Price Index Processor: CPI Compilation System

Consumer Price Index (CPI) Compilation System User Guide

Price Index Processor
Version IV: Consumer Price Index
(Updated and Revised Version)

Acknowledgements and disclaimer

The IMF Statistics Department (STA) and the Technology and General Services Department (TGS) developed the Price Index Processor Software (PIPS) to assist countries in improving their capabilities to process collected price observations and use them to compile price indices. Acknowledgements are due to Gangti Zhu for developing the software and Paul Armknecht for advising on the development of index number issues. The IMF has authorized the UNECE Statistical Division, with whom the IMF has no other affiliation, to distribute, modify, and maintain the software. While the IMF retains ownership rights to the original software, the IMF assumes no responsibility to users for support or maintenance. The IMF disclaims all liability for any errors that may exist in the software and for any other claims relating to the software.

This user guide is no more than a “guide” and **the user needs to become very familiar with the software by using a set of trial data, preferably from their own country, before adopting the software for use**. It is for the user to decide, based on such a trial, whether the software suits their needs. Neither the IMF nor UNECE Statistical Division are responsible for any errors of omission or commission in this documentation.

TABLE OF CONTENTS

Chapter

1

Introduction

A price index measures the proportionate, or percentage changes in a set of prices over time. Consumer Price Indices (CPIs) typically measure the changes in the price of a given basket of goods and services representing the purchases of a typical consumer. CPIs are used to measure the real purchasing power of money on various types of commodities and services. They are an important measure of the inflation experienced by households and changes in their cost of living. CPIs have become a key statistic for purposes of economic policy-making, especially monetary policy. It is often specified in legislation and in a wide variety of private contracts as the appropriate measure of inflation for purposes of adjusting payments, such as wages, rents, interests and social security benefits, for the effects of inflation.

Price Index Processor - Consumer Price Index System (CPI System) is an application software developed for assisting countries in compiling their CPIs. The system can also be used for teaching, training, and research purposes. Its main function is to compile CPIs for all items by product, by outlet,[1] and by geographic location.

The CPI System follows the guidelines of the *Manual – Consumer Price Index Manual: Theory and Practice (2004), International Labor Office, et al. URL:* http://www.ilo.org/public/english/bureau/stat/guides/cpi/, available in several languages, which describes the international standards for CPI compilation. These standards draw upon the collective experience and expertise accumulated in many countries.

[1] In "Resolution Concerning Consumer Price Indices, adopted by the Fourteenth International Conference of Labor Statisticians", "Outlet" is defined as a shop, market, service establishment, or other place, where goods and/or services are sold or provided to consumers for non-business use.

The CPI System calculates indices as weighted averages of the price changes for a specified set, or basket of consumer products, the weights reflecting their relative importance in household consumption in some period.

The CPI System consists of five parts: (1) *Data Entry, Editing; (2) Data Diagnosis and Validation*; (3) *Item Weights Creating, Editing and Distribution*; (4) *Index Imputation*; and (5) *Tabulation and Reporting*. They are integrated into a single database environment. The System works with COICOP-ICP - *Classification of Individual Consumption by Purpose (COICOP) - International Comparison Programme* (ICP) product classification. It allows the user to choose one of the two alternative formulas: Jevons (lower level)-Laspeyres (higher level) and Jevons(lower level)-Geometric Laspeyres(higher level) in combination with three basic compilation methods: compile by (a) product; (b) geographical region; and (c) outlet.

The System handles the consumption expenditure weights (typically obtained from household expenditure surveys) at any (COICOP-ICP) product classification) aggregation level. The detail of the levels of such a weights system can be mixed, catering to the different consumption patterns of different countries. The software has an embedded engine to distribute the item weights equally to the individual products and varieties by regions and by products that comprise them. Thus elementary “unweighted” price changes are ascribed equal weighting within each elementary index. The System also allows the user to revise the outlet, product and variety weights, after the item weights have been distributed, which will reflect the relative importance of the products and varieties. The System can calculate adjusted weights based on the outlet status and, if applicable, the probability group it belongs to, the result of which will be used in index imputation. If an outlet status is changed, the recalculation procedure can be triggered to re-distribute the item weights.

The weights inputted can be price updated to the reference period or can be actual weights. For the arithmetic option the resulting formulas would be a Lowe index (price-updated weights) and Young index (without price updating) should the weight reference period precede the base price reference period. For the geometric option the weights for a geometric Laspeyres, (also referred to as a Geometric Young or Cobb-Douglas index) should not be price updated. **Price updated weights should only be used with arithmetic aggregation.**

The System has flexible data input and output modules—the user can batch input and upload outlet information, item weights or price data using a spreadsheet and transfer them into the CPI system. Alternatively, the user can enter data into the system directly. Reports can be generated at different aggregation levels on an *ad hoc* basis or in time series format. All the compiled

results have a tabulation format and can be easily exported to Excel, HTML formats.

The System allows the user to define additional dimensions of product/variety in addition to COICOP-ICP and SPD (*Structured Product Description*) level. This variety list reflects the particular consumption pattern of a country and can be unique to a specific country. Each variety list has its own SPD and is closely integrated with COICOP-ICP and COICOP product classification systems. They are therefore internationally comparable, even though different countries will have different consumption patterns.

A geometric mean formula is used for the lower (elementary) level[2] calculation[3]. When used as an elementary aggregate, the geometric mean has superior axiomatic properties than alternative formulas. There is a choice between the Laspeyres or geometric Laspeyres formula for the upper level compilation of the index. Both versions take a "modified" or "two-stage" approach: the (Geo) Laspeyres formula is decomposed into three components: Short-Term Price Relatives ($STPR_t$) of the current period; Long-Term Price Relatives ($LTPR_{t-1}$) of the previous period, and Base Period Weight (W_0). The modified approach has several advantages over the standard formulation: it facilitates (i) the introduction of new varieties as soon as two successive price quotations are available; (ii) data verification; and (iii) more reasonable assumptions of similar short-run price changes for imputing missing values. The CPI is calculated using the equivalent of a recursive procedure, in which last period's cost weights (or base-weighted long-term price relatives, $p_{t-1}\, q_0$) are updated by the current period's price relatives (p_t/p_{t-1}) in order to obtain the corresponding cost weights ($p_t\, q_0$).

The price averages are calculated on the basis of "matched observations". Whenever a particular price observation is missing from either the previous month or the current month, the corresponding price observations are eliminated from the other period. This ensures a consistent sample of price quotations in each period. However, the system will be able to impute missing indices using next level data. Once the missing indices are estimated, the missing prices will be imputed and marked as red in the system.

The user can associate detailed meta data and notes (up to about 22 pages of an MS Word document) to an outlet, product/variety, or to a specific price observation. The CPI system has a flag to show whether the user has entered

---

[2] i.e. when information on weights is unavailable or assumption of equal weights used.

[3] The software only allows the Jevons index at this level. A ratio of arithmetic means (Dutot index) at the lower level may alternatively be used, but the CPI Manual recommends this only for strictly homogeneous goods and services. The arithmetic mean of price relatives (Carli index) is biased.

any meta data using different color coding. Each price observation can be flagged or un-flagged to indicate whether the price is imputed by the CPI system.

The user can also create a new outlet by cloning an existing one. Each outlet is associated with a label indicating its current status. The outlet status will determine whether the outlet will be included in the index calculation. This feature is important in conducting sample rotation and introducing new outlets. The user can also clone a variety, through which a quality adjustment is made possible. The software contains a facility for estimating the quality-adjusted base period price.

Both non-statistical checking and statistical checking routines are available for identifying the possible errors and outliers of input data. Non-statistical checking is implemented to check whether specific price data has increased more than a designated threshold, which is definable by the user. The recorded prices will be compared to the previous period's prices of the same items. While this procedure will detect unusual price changes, it is far from certain that all errors will be detected or that all unusual prices are errors.

Three statistical methods (or filters) have been implemented to detect the possible errors and outliers of short-term price relatives. The first one is based on Chebyshev's theorem that applies to all possible price relative distributions. This theorem predicts that at least 88.8 percent of all the observations in a data set will lie in the range of the mean plus or minus 3 standard deviations and at least 75 percent will fall within the mean plus or minus 2 standard deviations. Alternatively, observations with a z-score greater than 3 will be potential outliers. The second method is Box Plots, in which price relative data are re-sorted in ascending order, and the median and 1st and 3rd quartiles are calculated. Observations that fall outside of the outer quartiles are considered as possible outliers. The third method is to assume price relatives are log-normally distributed. Therefore, by transforming them into logarithmic form, the price relatives data will exhibit the normal distribution.[4] The intervals are calculated multiplicatively, and any price relatives that lie outside of the mean plus or minus 2 standard deviations will be identified as the possible outliers.

The System will impute price relatives for missing items and missing prices. If no price quotation is entered for any of the varieties covered by an item, its price relative is imputed using (geometric) average price relatives from the entire commodity group of the missing item. Missing prices of one or some varieties of a particular item are estimated as the previous period's price multiplied by the current period's price relative of that item.

---

[4] Price relatives of 1 (no price change) are excluded from the calculation.

Technical Overview

The CPI System is a Microsoft Visual Basic® application that runs in the Microsoft Windows environment. It works with Windows 2000, Windows Server 2003, Windows XP, Windows Vista, and Windows 7operating systems. The System stores both cross-sectional (e.g. outlets, products and varieties information) and time series data (e.g. price quotations and price indices) in a database. The System supports three database environments— Microsoft Access®, Microsoft SQL server® and Microsoft SQL Express® (formerly known as MSDE). (See the Appendix for the documentation on the transfer utility to convert an Access DB to a SQL DB for those users who already are using Access for CPI production.) The design makes use of a relational database architecture and object technology. A supplemental Guide is available for those users who choose an SQL DB from the start.

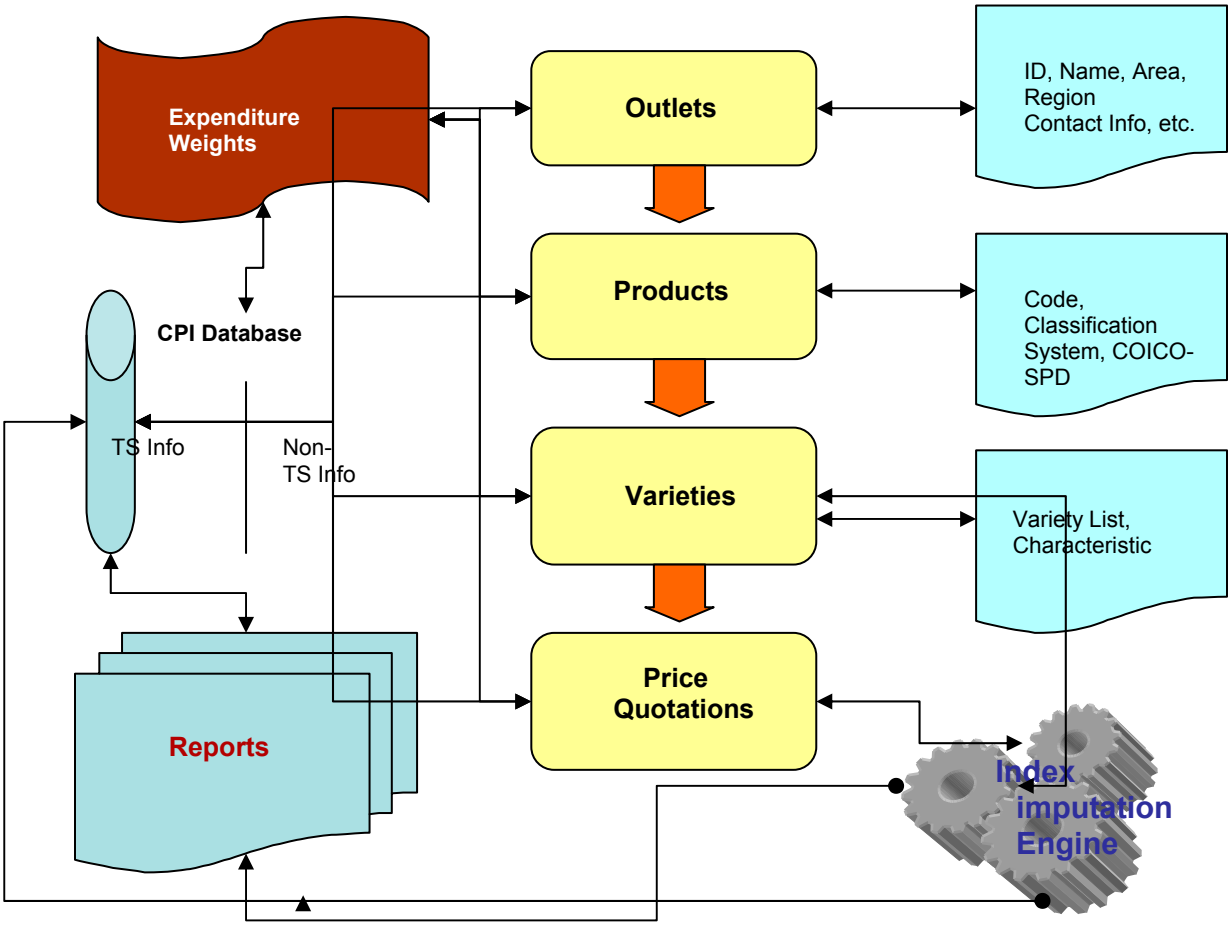
In the CPI System, several objects are generated representing outlet, product, variety and imputation. Each CPI System object contains a unique object name with its properties, attributes, and methods.

Depending on the situation, the CPI System engine creates links between the database where information is stored and an Excel spreadsheet where users can input, edit, and report information. Almost all data can be entered through either a Windows form screen or an Excel spreadsheet. The user can upload data in batch mode by using Excel macro-enabled spreadsheets generated by the system.

The CPI System brings together the familiarity of Excel with a powerful statistical and index compilation tool.

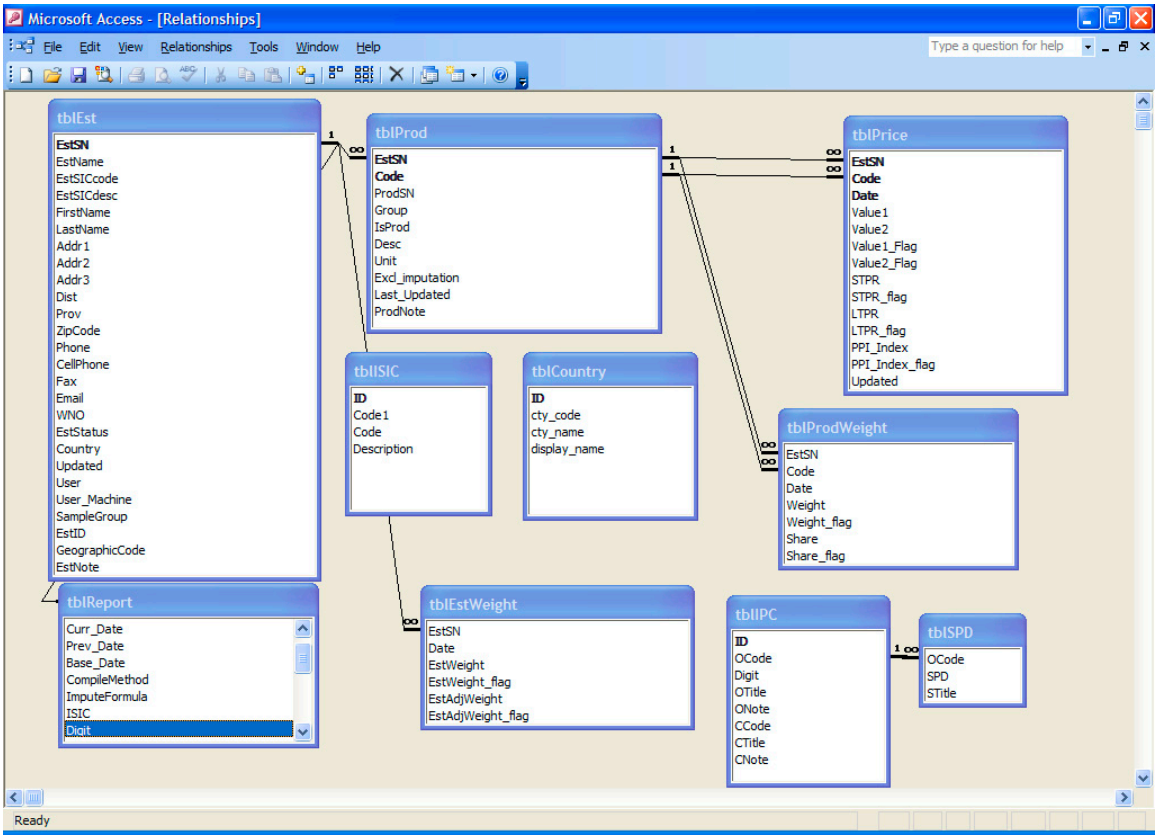
The CPI System Architecture

The diagram below shows the CPI System design and architect.



The CPI System Database Design

The CPI system allows the user to define, construct, and manipulate the underlying database using Microsoft ActiveX® Data Objects (ADO), the same type of interface and library used by Access to programmatically access data. the following diagram illustrates the database table design and relationships.



