nam, where agriculture contributed the highest and second-highest proportion of SD, respectively, and Bangladesh where private consumption accounted for the second-highest proportion of SD.

For all the countries studied in this note, estimates of investment are sourced mainly from annual and periodic surveys of business and industries. In estimating investment, national statistics offices frequently use data extrapolations and manipulations to approximate population from sample data and to transform survey results to national income accounts data.

For example, the value of additions and modifications to buildings and other physical structures are estimated using data from surveys of building and construction activities, building permits, and building and construction materials. In Korea, where the highest proportion of the SD is attributed to capital formation, part of the investment data is estimated from input-output tables, which are compiled only every 5 years. Further, the buildings component of gross fixed capital formation is obtained directly from the output of the construction industry after some adjustments such as deductions for repairs. Meanwhile, Thailand, aside from investment data from the government and business sectors, uses related indicators (such as tax revenues to estimate machinery and office equipment) to produce investment estimates. Also, the commodity flow method¹ is used to derive changes in stocks. Both of these require some interpolations and extrapolations, which contribute to measurement errors.

Although imports and exports are generally regarded as relatively well measured given that timely records of them are usually kept, the findings here indicate otherwise. Private consumption data tend to have fewer measurement errors than trade data, despite reliance on surveys and input-output tables, which are usually deemed less reliable than

4.1.4 Sources of statistical discrepancy

	Bangladesh	India	Korea, Rep. of	Philippines	Thailand	Viet Nam
Expenditure components	0.97	0.25	1.07	1.19	-1.82	0.98
Total consumption	0.38	-0.31	0.59	0.07	-1.52	-0.15
Private consumption	0.35	-0.35	0.47	0.03	-1.36	-0.13
Government consumption	0.03	0.04	0.12	0.04	-0.17	-0.02
Investment	0.00	0.38	1.83	0.35	-2.83	-0.02
Net exports	0.58	0.18	-1.35	0.77	2.53	1.15
Exports	-0.11	0.11	-0.83	0.73	0.13	
Imports	0.69	0.07	-0.52	0.04	2.40	
Production components	0.03	0.75	-0.07	-0.19	2.82	0.02
Agriculture	0.02	0.72	0.13	-0.05	0.10	-0.24
Industry	0.03	0.02	-0.22	-0.13	1.22	0.03
Services	-0.01	0.06	0.11	-0.01	1.51	0.23
Taxes and subsidies	-0.01	-0.04	-0.10			

Source: Authors' calculations.

government records such as customs trade data. Possibly small-sample problems, particularly for Bangladesh and Thailand, detract from the robustness of the results. Several factors have been cited to account for errors in trade data. These include underestimation of transactions, incomplete data in documents submitted by exporters and importers, failure of traders to submit required documents, undervaluation of export shipments to lessen or avoid imposition of taxes and other duties, overvaluation of the costs of imports to reduce income tax liabilities, and inability to capture some import and export flows.

A case in point is the Philippines, when in August 2005 large adjustments to its trade figures were made for the previous 3 years. The adjustments were due to intracompany import transactions as well as export transactions that were unreported (Habito 2005). A World Bank study, which looked into the trade statistics between India and Bangladesh, attributed the discrepancy between India's exports and Bangladesh's imports to " 'tax evasion' schemes where Indian exporters understate the value of their exports and Bangladeshi importers overstate the costs" (World Bank 2006). In Viet Nam, exports and imports of goods and services are derived from the balance-of-payments estimates, with imports adjusted for unrecorded cross-border trade. The International Monetary Fund has noted that the compilation of the national accounts suffers from poor data collection practices and a lack of coordination and communication between data collection agencies. Further, data on invisibles are based largely on banking records, which provide incomplete coverage and identification of the types of transactions.

In India, the agriculture data suffer from several shortcomings. First, survey results are available only after a lag of about 2 years. For example, land-use statistics for the computation of value added in agriculture are supplied with a lag of 2 or 3 years. Prices collected by state governments also are available with a lag of 1 or 2 years. Second, there is no consistent data collection methodology across the various agencies gathering information. Third, statistics for some subsectoral components, such as forestry, are not systematically obtained, and rely simply on irregular, ad hoc publications. All these factors contribute to the reduced precision of estimates for agriculture data.

In Viet Nam, where the agriculture sector contributes the second-highest proportion of SD, output is estimated using annual surveys of agricultural and livestock production. Other than the inherent sampling errors present in the surveys, a possible source of discrepancy lies in the seasonal differences in agricultural production. The difficulty lies in measuring output of crops in a continuous manner throughout the year given that the length and timing of crop production vary (the problem of measuring work in progress).

