

## Confidence interval for the Consumer Price Index

Jacek Białek<sup>1</sup>

### Abstract

The Consumer Price Index (CPI) is used as a basic measure of inflation. The index approximates changes of costs of households' consumption that provide the constant utility (COLI, Cost of Living Index). In practice, we use the Laspeyres price index in the CPI measurement. In the paper we present and discuss several methods of the construction of confidence intervals for the Laspeyres price index. We assume that prices of commodities have normal distribution and we consider independent prices. We compare our results to the confidence interval obtained from the simply regression model in a simulation study.

*Keywords:* CPI, the Laspeyres price index, the confidence interval

*JEL Classification:* E1, E2, E3

### 1. Introduction

Most countries use the Laspeyres price index as a basic measure of inflation (White, 1999). The Laspeyres price index is a function of a set of prices and quantities of the considered group of  $N$  commodities coming from the given moment  $t$  and the basic moment  $s$ . In reality, the price index formula is a quotient of random variables and thus it is really difficult to construct a confidence interval for that formula. The so called new stochastic approach (NSA) in the price index theory gives some solution for the above-mentioned problem. In this approach, a price index is a regression coefficient (unknown parameter  $\theta$ ) in a model explaining price variation. Having estimated sample variance ( $\hat{\sigma}_\theta^2$ ) we can build the  $(1-\alpha)$  confidence interval as  $\hat{\theta} \pm t_{1-\alpha/2, n-1} \hat{\sigma}_\theta$ , where  $n$  is the sample size and  $t_{1-\alpha/2, n-1}$  is the  $100(1-\alpha/2)$  percentile of the  $t$  distribution with  $n-1$  degrees of freedom (Von der Lippe, 2007). The recent resurrection of the stochastic approach to index number theory is due to (Balk, 1980), (Clements and Izan, 1981, 1987), (Bryan and Cecchetti, 1993), (Selvanathan and Prasada Rao, 1994), (Clements *et. al.*, 2006). In (Diewert, 1995) we can read: "The main attraction of the stochastic approach over competing approaches to index number theory is its ability to provide confidence intervals for the estimated inflation rates". However, the same paper seems to be a critical review of this approach. In fact, the new stochastic approach takes

---

<sup>1</sup> University of Łódź, Institute of Statistics and Demography, Department of Statistical Methods, address: 3/5 Polskiej Organizacji Wojskowej St., 90-255 Łódź, Poland, e-mail: jbialek@uni.lodz.pl.

standard error as an indication of the “reliability” or appropriateness of the index function  $\Theta$ , which means that it prefers  $\Theta_1$  to  $\Theta_2$  when the standard error of  $\Theta_1$  is smaller than of  $\Theta_2$  (Von der Lippe, 2007). This kind of reasoning seems to be wrong because strength or weakness of any index formula is evaluated through its properties (mathematical, statistical) rather than goodness of fit of the regression function in which this index is a coefficient. Some criticism of the NSA, which can be found in the above-mentioned literature, has been an argument for us to seek another way of obtaining confidence intervals for price indices. In the paper we do not assume the normality of the log-change in all prices, which is a very common assumption (see for example (Diewert, 1995), (Silver and Heravi, 2006)), but we assume that stochastic processes of prices are Gaussian. On one hand it leads to some technical problems with constructing of these confidence intervals (Marsaglia, 2006), but on the other hand it facilitates generalizations to the case of dependent prices. In our work we consider only the Laspeyres<sup>2</sup> index and its confidence intervals, but our remarks have common character and can be repeated for other price index formulas. Through the simulation study we compare the results of our estimation with confidence interval obtained from a simple regression model.

## 2. Confidence interval for the Laspeyres index in the case of independent prices.

In the following we consider the case where prices of commodities are stochastic processes but, according to NSA, quantities are scalars. Let us consider a group of  $N$  commodities observed during the time interval  $[0, T]$  and let us denote by  $P_i^\tau$  the price of  $i$ -th commodity at time  $\tau$  (a random variable for the given  $\tau$ ) and by  $q_i^\tau$  the quantity of  $i$ -th commodity at time  $\tau$ , where  $q_i^\tau \in R_+$ ,  $\tau \in [0, T]$ ,  $i \in \{1, 2, \dots, N\}$ . For fixed moments  $s, t \in [0, T]$ , where the moment  $s$  we consider as the basis, we assume the independence of variables  $P_i^s$  and  $P_i^t$  for any  $i \in \{1, 2, \dots, N\}$  and also the independence of variables  $P_i^\tau$  and  $P_j^\tau$ , for any  $i \neq j$  and  $\tau = s, t$ .

