VII Mechanical Projections

A. Introduction

7.1. This chapter presents some relatively simple techniques that can be used to fill information gaps with synthetic data using mechanical projections based on past trends. Note that this is a fundamentally different situation from the situation described in the previous chapter in that indicators are not available, although there are some similarities in the mathematics. Reliance on mechanical trend projection techniques is only justifiable if the gaps are few and minor because over-reliance on these techniques can easily impart a fictional character to the accounts and does not add any information about current trends. Furthermore, historic trends that are no longer relevant could muffle current trends that would be visible from other components calculated from actual direct or indirect indicators. Thus, as far as possible, quarterly national account (QNA) estimates should be based on direct observations of the relevant detailed accounting item, and QNA compilers should constantly be on the lookout for possibilities to improve the coverage of the economy with relevant source data.

7.2. Although great caution should be used in applying any of these techniques, there may be situations in which they are a last resort solution to covering gaps in the coverage of the economy. Even in the situation of well-established QNA that are underpinned by an extensive set of short-term data, there may be some economic activities for which no timely direct or indirect indicators are available. When that is the case, we can distinguish two situations: (a) no directly relevant short-term source data are available at all, and (b) an indicator becomes available with a time lag that bars its use in the compilation of the QNA. Obviously, the latter situation is more prominent for the first estimates of a quarter than for second or third estimates.

7.3. Compilation of national accounts requires that the whole economy be covered and thus all data gaps be filled, explicitly or implicitly. If QNA data are compiled from both the production and the expenditure side (which is a key recommendation of this manual), the confrontation of supply and demand can help in filling some gaps, and in fact this is recommended for estimating changes in inventories if no direct observations are available. Using the balancing process as an estimation process, however, diminishes the power of the plausibility checks that are such a strong advantage of the commodity flow method. Thus, it is recommended that estimates be generated for all elements of the commodity flow equation, even if some of the estimates are less than satisfactory. Obviously, the less satisfactory estimates are the first choice for making adjustments if the balancing process requires these, but having an estimate in place will support a well-considered decision.

7.4. To ensure control over the estimates, it is preferable to fill the gaps explicitly. Omitting an item from the estimation process means that implicitly the item is assumed either to be zero or to move in line with other parts of the aggregate of which the item is a part. For instance, compiling an output estimate based on the movements in the data for two months without making an explicit estimate of what the third month may look like is the same as forecasting the third month to be equal to an average of the two first months in the quarter. This may not be the most satisfactory way of forecasting (or nowcasting) the missing month. Thus, there is a need to produce an estimate to fill in the gap to ensure a comprehensive total, even if such an estimate is less than satisfactory.

7.5. Deriving estimates using projections based on past trends is particularly undesirable for current price data because, implicitly, current price data also depend on underlying price trends, which tend to be more volatile than volume trends. Thus, if possible, extrapolation based on past trends should be based on volume data combined with available price data. Relevant price data are often available. The timeliness of price statistics generally does not cause any problems, and if price data for the item are not collected, price indices for similar or related products may provide acceptable proxies.

7.6. There are two main QNA uses of projections based on past trends: one based on past trends in annual data and one based on past trends in monthly and quarterly data. Projections based on past trends in annual data are used to fill gaps in cases where no relevant quarterly information is available. Extrapolation based on past trends in monthly or quarterly data is used to mechanically extend indicator series that become available with a time lag that bars direct use.

B. Trend Projections Based on Annual Data

7.7. This section deals with the situation in which no short-term data are available at all and presents techniques that can be used to construct quarterly data based on past trends in annual data. The two main elements of constructing quarterly data based on past trends in annual data are (a) to extend the series of annual data to include forecasts or nowcasts for the current periods and (b) to fit a quarterly series through the annual totals. Extending the series with nowcasts can be achieved by using available forecasts (e.g., crop forecasts, forecasts based on econometrics models) or by simply assuming a continuation of the current trend in the data (e.g., expressed as a simple average of the growth in the series for the past years).

7.8. Fitting a quarterly series through annual totals should ideally be based on some actual information about the seasonal pattern of the series and the timing of any turning points in the series. In cases where data gaps have to be filled by trend projections based on annual data, however, information on the actual timing of possible turning points is normally not available. Although generally unknown, the seasonal pattern of the series may in some cases be broadly known from other information.