Chapter 2

Main Features

Outlet, Product, and Variety Model

The CPI System's design uses an outlet, product, and variety model. It stores the outlet information as the key element. Each outlet can carry multiple products; each product can have multiple varieties. The weights for outlets are derived form the weights for the products and varieties sampled within them. Such weights information is typically obtained from a household expenditure survey. The user can define and manipulate the weights. Every outlet is also classified into a specific geographic area for compilation purpose. Each outlet is associated with a list of properties including contact information for both the outlet and the data collector.

The user can easily add, edit, and delete an outlet. One can also input and edit the information on many outlets in a spreadsheet and upload it to the CPI database. The System also keeps track of the outlet's status and sample group information. The outlet's status indicates the current standing of the outlet as one of the following: (1) refuse to participate; (2) out of business; (3) could not locate; (4) initiated and reported; (5) resending for initiation; (6) no relevant product; (7) not yet initiated; and (8) unknown. Every outlet is classified into a sample group, selected either with certainty or with probability.

The user can quickly find a specific outlet by using the search feature. An outlet can be searched by its ID, name, keyword, area where it is located, or alphabetically.

Product and Variety Structure

The System allows the user to create a product list for each outlet by selecting from a COICOP-ICP product classification scheme, which is displayed in a hierarchical tree view structure, outlined below. The user can not only search for a particular product by keywords in the description field, but also search by

detailed notes associated with the product. The product list contains the aggregate 6-digit COICOP-ICP title and the 7-digit detailed structured product description (SPD) titles.

After the product is selected, the user can create a variety list of which is defined under this product. At this 8-digit level, the user has the freedom to create their own variety list, which can be specific to their country. The user is first given a list that has previously been created. If they find the variety that needs to be entered is in the list, they can select it. Alternatively, they can add a new variety that will be stored in the database and shared across the outlets for that product. In this way, the system provides an additional dimension for the user to reflect the country's own unique consumption pattern. At the same time, each variety is well integrated into a standard product classification system through COICOP.

Multiple Ways of Entering the Price Quotations

Once the products and varieties are defined, the user is ready for the price observation input. The system offers multiple ways for entering the price quotes. The user can enter the prices for (a) all varieties in one outlet for three periods (base, previous and current period); or (b) one variety for multiple periods in time series format; or (c) one variety across all outlets in one year. The user can also generate an Excel spreadsheet for price editing and inputting. This spreadsheet can be sent to a remote user for entering and validating prices and is reusable. Once validated data has been entered, it can be uploaded into the database.

For both time series and cross-outlet price entry, the system provides a threshold validation feature, in which the user can define a threshold value. The user defined threshold value will apply globally until the next time the program re-starts, which will reset to default value of 20%. When the price entered is greater or equal to the specified threshold, the PIPS shows that particular number in red to warn the user of a possible data entry error for the price quotation.

The system can also estimate the base period price of any particular variety using another calculated time series index as the deflator. The user can specify an imputed index series, select a particular period, and PIPS will use that index to estimate the base period price. This feature is very useful in the case when the user enters a new product or variety to the System where base period price is not observable.

Item Weights Design and Distribution

Information on item weights typically come from a household expenditure survey. The weights are used to reflect the relative importance of the goods and services as measured by their shares in the total consumption expenditure of households. The weight attached to each good or service determines the impact that its price change will have on the overall index.

The system requires weight data by area and by product. The user first has to determine the level of details of the weights that is available to the country, usually at the 6- (COICOP-ICP) or 7-digit (SPD) product classification level. Then the user has to select from a list the relevant items for that particular country. The PIPS will generate a spreadsheet template with different items/products in the rows and weight areas in the columns. The user has to key in the weight information in this matrix table and save it to the CPI database.

During the item weight distribution calculation, PIPS will distribute the item weight by area of a particular item/product to all varieties under this item equally. However, the user can modify the weights after distribution to specify the relative importance of varieties if necessary. Manually defined weights will replace/overwrite the auto-distributed weights and stored in the database for imputation.

Data Diagnosis and Error Checks

The System has a built-in module for the user to diagnose data errors and check for possible missing, critical variables. The System then generates a detailed report if it detects errors. By double clicking on the error item, the user goes directly to the screen where they can correct such errors. The user can also select the error item, right click to select the outlet, product, variety, weights or prices and then examine where the problem exists.

Data Status Indicator

The System has a graphical indicator to show the availability of all the outlet data for the compiling period. Indicators inform the compiler whether the price data are currently (1) available; (2) partially available; or (3) not yet ready. Data Status Indicator will also show outlet status (whether the outlet is initiated and reported or out of business, etc). By clicking on the outlet, the System will show the screen where the user can enter or update a price. This provides a useful and intuitive

tool for the statistical staff at the central office who can follow up any missing outlet price data and prepare for final compilation of the indices.

Product Classification, SPD Extension, and Varieties

The System uses COICOP-ICP as its default product classification scheme and displayed in hierarchical tree view structure for the user to select. The most detail level of COICOP-ICP has been extended by an additional level of SPD for details of structured products. The user can select a SPD which carries a code that combines the COICOP-ICP code and SPD extended code. The SPD is the parent code for the varieties, which can extend to one more SPD level. The user has complete freedom to build her own variety list with the constraint that each variety must have a parent that is a member of an SPD.

Calculating Average Prices

The CPI System has a separate module to calculate the average prices for the varieties across the outlets. The user has an option of using a geometric mean or arithmetic mean in such calculations. Since the Jevons index (geometric mean) is used for CPI compilation at the elementary level, the geometric option is recommended. Missing prices are imputed based on the average prices of identical varieties in the same unit of measurement. Such varieties are maintained in the variety list by the System.

Compilation Method and Elementary Formulae

The CPI System can compile price indices by (1) product; (2) geographic area; and (3) outlet . The System imputes missing price indices. The compilation results are stored in the database in time-series format.

Tabulation and Reporting

The System generates both *ad hoc* and time-series reports in a tabular format. *Short-term price relatives* (STPR), *long-term price relatives* (CPI indices), and *updated cost weights* are the three key indices produced by the System. It reports the most detailed indices for all components. The user can select the report detail at different aggregation levels (from level 1 to level 8). The reports can be exported to Microsoft Excel spreadsheets, HTML for web publication, and other formats.

Chapter 3

Methodology

Many countries use the standard Laspeyres-type arithmetic mean of price relatives to compile their consumer price indices. Although the term “Laspeyres” is often used to describe the formula, three points should be noted. First, Laspeyres requires that the weights reference period is the same as the price reference period, which is generally not the case. Weights may be from some prior survey period, say 2008, it taking some time to compile the weights for use with a price reference period of, say, January 2010. As noted below, the resulting index may more formally be a Young or Lowe index, depending on whether the weights are price-updated. The term “Laspeyres” or “Laspeyres-type” is used hereafter with this in mind. Second, the CPI is compiled in two stages, the elementary level using an equally-weighted geometric mean (Jevons) index and the weighted higher level using a Laspeyres or geometric Laspeyres.[5] Third, a modified/two-stage formulation will be used, as outlined below.

Elementary Index Formulas

The *CPI Manual* (Chapters 1 and 20) favors the use of geometric mean formula (Jevons index) on axiomatic grounds. A Jevons index formula is as followings:

$$(1) \quad P_J = \prod_{i=1}^{N} \left(\frac{p_i^t}{p_i^0} \right)^{1/N}$$

[5] Users should note that the distinction between the two levels in the software is dictated by the classification structure used, which is explained below. Aggregation at level 8 is the elementary level aggregation at which weights are equally distributed from some higher level, generally level 6 or 7. If the option is selected to use a geometric aggregator (Jevons) at the elementary level and an arithmetic aggregator (Laspeyres-type) at upper levels, the software will define the elementary level as level 8 and use Jevons and the higher levels as levels 7 to the overall CPI, and use a Laspeyres-type aggregator. The user should define products at different levels with their requirements for the use of these different formulas in mind.

The arithmetic mean of price relatives (Carli index) is biased, especially in a chained form, and the ratio of arithmetic means of prices (Dutot index) is only suitable for strictly homogeneous varieties.

The Standard Laspeyres-type Formula

The standard Laspeyres-type formula, applicable up to the most detailed level of weighted items in the CPI basket, compares the current period cost of the base period market basket (the numerator) with the cost of the base period market basket (the denominator) and can be written as:

$$(2) \quad I_{0\to t} = \frac{\sum_{i=1}^{N} q_i^0 p_i^t}{\sum_{i=1}^{N} q_i^0 p_i^t} = \sum_{i=1}^{N}\left[\frac{q_i^0 p_i^0}{\sum_{i=1}^{N} q_i^0 p_i^0} \times \left(\frac{p_i^t}{p_i^0}\right)\right]$$

where *i = 1, ... , n* stands for the products comprising the consumption basket and symbols *0,t* respectively designate the price reference period (or the base price period and the current price period). The symbols *p* and *q* designate the prices and the quantities of the products in question, respectively. The ratio ($p_{t,i}/p_{0,i}$) is the price relative to the base period for item *i* (sometimes called the long-term price relative).

By expressing the consumption expenditure's share for the item *i* as a ratio of the total expenditure during the base period as: [6]

$$(3) \quad w_i^0 = \frac{q_i^0 p_i^0}{\sum_{i=1}^{N} q_i^0 p_i^0}$$

Using the preceding expression formula (2) can be written in a slightly different form as:

$$(4) \quad I_{t\to 0} = \sum_{i=1}^{N} w_i^0 \left(\frac{p_i^t}{p_i^0}\right)$$

[6] The consumption expenditures may be adjusted for price changes occurring between the date of the household budget survey and that of the CPI rebasing, the first time these expenditures are used for the CPI compilation.

However, these versions of the Laspeyres formula do not provide the flexibility required for economies that are going through significant and rapid changes.

The Modified or Two-stage Laspeyres Approach

There are several reasons why the Modified Laspeyres Approach is superior to the standard formula. First, in the standard formula, we are comparing price relatives for the current period to the base period. In practice, the editing of the current period's price data is done by comparing the prices for the collection period for an item with those charged for the same item in the previous period. Any large variations falling outside predetermined range checks (e.g. 0.8000 to 1.1000) might indicate either the wrong item has been priced or some kind of error has been made in recording the price. With formula (4), this comparison cannot easily be made as it uses, for each item *i*, the price relatives of current period to the price reference period $\left(\frac{p_i^t}{p_i^0}\right)$.

Second, the standard formula involves a comparison of changes in prices for each item over long time periods, requiring the continuity of the priced item's specifications. In practice varieties become permanently missing or unrepresentative and need to be replaced with new varieties for which there is no price in the reference period 0 to compare with. In these circumstances, it is advisable to apply a modified version of the Laspeyres formula that makes use of the price relative to the previous period $\left(\frac{p_i^t}{p_i^{t-1}}\right)$ so that a new variety can be introduced as soon as two successive price quotes are available.

Third, when varieties are temporarily missing, imputed prices may be used based on the overall price change of the product group in question. Imputations over the short run are likely to be more reasonable than long-run ones.

The basic formula for computing the CPIs can be written as:

$$(5) \quad I_{0\to t} = \frac{\sum_{i=1}^{N}\left(\frac{p_i^t}{p_i^{t-1}}\right)\times q_i^0 p_i^{t-1}}{\sum_{i=1}^{N} q_i^0 p_i^0} \times 100$$

where $p_{t-1,i}q_{0,i} = q_i^0 p_i^0 \times \frac{p_i^1}{p_i^0} \times \frac{p_i^2}{p_i^1} \times \ldots \times \frac{p_i^{t-1}}{p_i^{t-2}}$

Formula (5), which is arithmetically equivalent to formulas (2) and (4), is considered more versatile than the formula using the long-term price relative to the base period, as the linking process used facilitates the introduction of new varieties and/or items or substitution when the need arises. It enables imputations that are more reasonable.

Formula (5) can also be rewritten as:

$$(6) \quad I_{0\to t} = \sum_{i=1}^{N} w_i^0 \times \left(\frac{p_i^t}{p_i^{t-1}}\right) \times \left(\frac{p_i^{t-1}}{p_i^0}\right)$$

which can be interpreted as:

$$(7) \quad I_{0\to t} = \sum_{i=1}^{N} w_i^{t-1} \times \frac{p_i^t}{p_i^{t-1}}$$

where $w_i^{t-1} = w_i^0 \times \frac{p_i^{t-1}}{p_i^0}$ is an updated weight sometimes referred to as a "cost weight" of item *i*.

In other words, to obtain the index for the current period *t*, the Modified Laspeyres Approach involves multiplying individual price relatives of the latest price compared period ($\frac{p_i^t}{p_i^{t-1}}$) by the previous period's updated weight (w_i^{t-1}), and then summing them.[7]

The Modified Laspeyres formula has obvious advantages over the standard Laspeyres formula when we consider the problems arising from permanently unobservable varieties, and the need in due course to bring in a new variety to replace the missing one. There is a need to impute a base period price if the

[7] Formula (6) can also be interpreted as:

$$I_{0\to t} = \sum_{i=1}^{n} w_{0,i} \times STPR_{t\to t-1,i} \times LTPR_{t-1\to 0,i}$$

where $STPR_{t\to t-1,i}$ is the short-term price relative of item *i* for current period (= p_i^t / p_i^{t-1}) and $LTPR_{t-1\to 0,i}$ is the long-term price relative of item *i* for previous period (= p_i^t / p_i^0).

standard Laspeyres formula is used. Such imputation is unnecessary while using the Modified Laspeyres formula, in which case the current period weight for the replacement item is obtained by simply multiplying the last updated weight for the replaced item by the current period's short-term price relative of the replacement item.

The system uses the modified Laspeyres approach to calculate the CPI based on monthly price quotations (or monthly average price quotations) and weights information. The price index is assigned a value of 100 in the base period and the value of the index for other periods of time, which indicate the average proportionate, or percentage, change in price levels.

Instead of holding expenditure weight reference period at 0, the CPI System allows the user to compile a CPI as a weighted geometric/arithmetic average of the individual price relatives holding constant the expenditure shares at period *b.*[8] The resulting index is a *Young* index. In this case, the formula is the following:

$$(8) \quad I_{0\to t} = \sum_{i=1}^{N} w_{b,i} \times \left(\frac{p_i^t}{p_i^{t-1}}\right) \times \left(\frac{p_i^{t-1}}{p_i^0}\right)$$

The weight reference period *b* is likely to precede the price reference period 0 because it takes time to collect and process the expenditure data. For example, a monthly CPI may run from January 2010 onwards, with January 2010 = 100, but the quantities may be derived from the annual expenditure survey made in, say, 2008 or June 2008 to July 2009.

In that case, we have the choice of assuming that *either* the quantities of period *b* remain constant *or* the expenditure shares in period *b* remain constant and equation (8) does the latter. A Lowe index holds quantities constant in period *b*, and the formula is as follows:

[8] As discussed in the *CPI Manual* (2004), "…any set of quantities could serve as the basket. The basket does not have to be restricted to the quantities purchased in one or other of the two periods compared, or indeed any actual period of time. … For practical reasons, the basket of quantities used for CPI purposes usually has to be based on a survey of household consumption expenditures conducted in an earlier period than either of the two periods whose prices are compared."

$$(9)\quad \frac{\sum_i p_i^b q_i^b \frac{p_i^o}{p_i^b} \frac{p_i^t}{p_i^0}}{\sum_i p_i^b q_i^b \frac{p_i^0}{p_i^b}} = \frac{\sum_i p_i^t q_i^b}{\sum_i p_i^0 q_i^b} = \sum_i w_i^{b,0} \frac{p_i^{t-1}}{p_i^0} \frac{p_i^t}{p_i^{t-10}}$$

Where $w_i^{b,0}$ are price-updated weights given by:

$$(10)\quad w_i^{b,0} = \frac{p_i^b q_i^b \frac{p_i^o}{p_i^b}}{\sum_i p_i^b q_i^b \frac{p_i^0}{p_i^b}}$$

Since the user is responsible for entering the weights, they can enter price-updated weights or weights without price updating as in (9) and (8), respectively.

The Geometric Laspeyres or Geometric Young Indices

In the geometric version of the modified Laspeyres index, a weighted geometric average is taken of the price relatives using the expenditure shares of period 0 as weights. It is defined as:

$$(11)\quad I_{0\to t} = \prod_{i=0}^{N} \left[\left(\frac{p_i^{t-1}}{p_i^0} \right) \times \left(\frac{p_i^t}{p_i^{t-1}} \right) \right]^{w_i^0}$$

Similarly, the geometric version of the *Young* Index if period *b* ≠ 0, that is, the expenditure shares are different from price reference period 0:

$$(12)\quad I_{0\to t} = \prod_{i=0}^{N} \left[\left(\frac{p_i^{t-1}}{p_i^0} \right) \times \left(\frac{p_i^t}{p_i^{t-1}} \right) \right]^{w_i^b}$$

Users should note that period *b* weights should not be price-updated to period 0 if using the Geometric formula, i.e. a geometric Young may be used but not Geometric Lowe.