---

<sup>2</sup> For example, the CPI (*Consumer Price Index*) is a Laspeyres-type index (White, 1999). But let us notice, that the actual CPI calculation is based on survey data. In last years increasing availability of bar-code scanning data has provided the opportunity to switch over the currently used Laspeyres formula to other price indices (Haan, 2002).

## 2.1. The case of Gaussian price processes

In the so called stochastic approach in the price index theory many authors assume that prices of commodities are lognormally distributed (Silver, Heravi, 2006), which means:

$$\text{Log}(P^\tau) \sim N(\mu^\tau, \sigma^\tau), \text{ for } \tau \in [T_1, T_2], \quad (1)$$

where  $P^\tau$  is treated as a random variable with realizations  $\{p_i^\tau : i = 1, 2, \dots, N\}$ . In our opinion assuming one, common distribution of prices at any moment  $\tau$  is too strong simplification. On the other hand the assumption that each  $P_i^\tau$  is lognormally distributed with the expected value  $p_i^\tau$  makes mathematical analysis unfeasible. As a solution of the afore-mentioned problem we propose to assume that for any  $i \in \{1, 2, \dots, N\}$  and  $\tau = s, t$  it holds that:

$$P_i^\tau = p_i^\tau + \varepsilon_i^\tau, \quad (2)$$

where  $p_i^\tau$  is the expected value of random variable  $P_i^\tau$  estimated from sample data and the random error is normally distributed with mean zero, namely  $\varepsilon_i^\tau \sim N(0, \sigma_i^\tau)$ . It means that:

$$P_i^\tau \sim N(p_i^\tau, \sigma_i^\tau), \quad (3)$$

where in practice,  $\sigma_i^\tau$  could be also estimated from sample data through the standard deviation of the price of  $i$ -th commodity at time  $\tau$ . Under above significations and considering moments  $s$  and  $t$ , the Laspeyres price index can be expressed as follows:

$$I_{La} = \frac{\sum_{i=1}^N q_i^s P_i^t}{\sum_{i=1}^N q_i^s P_i^s}, \quad (4)$$

where

$$\sum_{i=1}^N q_i^s P_i^t \sim N(\mu_1, \sigma_1), \quad (5)$$

$$\sum_{i=1}^N q_i^s P_i^s \sim N(\mu_2, \sigma_2), \quad (6)$$

with

$$\mu_1 = \sum_{i=1}^N q_i^s p_i^t, \quad \sigma_1 = \sqrt{\sum_{i=1}^N (q_i^s \sigma_i^t)^2}, \quad (7)$$

$$\mu_2 = \sum_{i=1}^N q_i^s p_i^s, \quad \sigma_2 = \sqrt{\sum_{i=1}^N (q_i^s \sigma_i^s)^2}. \quad (8)$$

Although the numerator and the denominator of the formula (4) are normally distributed as in (5) and (6), the distribution of the  $I_{La}$  index is still unknown. In the simplest case, when  $\mu_1 = \mu_2 = 0$ , we get the Cauchy distribution (under the normality and independence of the numerator and the denominator). However, when the two (even normal) random variables have non-zero mean then the distribution of their ratio is much more complicated. David Hinkley (Hinkley, 1969) found a formula for the above-mentioned distribution. Using his notation and under above assumptions and significations we get the following probability density of the random variable  $I_{La}$  ( $\phi$  denotes the standard normal distribution density):

$$f_{La}(x) = \frac{b(x)c(x)}{a^3(x)} \frac{1}{\sqrt{2\pi}\sigma_1\sigma_2} (2\phi(\frac{b(x)}{a(x)}) - 1) + \frac{1}{a^2(x)\pi\sigma_1\sigma_2} \exp(-\frac{\mu_1^2 + \mu_2^2}{2(\sigma_1^2 + \sigma_2^2)}), \quad (9)$$

where:

$$a(x) = \sqrt{\frac{1}{\sigma_1^2}x^2 + \frac{1}{\sigma_2^2}}, \quad (10)$$

$$b(x) = \frac{\mu_1}{\sigma_1^2}x + \frac{\mu_2}{\sigma_2^2}, \quad (11)$$

$$c(x) = \exp[\frac{1}{2}(\frac{b^2(x)}{a^2(x)} - \frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_2^2}{\sigma_2^2})]. \quad (12)$$

