Journal of Official Statistics, Vol. 14, No. 2, 1998, pp. 163–179

Estimating the Sampling Variance of the UK Index of Production

P.N. Kokic[1]

One of the main economic indicators produced by United Kingdom's Office for National Statistics is the Index of Production. It is a monthly index of the total volume of industrial output obtained by combining estimates from a number of survey sources. The purpose of this article is to construct an estimator of the sampling variance of the Index of Production. This variance estimator is obtained by linear approximation methods. The parametric bootstrap is employed to assess the adequacy of approximation made when deriving the variance estimator. Some of the advantages of using the parametric bootstrap to estimate the variance are briefly described. Both estimators have the advantage that they do not require revision whenever the methodology of any of the surveys is changed.

Key words: Sampling error; nonsampling error; parametric bootstrap.

1. Introduction

One of the main economic indicators produced by United Kingdom's Office for National Statistics (ONS) is the Index of Production (IoP). The IoP is a monthly index of the total volume of industrial output (or production). It covers the Mining, Manufacturing and Agricultural sectors of the economy and is currently based to 1990 prices. It is one of the main indicators of economic growth within the UK. It is reported monthly, and it receives much attention from both within and outside government.

The IoP is obtained by combining several different sources of data. By far the most significant source is ONS surveys. These include the Monthly Production Inquiry (MPI), Producer Price Index (PPI), and the Quarterly Stocks Inquiry (QSI). Other data used in its construction include the Export Price Deflator (EPD), which is currently derived from a combination of data collected by ONS and by Customs and Excise, and additional data on the oil, gas, electricity and mining industries from the Department of Trade and Industry, and on food production from the Ministry of Agriculture, Fisheries and Food.

The purpose of this article is to describe formulas for estimating the sampling variance of the IoP, and to present an alternative estimator based on the parametric bootstrap simulation method (Efron and Tibshirani 1993). The variance estimator is obtained by

[1] Department of Social Statistics, University of Southampton, Southampton SO17 1BJ, United Kingdom.
Acknowledgments: The author wishes to thank Chris Skinner of the University of Southampton, and Dave Watts and Mike Prestwood of the Office for National Statistics for their helpful comments on earlier versions of this article. The author is also grateful to the associate editor and three anonymous reviewers for their useful comments.

linear approximation methods; a common technique used for estimating variances of non-linear estimators in a single survey, see for example Andersson and Nordberg (1994). Both variance estimators have the advantage that they do not require revision whenever the methodology of any of the surveys is changed, and so either is suitable for use in a modular survey estimation system.

The advantage of the parametric bootstrap over using a conventional variance formula is the fact that it is more flexible in practice since it avoids the need for complex mathematical derivations. Its use is not restricted to the IoP but the method may be applied to any statistic which has been derived from several survey estimates. On the other hand it may be extremely difficult and time consuming obtaining variance estimators of these kind of statistics by Taylor series methods.

The motivation for undertaking this work has largely been the desire of ONS to make appropriate statements about the statistical significance of changes in the IoP. The need to quantify the accuracy of composite statistics like the national accounts and balance of payments has been clearly identified in the literature (UK Central Statistical Office 1992a, and Ramsay 1993). Given that the IoP is an official statistic of this type, the work in this article may be viewed as an initial response to the need above.

Also there has been a desire to determine in an objective fashion which sources of data are contributing most to the sampling error of the IoP, and as a consequence whether methodological procedures for certain data inputs require change. The analysis presented in this article goes a long way towards addressing the second of these two issues, but only part of the way to addressing the first. Given that ONS (and the other government departments providing statistical input to the IoP) will for some time not produce standard error estimates of month-on-month changes, it turns out, at least using the techniques presented in this article, to be impossible to produce accurate standard error estimates of short-term changes in the IoP. Given the considerable difficulties involved, it would seem that an appropriate initial goal is to attempt to produce accurate standard error estimates of the absolute level of the IoP itself.

The main reason that it becomes necessary to test the variance estimate of the IoP is that in practice the sampling variability of certain inputs are sometimes not available or can only be approximated. It must therefore be established that the assumptions underlying these approximations are valid in practice over a range of realistic alternatives. Another difficulty arises due to the limited amount of information available for simulation.

Valliant (1991, 1992) has considered the problem of variance estimation of price indexes derived from complex multistage surveys. As can be seen from the discussion above, effort here is concentrated instead on the derivation and testing of a fairly complicated index derived from several surveys which all have comparatively simple designs. Other articles of interest in this area include Andersson, Forsman and Wretman (1987), Balk and Kersten (1986), Dalén and Ohlsson (1995), Leaver (1990), and Leaver, Johnstone and Kenneth (1991).

This article begins with a brief overview of the construction of the IoP. In Section 3 an approximation to its sampling variance is derived. Results of a simulation study are presented in Section 4, and in Section 5 other potential sources of error in the measurement of the IoP are briefly overviewed. Finally conclusions are drawn in Section 6.

2. Construction of the IoP

The IoP is first constructed within industry groups at the 4-digit standard industry classification (SIC) level (UK Central Statistical Office 1992b). Let $I_{0t,h}$ be the IoP estimate for time period t relative to the reference base 0 in industry group $h = 1, 2, \ldots, 241$. Higher level estimates are produced by taking weighted averages of these IoP estimates, where the weights are determined by the gross value added in the base year (estimated from the Annual Census of Production survey). Thus the overall index I_{0t} is given by

$$I_{0t} = \left(\sum_h I_{0t,h} w_{0h}\right)\left(\sum_h w_{0h}\right)^{-1} \tag{1}$$

where w_{0h} is the value added weight in h. The relative change in the IoP between time periods r and t may be written as $I_{rt} = \Sigma_h I_{0t,h} w_{0h} / \Sigma_h I_{0r,h} w_{0h}$.

It is possible to view (1) as a modified form of a Laspeyres index (Allen 1975, p. 25). In general Laspeyres price and quantity indexes are defined as

$$\frac{\sum_h p_{th} q_{0h}}{\sum_h p_{0h} q_{0h}} \text{ and } \frac{\sum_h p_{0h} q_{th}}{\sum_h p_{0h} q_{0h}}$$

respectively, where q represents the quantity of product sold and p the price of the product. This may be written in the alternative form $\Sigma_h v_{0h} R_{th} / \Sigma_h v_{0h}$, where v_{0h} is the value of a product in the reference base period and R_{th} is either a price or production ratio of the current period (t) to reference base period (0). In the above formulation v_{0h} can be equated with w_{0h} and R_{th} with $I_{0t,h}$. The gross value added weights in (1) are fixed in the base year, currently 1990. However, as will be seen below $I_{0t,h}$ is neither a price or production ratio, but rather a deflated sales index adjusted for change in stocks.

From now on, except where necessary for clarity, we shall only make reference to the 4-digit industry IoP estimates $I_{0t,h}$, and so for simplicity the subscript h will be dropped. The process of construction can be broken down into a number of distinct steps.