Further, whether the index is a Geometric Laspeyres or Geometric Young depends on whether the user enters weights for period 0 or a preceding period *b*.

Whether the index is an arithmetic Laspeyres or Young or Lowe depends on whether the user enters weights for period 0 or a preceding period *b*, or price-updated weights from *b* to 0. The software uses the terminology “Laspeyres” of “Laspeyres-type” leaving it to the user to define the exact nature of the formula by virtue of the weights used. Countries generally use Laspeyres-type formulas at the higher level of aggregation with geometric means at the lower level. In spite of this, the geometric Laspeyres-type index has some advantages. Geometric means are preferred because they are:

(i) not as sensitive as arithmetic means to extreme values,
(ii) circular, i.e., fulfill a multi-period transitivity property that the product of the price index change going from a period 1 to a period 2 times the price index change going from period 2 to a period 3 should equal the price index going directly from period 1 to 3; and
(iii) more likely to lie between the Laspeyres and Paasche bounds, a desirable property.

Matched Price Observations

An average price is calculated in each of the geographical areas covered and for each variety comprising the CPI basket. The system allows a different number of areas and variety structure. The calculation of average prices would be simple if a set of price quotations were available for the current and previous month. In reality, this does not always happen. Quite often, some of the respondents are unable to quote a price for a particular variety because it is out of stock. Whenever a particular price observation is missing from either the previous month or the current month, the CPI system eliminates the corresponding price observations from the other period. This is equivalent to imputing the price of variety 1 in period *t* by the short-run price change of the other varieties in the product group. This ensures that the price averages are calculated based on “matched observations”, i.e., a consistent sample of price quotations in each period.

In the following example we consider that item’s prices are collected for four representative varieties *1*,*2*,*3*, and *4*. In the current month *variety 1*'s price cannot be collected (is missing).

Variety *v* of item *I*	Month *t-1*	Month *t*	Price relative in *t*
Variety 1	1.50	-	-
Variety 2	1.25	1.25	1.0000
Variety 3	1.25	1.50	1.2000
Variety 4	1.50	1.50	1.0000
Geometric average price of all varieties	**1.3693**	**1.4116**	
Geometric average price relative for matched observations (Variety 2, Variety 3, Variety 4)	**-**	**-**	**1.0627**
Geometric average of matched prices	**1.3283**	**1.4116**	
Short-term relative for item *i* using matched sample average prices	**1.0627 (= 1.4116/1.3283)**		

For the CPI calculation of month *t*, the geometric average price relative of 1.0627 should be calculated based on matched observations as $(1.0000*1.2000*1.0000)^{(1/3)}$ = ***1.0627***. We also get the same result by taking the ration of matched sample average prices (1.4116 / 1.3283 = 1.0627)

The month's short-term price relative for item *i* is then used to impute the missing price for Variety 1: 1.0627 x 1.5 = 1.594.

Impute Missing Indices & Prices

The CPI System will impute missing STPRs, LTPRs, and updated weights based on the information available from other price quotations of varieties in that commodity group. Imputed indices and prices will be stored in the database with a flag. These imputed indices and prices appear in the tabulation in **red** color.

Missing price indexes are estimated using its parent index as the proxy, i.e., if a specific variety's index is missing due to the missing prices, the index of the product or item it belongs to will be taken to be representative. The system always uses the next available level index data in the same group or item for the missing index, e.g., if the level 7 index is missing, the level 6 index will be used; if level 6 is missing, then the level 5 index will be used and so on. Holding missing prices for a variety constant by carrying the last observation forward (i.e., making the short-term price relative for that variety equal to 1.0) during a period of high inflation would cause short-term distortion in the index. This would understate inflation while the variety was unavailable and then show a large increase in the index when the variety became available. *The system therefore does not use carry-forward prices.* We do not use the price for the same variety in another outlet to represent the missing price in this outlet. Thus if a Coca-Cola price is

missing in one particular outlet, the system will not take the other outlet's Coca-Cola price as a proxy, instead it will take soft drink index in the same outlet to represent that of Coca-Cola. For some products such as tuition, insurance, and rents the prices change very infrequently and the sample of observations is only collected periodically. In these situations the user should enter the last available price in the system as if it were actually collected rather than have the system impute a price.

Since price changes for the parent group are always calculated as geometric mean changes, imputations are based on geometric means.

Missing prices of one or some varieties then are estimated by multiplying the previous period's price by the current period's short-term price relative of that variety, which in turn was estimated using the index of the item/group. If the previous period's price is not available, the System estimates the missing price by multiplying the reference/base period price by the LTPRs. If both previous period price *and* base period price are not available, the missing price cannot be imputed.

If no price is collected for any variety covered by a product (the prices for whole product is missing), Its price relatives will be imputed using the geometric average price relatives from the item group of the missing price.

Detection of Outliers

What is an outlier?

An observation that is unusually large or small relative to the other values in a price relative data set is called an outlier. Outliers are the observations that appear to be inconsistent with the remainder of the collected data.

There are several possible sources for outliers:

1. The price quotation of a transaction or variety is observed, recorded, or entered into the computer incorrectly.

2. The price quotation come from a different population, or quality of that transaction/variety has been changed.

3. The price quotation entered is correct, but represents a rare event or novel phenomenon.

Outliers occur when the relative frequency distribution of the data set is extremely skewed. Such distributions have a tendency to include extremely large or small observations.

The PPI/CPI software implemented two types of procedures to identify the possible errors and outliers. The first one is a non-statistical procedure, which is to find whether specific price observation falls outside some pre-specified acceptance interval. In the “Input Price in Time Series Format” screen, the user can specify a threshold value (default is 20%), for any price change is greater than the threshold, the increased percentage will be shown in red and bold which indicates the possibility of errors or outliers. The second set provides three statistical procedures to use. In all cases, outlier detection should *not* result in automatic deletion. Often price changes are undertaken after some time and the “pent-up” price changes are unusually large. To delete them would bias the index downward. The outlier detection is to alert the compiler about a possible error that needs further investigation.

Z-score Method:

In a z-score test, the mean and standard deviation of the entire data set are used to obtain a z score for each data point, according to following formula:

$$Z_i = \frac{(x_i - \bar{x})}{s}$$

Where

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

If the observations have a bell shaped distribution (standard normal distribution), the interval from $\bar{x} - s$ to $\bar{x} + s$ will contain approximately 68% of the measurements. If the interval from $\bar{x} - 2s$ to $\bar{x} + 2s$ will contain approximately 95% of the measurements, and the interval from $\bar{x} - 3s$ to $\bar{x} + 3s$ will contain approximately all of the measurements.

In the case of price relatives, we do not know the underlying distribution of the data set. Many studies of price change show that price relatives are not normally distributed. Thus, we appeal to Chebyshev’s theorem that applies to all possible distributions. According to Chebyshev’s theorem, for any set of measurements and any number $k \geq 1$, *the* interval from $\bar{x} - s$ to $\bar{x} + s$ will contain at least (1 - 1/*k*2)*100 percent of the measurements.

Thus, at least 88.8 percent of all the observations in a data set will have a Z score of less than 3 in absolute value, i.e., it will fall into the interval $(\bar{x}-3s, \bar{x}+3s)$, and at least 75 percent will fall within 2 standard deviations, where $\bar{x}$ is the mean and *s* is the standard deviation of the sample. Therefore, the observations with a Z score greater than 3 will be potential outliers.

Example

The short-term price relatives measured by dividing current period prices by previous period prices are recorded in the following table.

100	98	103	**121**	104	102
101	102	101	102	105	105
103	110	100	104	106	102
85	102	101	103	110	101
101	104	105	106	107	100

Sample Data for Short-Term Price Relatives

For this data set, $\bar{x}$ =103.1333, s= 5.4818, 3s=16.4454. The Z score of the observation of 121 is: (121-103.1333)/5.4818=3.2593.

The Z score of 85 is (85-103.1333)/5.4818 = -3.3079.

Since the absolute values of Z score of 121 and 85 are more than 3, these price relatives are outliers in the data set.

The Z-score method is biased in that the outliers affect both the mean and standard deviation.

Box Plots Method

Another procedure for detecting outliers is to construct box plots of the price relatives data. They make no distributional assumptions and, since they rely on the median and quartiles as parameters, the outliers themselves do not influence the method of detection. Below are the steps implemented in constructing the box plots for the software.

- The median M, lower and upper quartiles, QL and QU, and the inter-quartile range, IQR= QU - QL are calculated for the data set.

- The limits on the box plot are constructed with fences located a distance at QL and QU;

Suspect outliers are the observations below QL and above QU..

Locate the suspect outliers on the box plot using asterisks (*).Observations that fall outside the outer fences are called highly suspect outliers.

How the Quartile is Calculated?

The Quartile calculation depends on the percentile definitions. The First quartile is the 25th percentile (noted Q1), the Median value is the 50th percentile (noted Median), and the Third quartile is the 75th percentile (noted Q3). The method to calculate the quartiles in CPI/PPI application is same as that used in Excel. It uses n-1 instead of n.

The p-th percentile is defined by:

y = (1-g)* x(j+1) + g*x(j+2) where (n-1)*p=j+g (and x(0) is taken to be x(1)).

Let n be the number of observations in a data set (here n=4), and X(1)…X(n) the ordered values of a data set. Let p be the p-th percentile we want to calculate (e.g. p=0.25, 0.5, or 0.75). We'll calculate the product n*p; the product n*p can be split up between j and g, where j is the integer part of n*p and g is the decimal part of n*p.

Example

To better understand this method, we'll apply them on an simple example.
The data set studied is:

Variable	X1	X2	X3	X4
Value	2	1	4	3

Once ordered it becomes:

Variable	X1	X2	X3	X4
Value	1	2	3	4

In this example, for Q1, p=0.25, n=4, (n-1)*p=3*0.25=0.75*X(2) = 0.25*1+0.75*2 = 1.75, thus the 25th percentile is 1.75 with this method.

Log-normal Method

Another procedure that has been implemented into the software to identify the possible errors and outliers is to use a 2 standard deviation upper and lower limit and assume a log-normal distribution of the price relatives after excluding price relatives of 100 (no change of prices from the previous to current period).[9] The standard deviation and mean of the natural logarithms of the remaining price relatives in the sample are calculated. Those price relatives that fall outside of 2 standard deviations (with 95% confidence level) are considered as possible outliers.

Calculating Adjusted Weights

If sampling outlets so that some are selected with certainty, say as a cut-off sample, and some are selected to be representative of the remaining outlets the weights of each outlet in the latter need to be adjusted. Further, outlets may disappear from the active sample and it may be necessary to redistribute the weight across the active sample. These two effects are picked up in an adjustment routine for the weights. The outlet-adjusted weight is calculated based on the sample group to which the outlet is classified. For the sample group selected with certainty, an outlet-adjusted weight is equal to the outlet assigned weight. They only represent themselves. However, if outlet(s) disappear from the sample, though still sell goods and services in reality, the weights of the remaining establishments can be adjusted so that those still active get allocated a *prorata* share of those that are "inactive" within its sample segment. If there were 5 selects initially and they each had a value weight of 20 and one disappears and 4 remain, each would get a weight of 25. An outlet adjusted value weight is equal to an outlet assigned value weight (20) divided by the total value weight of outlets that are active in the product group (80), then times the total value weight of outlets (both active and inactive outlets) of the product group (100) i.e.

$$W_i^{adj} = \frac{W_i^{Assigned}}{\sum_i^n W_i^{active}} \times \sum_i W_i^{Total}$$

in which W_i^{adj} is adjusted weight of an outlet *i*; $W_i^{Assigned}$ is assigned weight of an outlet *i*, $\sum_{i=1}^{n} W_i^{active}$ is the total weight of active outlets and $\sum_{i=1}^{n} W_i^{Total}$ is the total

[9] The reason to exclude those price relative with the value of 100 is that, presumably, there are many "no change" price relatives. By including them, we will observe a bimodal distribution. Since it is no longer normally distributed, we cannot apply a 2-sigma limit with 95% confidence to detect outliers.

weights of all the outlets. Thus the System will take and redistribute the weight of the inactive outlets to the active outlet based on its share in total active weights.

In the probability-selected group, the adjusted weight for an outlet being selected with probability to represent others is:

$$\frac{\bar{W}^{Active}}{\sum_i W_i^{Active}} \times \sum_i W_i^{Total} = \frac{1}{n} \sum_i W_i^{Total}$$ and the adjusted weight to represent outlets no longer active is:

$$W_i^{adj} = \frac{\frac{1}{n} \sum_i W_i^{Total}}{\sum_i W_i^{active}} \times \sum_i W_i^{Total}$$

Hierarchical Structure of the Product Classification

The hierarchical levels of the product classification used in the CPI System are as follows, moving from the general to the more detailed level. The structure is that used for the 2003-2006 ICP round. The basis of this structure and its relationship to COICOP appears in the ICP Handbook, Chapter 5.[10] Examples of the six digit "basic headings" code are *Fresh milk* 11.01.14.1; *Shoes and other footwear* 11.03.21.1, *Bread* 11.01.11.3, and *Rice* 11.01.11.1. This is the level at which weights are generally available. An example of the structure is given below. Note that level 7 are "product clusters" and may be more detailed than level 6, for example, *Bread* at level 6 may be defined as *White Bread* and *Bread Other Than White* at level 7. However, as in the example below for Rice, levels 6 and 7 may be the same. The main distinction between level 6 and 7 in such cases is that the latter contains the structured product descriptions (SPDs) – a listing of possible characteristics that can be used to precisely define the items. Level 8 is the result of including the precise characteristics of the varieties of individual products for which prices are to be collected in the SPDs. There may be more than one defined product/variety specification at level 8. The compiler can adjust this assortment to reflect changes in product supply and consumer behavior.

---

[10] Available at: http://web.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/ICPEXT/0,,contentMDK:20962711~menuPK:2666036~pagePK:60002244~piPK:62002388~theSitePK:270065,00.html. See also www.unece.org/stats/documents/ece/ces/ge.22/2006/mtg2/12.e.pdf

Weights should be entered at level 6 or a more detailed level where possible and be derived from actual data.

Level	COICOP-ICP Code	Description
0	0	All Products
1	1	GROSS DOMESTIC PRODUCT
2	11	INDIVIDUAL CONSUMPTION EXPENDITURE BY HOUSEHOLDS
3	11.01	FOOD AND NON-ALCOHOLIC BEVERAGES
4	11.01.1	FOOD
5	11.01.11	Bread and cereals
6	11.01.11.1	Rice
7	11.01.11.1_01	Rice/RICE
8	11.01.11.1_01a	Imported Uncle Ben's long-grain rice 1lb box in supermarket
8	11.01.11.1_01c	Domestic brown medium-grain rice 1lb sold loose in open market

Chapter

4

Installation of the Software

The Basic Information

The CPI System was developed using Microsoft Visual Basic ® 6.0. Both cross-section data and time-series data are stored in Microsoft Access format. The system uses Microsoft ActiveX Data Objects (ADO 2.8) for the data access and database management that is included in Microsoft Data Access Components (MDAC). The system uses VideoSoft ActiveX Controls® VSFlexGrid® 7.0 and Formula One® for the tabulation, Wise InstallBuilder® 8.03 for software packaging and RoboHelp® for generating standard help file.

Package Contents

The installation package on the CPI distribution web page includes all necessary files and ActiveX components for the system to operate. The MDAC 2.8 setup file is also included in the installation package and will be installed into the user machine.

The package contains standard COICOP-ICP and COICOP product classification together with an SPD list.

System Requirements

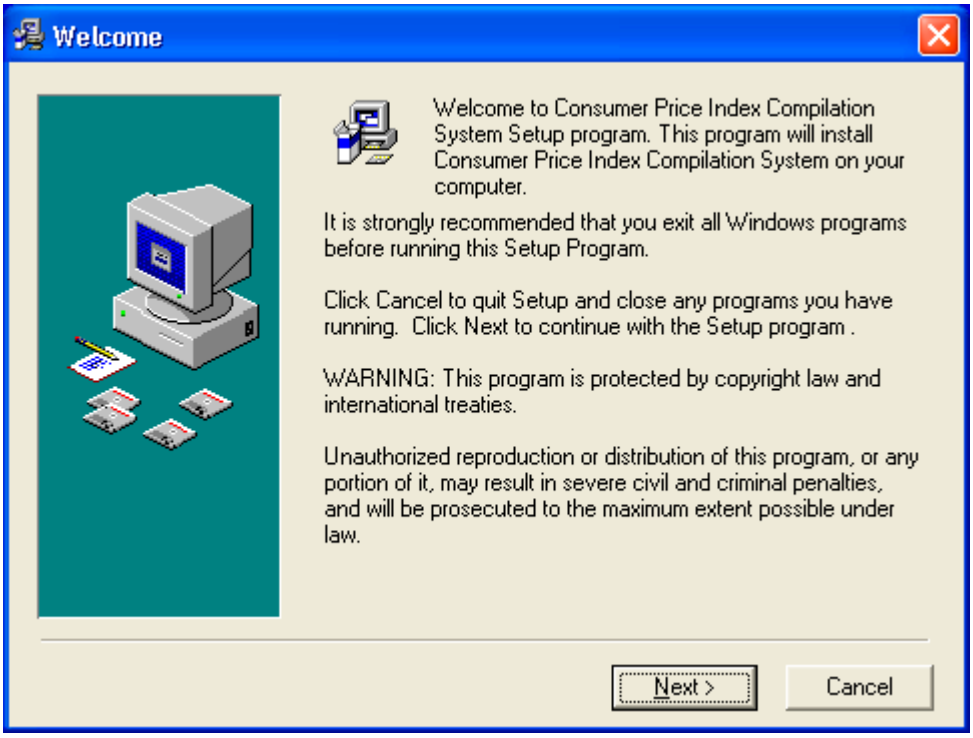
The system requirements are as follows:

a) Operating Systems: Windows 98, Windows NT 4.0, Windows 2000, Windows ME, Windows XP, Windows Vista, or Windows 7.
b) Microsoft Office 97 or above.
c) CPU: Pentium-600 MHz or higher
d) 512 MB RAM or more
e) 60 MB free disk space
f) VGA–True Color video mode, displays at 800 x 600 or 1024 x 768 pixels.