Although the formula (9) seems to be complicated it does not bring problems with the numerical calculation of mean, variation and any other parameter of the random variable  $I_{La}$  (see the simulation study). Thus, having any  $\alpha$ -quantile ( $q_{La}^\alpha$ ) of the distribution described by (9) we can construct the  $(1-\alpha)$  confidence interval for the Laspeyres index as follows:  $CI = (q_{La}^{\alpha/2}; q_{La}^{1-\alpha/2})$ . But let us notice that the  $(1-\alpha)$  confidence interval can be also expressed by  $(q_{La}^{c\alpha}, q_{La}^{1-\alpha+c\alpha})$ , where  $c \in (0,1)$ . When we expect the confidence interval that has minimum length ( $CI_{\min}$ ) we should find numerically:

$$c_0 = \min_{c \in (0,1)} (q_{La}^{1-\alpha+c\alpha} - q_{La}^{c\alpha}), \quad (13)$$

to obtain:

$$CI_{\min} = (q_{La}^{c_0\alpha}, q_{La}^{1-\alpha+c_0\alpha}). \quad (14)$$

If we intend to get the symmetrical confidence interval ( $CI_{sym}$ ) we should find the expected value of the Laspeyres index ( $\mu_{La}$ ) and solve numerically (with respect to  $x$ ):

$$\int_{\mu_{La}^{-x}}^{\mu_{La}^{+x}} f_{La}(t) dt = 1 - \alpha, \quad (15)$$

to obtain:

$$CI_{symm} = (\mu_{La} - x, \mu_{La} + x). \quad (16)$$

## 2.2. A simple regression model

The first stochastic index number model comes from (Selvanathan and Prasada Rao, 1994) and is given by the following equations<sup>3</sup> for  $t = s, s + 1, s + 2, \dots, T$

$$\frac{P_i^t}{P_i^s} = \theta_t + \varepsilon_i^t, \quad i = 1, 2, \dots, N, \quad (17)$$

where  $\theta_t$  represents systematic part of the price change from period  $s$  to  $t$  and the independently distributed uncorrelated random variables  $\varepsilon_i^t$  satisfy the following assumptions

$$E\varepsilon_i^t = 0, \quad Var\varepsilon_i^t = \sigma_t^2. \quad (18)$$

The least squares and maximum likelihood estimator for  $\theta_t$  is the Carli (1764) price index (Diewert, 1995):

$$\hat{\theta}_t = \frac{1}{N} \sum_{i=1}^N \frac{P_i^t}{P_i^s}. \quad (19)$$

In the so called New Stochastic Approach (NSA) the assumptions (18) are replaced by the following assumptions:

$$E\varepsilon_i^t = 0, \quad Var\varepsilon_i^t = \sigma_t^2 / w_i, \quad i = 1, 2, \dots, N, \quad (20)$$

where the  $w_i$  are nonrandom fixed shares that satisfy:

$$w_i > 0, \quad \sum_{i=1}^N w_i = 1. \quad (21)$$

Since the  $w_i$  are positive, we can multiply both sides of (17) by  $\sqrt{w_i}$ , in order to obtain homoscedastic errors. The resulting least squares and maximum likelihood estimator for  $\theta_t$  is:

$$\hat{\theta}_t = \sum_{i=1}^N w_i \frac{P_i^t}{P_i^s}, \quad (22)$$

thus it can be seen that  $\hat{\theta}_t$  is an unbiased estimator of  $\theta_t$  (Diewert, 1995) and its variance is as follows:

$$Var\hat{\theta}_t = \sum_{i=1}^N w_i^2 \frac{\sigma_t^2}{w_i} = \sigma_t^2. \quad (23)$$

---

<sup>3</sup> In 2.2. we treat  $p_i^s$ ,  $q_i^s$  and  $q_i^t$  as some fixed positive numbers and only  $P_i^t$  as a random variable.