7.9. In cases where no information is available about a series' seasonal pattern, the only available option is to use the trend in the annual data to construct a quarterly series without any seasonal pattern that equals the annual totals. Such a series should be as smooth as possible to ensure that its impact on the period-to-period change in the aggregates is minimized.

7.10. A large number of disaggregation methods, with different degrees of sophistication, have been proposed in the academic literature. In general, most of these methods produce similar results. The main goal in these circumstances is to select a method to fill the gaps that is simple and can be implemented easily.

7.11. It is important to emphasize that quarterly distribution without any related series produces purely synthetic numbers that may not be indicative of the real developments. In particular, such numbers do not contain any information about the precise timing of turning points. Because of this, quarterly distributed data may also deviate substantially from estimates of the underlying trend in subannual data produced by standard seasonal adjustment packages.

7.12. In cases where the seasonal pattern of the series is broadly known, the distribution procedure can be improved by superimposing this known seasonal pattern on the derived quarterly series.

7.13. In this chapter, we look at two methods to construct synthetic quarterly data based on past trends in annual data that are reasonably simple and give similar results, as illustrated in Example 7.1. Both are used by several countries. The first is a purely numerical disaggregation technique proposed by Lisman and Sandee, while the second is based on the least-squares techniques discussed in Chapter VI.[1] The latter can, as will be shown, easily be extended to incorporate a known seasonal pattern into the estimates.

1. The Lisman and Sandee Quarterly Distribution Formula

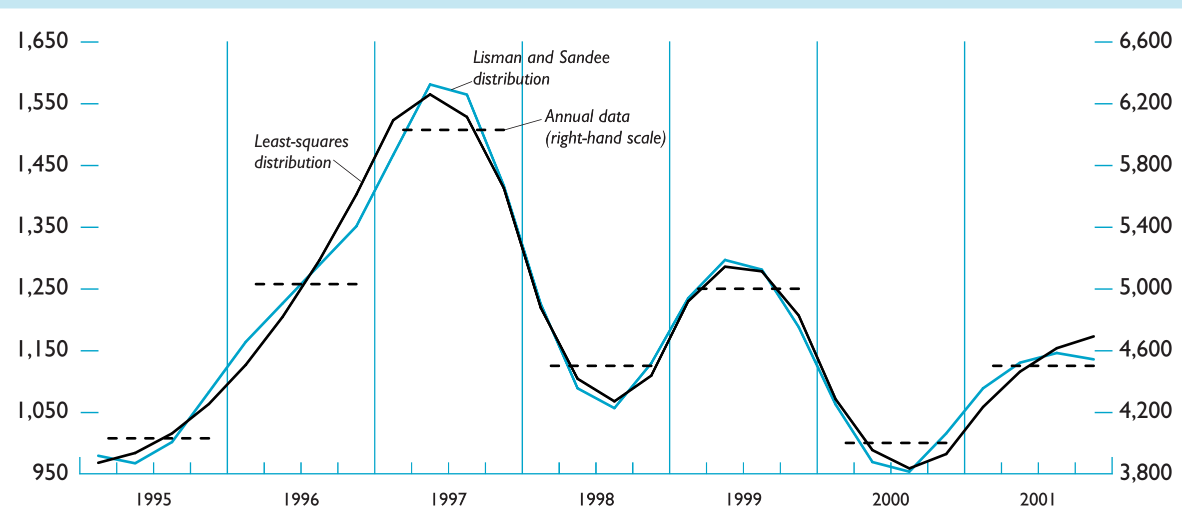
7.14. Lisman and Sandee (1964) proposed a purely numerical technique for constructing synthetic quarterly data based on past trends in annual data. It works as follows:

(i) Make a forecast of the annual data for the current year ($A_{\beta+1}$) and for the next year ($A_{\beta+2}$).

[1]Some of the alternatives to the two methods presented in this chapter include the autoregressive integrated moving average (ARIMA) model-based procedure proposed in Stram and Wei (1986) and Wei and Stram (1990); and the state space modeling procedure proposed in Al-Osh (1989). While generally producing similar results to the two presented in this chapter, these alternative methods are substantially more complicated.