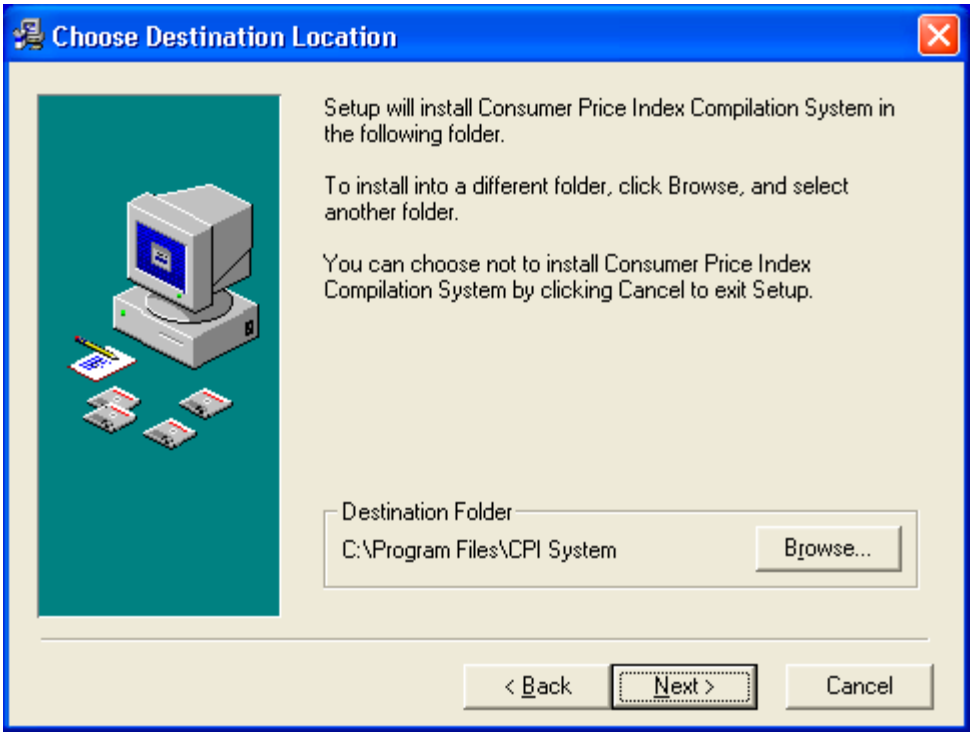
The installation package for the CPI has all necessary files, including ActiveX components for the system to operate and MDAC 2.8 setup file.

Installation Procedures

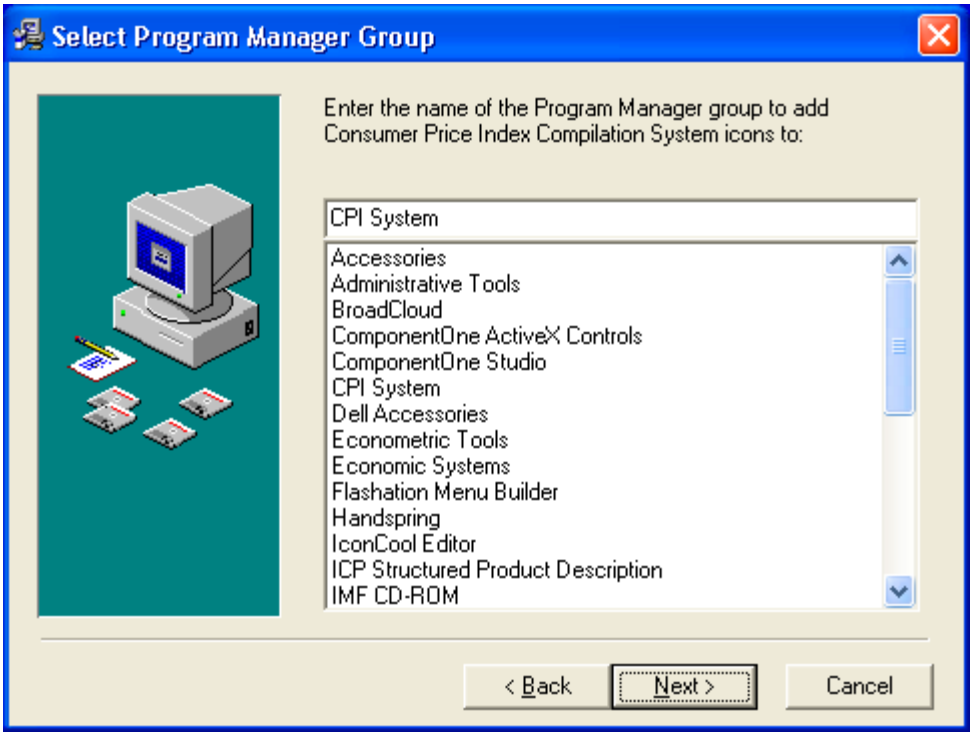
Go to the UNECE website, which has the following URL, (http://www.unece.org/stats/downloads/SW_CPI_PPI/pips.html) and click on "Download the CPI Compilation System" and when the download box appears, select "run". The download will take some time depending on the download speed of your Internet service provider. (You are downloading 42MB.) After the download, a second box may appear noting that the file is from an unknown publisher; also click on "run". You may be asked to use administrator rights at this stage by your local network. The user will see the following screen:



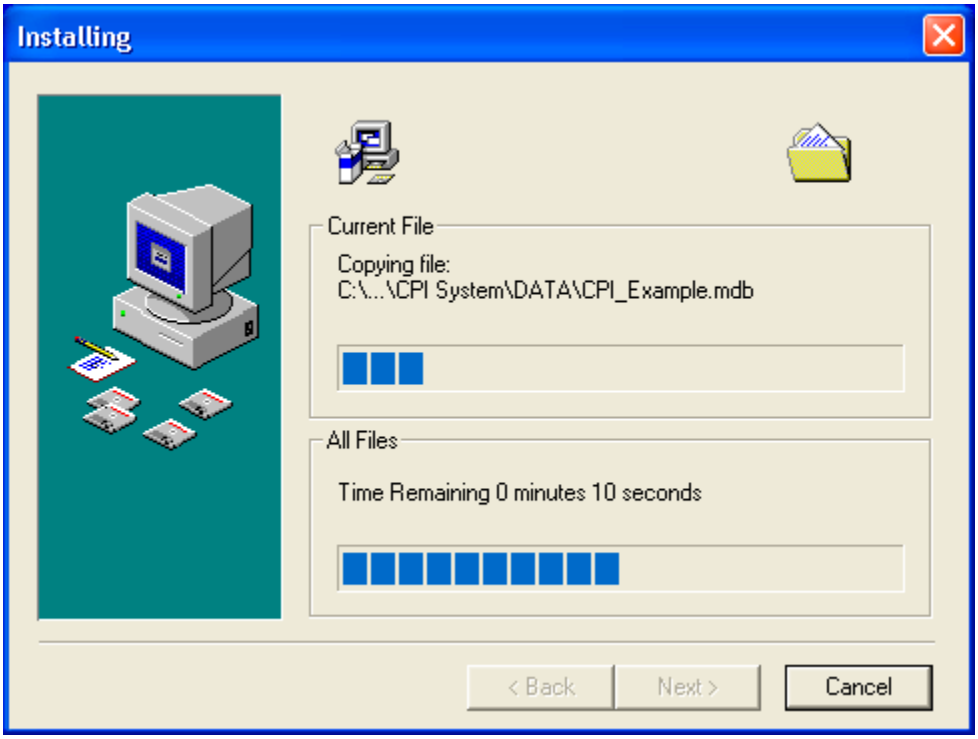
Click **Next** and go to the following screen.



Click **Next** if this is the target folder**. If your corporate setup does not allow users to read and write to files in the Program Files folder, you need to select the "Browse" button and establish a different file location for the program. Consider creating a "C:\users\name\PIPS\" folder to store the CPI system.**



Click **Next.** Wait until the installation process is complete.



Click Cancel to exit.

Click the **Finish** button to complete the installation process.
The Installation procedure will create a CPI System shortcut icon on the Desktop and create an icon in the Program Manager Group.

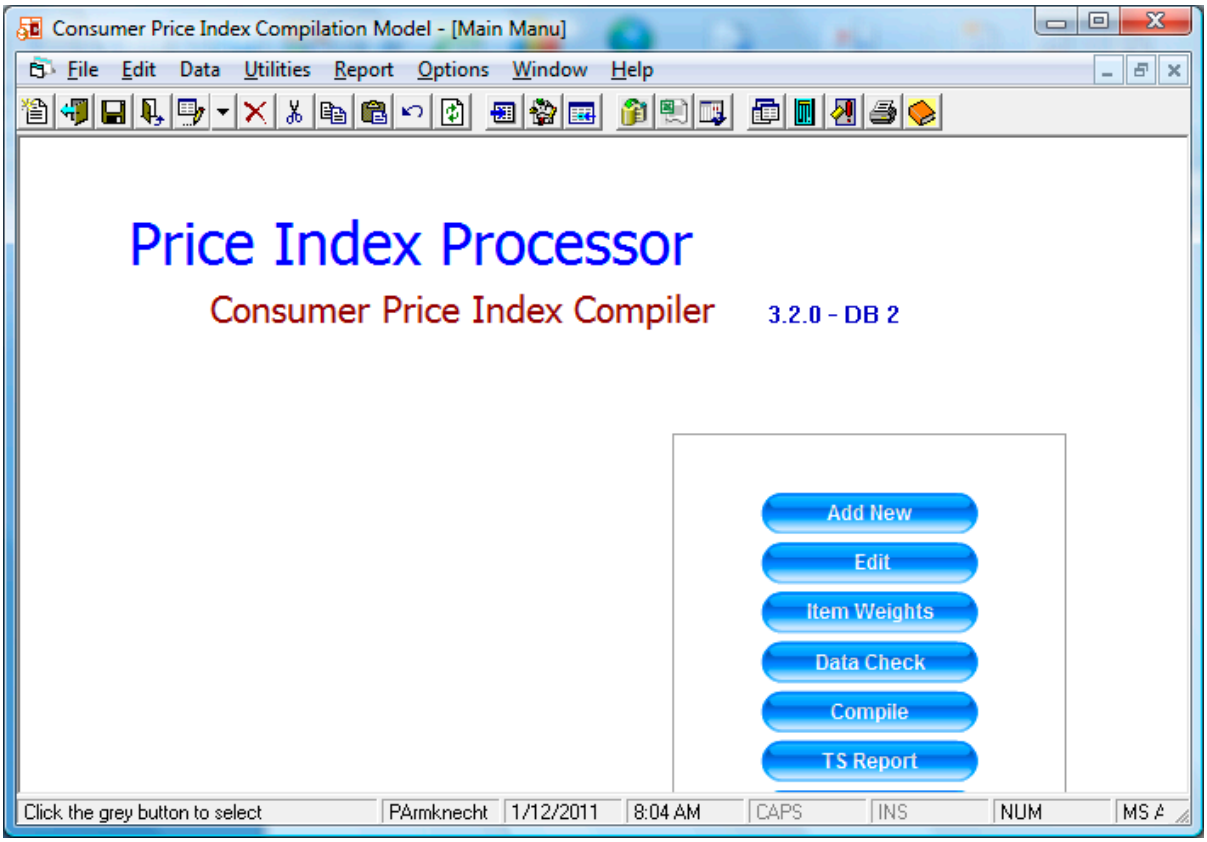
To convert an existing PIPS ACCESS DB to SQL, follow the instructions in Appendix I: Transfer Utility for Converting Databases between Access and SQL.

To setup the PIPS DB on an SQL server, follow the instructions in Appendix II: Enable Remote Connections on SQL Server.

Chapter 5

The Basics

Click **Start**, **All Programs**, **CPI System,** then click  icon or click CPI System icon from the desktop to start the CPI program.



“3.2.0” is the application version number and “DB2” is the database version.

System Configuration

The user can configure the CPI System work environment by clicking System Config button in the main screen or clicking Options, Configure in the toolbar. The user first must identify the location of the CPI DB source:

Consumer Price Index Compilation Model - [Configure CPI Working Space]

File Edit Data Utilities Report Options Window Help

Configure Default Database Provider

MS Access | MS SQL Server | DB Status: On Line | Test Connection | Configure DataSource | Create Access DB

Connection String: Provider=Microsoft.Jet.OLEDB.4.0;Data Source=C:\Users\parmknecht\PIPS_3_3_1\CPI System\DATA\Antigua.mdb;Persist Security Info=False

Configure CPI Working Environment

Defalt CPI Compilation Country: Antigua and Barbuda

Default Database File Location: C:\Users\parmknecht\PIPS_3_3_1\CPI System\DATA\

Default CPI Database (e.g., CPI.MDB): Antigua.mdb

Create New CPI Database: (e.g. newppi.mdb) | Set it as Default DB

Change the weight date 1/2006 to 1/2006

Set Compilation Date | DB Repair + Compact | Update DB Table

Default Imputation Method: By Product, By Region, By Outlet

Default Estimation Formula: Jevons:Laspeyres, Jevons:Geo-Laspeyres, Dutot:Laspeyres, Fisher Index

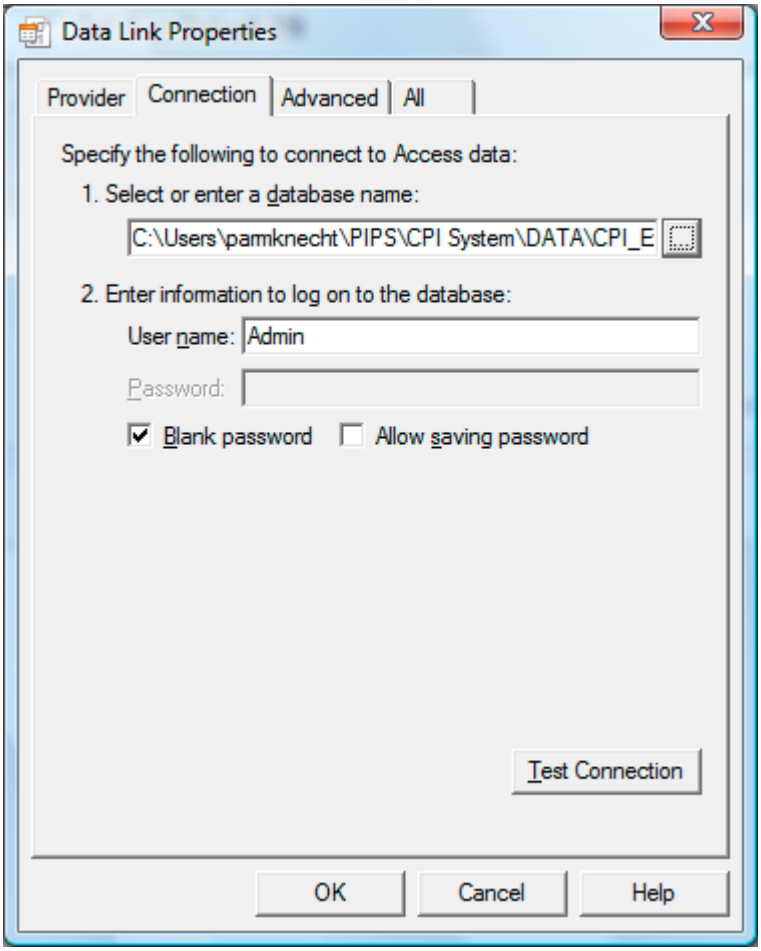
Other Settings: Include Imputed Index and Price, Two-Stage Index, Two-Step Imputation, Apply Constant Weight Structure

Apply | Refresh | Cancel

1) Click on the Configure Data Source button to open the data links screen

2) Select the default database location using the browse feature or enter identifying string directly in box 1.;

3) Leave the information in box 2 as shown for the defaults unless changed by your IT staff. Select OK



4) At the first application the user will have to “Create New CPI database” by entering a name is this field such as “countryname.mdb”, click on the “set it as default DB” box, and then click the “Create Access DB” button at the upper right of the screen near the tool bar. During later usage, select the DB from among those saved in the CPI Systems Data folder.