An unbiased estimator of  $\sigma_t^2$  is (Diewert, 1995):

$$\hat{\sigma}_t^2 = \frac{1}{N-1} \sum_{i=1}^N w_i \left( \frac{P_i^t}{P_i^s} - \hat{\theta}_t \right)^2. \quad (24)$$

In (Salvanathan and Prasada Rao, 1994) authors consider, for example, the following case for the  $w_i$ :

$$w_i = \frac{p_i^s q_i^s}{\sum_{k=1}^N p_k^s q_k^s}, \quad (25)$$

and it causes  $\hat{\theta}_t$  to become the fixed based Laspeyres price index. In fact, in equations (19) the current period prices  $p_i^t$  are the only random variables. Under the additional assumption that the residuals are normally distributed, (22), (23) and (24) may be used to obtain  $1-\alpha$  confidence interval for the Laspeyres price index  $\hat{\theta}_t$  (Von der Lippe, 2007):

$$\hat{\theta}_t \pm t_{1-\alpha/2, n-1} \hat{\sigma}_t, \quad \hat{\theta}_t = I_{La}. \quad (26)$$

As it was above-mentioned, in most countries the CPI takes the Laspeyres form (Clements *et al.*, 2006). This index formula does not change with the passage of time as new information on subsequent price values becomes available. Many other popular index numbers also share this property. Moreover, stochastic index numbers can be treated as regression coefficients and thus are subject to aging (they will take different values when we obtain an additional data point and re-estimate the parameters of the model). It should be emphasized, that we entirely dismiss such criticisms in this paper. Like other approaches, NSA has both strengths and weaknesses and we use it for constructing the confidence interval of the price index.

### 3. Simulation study

Let us take into consideration a group of  $N = 10$  commodities, where random vectors of independent prices and deterministic vector of quantities are as follows:

$$\begin{aligned} P^s &= [N(8;0,5), N(200,10), N(30,2), N(300,20), N(5;0,2), N(110,5), N(15,1), N(210,7), N(9;1,2), N(60,5)]'; \\ P^t &= [N(9;0,3), N(220,10), N(28,1), N(350,10), N(6;0,1), N(100,3), N(14,1), N(240,5), N(9;1,1), N(70,7)]'; \\ Q^s &= [30,100,2000,5,800,40,600,15,200,5000]'; \end{aligned}$$

where  $N(\mu, \sigma)$  - denotes a random variable with a normal (Gaussian) distribution with a mean  $\mu$  and a standard deviation  $\sigma$ .

We calculate  $CI$ ,  $CI_{\min}$  and  $CI_{\text{symm}}$  and compare the confidence interval for the Laspeyres price index calculated from (24) in a simulation study. For this purpose, let us calculate firstly (see formulas (7) – (12)):

$$\mu_1 = 452620 \quad \sigma_1 = 35077.6,$$

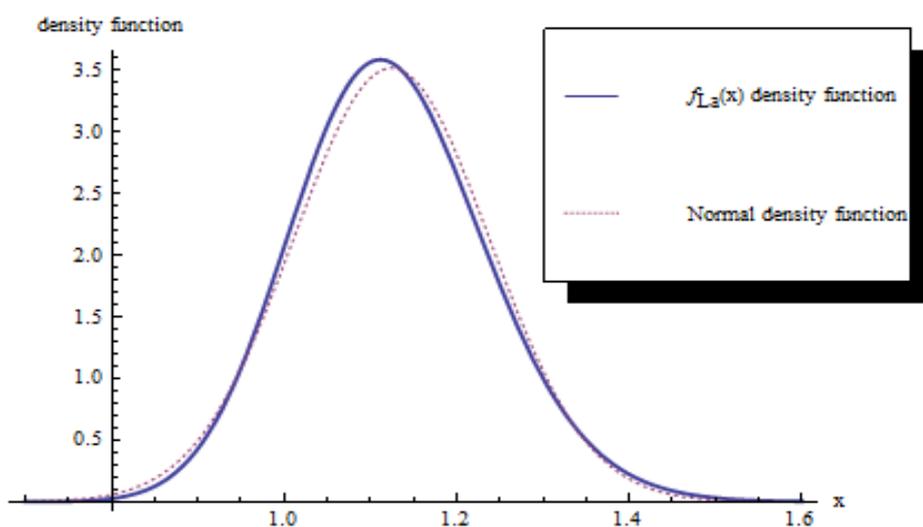
$$\mu_2 = 404090 \quad \sigma_2 = 25347.7,$$

$$a(x) = \sqrt{1.55641 \cdot 10^{-9} + 8.12719 \cdot 10^{-10} \cdot x^2},$$

$$b(x) = 628929 \cdot 10^{-4} + 367853 \cdot 10^{-4} \cdot x,$$

$$c(x) = \exp(-210.321 + \frac{0.5 \cdot (628929 \cdot 10^{-4} + 367853 \cdot 10^{-4} \cdot x)}{1.55641 \cdot 10^{-9} + 8.12719 \cdot 10^{-10} \cdot x^2}).$$

The density function  $f_{La}(x)$  and its normal approximation are presented in Figure 1.