Errors in the measurement of private consumption can be attributed to the fact that information is based on various surveys and administrative record systems and on a set of indicators. Household income and expenditure surveys, which provide benchmark estimates of private consumption, are carried out every 3 to 5 years. Data derived from the more frequent surveys are then used to interpolate and extrapolate the benchmark estimates. India, Korea, and Thailand are examples of countries that apply the commodity flow method to survey data to

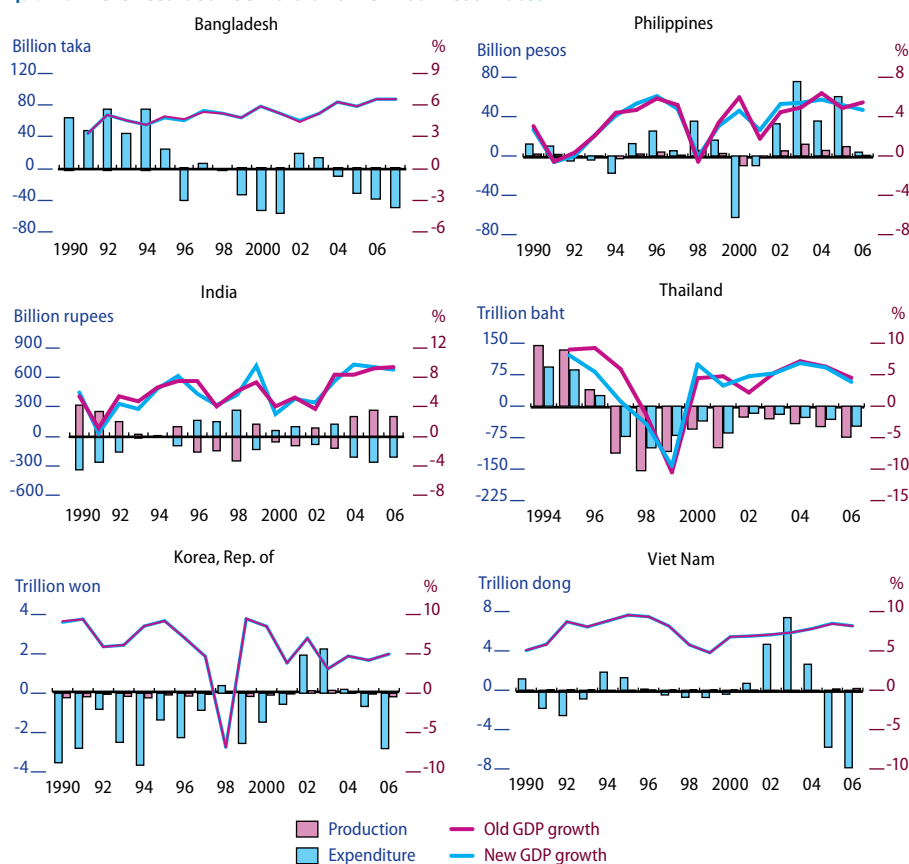
get estimates of personal consumption as well as other expenditure components of GDP.

As to the argument that data sourced primarily from government records tend to be more accurate than nongovernment sources, results do support this view. Government spending accounts for at most 4% of the measurement error.

In general, the SD is attributed more to expenditure components than to production. This supports the common notion that data support is stronger in the production accounts. Usually, estimates of GDP levels are based on the income estimates from the production side of the national accounts.

Figure 4.1.2 plots the differences between the calculated GDP levels and the old observed values. Positive values indicate underestimation of original observations. In the main, GDP estimates via the expenditure approach appear to be underrecorded for Bangladesh and the Philippines, while those for India, Korea, Thailand, and Viet Nam appear to be overestimated. True GDP values for Bangladesh and India seem to lie between observed values from the expenditure and production approaches. For Korea, Thailand, and Viet Nam, new GDP estimates are generally less than both the income and expenditure estimates. The reverse is true for the Philippines. However, while large differences in levels are apparent, revised growth rates are in line with old figures.

4.1.2 Differences between old and new GDP estimates



Sources: Asian Development Outlook database; authors' calculations.

[Click here for figure data](#)

Conclusion

This note has characterized SDs in six developing Asian countries. It has provided a short description of why SDs exist in national income accounts and has analyzed which of the various components of GDP have contributed most to SDs. This has helped provide a clearer sense of how each economy performed and how each sector contributed to growth.