Step 1. The first step in the process is to construct a combined price deflator. Price deflators for home sales (that is domestic sales) are estimated for the current month from PPI data, and for export sales from EPD data. Note that individual estimates of these two price deflators and all subsequent survey estimates used in constructing the IoP are produced for each of the 241 4-digit industries mentioned above. The inverse of these deflators estimate the average price increase from the base year for commodities produced and sold by all contributors in a given industry. The combined deflator is a harmonic mean of the home and export price deflators weighted together by total home sales and total export sales (estimated from MPI data in the current month). Specifically, if $\hat{D}_{0t,1}$ is the PPI home price deflator, $\hat{D}_{0t,2}$ is the export price deflator, $\hat{S}_{t1}$ is home sales and $\hat{S}_{t2}$ is export sales, then the combined deflator $\hat{D}_{0t}$ satisfies

$$\frac{1}{\hat{D}_{0t}} = \frac{1}{\hat{D}_{0t,1}} \frac{\hat{S}_{t1}}{\hat{S}_t} + \frac{1}{\hat{D}_{0t,2}} \frac{\hat{S}_{t2}}{\hat{S}_t} \tag{2}$$

where $\hat{S}_t = \hat{S}_{t1} + \hat{S}_{t2}$

Step 2. The next step is to construct the deflated weighted sales index. This index represents the relative increase in real terms of sales in the current month compared to the base year. Total sales in the current month is first deflated by the combined deflator then divided by the average total sales over the 12-month period in the base year (estimated from MPI data). This divisor is referred to as the published group divisor. The derivation of the index is actually slightly more complicated than this as merchanted goods are treated separately in the process. Merchanted goods are products sold on by a business without being subjected to a manufacturing process. If $\hat{M}_t$ is sales of merchanted goods, $\hat{g}_{01}$ is the monthly average total sales less merchanted goods in the base year and $\hat{g}_{02}$ is the monthly average of merchanted goods in the base year, then the deflated weighted sales index is

$$I_{0t,0} = \left\{ (\hat{S}_t - \hat{M}_t) \frac{w_{01}}{\hat{g}_{01}} + \hat{M}_t \frac{(1 - w_{01})}{\hat{g}_{02}} \right\} \frac{1}{\hat{D}_{0t}} \tag{3}$$

where w_{01} is the proportion of total sales which are non-merchanted goods.

Step 3. A benchmark sales index suitable for seasonal adjustment is created by multiplying the deflated weighted sales index by a constraining factor, rebasing and then adding tuning constants. The purpose of the constraining factor is to make the IoP estimates meet certain (externally imposed) constraints for publication, and tuning constants are used for minor adjustments when the IoP does not follow patterns expected in the relevant industry. Thus, if c_t is the constraining factor, d_0 the monthly average of the deflated weighted sales index in the base year, and a_t is the tuning constant, then the benchmark sales index is

$$I_{0t,1} = I_{0t,0} \frac{c_t}{d_0} + a_t \tag{4}$$

Step 4. The next step is to seasonally adjust the benchmark sales index using the X11-ARIMA algorithm. However, since our concern here is to measure the sampling variability of the non-seasonally adjusted series, we shall move straight on to the next step in the process, which is stock adjustment. Since goods are often produced in one month and sold in another, it becomes necessary to introduce a stock adjustment, $\hat{A}_t$ say. The stock adjustment is actually estimated from QSI data on the basis of average stock changes within each quarter, and then the same factor is applied equally to each month within the quarter. Thus, the non-seasonally adjusted IoP is

$$I_{0t} = I_{0t,1} + \hat{A}_t \tag{5}$$

An approximation to I_{0t} suitable for deriving a sampling variance will be given in Section 3.

3. Estimating the Sampling Variance of the IoP

3.1. An approximation to the IoP

In order to estimate the sampling variance of the IoP we make a number of approximations. The first approximation is that $d_0 \cong 1$. That is, the second rebasing has virtually no effect on the index. Since the combined deflator $\hat{D}_{0t}$ is close to 1 over the 12-month base period, and both $(\hat{S}_t - \hat{M}_g)/\hat{g}_{01}$ and $\hat{M}_t/\hat{g}_{02}$ have mean one over the base year, such an approximation would seem entirely reasonable. Note that if $\hat{g}_{01}$ and $\hat{g}_{02}$ had been defined, respectively, as the average of the price-adjusted total sales and merchanted-good sales figures instead of the unadjusted version, then d_0 would in fact be exactly equal to one.

The second approximation is that $w_{01} \cong \hat{g}_{01}/\hat{g}_0$, where $\hat{g}_0 = \hat{g}_{01} + \hat{g}_{02}$ is the 12-month average of total sales over the base year, or that $(\hat{S}_t - \hat{M}_t)/\hat{S}_t \cong \hat{g}_{01}/\hat{g}_0$. Under either approximation, it follows from (3) that

$$I_{0t,0} \cong \frac{\hat{S}_t}{\hat{g}_0 \hat{D}_{0t}} \tag{6}$$

Furthermore, in most industries merchanted goods are a fairly minor contributor to total sales so (6) would normally be a good approximation. To see this it is necessary to take account of the practical situation and consider what a likely worst-case scenario would be. First note that by (3), an upper bound for the standard error (SE) of $I_{0t,0}$ is

$$\mathrm{SE}(I_{0t,0}) \leq \mathrm{SE}\left(\frac{(\hat{S}_t - \hat{M}_t)w_{01}}{\hat{g}_{01}\hat{D}_{0t}}\right) + \mathrm{SE}\left(\frac{\hat{M}_t(1 - w_{01})}{\hat{g}_{02}\hat{D}_{0t}}\right) \tag{7}$$

For almost all industries merchanted goods makes up less than five per cent of total sales. Thus the SE of $(\hat{S}_t - \hat{M}_t)/(\hat{g}_{01}\hat{D}_{0t})$ should be close to that of $\hat{S}_t/(\hat{g}_0\hat{D}_{0t})$, whereas the SE of $\hat{M}_t/(\hat{g}_{02}\hat{D}_{0t})$ is likely to be up to twice this value in a worst-case situation. Also $w_{01} \geq 0.95$ for most industries. Thus from (7), the SE of $I_{0t,0}$ will be at most five per cent larger than that of $\hat{S}_t/(\hat{g}_0\hat{D}_{0t})$, and indeed it would be much closer to the SE of $\hat{S}_t/(\hat{g}_0\hat{D}_{0t})$ in most 4-digit industries. Thus the degree of under-estimation of sampling error introduced by (6) will almost certainly be negligible.

Under the two approximations $d_0 \cong 1$ and (6), it follows from (3), (4), and (5) that

$$\begin{aligned} I_{0t} &\cong \frac{\hat{S}_t c_t}{\hat{g}_0 \hat{D}_{0t}} + a_t + \hat{A}_t \\ &= I_{0t,2} c_t + a_t + \hat{A}_t \end{aligned} \tag{8}$$

where $I_{0t,2} = \hat{S}_t/(\hat{g}_0\hat{D}_{0t})$. This approximation is used in the following subsection to obtain an estimator for the sampling variance of the IoP.