5) Enter a country name in the “Default CPI Country Compilation” box from the drop-down menu.

6) Enter the “weight date” for the weight reference period

7) Select the default imputation method in the lower left screen (By Product is standard method).

8) Select the default formulas that to use—Laspeyres is arithmetic aggregation and Geo-Laspeyres is geometric aggregation. (Consult Chapter 3 Methodology for discussions of the differences.)

9) Select the other settings—all of these boxes should be checked if one uses the two-stage (short-term price relative, or chained) calculation method.

In the Other Settings box, the "**Include imputed prices**" allow imputed prices from the previous period to be used to calculate a short-term price relative. The "**Two-stage**" refers to using the short-term relative method of calculation (vs. the long-term relative from the base month). The "**Two-step Imputation**" has to be used with the two-stage calculations for making imputations: during the first pass preliminary imputations are made at the variety and outlet level, at the second stage imputations are made at higher levels. The "**Apply Constant Weight Structure**" uses the fixed aggregation weights to the imputed indexes rather than just the sample weights. This approach should be used all the time.

Click **Apply** to save the changes.. Click **Refresh** to view the screen with the recent changes.

Click **DB Repair + Compact** button to repair possible database error and to compact the Access database. This should be invoked if the user gets and error message during compilation of the index stating that the compilation array will create duplicate entries. In this case the DB sequencing has gotten out of proper sequencing and needs to be fixed.

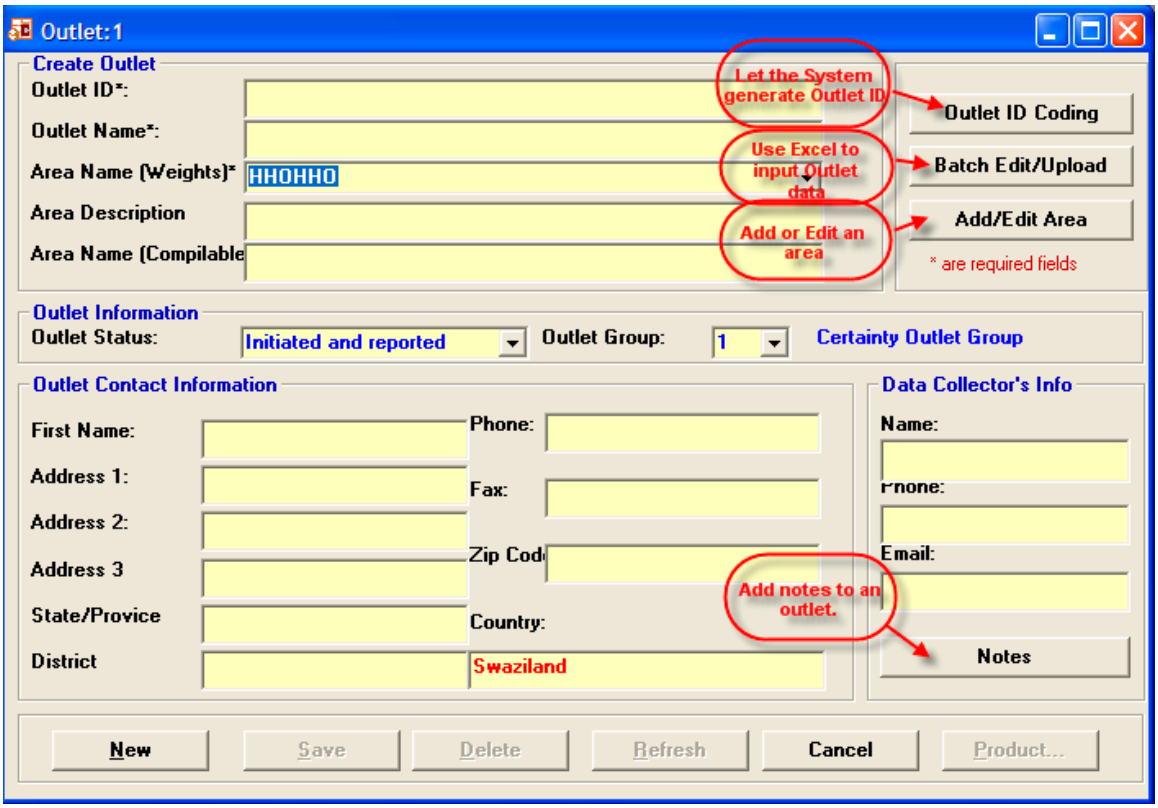
Click **Set Compilation Date** to specify the compilation date. (You are compiling the index for this month; however, it does not need to be set until all data have been entered for the outlets, products, varieties, and prices.) The base price date can be set on the compile screen.

This provides the initial set up for using PIPS. Now begin the process of entering the data for the CPI outlet sample.

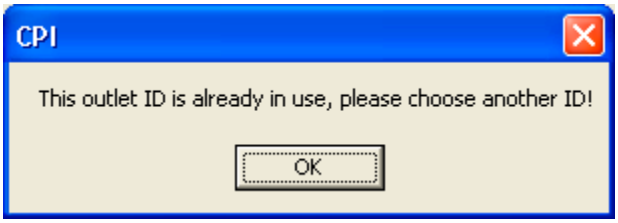
Add New Outlet

Go to the Main Menu. Click the 

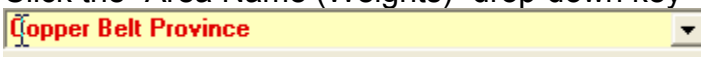
button to add a new outlet. Enter the outlet information in each field. Fields with * indicates a required field.



Outlet ID is unique field. If the outlet ID is already taken, the following message will be displayed:



Click the “Area Name (Weights)” drop-down key



to select an area. If no area has been defined,

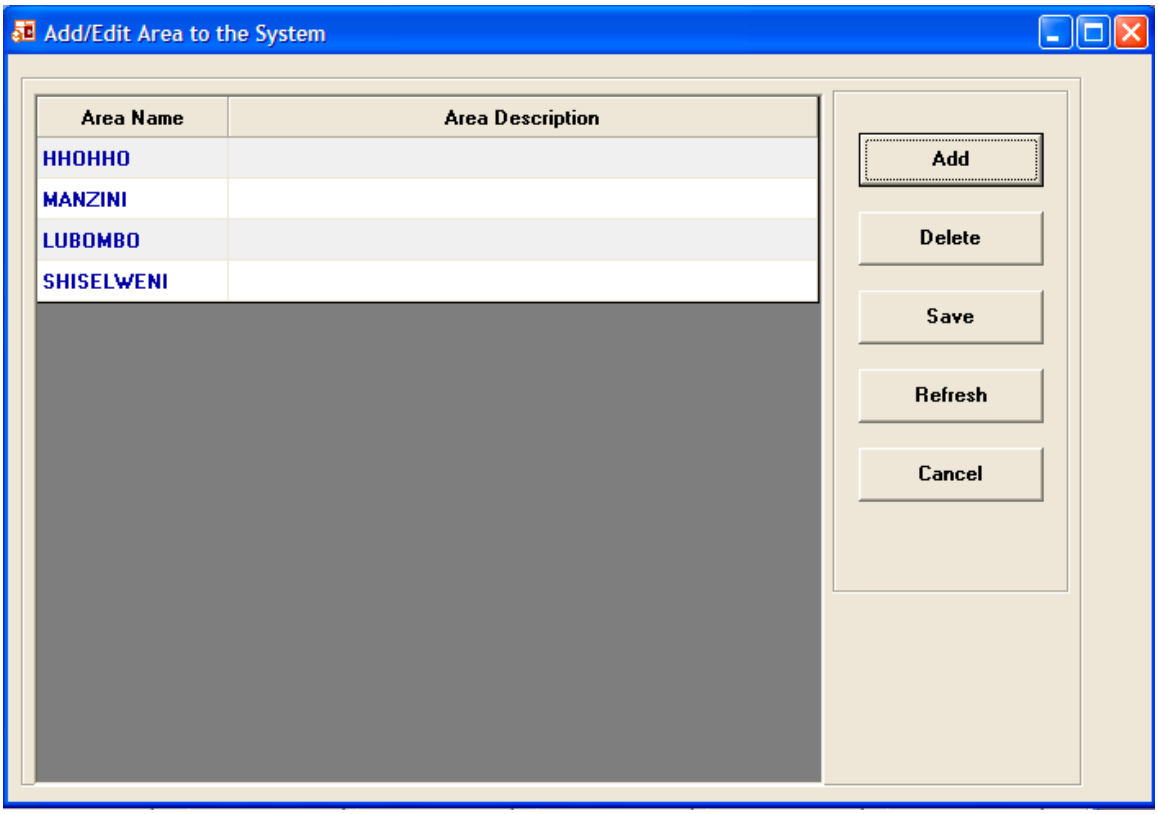
will be displayed here.

Select the “Add/Edit Area” button on the right of the screen to open a table for adding the area names and their description. If you compile the CPI by geographic area or

region, the area name must appear here as well as in the “Area Name (Compilable)” field.

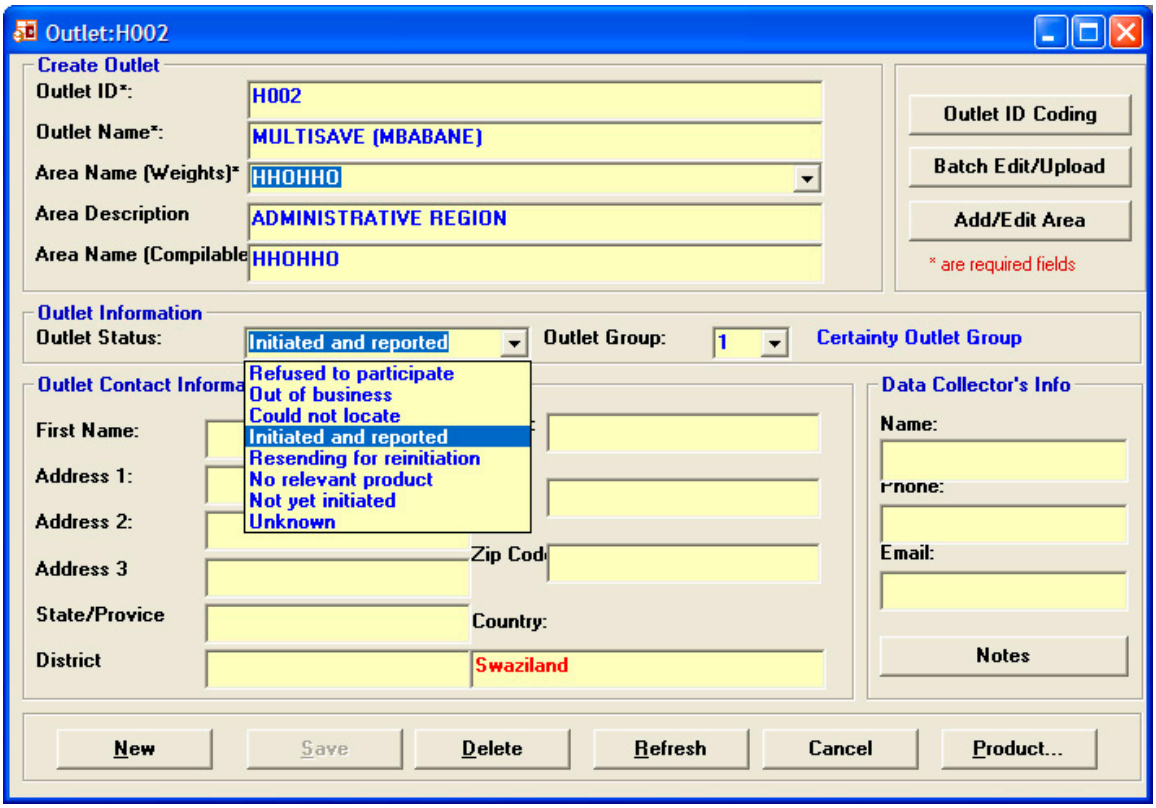
Click 

button to add or edit area.



The user can also manage the area list here. It is important to note that the area defined here corresponds to the geographical area where household expenditure survey is conducted and where weights information is available. *If regional weights are not available, national weights by product group will be distributed equally across the regions that have been defined and by variety.*

Click the **Outlet Status** dropdown list to define the outlet status (default value is “Initiated and Reported”. Click Outlet Group to define the group that outlet belongs to (default value is “Group I - Certainty Group”). Outlet status will affect the distribution of expenditure weights and calculation of indices. An outlet that is defined as “Out of Business” or “Refuse to Participate” will be suppressed from calculation even though they physically exist in the database.



Click **Batch Edit/Upload** button to input or edit multiple outlets at a time. The System can generate an Excel template for you to enter multiple outlets with the required fields. However, you cannot enter information in the spreadsheet for the Outlet Contact and Data Collector Info appearing on the Add Outlet screen. You can save the template and upload the data later, but remember to save it as a macro-enabled spreadsheet (required for the upload feature to operate, particularly in Excel 2007 and above).

Consumer Price Index Compilation Model - [Edit Reveiw Item Weights]

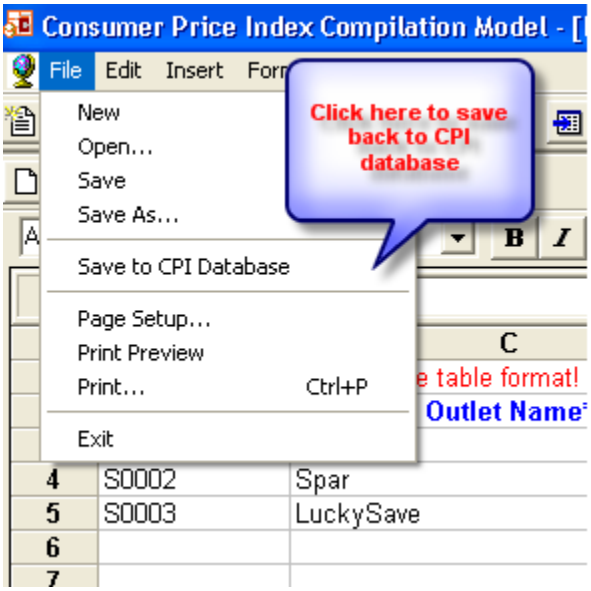
File Edit Insert Format Tools Help

***Please do not change the table format! You can add rows at the end of the worksheet! Click 'Save to CPI Database' once you finished.

Outlet ID*	Outlet Name*	Area *	Description	Area Compilabl	Outlet Status
H001	THE MALL SPAR	HHOHHO	ADMINISTRATION REGION		5
H002	MULTISAVE (MBABANE)	HHOHHO	ADMINISTRATIVE REGION	HHOHHO	5
H003	SHOPRITE	HHOHHO			5
H004	CLICKS	HHOHHO		HHOHHO	5
H005	JET/SALESHOUSE	HHOHHO			5
L001	SCORE_SIMUNYE	LUBOMBO		LUBOMBO	5
S001	TARGET_NHLANGANO	SHISELWENI		SHISELWENI	5
H006	ACKERMANS	HHOHHO		HHOHHO	5
H007	PHOENIX SPURS	HHOHHO		HHOHHO	5
S002	PIERTESE BUTCHERY	SHISELWENI		SHISELWENI	5
H008	SCORE SUPERMARKET	HHOHHO	PIGGSPEAK	HHOHHO	5
H009	DUNNS STORES	HHOHHO	PIGGSPEAK	HHOHHO	5
H010	D.B SUPERMARKET	HHOHHO	PIGGSPEAK	HHOHHO	5
S004	SCORE SUPERMARKET	SHISELWENI	HLATIKHULU	SHISELWENI	5
L002	SCORE SUPERMARKET	LUBOMBO	SIMUNYE	LUBOMBO	5
L003	SIMUNYE GARAGE	LUBOMBO	SIMUNYE	LUBOMBO	5
L006	SIMUNYE CHEM	LUBOMBO	SIMUNYE	LUBOMBO	5
M002	SUPREME FURNITURES	MANZINI	MANZINI	MANZINI	5
H011	SUPREME FURNITURES	HHOHHO	MBABANE	HHOHHO	5
S005	SPAR SUPERMARKET	SHISELWENI	NHLANGANO	SHISELWENI	5
M003	HUB SPAR	MANZINI		MANZINI	5
M004	THE IDEAL CLOTHIER STORE	MANZINI		MANZINI	5
M005	THEKWINI W.S	MANZINI	MANZINI	MANZINI	5
M006	TOWN TALK	MANZINI	MANZINI	MANZINI	5
H012	TRUE PEP	HHOHHO	MBABANE	HHOHHO	5
L004	VUKUZENZELE, SITEKI	LUBOMBO		LUBOMBO	5
H013	CENTRAL BUTCHERY, MBABANE	HHOHHO		HHOHHO	5

* indicates required fields. Do not change the table format. You can add records at the bottom of the worksheet. 9/27/2005 9:18 AM