Example 7.1. Quarterly Distribution of Annual Data Without a Related Series

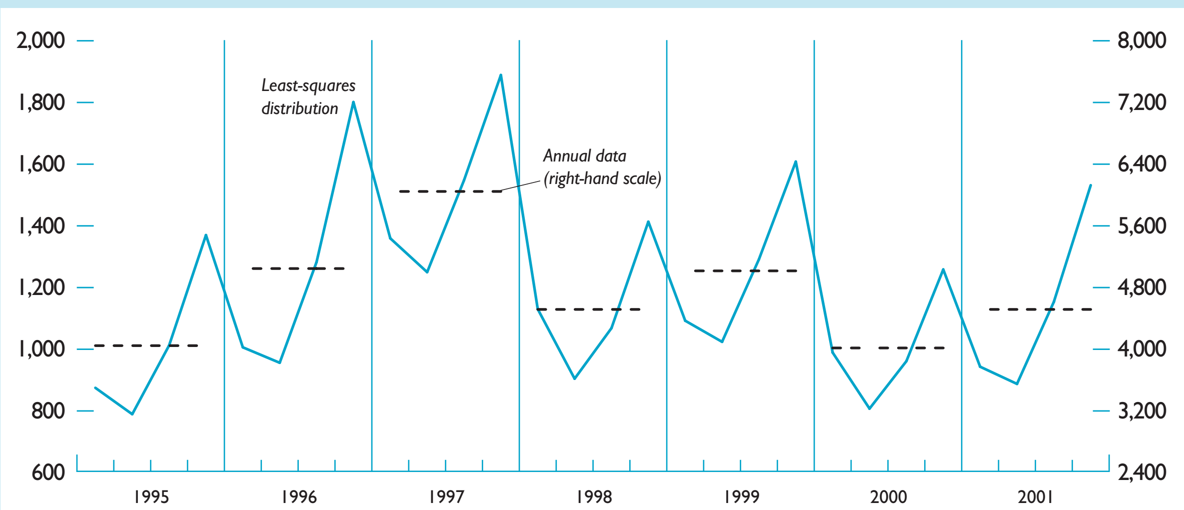
Date	Annual Data	Least-Squares Distribution	Lisman & Sandee Distribution
1994	3,930.0		
q1 1995		967.8	979.2
q2 1995		983.7	967.0
q3 1995		1,015.4	1,001.4
q4 1995	4,030.0	1,063.1	1,082.4
q1 1996		1,126.6	1,163.8
q2 1996		1,204.4	1,226.3
q3 1996		1,296.4	1,288.8
q4 1996	5,030.0	1,402.7	1,351.2
q1 1997		1,523.2	1,466.9
q2 1997		1,565.1	1,581.2
q3 1997		1,528.5	1,564.7
q4 1997	6,030.0	1,413.2	1,417.2
q1 1998		1,219.4	1,225.8
q2 1998		1,104.1	1,088.6
q3 1998		1,067.4	1,056.4
q4 1998	4,500.0	1,109.1	1,129.2
q1 1999		1,229.5	1,234.6
q2 1999		1,285.8	1,296.6
q3 1999		1,278.2	1,281.0
q4 1999	5,000.0	1,206.6	1,187.8
q1 2000		1,071.0	1,062.3
q2 2000		988.3	969.0
q3 2000		958.7	953.4
q4 2000	4,000.0	982.0	1,015.4
q1 2001		1,058.3	1,088.6
q2 2001		1,115.5	1,130.1
q3 2001		1,153.6	1,145.8
q4 2001	4,500.0	1,172.7	1,135.5
2002	4,500.0		



As can be seen, the two alternative procedures for quarterly distribution of annual data without using a related series give very similar results.