3.2. *An approximate estimator of the sampling variance*

In the appendix, linearization is applied to Equation (8) to obtain an approximate equation for the sampling variance of I_{0t}. In deriving this equation it is assumed that the constraining factor c_t, and the tuning constant a_t, are fixed, as both these terms do not contribute to the sampling variability of I_{0t}, see Subsection 5.2.

Let $\hat{v}(.)$ denote the estimate of variance of its argument and initially assume that $\hat{v}(\hat{S}_t)$, $\hat{v}(\hat{S}_{t1})$, $\hat{v}(\hat{S}_{t2})$, $\hat{v}(\hat{D}_{0t,1})$, $\hat{v}(\hat{D}_{0t,2})$, $\hat{v}(\hat{g}_0)$, $\hat{v}(\hat{A}_t)$ are all available from external sources, see Subsection 3.4 below. Then results derived in the appendix suggest the following estimator:

$$\hat{v}(I_{0t}) = c_t^2 I_{0t,2}^2 \left\{ \frac{\hat{v}(\hat{g}_0)}{\hat{g}_0^2} + \frac{\hat{D}_{0t}^2}{\hat{D}_{0t,1}} \left(\frac{1}{\hat{D}_{0t,1}} - \frac{1}{\hat{D}_{0t,2}} \right) \frac{\hat{v}(\hat{S}_{t1})}{\hat{S}_t^2} + \frac{\hat{D}_{0t}^2}{\hat{D}_{0t,2}} \left(\frac{1}{\hat{D}_{0t,2}} - \frac{1}{\hat{D}_{0t,1}} \right) \frac{\hat{v}(\hat{S}_{t2})}{\hat{S}_t^2} \right.$$
$$\left. + \left(\frac{\hat{S}_{t1}\hat{D}_{0t}}{\hat{S}_t \hat{D}_{0t,1}} \right)^2 \frac{\hat{v}(\hat{D}_{0t,1})}{\hat{D}_{0t,1}^2} + \left(\frac{\hat{S}_{t2}\hat{D}_{0t}}{\hat{S}_t \hat{D}_{0t,2}} \right)^2 \frac{\hat{v}(\hat{D}_{0t,2})}{\hat{D}_{0t,2}^2} + \frac{\hat{D}_{0t}^2}{\hat{D}_{0t,1}\hat{D}_{0t,2}} \frac{\hat{v}(\hat{S}_t)}{\hat{S}_t^2} \right\} + \hat{v}(\hat{A}_t) \qquad (9)$$

As not all the variance estimates contributing to (9) are readily available, assumptions must be made in order to apply the estimator in practice. These assumptions are as follows.

(a) There is no correlation between the group divisor $\hat{g}_0$ and $\hat{S}_t$, $\hat{S}_{t1}$ or $\hat{S}_{t2}$. This will normally be the case as usually the reference time point and base year are more than 15 months apart (this corresponding to the time for complete rotation of the MPI sample). The magnitude of the correlation for estimates 12 months apart should also be negligible.

(b) Since the EPD is estimated from a completely enumerated cut-off sample the sampling variance of $\hat{D}_{0t,2}$ is zero. That is, the units on the sampling frame are sorted by size and the largest are selected for inclusion in the sample. What has been done is that the sampling variance of the EPD has been reduced to zero by effectively replacing it with an unmeasurable (sampling) bias. Suppose that the resulting mean squared error of the EPD is about the same as the sampling variance of the PPI estimates, which is not unreasonable given that both estimates play a similar role in the IoP. Then a bound on its effect on the precision of the IoP would be obtained by setting $\hat{v}(\hat{D}_{0t,2}) = \hat{v}(\hat{D}_{0t,1})$. In Section 4 it is established through simulation that the contribution of the EPD to the sampling variability of the IoP is insignificant even when the assumption above is incorrect by a large degree. Setting $\hat{v}(\hat{D}_{0t,2})$ to a realistic non-zero value will also enable us to assess whether (9) will continue to work if in the future a sampling variance can be produced.

(c) The PPI index is produced from a fixed panel of units selected from the 1990 survey. Variance estimates are currently produced but these are based on various approximations to the sample design, see Subsection 3.4 below. For reasons similar to those valid for the EPD it is important that (9) works well for a range of values since there are plans afoot to change the sampling methodology of the PPI.

(d) Since the MPI was a cut-off sample in 1990, there is no sampling error associated with $\hat{g}_0$. However, for reasons similar to those pertaining to the EPD, it would still be wise to use some positive value for its variance. Given that the base will soon be moved forward, it would be preferable to estimate the variance of $\hat{g}_0$ using the current MPI sampling scheme. Using this approach and noting that the rotation period for the MPI is currently 15 months, it is likely that the relative variance of the group divisor is significantly less than half the relative variance of total sales in any particular month. (It is straight-forward to establish this simple upper bound under simplifying assumptions, and so for brevity its proof has not been included here.) Therefore, assume that the relative variance of $\hat{g}_0$ is half the relative variance of $\hat{S}_t$.

(e) Assume that the weights used to aggregate the index have no sampling variance. It would be possible to incorporate the additional variability from this source; however, considerable additional complexity is involved, and its contribution to total variance is expected to be relatively minor. Furthermore, these weights are essentially treated as fixed known constants when computing the IoP index. Thus it would make sense to treat them in a similar way when deriving an estimate of the variance.

(f) Even though it can potentially be produced no estimate of sampling variance is currently available for the stocks adjustment and it is difficult to judge its magnitude. Thus it is necessary to assume that its variance is zero. Since $\hat{A}_t$ is statistically independent of the other terms in the index, provided Equation (9) estimates the variance of I_{0t} well with $\hat{A}_t = 0$, then it should also work when an estimate of the sampling variance is available for the stocks adjustment.

Assumptions (b)–(d) will be tested by simulation in the following section, while (a), (e) and (f) will be taken as facts. Indeed, the first assumption has already been imposed when deriving (9). Under (a)–(f) the variance estimate (at the 4-digit industry level) simplifies slightly to

$$\hat{v}(I_{0t}) = c_t^2 I_{0t,2}^2 \left[\frac{\hat{D}_{0t}^2}{\hat{D}_{0t,1}} \left(\frac{1}{\hat{D}_{0t,1}} - \frac{1}{\hat{D}_{0t,2}} \right) \frac{\hat{v}(\hat{S}_{t1})}{\hat{S}_t^2} + \frac{\hat{D}_{0t}^2}{\hat{D}_{0t,2}} \left(\frac{1}{\hat{D}_{0t,2}} - \frac{1}{\hat{D}_{0t,1}} \right) \frac{\hat{v}(\hat{S}_{t2})}{\hat{S}_t^2} \right.$$
$$\left. + \left(\frac{\hat{D}_{0t}}{\hat{S}_t} \right)^2 \left\{ \left(\frac{\hat{S}_{t1}}{\hat{D}_{0t,1}^2} \right)^2 + \left(\frac{\hat{S}_{t2}}{\hat{D}_{0t,2}^2} \right)^2 \right\} \hat{v}(\hat{D}_{0t,1}) + \left(\frac{1}{2} + \frac{\hat{D}_{0t}^2}{\hat{D}_{0t,1}\hat{D}_{0t,2}} \right) \frac{\hat{v}(\hat{S}_t)}{\hat{S}_t^2} \right] \tag{10}$$

At higher levels

$$\hat{v}(I_{0t}) = \left\{ \sum_h w_h^2 \hat{v}(I_{0t,h}) \right\} \left(\sum_h w_h \right)^{-2}$$

3.3. *Estimating the variance using the parametric bootstrap*

An alternative method of estimating the sampling variance of the IoP using the parametric bootstrap (Efron and Tibshirani 1993, p. 53) is presented below. We begin with a brief description of the bootstrap technique.