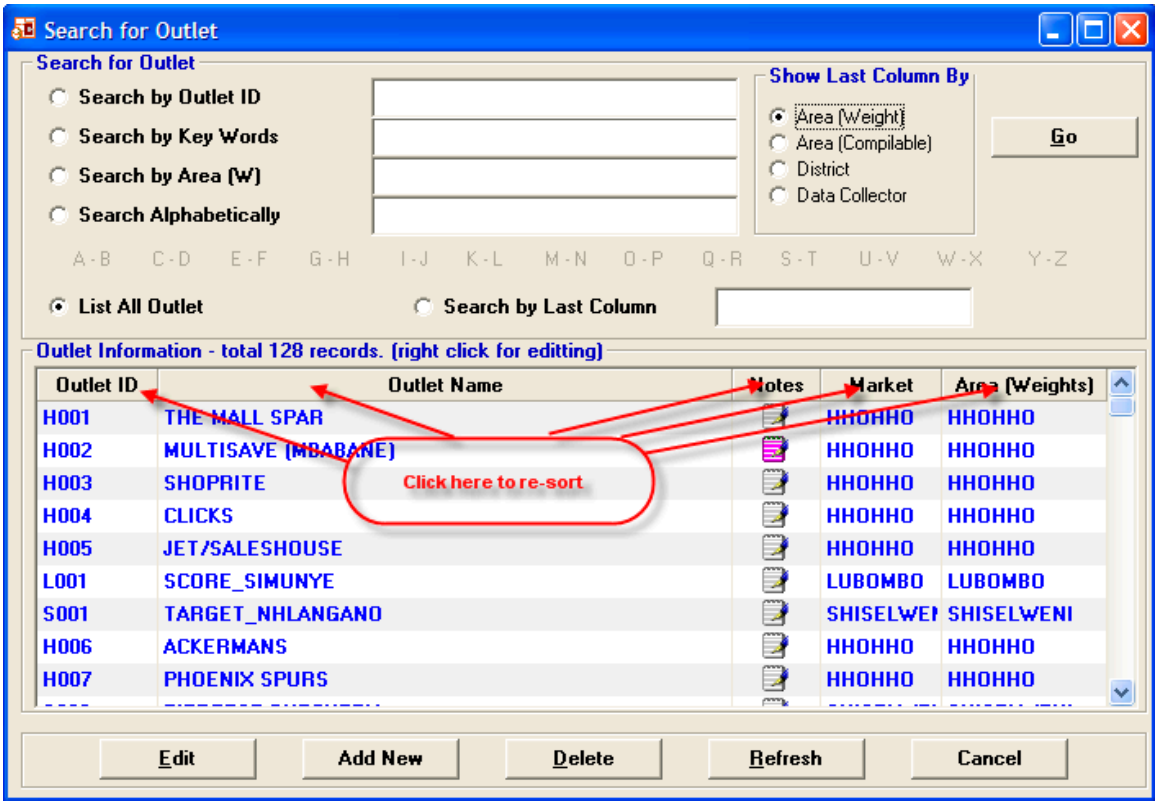
Ready... supergzhu 9/27/2005 9:18 AM CAPS



Edit Outlet, Product or Variety

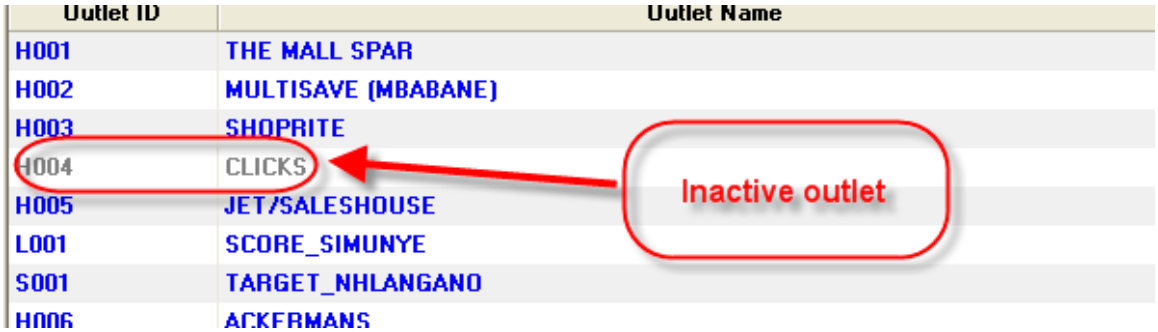
Click 

button to edit outlet information.

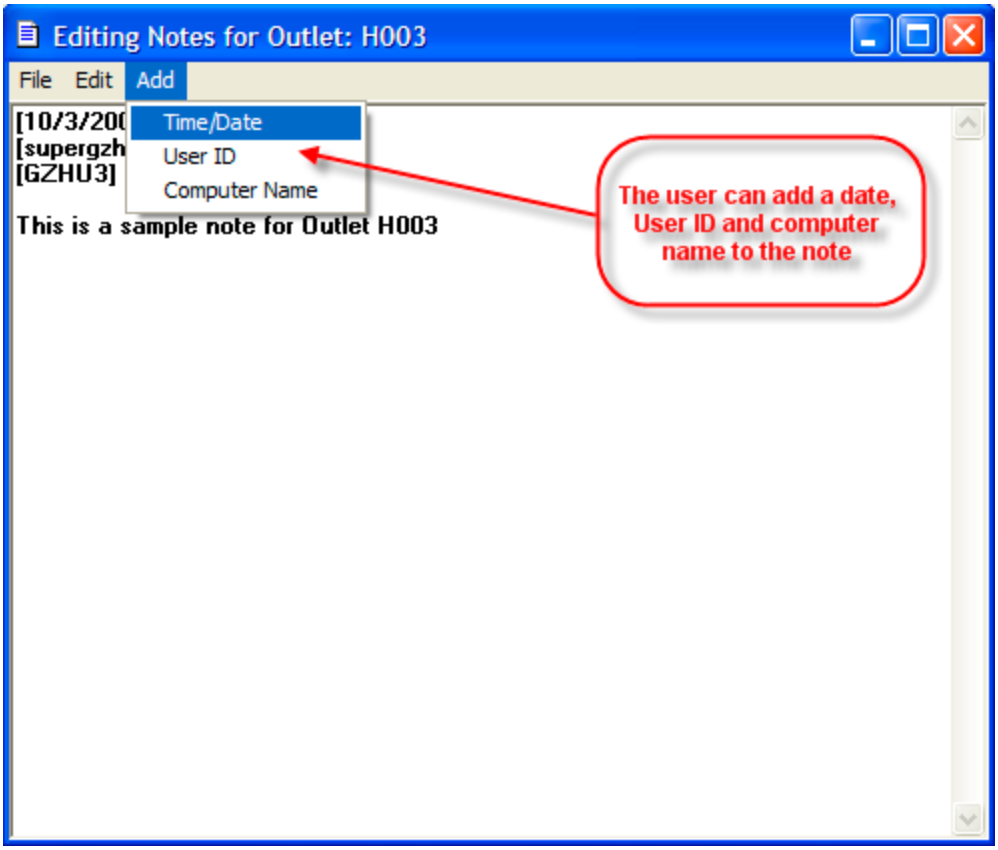


Click the first row of the respective column to re-sort the information in ascending or descending alphabetic order. Select a specific outlet (by clicking once on the particular row), right click (once) to invoke an edit Menu. It is important to note that you can change an outlet ID provided that ID is not already in use. It the user wants to create a new outlet using an ID that already exist, they can either delete that ID first, or hit the save button three times. In that way, the System will overwrite the existing outlet information in the database with the information newly inputted. The user can edit outlet, product, or price information by clicking the appropriate button.

Those inactive outlets (including those of out of business, not yet initiated, or refuse to participate) will show in grey.

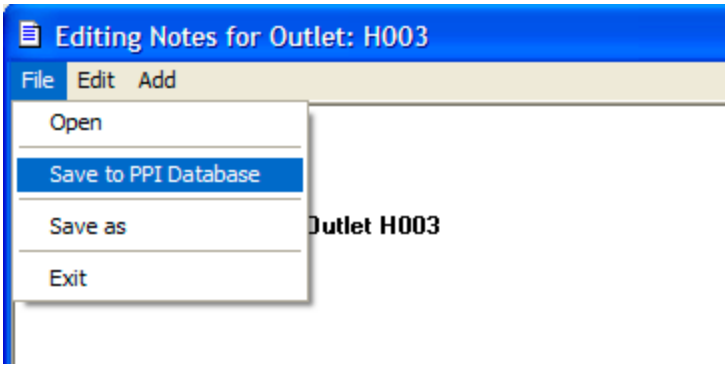


The user can add notes to an outlet by double clicking the icon in Notes column. An icon indicates that there is a note associated with this outlet.

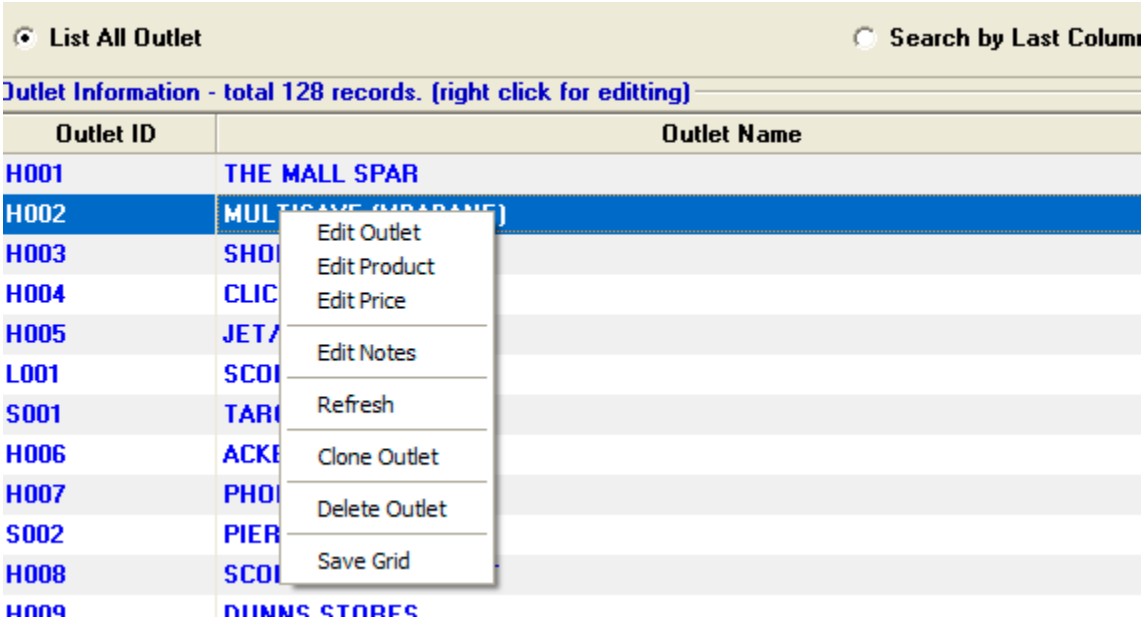


It is important to note that the metadata text length should not exceed 65353 characters (a memo data type in the database)

After adding a note, the user can save it to the CPI database or save it externally for future reference.



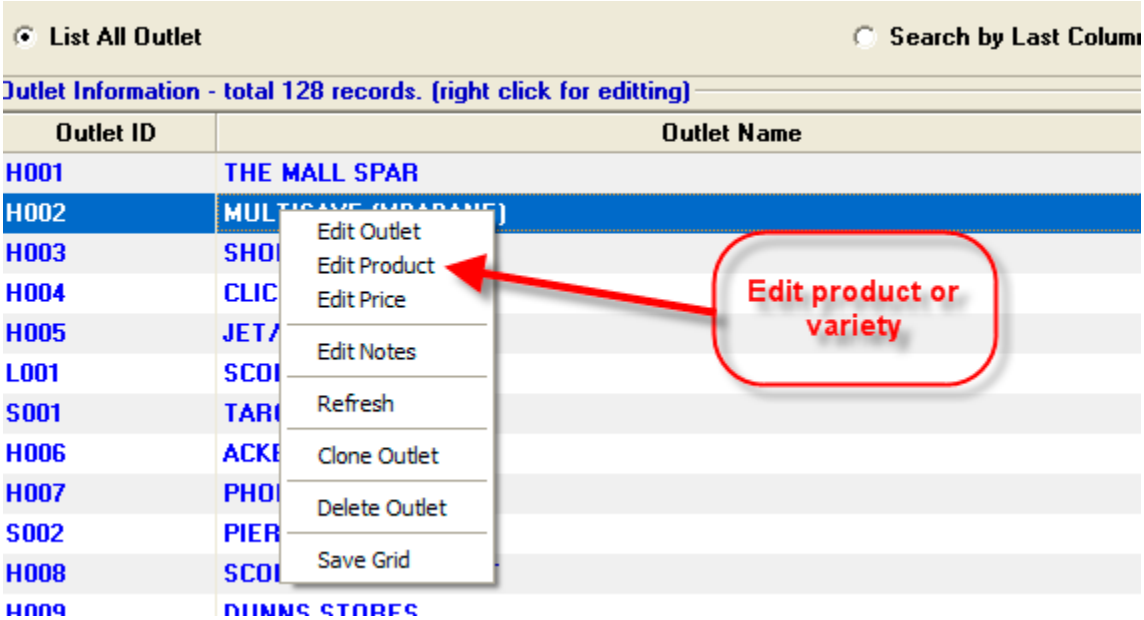
The user has the choice of editing outlet, editing a product, editing a price, editing notes, cloning an outlet, deleting an outlet or saving the contents of the grid to a spreadsheet.



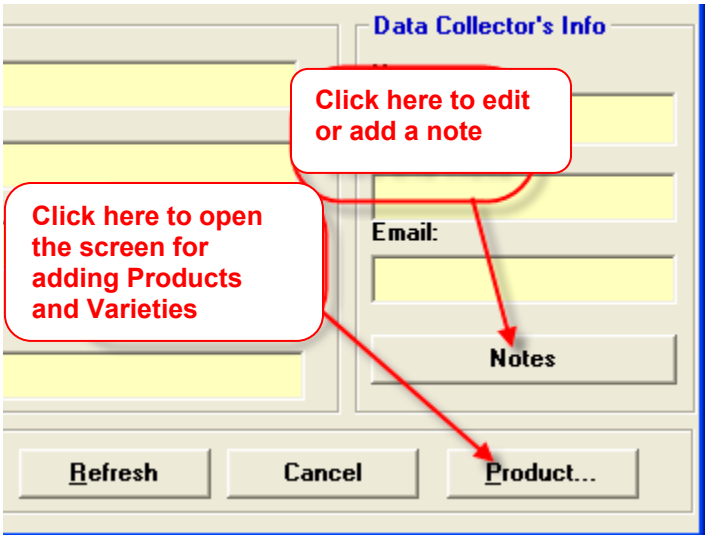
Cloning an outlet is to duplicate the complete outlet information, including the product structure. This is to provide the user with a quick way to create an outlet record without typing every piece of information all over again. This is very helpful when entering product and variety data for supermarkets or grocery stores that often carry the same products. It is also useful in the sample replacement exercise, where newly selected outlets replace the existing one.

The “Save Grid” feature allows the user to save outlet grid data into a spreadsheet. The last column of the saved spreadsheet contains the establishment sequence number (EstSN) that is an important identifier in the database tables.

Add a New Product and Variety



To add product or variety information, the user can select an outlet (single click on the outlet name), right Click, then select the **Edit Product** button, or click the **Product** button on the outlet edit screen.