Example 7.2. Quarterly Distribution of Annual Data with a Superimposed Seasonal Pattern

Date	Assumed Seasonal Pattern	Annual Data	Least-Squares Distribution
q1 1995		3,930.0	979.2
q1 1995	0.9		870.7
q2 1995	0.8		785.2
q3 1995	1.0		1,008.2
q4 1995	1.3	4,030.0	1,365.9
q1 1996	0.9		1,002.1
q2 1996	0.8		952.0
q3 1996	1.0		1,278.6
q4 1996	1.3	5,030.0	1,797.3
q1 1997	0.9		1,355.5
q2 1997	0.8		1,245.8
q3 1997	1.0		1,543.8
q4 1997	1.3	6,030.0	1,884.9
q1 1998	0.9		1,126.1
q2 1998	0.8		900.3
q3 1998	1.0		1,064.3
q4 1998	1.3	4,500.0	1,409.4
q1 1999	0.9		1,088.4
q2 1999	0.8		1,019.9
q3 1999	1.0		1,287.5
q4 1999	1.3	5,000.0	1,604.2
q1 2000	0.9		985.1
q2 2000	0.8		803.3
q3 2000	1.0		957.2
q4 2000	1.3	4,000.0	254.4
q1 2001	0.9		939.2
q2 2001	0.8		883.5
q3 2001	1.0		1,149.6
q4 2001	1.3		1,527.7



(ii) Derive a smooth continuous quarterly time series from the annual data using the following disaggregation formula:

$$X_{1,y} = 1/4(0.291 \cdot A_{y-1} + 0.793 \cdot A_y - 0.084 \cdot A_{y+1}) \quad (7.1)$$

$$X_{2,y} = 1/4(-0.041 \cdot A_{y-1} + 1.207 \cdot A_y - 0.166 \cdot A_{y+1})$$

$$X_{3,y} = 1/4(-0.166 \cdot A_{y-1} + 1.207 \cdot A_y - 0.041 \cdot A_{y+1})$$

$$X_{4,y} = 1/4(-0.084 \cdot A_{y-1} + 0.793 \cdot A_y + 0.291 \cdot A_{y+1})$$

where

$X_{q,y}$ is the derived quarterly estimate for quarter q in year y,

A_y is the annual estimate for year y, and

β is the last year for which annual data are available.

7.15. The coefficients in the Lisman and Sandee disaggregation formula were derived by imposing a number of restrictions; for example, when the annual data for three consecutive years $y-1$, y, and $y+1$ are not on a straight line, they are assumed to lie on a sine curve.

2. Least-Squares Distribution

7.16. Boot, Feibes, and Lisman (1967) proposed a least-squares-based technique for constructing synthetic quarterly data based on past trends in annual data. It works as follows:

(i) make a forecast of the annual data for the current year ($A_{\beta+1}$).

(ii) Derive a smooth continuous quarterly time series from the annual data using a least-squares minimization technique, as follows:

$$\min_{(X_1,\ldots,X_{4y})} \sum_{t=2}^{4y} [X_t - X_{t-1}]^2, \quad (7.2)$$

$$t \in \{1,\ldots.(4\beta+1)\} \quad y \in \{1,\ldots.(\beta+1)\}$$

under the restriction that

$$\sum_{t=4y-3}^{4y} X_t = A_y$$

(that is, the sum of the quarterized data should be equal to the observed annual data)

where

t is used as a generic symbol for time ($t = q,y$) (e.g., $t = 4y - 3$ is equal to the first quarter of year y, and $4y$ the fourth quarter of year y);

X_t is the derived quarterly estimate for quarter t;

A_y is the annual estimate for year y; and

β is the last year for which any annual observations are available.

7.17. This least-squares-based technique can be extended to incorporate a known seasonal pattern into the estimates by replacing the least-squares expression in step (ii) above with the following expression:[2]

$$\min_{(X_1,\ldots,X_{4y})} \sum_{t=2}^{4y} \left[\frac{X_t}{SF_t} - \frac{X_{t-1}}{SF_{t-1}}\right]^2, \quad (7.3)$$

$$t \in \{1,\ldots.(4\beta+1)\} \quad y \in \{1,\ldots.(\beta+1)\}$$

under the restriction that

$$\sum_{t=4y-3}^{4y} X_t = A_y$$

(that is, the sum of the quarterized data should be equal to the observed annual data)

where

SF_t is a time series with assumed seasonal factors.