Suppose that $\hat{\theta}$ is an estimator of some parameter θ based on a sample of independent and identically distributed variables $X_1, \ldots, X_n$. In the ordinary bootstrap B independent random samples of size n are drawn from the empirical distribution function of $X_1, \ldots, X_n$ and a new estimate $\hat{\theta}_{(b)}$ is constructed for each of the samples, $b = 1, 2, \ldots, B$. The empirical variance of $\hat{\theta}_{(1)}, \ldots, \hat{\theta}_{(B)}$, which is called the bootstrap variance of $\hat{\theta}$, is an estimate of the true variance of $\hat{\theta}$. Efron and Tibshirani (1993, p. 52) recommend that $B = 200$ will be sufficiently large in most cases. The parametric bootstrap differs from the ordinary bootstrap in so far as each bootstrap sample is drawn from some parametric estimate of the true underlying distribution function of $X_1, \ldots, X_n$ rather than from the empirical distribution function of the data.

Our intention is to apply this technique to each of the survey inputs of the IoP at the 4-digit industry level and not directly to the survey data itself. At the 4-digit level estimates and sometimes their corresponding variances are available. These may be used to estimate parametric distributions for each of the survey inputs. Generally normal distributions (Bickel and Doksum 1977, p. 458) are fitted with mean and variance set equal to the survey estimate and its variance estimate, respectively. For most industries the accuracy of this normal approximation should be fairly good. For industries where a normal approximation is not sufficiently accurate samples can be drawn from a more 'heavy-tailed' or skewed distribution such as a χ^2 or Student's t distribution (Bickel and Doksum 1977, p. 16). Note that the IoP depends upon two variables from the MSI, home and export sales and so it is necessary to account for the correlation, ρ_t say, between these two variables. This correlation can be estimated in a straightforward manner from the variances of each of home and export sales, and total sales.

To be precise, the parametric bootstrap technique operates as follows for the IoP. As for (10) set $\hat{A}_t = 0$. The bootstrap estimates for the bth simulation in industry group h are

$$\hat{D}_{0t,1h(b)} \sim^{IID} Normal(\hat{D}_{0t,1h}, \hat{v}(\hat{D}_{0t,1h}))$$

$$\hat{D}_{0t,2h(b)} \sim^{IID} Normal(\hat{D}_{0t,2h}, \hat{v}(\hat{D}_{0t,2h}))$$

$$(\hat{S}_{t1h(b)}, \hat{S}_{t2h(b)}) \sim^{IID} Normal\left(\begin{pmatrix}\hat{S}_{t1h}\\ \hat{S}_{t2h}\end{pmatrix}, \begin{pmatrix}\hat{v}(\hat{S}_{t1h}) & \hat{\rho}_{th}\sqrt{\hat{v}(\hat{S}_{t1h})\hat{v}(\hat{S}_{t2h})}\\ \hat{\rho}_{th}\sqrt{\hat{v}(\hat{S}_{t1h})\hat{v}(\hat{S}_{t2h})} & \hat{v}(\hat{S}_{t2h})\end{pmatrix}\right) \text{ and}$$

$$\hat{g}_{0h(b)} \sim^{IID} Normal(\hat{g}_{0h}, \hat{v}(\hat{g}_{0h})), \tag{11}$$

where the notation $\sim^{IID}$ indicates that a random observation was drawn from the distribution function indicated. These simulated values are combined according to Equations (1) and (8) to obtain simulated values of $I_{0t,h(b)}$ and $I_{0t(b)}$. The bootstrap variance estimate of the IoP is simply

$$\hat{v}_B(I) = (B-1)^{-1}\sum_{b=1}^{B}(I_{0t(b)} - B^{-1}\sum_{c=1}^{B} I_{0t(c)})^2$$

As mentioned above this variance estimate should be close to the true sampling variance of the IoP and so is useful for assessing the precision of the IoP variance estimate (10).

The advantage of the parametric bootstrap over using a conventional Taylor series approach is the fact that it avoids the need for complex mathematical derivations.

3.4. *Input survey variance estimates*

Both methods of variance estimation described in the previous two subsections require estimates of variances for each survey input to the IoP. As already described a number of assumptions have had to be made since no variance estimates are available for some of these survey inputs. There are two primary variance sources entering expressions (10).

The first of these are the variance components from MPI. This survey has a single-stage design and is stratified by 4-digit industry and turnover size groups. Within stratum ratio estimation with a turnover auxiliary variable is used to estimate home and export sales and thus the variances of these are estimated in a straightforward manner using the standard ratio variance formula, see Cochran (1977, p. 155).

The second component in (10) is the variance of the PPI. Purdon (1994) constructed an estimator of this variance by approximating the PPI selection process. Price quotes in the

PPI are obtained from businesses under various product groups (i.e., 6-digit SIC industry categories). The businesses are selected from units included in ONS's annual and quarterly surveys of manufacturing sales. Most of these units were selected from the 1990 surveys, but some updating has occurred from later surveys. Within any product group businesses selected for the PPI are those with the largest sales for that particular group, and then a number of items are selected that are 'representative' of the price movements within the product group for the business. The PPI is a Laspeyres price index based on these sales and price movement data.

Purdon (1994) argued that for the purposes of estimating the PPI variance this sample selection process would be approximated by a three stage design. The first stage was a stratified random selection of product groups. The strata were defined in terms of total sales in the product group and selection probabilities within each stratum were based on the observed frequency of selection of product groups. The second stage was the selection of businesses within a product group, which was modelled using a probability proportional to size of sales selection process. The third stage was selection of items within any business, which was approximated by a simple random sampling procedure. With this well-defined selection procedure, Purdon was able to use standard conditioning arguments to derive a variance estimator for the PPI.