The presence of SDs highlights the need to strengthen the statistical capability of government data collection agencies. The results suggest possible areas where the quality of data is poor and where government efforts to improve data quality need to be directed.

Endnote

1. A method of compiling national accounts in which the total supplies and uses of individual types of goods and services have to be balanced with each other (United Nations 2001).

References

- Bajada, C. 2001. "An Examination of the Statistical Discrepancy and Private Investment Expenditure." *Journal of Applied Economics* 4(1):27-61.
- Habito, C. 2005. "Why the trade deficit was understated." *Philippine Daily Inquirer*. 8 August. Available: http://www.census.gov.ph/data/technotes/Habito_article.html.
- International Monetary Fund. Special Data Dissemination Standards, various countries.
- Organisation for Economic Co-operation and Development. 2001. "Quarterly National Accounts in Asia: Sources and Methods." Proceedings of Joint OECD-ADB-ESCAP Workshop on Quarterly National Accounts. Bangkok, October.
- United Nations. 2001. *The 1993 System of National Accounts*. Available: <http://unstats.un.org/unsd/sna1993/introduction.asp>.
- Weale, M. 1992. "Estimation of Data Measured with error and Subject to Linear Restrictions." *Journal of Applied Econometrics* 7(7):167-74.
- World Bank. 2006. "India-Bangladesh Bilateral Trade and Potential Free Trade Agreement." *Bangladesh Development Series*. Paper No. 13. World Bank Office, Dhaka. Available: <http://siteresources.worldbank.org/INTBANGLADESH/Resources/Trade.pdf>.

Appendix: Sources of statistical discrepancy

Statistical discrepancy (SD) can result from the selection of data sources, the timing of data, or the methods used in estimating the components.

Data sources

The table lists the basic sources of data for the different components of gross domestic product (GDP). The sources can be broadly categorized into three: administrative and financial records/reports from the public and private sectors, sample surveys, and censuses. Data from these sources can either be used directly in the estimation of the GDP components or indirectly as indicators to be fed into a mathematical or statistical model. The expenditure components of GDP rely primarily on administrative/financial data while the production components are mostly drawn from sector- or establishment-based surveys. Recording errors and deficiencies commonly plague administrative-based data while sampling and nonsampling errors are inherent in survey data.

Timing of data

Discrepancies can also be attributed to the difference in the timing of data release from the different sources. Administrative-based data are more frequently released and with shorter lags than survey data. The timing issue not only affects the use of survey results (on which estimates are based) but also the recording of actual transactions. Production, especially in the agriculture sector, may span several accounting periods, while sales are recorded continuously throughout the year and can be confined to the particular year being measured.

Methodological issues

Surveys and censuses are primarily intended to obtain trends and reveal structures, and are not originally designed for national accounts purposes. Since these data may not conform exactly to national accounts definitions nor cover all the items to be measured in the accounts, measurement errors are expected. Such would be the usual case for the measurement of gross capital formation, for which information is culled from administrative and financial reports, as well as from industry surveys.

In many cases, extrapolations from sample survey data are used to approximate population data that are based on a benchmark census and value-added data that are based on a benchmark year. These extrapolations are then adopted directly in the estimation procedure or as parameters that are fed into a statistical or mathematical model. For example, data on retail trade sales and turnover are often used to estimate personal consumption expenditures. In some cases, technical coefficients from the input-output tables, which give ratios of value added to output or of intermediate consumption to value added, are also applied, especially when information is very limited. Necessarily, these methods of using survey data will have measurement errors. The input-output tables, which are revised every 5 or 10 years or so, may likewise fail to capture changes in productivity and technology.

Conversion of current price estimates to constant price estimates also contributes to differences in GDP between the production and expenditure sides. Deflators are not used consistently. For example, the wholesale price index is used to deflate manufacturing output, the consumer price index—personal consumption expenditures, and the producers' price index—agricultural output.