**Fig. 1.** Density function  $f_{La}(x)$  and its normal approximation.

After subsequent calculations we obtain numerically the expected value of the  $I_{La}$  index and its standard deviation as follows:

$$\mu_{La} = \int_R x f_{La}(x) dx = 1.12431, \quad (27)$$

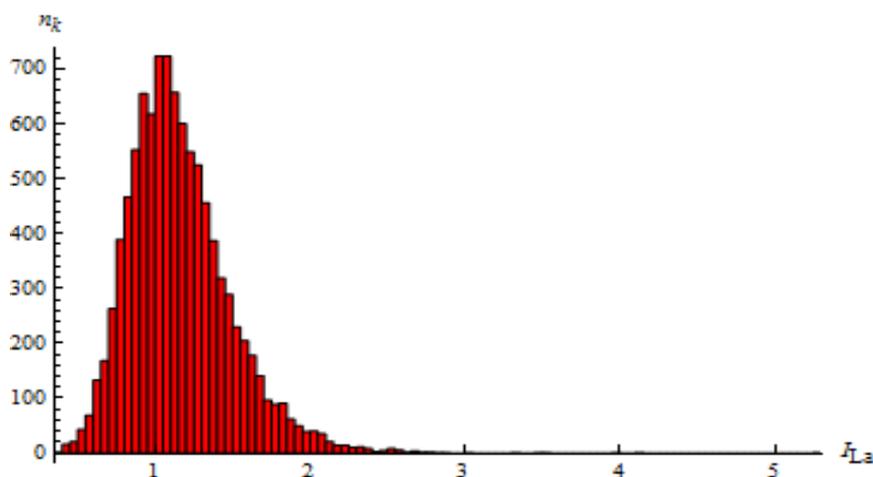
$$\sigma_{La} = \sqrt{\int_R (x - E(I_{La}))^2 f_{La}(x) dx} = 0.113494. \quad (28)$$

It is worth to be added that there are some approximations for  $\mu_{La}$  and  $\sigma_{La}$  in the literature. In (Hayya et al., 1975) we can meet second order Taylor approximations as follows:

$$\mu_{La} \approx (\mu_1 / \mu_2) + \sigma_2^2 \mu_1 / \mu_2^3, \quad (29)$$

$$\sigma_{La} = \sqrt{\sigma_2^2 \mu_1^2 / \mu_2^4 + \sigma_1^2 / \mu_2^2}. \quad (30)$$

In our case we get the approximated values:  $\mu_{La} \approx 1.1245$  and  $\sigma_{La} \approx 0.1116$ . If we assume that the distribution described by  $f_{La}(x)$  can be approximated by normal distribution (see Fig. 1) we can build the following  $(1-\alpha)$  confidence interval for the Laspeyres price index:  $CI_{Norm} = (\mu_{La} - u_{1-\alpha/2}\sigma_{La}, \mu_{La} + u_{1-\alpha/2}\sigma_{La})$ , where  $\phi(u_{1-\alpha/2}) = 1-\alpha/2$ . We generate values of vectors  $P^s$  and  $P^t$  in  $k = 10000$  repetitions and for each repetition the  $(1-\alpha)$  confidence interval, according to (24), is built. Taking means of confidence intervals' bounds we obtain the final confidence interval for  $I_{La}$ , here denoted by  $CI_{sim}$ . A histogram for the generated values of the Laspeyres index is presented in Figure 2 with its basic characteristics presented in Table 1. All results of confidence intervals obtained from different, discussed methods are in Table 2.



**Fig. 2.** Histogram for the generated values of the Laspeyres price index.

Parameter	Mean	Standard Deviation	Skewness	Median	Quartile Deviation
Value	1.2835	0.3575	$-5.61 \cdot 10^{-36}$	1.1215	0.1539

**Table 1** Basic characteristics for the simulated Laspeyres price index.

We can notice that taking into consideration median and quartile deviation from Tab.1 and thus ruling out some extreme values of the generated  $I_{La}$  index, we obtain results which are similar to calculated values of  $\mu_{La}$  and  $\sigma_{La}$ . Moreover, the calculated skewness virtually equals zero, thus the distribution presented in Fig. 2 is almost symmetrical. Nevertheless, the dispersion of the Laspeyres index calculated in our simulation study seems to be greater than

the dispersion obtained from numerical calculations based on the density function  $f_{La}(x)$ . It can be one of the reasons explaining the differences in confidence intervals presented in Tab. 2.