4. Evaluation of the Estimator by Simulation

To assess the bias of the variance estimator (10) and to test the assumptions made in the

Table 1. Description of the various scenarios used when simulating the IoP

Scenario	Description
0 (Base scenario)	$Var(EPD) = var(PPI), rel.var(\hat{g}_0) = rel.var(\text{total sales}) \div 2$, 1,000 bootstrap simulations in each 4-digit SIC from a normal distribution as described in Subsection 3.3, and stock adjustment is zero.
1	Same as the base scenario except that $var(EPD) = var(PPI) \div 2$.
2	Same as the base scenario except that $var(EPD) = 2 \times var(PPI)$.
3	Same as the base scenario except that $rel.var(\hat{g}_0) = rel.var(\text{total sales}) \div 4$.
4	Same as the base scenario except that $rel.var(\hat{g}_0) = rel.var(\text{total sales})$.
5	Same as Scenario 3 except that for 2,000 simulations in each 4-digit SIC from a Student's *t* distribution with 5 degrees of freedom (Bickel and Doksum 1977, p. 16) was mixed with a normal distribution. The mixing proportion was 10 per cent. That is, values were drawn from the standard normal distribution with 90 per cent probability, or from a Student's *t* distribution with 5 degrees of freedom with 10 per cent probability. These values were then linearly transformed to obtain the appropriate bootstrap values in place of those at (11).
6	Same as the base scenario except that $var(PPI)$ was doubled without changing the value of $var(EPD)$.
7	Same as the base scenario except that $var(\text{total sales})$, $var(\text{home sales})$ and $var(\text{export sales})$ are doubled, without changing the value of $var(\hat{g}_0)$.

previous section, a sensitivity analysis was undertaken using the bootstrap simulation technique. The scenarios that were tested are described in Table 1. For simplicity and without loss of generality we set $a_t = 0$, $\hat{A}_t = 0$ and $c_t = 1$, see the discussion in Subsections 3.2 and 5.2.

The assumption made under the Base scenario (Scenario 0) are exactly the same as were made when deriving the IoP variance estimate (10). This scenario is therefore useful for testing the degree of bias introduced by the linear approximations made when deriving the variance.

Scenario 5 should capture the likely degree of variance inflation due to a slight, but at the same time fairly realistic departure from normality. Other mixtures of distributions could have been used; for example a mixture of a normal and χ^2 distribution would have been appropriate if the underlying data were right-skewed. However, the likely effects on the variance of the IoP should be of a similar degree to those under Scenario 5.

The data used to perform the simulations were for the December 1995 IoP. There are a total of $H = 192$ 4-digit industries for which complete data were available. This covers most of the Manufacturing sector and the Mining and Quarrying except Energy Producing Materials Subsection, see UK Central Statistical Office (1992b). They can be classified into 14 2-letter industry groups. The 31 out of 218 4-digit industries not covered in the Manufacturing sector were excluded from the analysis because data for these industries are supplied to ONS by other government departments in strict confidence. The effects of this slight under-representation was a relatively small increase in the bootstrap variances of the All-Manufacturing and of some 2-letter industry IoP estimates. However, in most cases estimates of the relative bias of the SEs should be virtually unaffected. Currently the IoP is published at the 2-letter level and at the aggregate All-Manufacturing level, although it is provided to some users at the finer 4-digit level. Thus, most results presented in this article concentrate on the 2-letter level and higher.

Most results in the subsequent tables refer to the IoP 12 months after the base period, that is where $r = 0$ and $t = 12$ in I_{rt}. In this case the index compares production in December 1991 to the average production in 1990. The IoP data above will be treated as if it were from December 1991. Bootstrap SE estimates will also be presented for the case $r > 0$ and $t = r + 12$, that is for a 12-month change in the index.

Table 2 shows the relative standard errors (RSEs) of the two primary inputs to the IoP: total sales and the PPI, and the simulated SE and SE estimates of the IoP at the 2-letter and All-Manufacturing levels under the Base scenario. It also shows the bias of the SE estimate relative to the simulated SE.

The simulated SE at the All-Manufacturing level is 0.82 while the SE estimate itself is 6 per cent less than this figure. If the true SE is indeed 0.82, then a 95 per cent confidence interval calculated using the SE estimate would be roughly a 93 per cent confidence interval in practice. Clearly this degree of error is fairly minor. The relative difference of 6 per cent is in fact due to the linear approximations made when deriving (10). The simulated SEs at the 2-letter level range from about 1.3 to 5.7 per cent while the absolute relative bias of the SE estimate is no larger than 9.2 per cent.

A comparison of the relative biases for the first six scenarios at the 2-letter and All-Manufacturing levels is given in Table 3. It should be noted that the SE estimate is the same under these scenarios allowing easy comparison using the relative bias alone. The

Table 2. Summary statistics, simulated SE, estimated SE of the IoP at the 2-letter SIC level. The current time point (t) is 12 months after the reference base period ($r = 0$)

SIC	RSE of total sales (%)	RSE of the PPI (%)	Simulated SE of IoP	IoP SE estimate	Relative bias (%)
All Manufacturing[1]	0.56	0.16	0.82	0.78	−6.00
CB	3.48	0.65	2.74	2.58	−5.96
DA	1.27	0.58	2.56	2.46	−4.13
DB	2.25	0.32	2.24	2.10	−6.02
DC	6.35	0.68	5.69	5.33	−6.21
DD	4.59	0.72	3.97	3.74	−5.71
DE	1.58	1.15	2.82	2.56	−9.19
DG	0.79	0.50	1.25	1.19	−4.79
DH	2.66	0.49	3.66	3.47	−5.20
DI	2.13	0.36	2.05	1.93	−5.82
DJ	2.13	0.43	2.10	1.97	−6.22
DK	1.95	0.27	2.41	2.27	−5.83
DL	1.68	0.41	3.58	3.40	−5.10
DM	1.78	0.49	1.91	1.81	−4.85
DN	2.98	0.38	3.00	2.83	−5.61

[1]Excluding Subsection CB and 25 4-digit industries within the Manufacturing sector.

relative biases for Scenarios 1 and 2 are almost the same as for Scenario 0. This indicates that the variance of the IoP is fairly insensitive to the assumptions made about the variance of the EPD.

This continues to be the case at 4-digit level, see Table 4. Thus assumption (b), see Section 3, concerning the variance of the EPD should be suitable.

Table 3. Relative bias (%) of the IoP SE estimate at the 2-letter SIC level. The current time point (t) is 12 months after the reference base period ($r = 0$)

SIC	Scenario 0	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
All Manufacturing[1]	−6.00	−5.35	−7.36	1.58	−17.30	1.30
CB	−5.96	−5.87	−6.14	3.71	−19.94	3.33
DA	−4.13	−4.11	−4.18	1.89	−13.54	−1.10
DB	−6.02	−6.01	−6.04	3.96	−20.14	4.74
DC	−6.21	−6.20	−6.24	4.31	−20.87	5.82
DD	−5.71	−5.70	−5.71	4.13	−19.55	4.21
DE	−9.19	−7.81	−12.37	−5.30	−15.79	−6.92
DG	−4.79	−3.35	−7.46	3.16	−16.43	3.12
DH	−5.20	−5.17	−5.28	4.26	−18.39	4.62
DI	−5.82	−5.78	−5.89	3.80	−19.66	3.61
DJ	−6.22	−6.17	−6.30	3.39	−20.28	3.46
DK	−5.83	−5.80	−5.89	3.65	−19.34	4.47
DL	−5.10	−4.23	−6.74	3.23	−17.08	3.79
DM	−4.85	−4.63	−5.28	4.37	−17.73	5.62
DN	−5.61	−5.61	−5.63	4.03	−19.12	4.93

[1]Excluding Subsection CB and 25 4-digit industries within the Manufacturing sector.