1 Nature of basic data					
	Bangladesh		India		Korea, Rep. of
	Frequency	Source of basic data	Frequency	Source of basic data	
Private consumption	Every 4–5 years	Household income and expenditure survey	Every 5 years	Input-output table and household surveys	Input-output table
Government consumption	M	Public administrative and financial reports	M	Public administrative and financial reports	Public administrative and financial reports
	M	Public administrative and financial reports	M	Public administrative and financial reports	Consumption surveys
		Industry sector surveys	M	Private corporate administrative and financial reports	Public administrative and financial reports
		Foreign trade statistics	Every 5 years	Household surveys	Industry sector surveys
Trade	M	Foreign trade statistics (balance-of-payments-based)	Every 5 years	Input-output tables	Input-output tables
Agriculture		Agriculture sector surveys	A	Foreign trade statistics (customs-based)	Foreign trade statistics (balance-of-payments-based)
		Census of Agriculture	A	Agriculture sector surveys	Agriculture sector surveys
		Farm forest and HH expenditure survey	A	Fisheries sector surveys	Fisheries sector surveys
		Public administrative and financial reports	M	Cost of cultivation studies	Price surveys
Industry	M	Price surveys	M	Public administrative and financial reports	
	M	Foreign trade statistics	M	Price surveys	
	M	Input-output tables	A	Industry sector surveys	Industry sector surveys
		Public administrative and financial reports	M	Public administrative and financial reports	Public administrative and financial reports
Services		Private corporate administrative and financial reports	M	Private corporate administrative and financial reports	Private corporate administrative and financial reports
	A	Price surveys	M	Price surveys	
	A	Industry sector census			
	M	Input-output table	M	Employee compensation surveys	Services sector surveys
		Services sector surveys	M	Public administrative and financial reports	Public administrative and financial reports
		Public administrative and financial reports		Population census	Private corporate administrative and financial reports
		Population census			
		Input-output table			
		Labor force surveys			

1 Nature of basic data (continued)

	Philippines			Thailand		Viet Nam	
	Frequency	Source of basic data	Frequency	Source of basic data	Frequency	Source of basic data	
Private consumption	Q, A	Industry sector surveys	Every 5 years	Input-output table	M	Retail trade turnover	
	M	Public administrative and financial reports	Every 2 years	Household socioeconomic surveys	M	Balance of agricultural production and use	
	Every 3 years	Family Income and Expenditure Survey			Every 2–4 years	Household living standards survey	
	M	Public administrative and financial reports	M	Public administrative and financial reports	M	Public administrative and financial reports	
Government consumption	Q, A	Industry sector surveys	A	Public administrative and financial reports	M	Public administrative and financial reports	
		Public administrative and financial reports		Private corporate administrative and financial reports		Survey of enterprises with foreign direct investment	
			M	Foreign trade statistics		Survey of private construction companies	
			Q	Socio-economic survey		Balance of agricultural production and use	
Trade	M	Foreign trade statistics (balance-of-payments-based)	M	Foreign trade statistics (balance-of-payments-based)	M	Foreign trade statistics (balance of payments-based)	
Agriculture	M	Agriculture sector surveys		Agriculture sector surveys	2–3 times a year	Agriculture sector surveys	
	M	Fisheries sector surveys	Every 10 years	Census on the agricultural sector	A	Fisheries sector surveys	
	M	Price surveys	Q, A	Industry sector surveys	Q, A	Forestry sector surveys	
	Q,A	Industry sector surveys	Every 10 years	Industry sector census	M, Q	Industry sector surveys	
Industry	M	Public administrative and financial reports			Every 5 years	Industry sector census	
	M	Private corporate administrative and financial reports	M	Price surveys	M	Price surveys	
		Price surveys	M	Public administrative and financial reports	M	Public administrative and financial reports	
	Every 5 years	Industry sector census	M	Private corporate administrative and financial reports	M	Private corporate administrative and financial reports	
Services	M, Q, A	Input-output tables	Every 5 years	Input-output tables	M	Foreign trade statistics	
	M	Services sector surveys	Every 2 years	Services sector surveys	M, Q	Services sector surveys	
		Public administrative and financial reports	M	Public administrative and financial reports	M	Public administrative and financial reports	
	M	Private corporate administrative and financial reports	M	Private corporate administrative and financial reports	M	Private corporate administrative and financial reports	
	Every 5 years	Services sector census	A	Labor force surveys			
		Population census		Services sector census			

A = annually; M = monthly; Q = quarterly.

Sources: International Monetary Fund, Dissemination Standards Bulletin Board, available: <http://dsbb.imf.org>; National Statistical Coordination Board, available: <http://www.nscb.gov.ph/technotes/sna/default.asp>; Organisation for Economic Co-operation and Development, 2001, "Quarterly National Accounts in Asia: Sources and Methods," Proceedings of Joint OECD-ADB-ESCAP Workshop on Quarterly National Accounts, Bangkok, October; Ministry of Statistics and Programme Implementation, National Accounts Statistics—Sources and Uses 1989, available: <http://mospi.nic.in>.