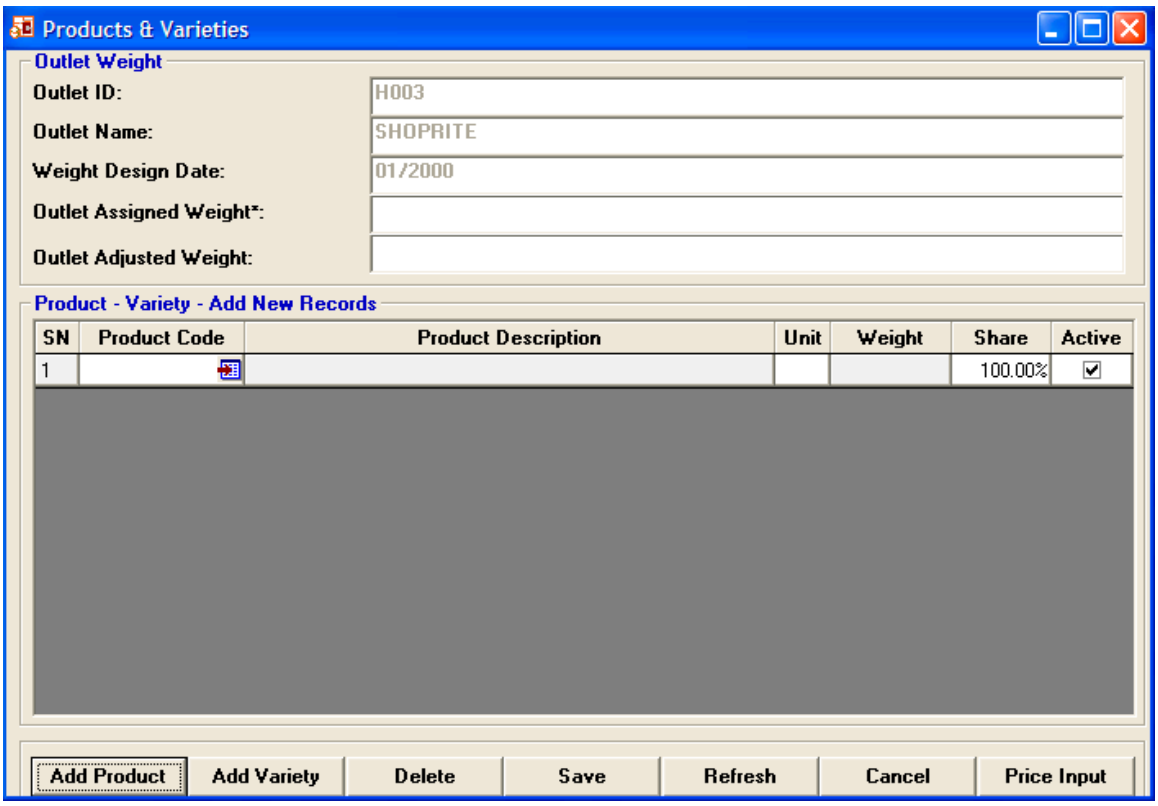


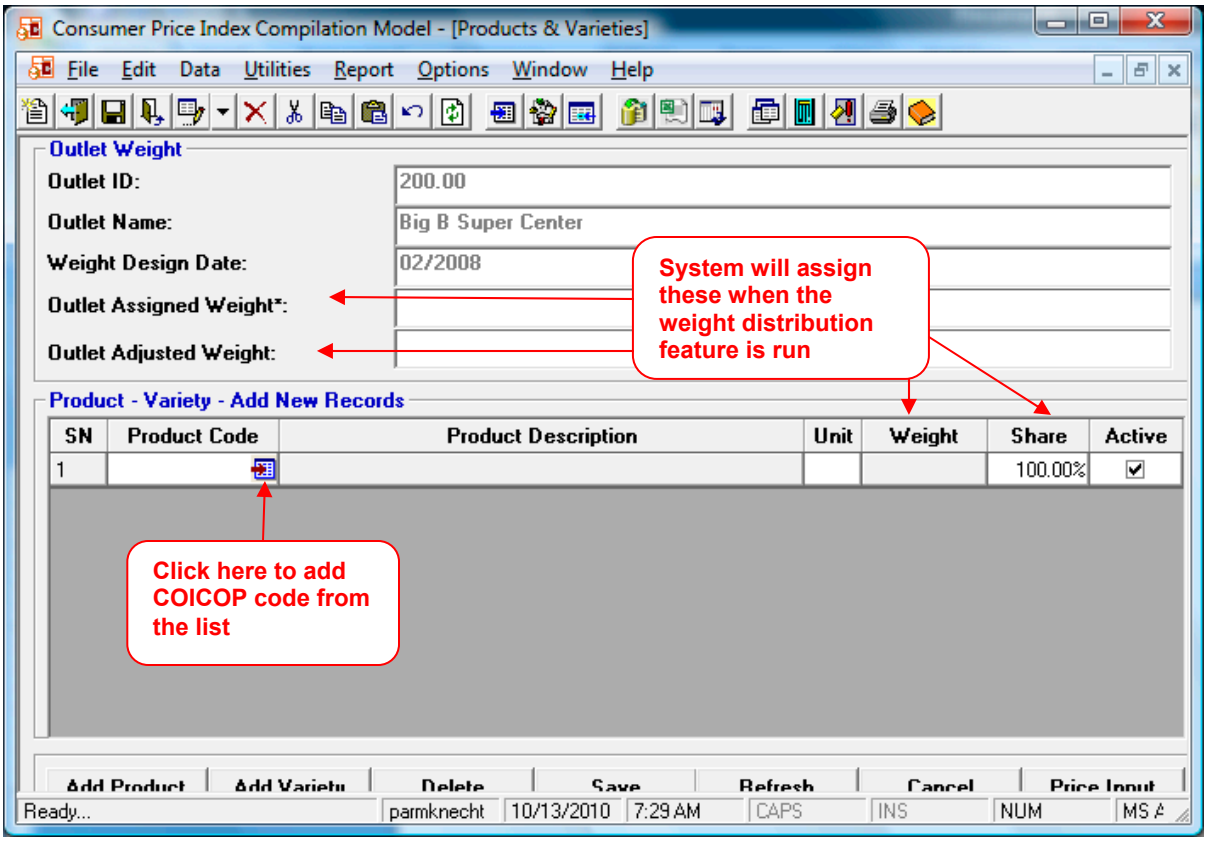
Also, Click the **Product** button at the bottom of the Edit Outlet screen to add a product. A blank row will be added for the user to input the product information. If the **Outlet**

Assigned Weight is not available for that outlet, the user can leave it blank. The System will calculate this once the area expenditure weights for the product become available. Outlet Adjusted Weight is an imputed value based on the Outlet Assigned Weight, Outlet Status, and Sample Group it belongs to.

Weight and Share fields for a product will be calculated once the Outlet Adjusted Weight becomes available.

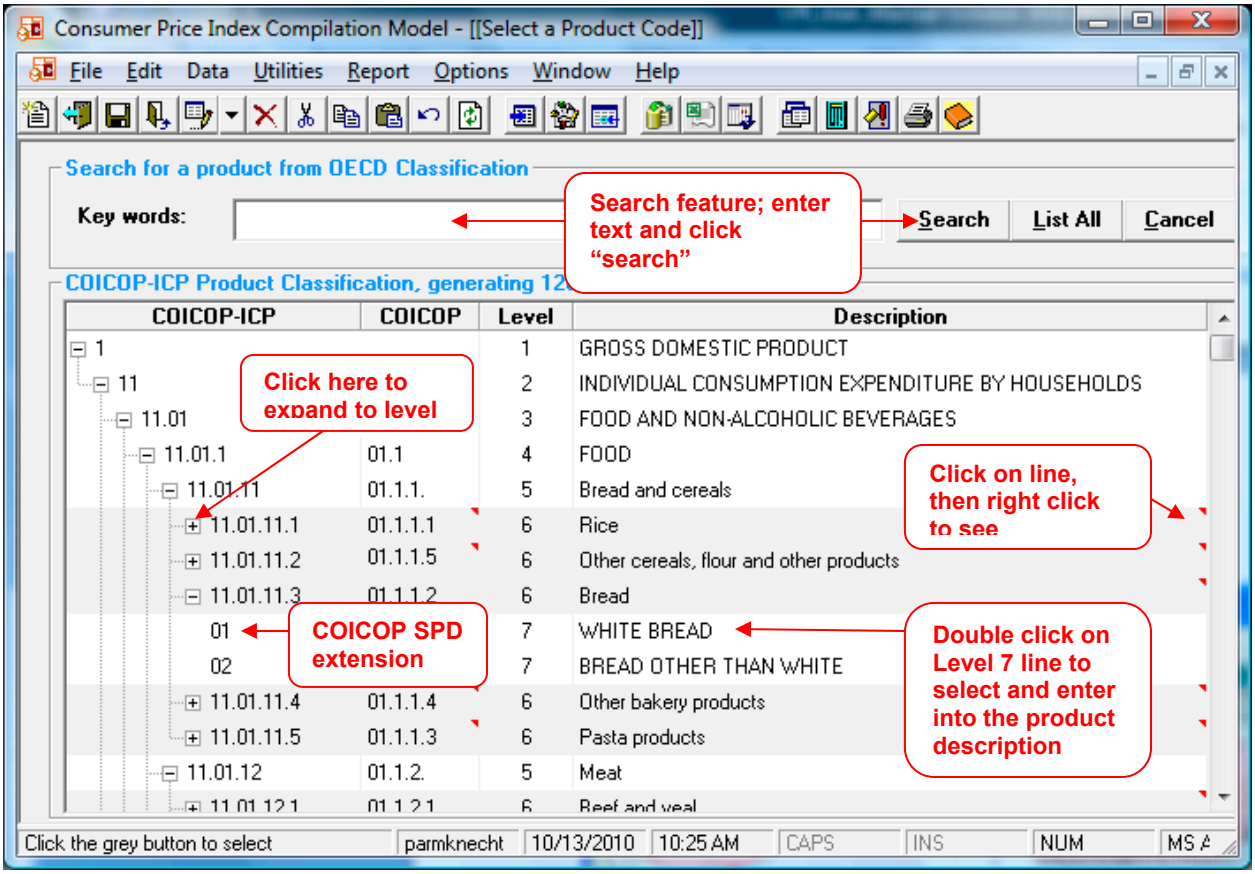
The following screen will allow the user to input product and variety information: Leave the unknown field blank, define the product structure by selecting from the COICOP-ICP classification system.



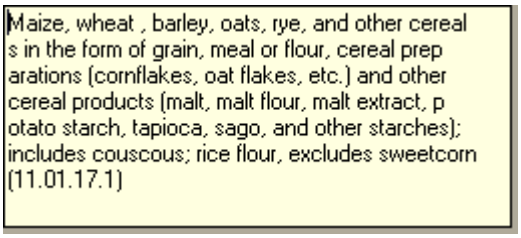


Click button to select a product from a pre-defined product classification system. In the CPI System the default product classification is the COICOP-ICP system with the SPD extension.

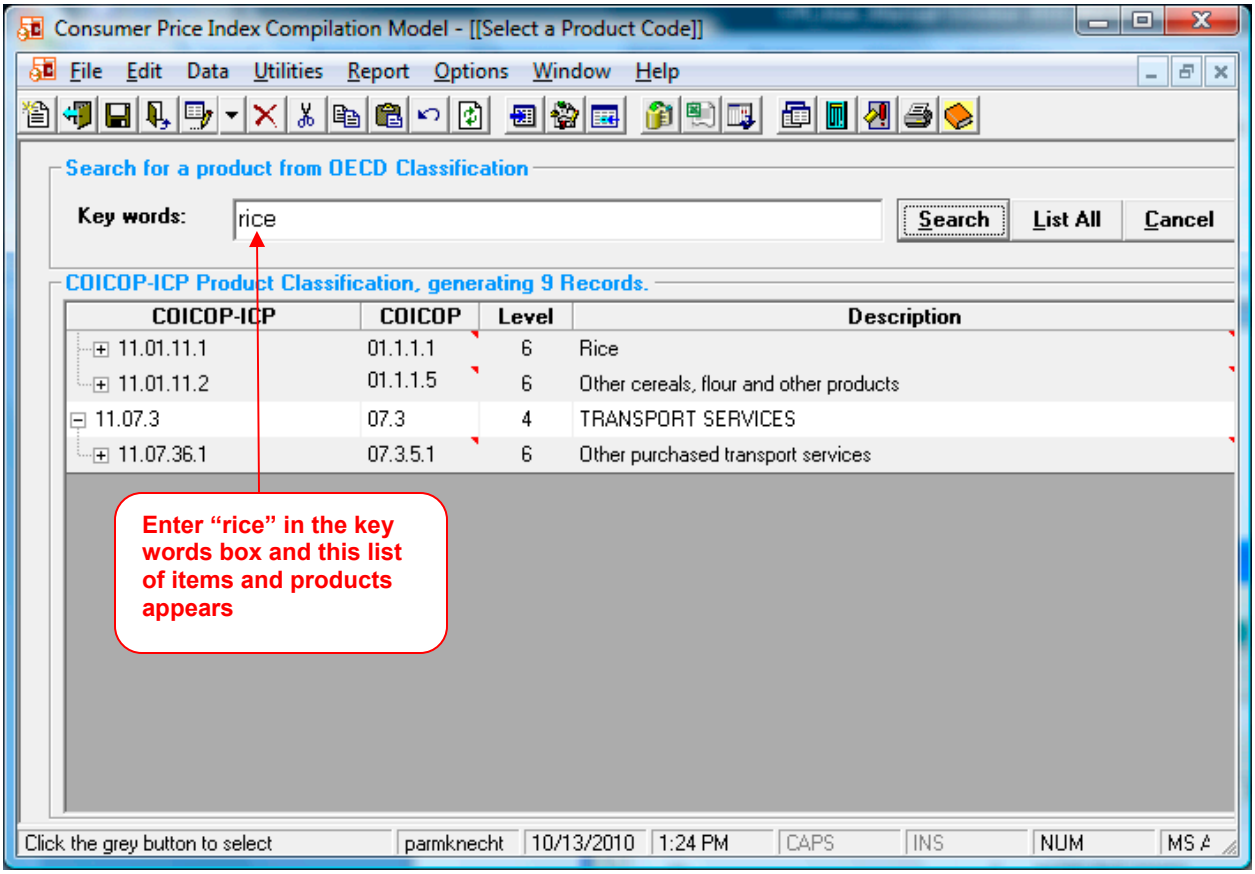
The user should double click to select a product from the most detailed level (in this case, at the 7-digit level, which is also the SPD level) from the product list. The screen will open at the 6-digit level. The user will must select the 6-digit COICOP description they want to expand by clicking on the ⊕ box preceding the description. The System has a search feature for the user to locate quickly the product they want to find. The user can type the complete word or part of that word; the System will perform the search not only on the description but also on the notes field.



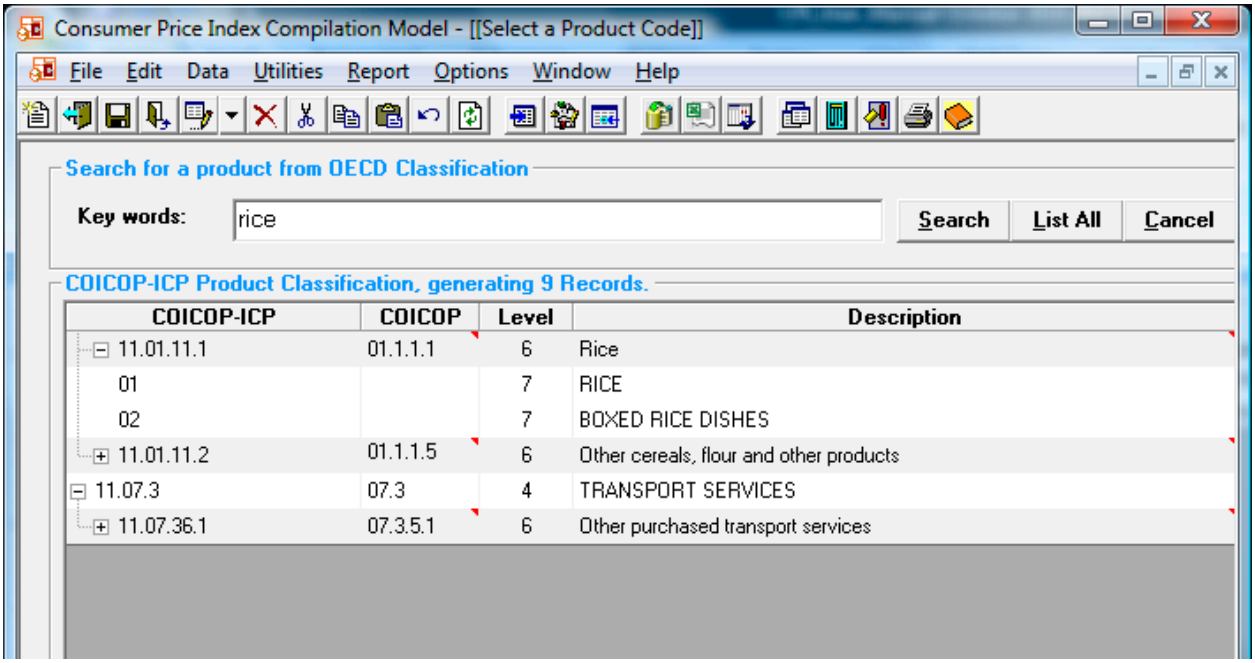
Additional information is provided in the explanatory notes which are indicated by a red triangle at the upper-right of the cell. The following note is associated with product "11.01.11.2, Other cereals flours and other product".