Example 7.2 shows the results of using equation (7.3) to superimpose a seasonal pattern on the annual data used in Example 7.1.

7.18. A small problem with the Boot-Feibes-Lisman method, as well as other methods of distribution that use least squares, is a tendency of the derived series to flatten out at endpoints[3] (as can be seen from Example 7.1). This problem can be alleviated by projecting the annual series for two years in both directions and distributing the extended series.

C. Projection Based on Monthly or Quarterly Data

7.19. This section presents some simple techniques that can be used to mechanically extend data series that are not sufficiently timely to be used when the first QNA estimates for a particular quarter are compiled. The monthly and quarterly source data commonly become available with varying delays. Some quarterly and monthly source data may be available

[2]As proposed in for example Cholette (1998a).

[3]This is not a problem for using least squares for benchmarking as discussed in Chapter VI. In that case, the implied flattening out at the endpoints of the quarterly benchmark-indicator (BI) ratios helps reduce the potential wagging tail problem discussed in Annex 6.2.

within the first month after the end of the reference period (e.g., price statistics and industrial production indices), while other data may only be available with a delay of more than three months. Thus, when preparing the first estimates, for some series only data for two months of the last quarter may be available, while for other series data may be missing altogether.

7.20. If no related indicator is available to support an extrapolation, several options can be considered, depending on the strength of the underlying trend in the series and the importance of seasonality in the series. One generally applicable option would be to use ARIMA[4] time-series modeling techniques, which in many cases have proved to produce reasonable forecasts for one or two periods ahead. ARIMA modeling, however, is complicated and time-consuming, and requires sophisticated statistical knowledge. Also, ARIMA models are basically not able to forecast changes in the underlying trend in the series. Their good forecasting reputation stems mainly from their ability to pick up repeated patterns of the series, such as seasonality.

7.21. Thus, if there is strong seasonal variation and trend in the series, a substantially less demanding, and potentially better, solution would be the following three-step procedure:

- First, use standard seasonal adjustment software (e.g., X-11-ARIMA or X-12-ARIMA) to seasonally adjust the series and to estimate the trend component of the series. For this particular purpose, only a basic knowledge of seasonal adjustment is required, and knowledge of ARIMA modeling is not necessary.
- Second, extend the trend component of the series based on judgment, forecasts, or annual data, or by projecting the current trend using the simple trend formula in equation (7.5) below.
- Third, multiply the trend forecast with the seasonal and irregular factors computed by the program.

7.22. In many cases, the following, much simpler, approaches may prove sufficient:

- If there is no clear trend or seasonality in the movements of the series (either in volume or price), one may simply repeat the last observation or set the value for the missing period equal to a simple average of, for example, the last two observations.
- With strong seasonal variation in the series but no clear underlying trend in the series' movements, one may simply repeat the value of the variable in the same period of the previous year or set the value for the missing observation equal to the average for the same period in several of the previous years.
- If there is a clear trend in the series but no pronounced seasonal variation, the past trend may be projected using a weighted average of the period-to-period rates of change for the last observations, for example, by using a weighted average for three last observations as follows:

$$X_{T+t} = X_{T+t-1} \cdot \left[\frac{3}{6} \cdot \frac{X_T}{X_{T-1}} + \frac{2}{6} \cdot \frac{X_{T-1}}{X_{T-2}} + \frac{1}{6} \cdot \frac{X_{T-2}}{X_{T-3}} \right] \quad (7.4)$$

- With both a clear trend and strong seasonal variation in the series, one simple option may be to extrapolate the value of the series in the same period in the previous year, using a weighted average of the rates of change from the same period in the previous year for the last observations as an extrapolator, for example, by using a weighted average for three last observations as follows:

$$X_{T+t} = X_{T+t-s} \cdot \left[\frac{3}{6} \cdot \frac{X_T}{X_{T-s}} + \frac{2}{6} \cdot \frac{X_{T-1}}{X_{T-s-1}} + \frac{1}{6} \cdot \frac{X_{T-2}}{X_{T-s-2}} \right] \quad (7.5)$$

In this formula, s is the periodicity of the series, X_T is the level of the last observation, and t is the number of periods to be projected.

[4]Autoregressive integrated moving average.