Table 4. Percentiles of the relative bias (%) of the IoP SE estimate at the 4-digit SIC level. The current time point (t) is 12 months after the reference base period (r = 0)

Percentile	Scenario 0	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
97.5	−2.402	1.883	−3.669	5.143	−7.437	7.539
90.0	−3.943	−2.440	−4.552	4.473	−13.432	6.550
75.0	−4.677	−4.450	−5.116	4.149	−16.981	5.380
50.0	−5.302	−5.279	−5.623	3.795	−18.366	4.444
25.0	−5.781	−5.750	−6.276	3.237	−19.465	3.176
10.0	−6.437	−6.349	−8.352	1.837	−20.953	0.379
2.5	−8.413	−8.411	−12.311	−0.193	−24.647	−5.236

However, the choice made for the variance of the group divisor is important (assumption (d)) as shown by the results for Scenarios 3 and 4 in Tables 3 and 4. It is therefore preferable to err on the side of caution and, the choice that the relative variance of the group divisor is about half the relative variance of total sales in any particular month is fairly conservative. In Scenario 3 the SE estimate nearly always over-estimates the simulated SE at both the 2-letter and All-Manufacturing levels.

The most realistic assumptions out of all the scenarios tested were made in Scenario 5. Comparing the results for Scenario 5 with those for Scenario 3, it can be seen that contamination of the assumed normal distribution with the heavy-tailed Student's t distribution has only led to a relatively minor inflation in the simulated variance of the

Table 5. Simulated SE's for Scenarios 6 and 7, and differences relative to Scenario 0. The current time point (t) and comparison time point (r) are 12 months apart

SIC	$r = 0$					$r > 0$	
	Scenario 0	Scenario 6	Difference (%)	Scenario 7	Difference (%)	SE	Difference relative to Scenario 3 (%)
All[1]	0.82	0.92	12.20	1.03	25.61	1.05	38.68
CB	2.74	2.78	1.46	3.49	27.37	4.32	74.31
DA	2.56	3.00	17.19	3.06	19.53	2.72	12.90
DB	2.24	2.24	0.00	2.87	28.12	3.06	51.63
DC	5.69	5.69	0.00	7.27	27.77	9.58	87.55
DD	3.97	3.98	0.25	5.09	28.21	6.88	91.67
DE	2.82	3.90	38.30	3.24	14.89	3.58	32.14
DG	1.25	1.30	4.00	1.56	24.80	1.23	6.96
DH	3.66	3.68	0.55	4.69	28.14	4.11	23.51
DI	2.05	2.09	1.95	2.60	26.83	3.29	77.10
DJ	2.10	2.15	2.38	2.68	27.62	3.11	62.93
DK	2.41	2.42	0.41	3.10	28.63	2.90	32.28
DL	3.58	3.67	2.51	4.54	26.82	2.87	−12.77
DM	1.91	1.93	1.05	2.44	27.75	2.37	36.15
DN	3.00	3.01	0.33	3.85	28.33	4.55	67.46

IoP. Results for Scenario 5 in Table 3 indicate that the variance estimate is slightly conservative for most 2-letter industries. Similarly, at the 4-digit level, the SE equation (10) will rarely under-estimate and at the same time it will not severely over-estimate the true SE. In summary, the SE estimator derived in Section 3 appears to operate quite well in practice.

The purpose of Scenarios 6 and 7 was not to test the precision of the variance equation (10), but rather to assess the sensitivity of the SE of the IoP to changes in the variances of the PPI and MPI inputs. As can be seen from the results in Table 5 ($r = 0$), doubling the variance of the PPI has in most cases had little effect on the simulated SE of the IoP. Thus assumption (c) in Section 3 should be suitable, whereas doubling the variance of the MPI inputs has had a significant effect.

These results, along with those for Scenarios 1–4, indicate that from the perspective of improving the precision of the IoP, the most benefit would be obtained by adopting procedures which significantly increase the precision of the MPI inputs rather than the other survey inputs to the IoP.

Let us now briefly consider the case of estimating the SE of a 12-month change in the IoP, that is when $r > 0$. Note that I_{rt} will in general be more complicated in this case than when $r = 0$ and so it would be considerably more difficult to obtain a variance estimate by analytical methods. However, it is relatively straightforward to produce a parametric bootstrap variance estimate. The statistic I_{rt} was simulated in the case $r =$ December 1990 and $t =$ December 1991 by using the December 1995 IoP at both time points (real data for only one month was available for inclusion in the simulation study). For simplicity $D_{rt,1h}$ and $D_{rt,2h}$ were set equal to 1 in all strata and so to generate a bootstrap value of $\hat{D}_{0t,1h}$, for example, values of $\hat{D}_{0r,1h}$ and $\hat{D}_{rt,1h}$ were simulated independently from normal distributions both with variance $\hat{v}(\hat{D}_{0r,1h})$, but with means $\hat{D}_{0r,1h}$ and 1 respectively, then the values were multiplied together. Also $\hat{S}_{rh}$ and $\hat{S}_{th}$ were simulated independently from the same normal distribution, and the same assumptions as in Scenario 3 were made for the variance of the group divisor. The resulting bootstrap SEs of I_{rt} are presented in the final two columns of Table 5. As can be seen from Table 5, due to the fact that the divisor in I_{rt} is not as well estimated as in the case $r = 0$, the SE of I_{rt} is considerably larger than the SE of I_{0r}, often by more than 30 per cent.

Finally, it is worth noting that all the bootstrap simulations performed in this article required a relatively small amount of computational resources and were carried out efficiently on a desktop computer.

5. Other Sources of Error

5.1. Model errors in the IoP

Under the design-based paradigm an index obtained by combining several survey estimates (such as the IoP) can be viewed as an estimate of the population counterpart that would have been obtained if a complete census had been undertaken for each survey input used in deriving the index. This was the philosophical approach adopted in this article. However, it is also possible to view the population values themselves as being generated by some super-population model. In this case there is an additional source of error, which is not measured by the sampling error alone, and which may explain the movement in an estimate from one time period to the next.

For example, when the index is seasonally adjusted, one school of thought would say that a superpopulation model is at least implicitly being fitted to the data. Another viewpoint is that seasonal adjustment is just a linear filter (with known weights) applied to the time series data and so it continues to be sufficient to estimate the sampling error alone. For a discussion of this and related issues see Pfefferman (1994). However, this philosophical issue has never really become crucial in this article since, for simplicity, effort was concentrated solely on the non-seasonally adjusted constant price series. It would actually be possible to adapt the methods developed by Pfefferman to the more complicated situation of estimating the sampling variance of the seasonally adjusted IoP series, but these estimates would depend on quantities that are currently unavailable.