Confidence intervals	Confidence levels		
	0.9	0.95	0.99
$CI$	(0.947, 1.315) 0.368	(0.916, 1.358) 0.442	(0.857, 1.445) 0.588
$CI_{min}$	(0.948, 1.311) 0.363	(0.917, 1.356) 0.439	(0.860, 1.443) 0.583
$CI_{symm}$	(0.939, 1.309) 0.370	(0.903, 1.345) 0.442	(0.829, 1.419) 0.590
$CI_{Norm}$	(0.937, 1.310) 0.373	(0.902, 1.346) 0.444	(0.832, 1.416) 0.584
$CI_{sim}$	(1.020, 1.304) 0.2844	(0.987, 1.338) 0.351	(0.910, 1.414) 0.504

**Table 2** Confidence intervals for the Laspeyres price index with their length.

## Conclusions

In our simulation, when prices are not correlated, we observe that the confidence intervals obtained from the regression models  $CI_{sim}$  are shortest (see Tab. 2). However, we should remember that values of these confidence intervals are calculated as means of 10000 generated values. Moreover, the intervals  $CI_{sim}$  are not symmetric with regard  $\mu_{La}$  although the distribution of  $I_{La}$  seems to be symmetric (see Fig. 2 and Tab. 1). Taking into account the remaining confidence intervals we can notice that their lengths are similar for each confidence level. Having  $\mu_{La}$  and  $\sigma_{La}$  values (in practice the values can be estimated through the approximations) the calculation of  $CI_{Norm}$  is very quick but calculating  $CI_{min}$  or  $CI_{symm}$  still lasts long<sup>4</sup>. We should also add that  $CI$  and  $CI_{min}$  are slightly moved to the right in comparison with  $CI_{symm}$ .

<sup>4</sup> For example, the calculation of  $CI_{min}$  lasted over 45 minutes on our computer.

## References

- Balk, B. M. (1980). A Method for Constructing Price Indices for Seasonal Commodities. *Journal of the Royal Statistical Society A*, 143, 68-75.
- Bryan, M. F., & Cecchetti, S. G. (1993). The Consumer Price Index as a Measure of Inflation. *Economic Review, Federal Reserve Bank of Cleveland*, 29, 15-24.
- Clements, K. W., & Izan, H. Y. (1981). A Note on Estimating Divisia Index Numbers. *International Economics Review*, 22, 745-747.
- Clements, K. W., & Izan, H. Y. (1987). The Measurement of Inflation: A Stochastic Approach. *Journal of Business and Economic Statistics*, 5, 339-350.
- Clements, K. W., Izan, H. Y., & Selvanathan, E. A. (2006). Stochastic Index Numbers: A Review. *International Statistical Review*, 74(2), 235-270.
- De Haan, J. (2002). Generalised Fisher Price Indexes and the Use of Scanner Data in the Consumer Price Index (CPI). *Journal of Official Statistics*, 18(1), 61-85.
- Diewert, W. E., (1995). On the Stochastic Approach to Index Numbers. *Discussion Paper No. 05-31*, Department of Economics, University of British Columbia.
- Hayya, J., Armstrong, D., & Gressis, N. (1975). A note on the ratio of two normally distributed variables. *Management Science*, 21(11), 1338-1341.
- Hinkley, D. (1969). On the Ratio of Two Correlated Normal Random Variables. *Biometrika*, 56, 635-639.
- Marsaglia, G. (1965). Ratios of Normal Variables and Ratios of Sums of Uniform Variables. *Journal of the American Statistical Association*, 60, 193-204.
- Maraglia, G. (2004). Evaluating the Normal Distribution. *J. Stat. Softw.*, 11(4), 1-11.
- Marsaglia, G. (2006). Ratios of Normal Variables. *J. Stat. Softw.*, 16(4), 1-10.
- Selvanathan, E. A., & Prasada Rao, D. S. (1994). *Index Numbers: A Stochastic Approach*. Ann Arbor: The University of Michigan Press.
- Silver, M., & Heravi, S. (2007). Why elementary price index number formulas differ: Evidence on price dispersion. *Journal of Econometrics*, 140, 874-883.
- Von der Lippe, P. (2007). *Index Theory and Price Statistics*. Peter Lang, Frankfurt, Germany.
- White, A. G. (1999). Measurement Biases in Consumer Price Indexes. *International Statistical Review*, 67(3), 301-325.