5.2. Nonsampling sources of error in the IoP

Although not incorporated in the variance estimates developed in Section 3, there are various additional sources of (nonsampling) error in the IoP. In general it is difficult to assess the magnitude of error they introduce without additional information which is currently not available.

One source of error is the tuning constant mentioned in Step 4 of the construction of the IoP above. It is set by human judgement. An examination of its historical values suggests that it is a fairly minor adjustment. One possibility is that it plays a smoothing role in so far that it is used to adjust for random fluctuations in the data that happen to go against expected trends. If this is the case then by ignoring its effects a conservative estimate of total error would be produced. The constraining factor c_t is also set according to human judgement, but in part it depends on what the IoP estimates turn out to be. Historical information shows that it is usually set to 1, and its effect on total variability is unclear.

A second more difficult issue is the process of revisions of the IoP, due to preliminary stock and MPI estimates that are made before the ‘final’ estimates of the IoP are produced. This process of revisions can spread out over a period anywhere from three up to six months. Measurement of the additional (nonsampling) error in the preliminary monthly estimate would be difficult to produce and its magnitude cannot be currently assessed. Thus the estimates produced refer only to ‘final’ IoP estimates. However, it may be possible, through examining the historical differences between preliminary and ‘final’ stocks and MPI estimates, to incorporate the additional uncertainty in a preliminary IoP sampling variance estimate. In addition there is further nonsampling error introduced through the fact that the stock adjustment refers to a quarterly period rather than a month.

6. Conclusions

In this article it was demonstrated how the parametric bootstrap method may be used to estimate the variance of the IoP. The advantage of this approach over using conventional Taylor series methods is the fact that it is more flexible in practice and avoids the need for complex mathematical derivations of variance formulas. Furthermore, it is possible to perform the bootstrap simulations quickly and efficiently without much computing resources.

Despite the number of approximations made in deriving the variance estimator for the IoP, bootstrap simulation results indicate that it will work well in most practical situations. One of its properties is that it only depends on the estimates used in constructing the IoP and their corresponding variances. Thus the equation has the advantage that it does not need to be revised whenever the methodology of any particular survey input is changed, provided estimates of variance continue to be produced. Simulation results presented in this article suggest that the variance equation continues to work well even when the input variance parameters are altered dramatically.

Another important conclusion is that the main survey input influencing the precision of IoP is sales figures from the MPI. That is to say, under the current method of constructing the IoP, if more resources were available to improving its precision, then these would be best directed towards improving the precision of the MPI inputs. Of course there are some doubts about this conclusion given the unknown extent of certain nonsampling errors, and since the variance of some survey inputs can only be approximately estimated.

Appendix

Derivation of the Variance Estimate of I

Let E(.) denote expectation, $v(.)$ variance and $c(.,.)$ covariance. The method that will be used is to linearize $I_{0t,2} = \hat{S}_t/(\hat{g}_0\hat{D}_{0t})$ using Taylor series techniques and then use the linear approximation to obtain a variance estimate for I_{0t}. Now

$$I_{0t,2} = \frac{\hat{S}_t}{\hat{g}_0\hat{D}_{0t}} = \frac{1}{\hat{g}_0}\left(\frac{\hat{S}_{t1}}{\hat{D}_{0t,1}} + \frac{\hat{S}_{t2}}{\hat{D}_{0t,2}}\right) \tag{12}$$

is a function of the random variable $\mathbf{x} = (\hat{g}_0, \hat{S}_{t1}, \hat{S}_{t2}, \hat{D}_{0t,1}, \hat{D}_{0t,2})'$. Expanding $I_{0t,2} = I_{0t,2}(\mathbf{x})$ around $\mu_x = E(\mathbf{x})$ we find that

$$\begin{aligned} I_{0t,2}(\mathbf{x}) \cong{} & I_{0t,2}(\mu_x) + \frac{\partial I_{0t,2}(\mu_x)}{\partial \hat{g}_0}\{\hat{g}_0 - E(\hat{g}_0)\} + \frac{\partial I_{0t,2}(\mu_x)}{\partial \hat{S}_{t1}}\{\hat{S}_{t1} - E(\hat{S}_{t1})\} \\ & + \frac{\partial I_{0t,2}(\mu_x)}{\partial \hat{S}_{t2}}\{\hat{S}_{t2} - E(\hat{S}_{t2})\} + \frac{\partial I_{0t,2}(\mu_x)}{\partial \hat{D}_{0t,1}}\{\hat{D}_{0t,1} - E(\hat{D}_{0t,1})\} \\ & + \frac{\partial I_{0t,2}(\mu_x)}{\partial \hat{D}_{0t,2}}\{\hat{D}_{0t,2} - E(\hat{D}_{0t,2})\} \end{aligned} \tag{13}$$

where

$$\begin{aligned} & \frac{\partial I_{0t,2}(\mathbf{x})}{\partial \hat{g}_0} = -\frac{\hat{S}_t}{\hat{g}_0^2\hat{D}_{0t}}, \frac{\partial I_{0t,2}(\mathbf{x})}{\partial \hat{S}_{t1}} = \frac{1}{\hat{g}_0\hat{D}_{0t,1}}, \frac{\partial I_{0t,2}(\mathbf{x})}{\partial \hat{S}_{t2}} = \frac{1}{\hat{g}_0\hat{D}_{0t,2}}, \\ & \frac{\partial I_{0t,2}(\mathbf{x})}{\partial \hat{D}_{0t,1}} = -\frac{\hat{S}_{t1}}{\hat{g}_0\hat{D}_{0t,1}^2} \text{ and } \frac{\partial I_{0t,2}(\mathbf{x})}{\partial \hat{D}_{0t,2}} = -\frac{\hat{S}_{t2}}{\hat{g}_0\hat{D}_{0t,2}^2} \end{aligned} \tag{14}$$

Hence by the approximation $E(I_{0t,2}(\mathbf{x})) \cong I_{0t,2}(\mu_x)$, (13) and (14), and since $\hat{g}_0$, $\hat{D}_{0t,1}$, $\hat{D}_{0t,2}$, and $\hat{S}_t$ are uncorrelated,

$$\begin{aligned}
v(I_{0t,2}) &\cong E(I_{0t,2}(\mathbf{x}) - I_{0t,2}(\mu_x))^2 \\
&\cong \left\{\frac{\partial I_{0t,2}(\mu_x)}{\partial \hat{g}_0}\right\}^2 v(\hat{g}_0) + \left\{\frac{\partial I_{0t,2}(\mu_x)}{\partial \hat{S}_{t1}}\right\}^2 v(\hat{S}_{t1}) + \left\{\frac{\partial I_{0t,2}(\mu_x)}{\partial \hat{S}_{t2}}\right\}^2 v(\hat{S}_{t2}) \\
&\quad + \left\{\frac{\partial I_{0t,2}(\mu_x)}{\partial \hat{D}_{0t,1}}\right\}^2 v(\hat{D}_{0t,1}) + \left\{\frac{\partial I_{0t,2}(\mu_x)}{\partial \hat{D}_{0t,2}}\right\}^2 v(\hat{D}_{0t,2}) \\
&\quad + 2\frac{\partial I_{0t,2}(\mu_x)}{\partial \hat{S}_{t1}}\frac{\partial I_{0t2,}(\mu_x)}{\partial \hat{S}_{t2}} c(\hat{S}_{t1}, \hat{S}_{t2}) \qquad (15)
\end{aligned}$$

The covariance in this expression may alternatively be written as $\frac{1}{2}(v(\hat{S}_t) - v(\hat{S}_{t1}) - v(\hat{S}_{t2}))$. An estimate of the variance of $I_{0t,2}$ can be constructed from (15) by using $\partial I_{0t,2}(\mathbf{x})/\partial \hat{g}_0$ as an estimate of $\partial I_{0t,2}(\mu_x)/\partial \hat{g}_0$, etc. Thus from (14) and (15),

$$\begin{aligned}
\hat{v}(I_{0t,2}) &= \left\{-\frac{\hat{S}_t}{\hat{g}_0^2 \hat{D}_{0t}}\right\}^2 \hat{v}(\hat{g}_0) + \left\{\frac{1}{\hat{g}_0 \hat{D}_{0t,1}}\right\}^2 \hat{v}(\hat{S}_{t1}) + \left\{\frac{1}{\hat{g}_0 \hat{D}_{0t,2}}\right\}^2 \hat{v}(\hat{S}_{t2}) \\
&\quad + \left\{-\frac{\hat{S}_{t1}}{\hat{g}_0 \hat{D}_{0t,1}^2}\right\}^2 \hat{v}(\hat{D}_{0t,1}) + \left\{-\frac{\hat{S}_{t2}}{\hat{g}_0 \hat{D}_{0t,2}^2}\right\}^2 \hat{v}(\hat{D}_{0t,2}) \\
&\quad + \frac{1}{\hat{g}_0 \hat{D}_{0t,1}} \frac{1}{\hat{g}_0 \hat{D}_{0t,2}} (\hat{v}(\hat{S}_t) - \hat{v}(\hat{S}_{t1}) - \hat{v}(\hat{S}_{t2})) \\
&= I_{0t,2}^2 \left\{\frac{\hat{v}(\hat{g}_0)}{\hat{g}_0^2} + \frac{\hat{D}_{0t}^2}{\hat{D}_{0t,1}}\left(\frac{1}{\hat{D}_{0t,1}} - \frac{1}{\hat{D}_{0t,2}}\right)\frac{\hat{v}(\hat{S}_{t1})}{\hat{S}_t^2} + \frac{\hat{D}_{0t}^2}{\hat{D}_{0t,2}}\left(\frac{1}{\hat{D}_{0t,2}} - \frac{1}{\hat{D}_{0t,1}}\right)\frac{\hat{v}(\hat{S}_{t2})}{\hat{S}_t^2}\right. \\
&\quad \left. + \left(\frac{\hat{S}_{t1}\hat{D}_{0t}}{\hat{S}_t \hat{D}_{0t,1}}\right)^2 \frac{\hat{v}(\hat{D}_{0t,1})}{\hat{D}_{0t,1}^2} + \left(\frac{\hat{S}_{t2}\hat{D}_{0t}}{\hat{S}_t \hat{D}_{0t,2}}\right)^2 \frac{\hat{v}(\hat{D}_{0t,2})}{\hat{D}_{0t,2}^2} + \frac{\hat{D}_{0t}^2}{\hat{D}_{0t,1}\hat{D}_{0t,2}} \frac{\hat{v}(\hat{S}_t)}{\hat{S}_t^2}\right\}
\end{aligned}$$

The variance estimate for I_{0t} at (9) follows from this expression and (8).

7. References

Allen, R.G.D. (1975). Index Numbers in Theory and Practice. Macmillan Publishing Co. Inc., New York, London.

Andersson, C., Forsman, C., and Wretman, J. (1987). Estimating the Variance of Complex Statistics: A Monte Carlo Study of Some Approximate Techniques. Journal of Official Statistics, 3, 251–265.

Andersson, C. and Nordberg, L. (1994). A Method for Variance Estimation of Non-Linear Functions of Totals in Surveys – Theory and Software Implementation. Journal of Official Statistics, 10, 395–405.

Balk, B. and Kersten, H.M.P. (1986). On the Precision of Consumer Price Indices Caused by the Sampling Variability of Budget Surveys. Journal of Economic and Social Measurement, 14, 19–35.

Bickel, P.J. and Doksum, K.A. (1977). Mathematical Statistics: Basic Ideas and Selected Topics. Holden-Day Inc., San Francisco.

Cochran, W.G. (1977). Sampling Techniques (3rd ed.). New York: John Wiley.

Dalén, J. and Ohlsson, E. (1995). Variance Estimation in the Swedish Consumer Price Index. Journal of Business and Economic Statistics, 13, 347–356.

Efron, B. and Tibshirani, R.J. (1993). An Introduction to the Bootstrap. Chapman and Hall, New York, London.

Leaver, S. (1990). Estimating Variances for the U.S. Consumer Price Index for 1978–1986. In Proceedings of the American Statistical Association, Section on Survey Research and Methods, 290–295.

Leaver, S.G., Johnstone, J.E., and Kenneth, P.A. (1991). Estimating Unconditional Variances for the U.S. Consumer Price Index for 1978–1986. In Proceedings of the American Statistical Association, Section on Survey Research Methods, 614–619.

Pfeffermann, D. (1994). A General Method for Estimating the Variances of X-11 Seasonally Adjusted Estimators. Journal of Time Series Analysis, 15, 85–116.

Purdon, S. (1994). Producer Price Indices: Variance Estimation and Optimal Sample Allocation. SCPR report Ref: P1400, Social and Community Planning Research, London, UK.

Ramsay, D. (1993). Robust Official Statistics and Economic Deregulation in New Zealand. Proceedings of the ISI 49th Session, Firenze, Italy, 363–379.

UK Central Statistical Office (1992a). Measuring the Accuracy of the National Accounts. OECD Meeting of National Accounts Experts. Paris.

UK Central Statistical Office (1992b). Standard Industrial Classification of Economic Activity. Government Statistical Service, UK.

Valliant, R. (1991). Variance Estimation for Price Indexes from a Two-Stage Sample with Rotating Panels. Journal of Business and Economic Statistics, 9, 409–422.

Valliant, R. (1992). Smoothing Variance Estimates for Price Indexes Over Time. Journal of Official Statistics, 8, 433–444.

Received January 1997
Revised September 1997