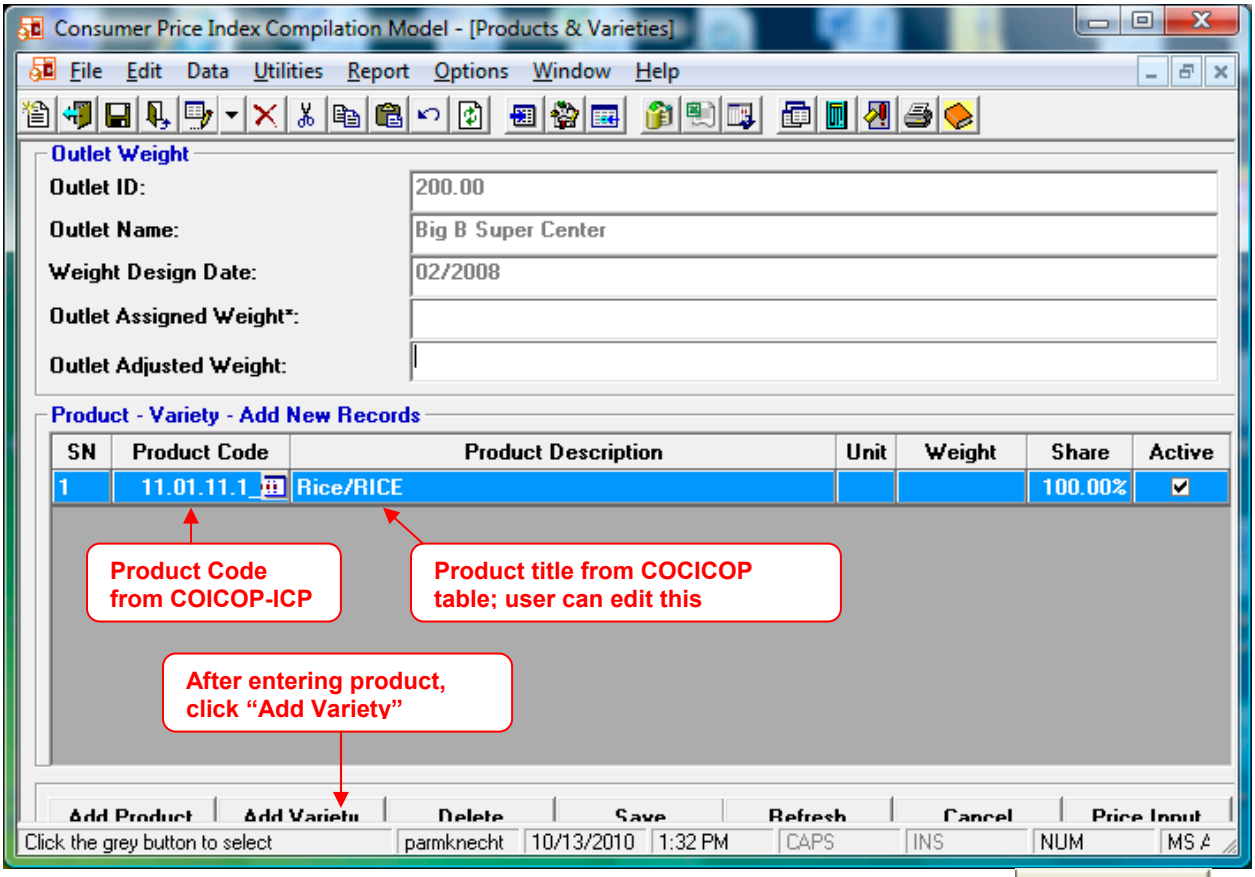


The CPI System provides the user with an intelligent search facility. Let us assume the user wants to search for rice-related products, they enter "rice" in search textbox, click the **Search** button, and the following screen will be shown.



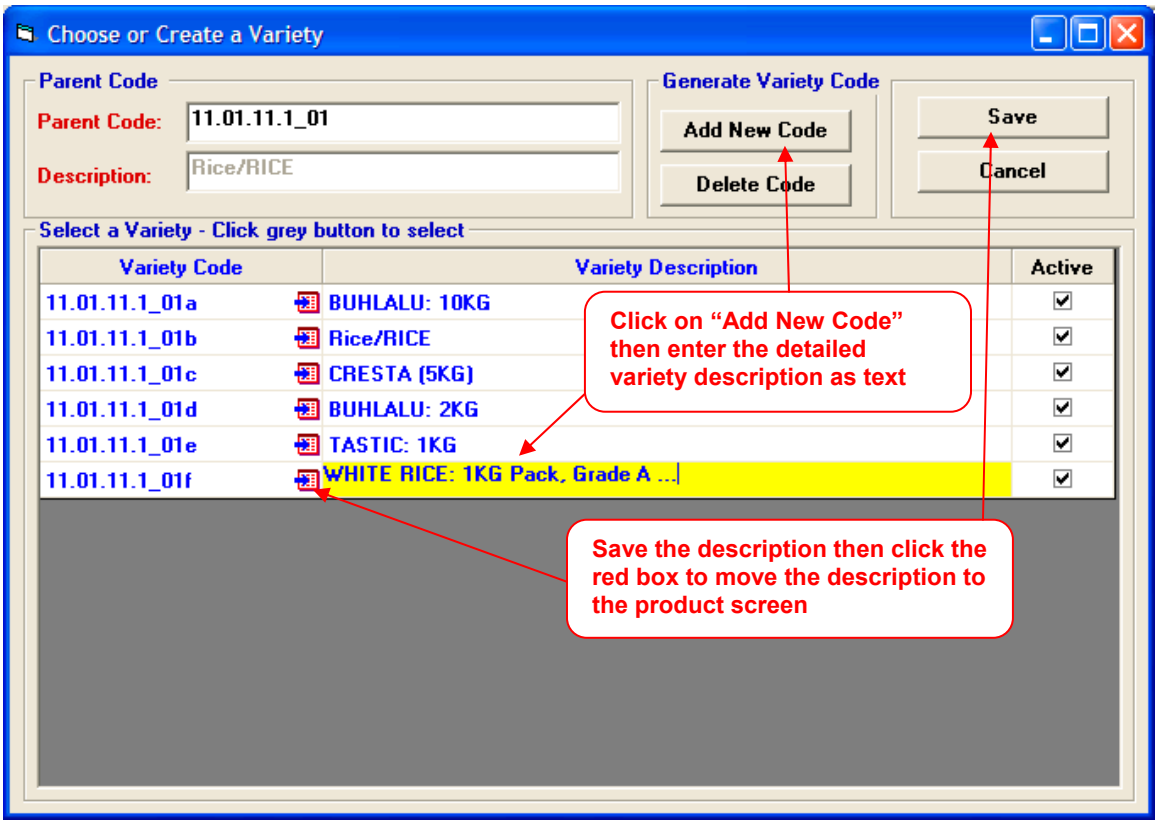
Click on the box next to Rice in the COICOP-ICP column and the product (7-digit level) appears. By double clicking the product (level 7) the user wants to select, the product code with its description will be extracted and displayed on the product selection screen.





The user then can add a variety to this product by clicking **Add Variety** button.

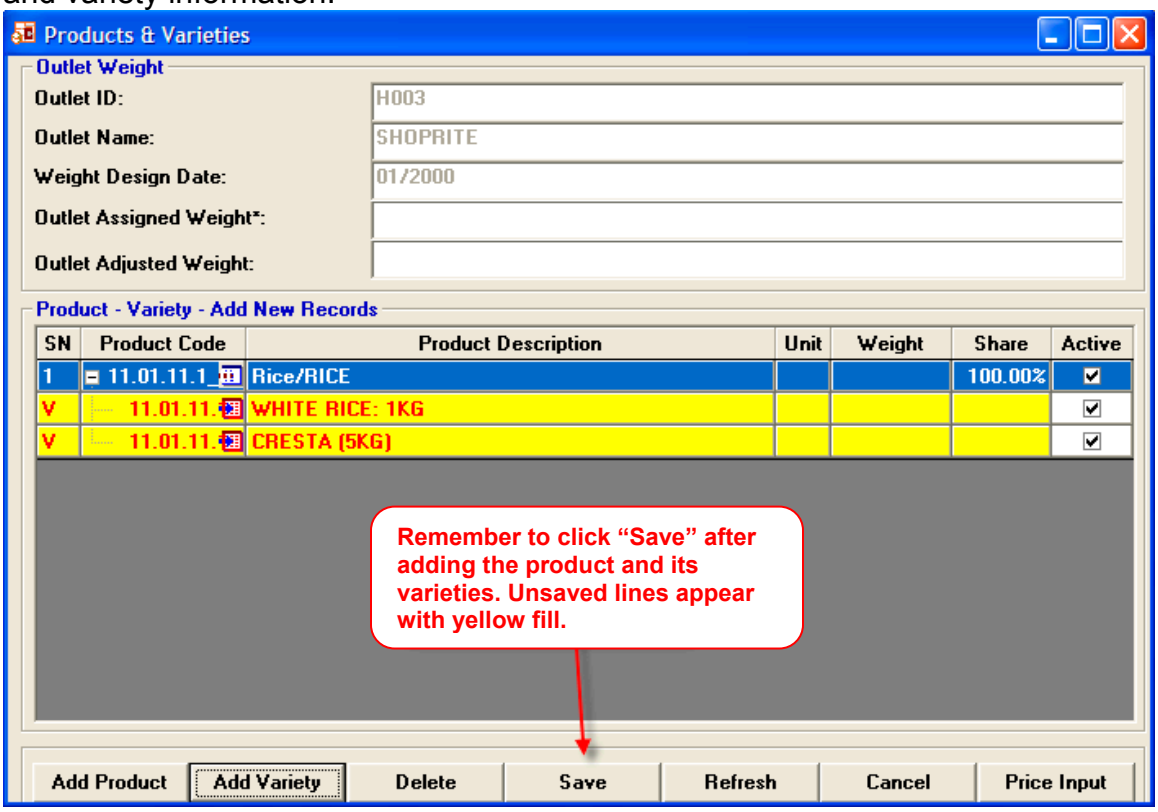
The variety list is a user-defined list that extends the international standard product classification defined by the COICOP-ICP. The Variety list comprises many different varieties that users may price for the product. Each variety of the product can have a user-definable, unique code and description with an SPD checklist type of detail in the Description field, such as package type, serving, number of units in package, size of unit, unit of measure, origin, seasonal availability, product characteristics, etc. The variety code has at its core the parent product code with an additional one or two alphabetical letters. (Do not use numeric values for the extension as the system expects letters.) The system will generate such code when the user presses the **Add New Code** button and the user can change this code if they want. The variety list allows each country to maintain an additional dimension to the product/variety information that reflects its own specific consumption patterns. Each variety list has a parent, which is listed in the standard COICOP-ICP/SPD product classification. **In this way, the system maintains an international standard while providing each country with the flexibility of creating additional dimensions of variety with their unique characteristics.**



It is important to note that all outlets share the same variety list. If one particular variety is already used by another outlet, it can not be deleted from database.

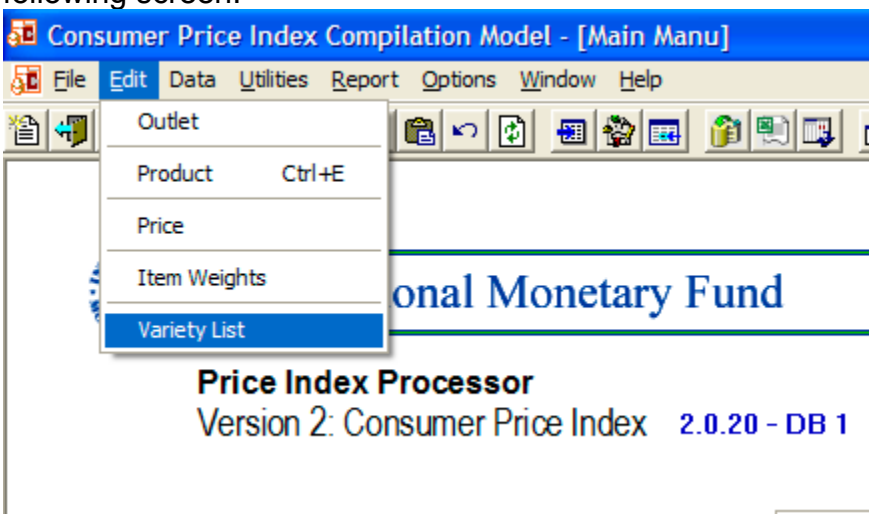
The user should look at the variety list carefully to decide whether an existing variety code should be used or a new variety should be added and defined. Click the small red button at right side of the Variety Code column to select a variety. Once a variety is selected, the "Choose or Add Variety" screen will disappear and the variety information

will show in the product screen. The user has to click **Save** button to store the product and variety information.



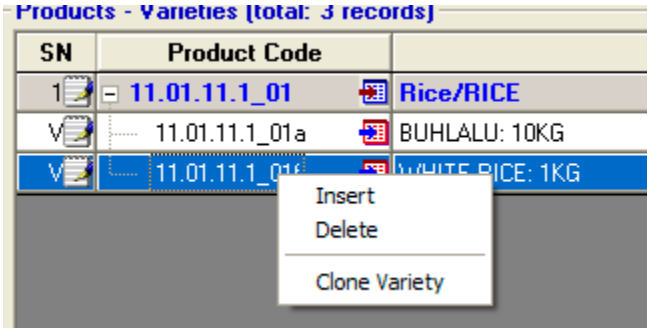
Please note, if there are multiple products in one outlet, the user has to select the product to which they want to add a variety. The user cannot add a variety directly under another variety.

Variety lists can be managed by clicking the **Edit**, **Variety List** button as shown in following screen.



A variety can be “cloned”, i.e., have all the information for that variety copied. Select a Product Code (move the mouse to Product Code column, then single click), then right click. The user can Insert, Delete or Clone Variety. If the user selects to Clone, a new variety is created with the same description but having a new variety code. If there is any price data associated with the variety being cloned, it will also appear as price data for the new variety. The use of the clone feature is a topic in the section on Substitution and Quality Adjustment.

Click icon, the user can add a note to the product or variety. A red notepad icon, , indicates that there is already a note for this product or variety. You can add more text to the existing note.

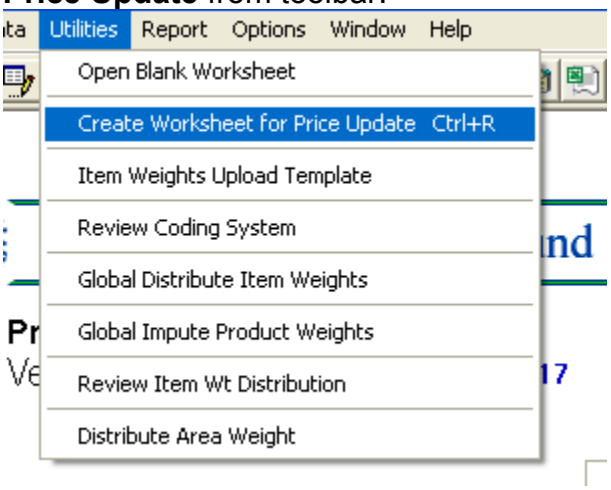


Input Price Quotations

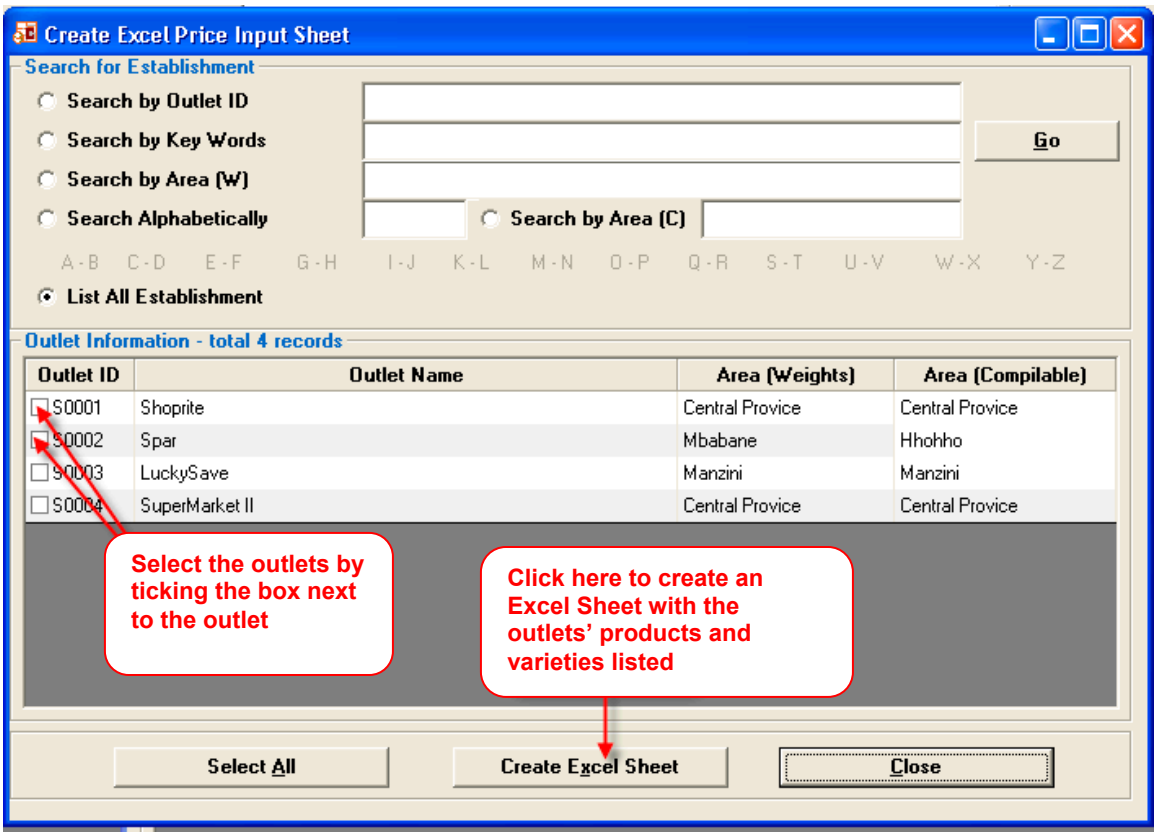
After defining the products and varieties structure, the user is ready to input price quotations. The user has several methods to input prices:

Method 1, Batch Price Data Input

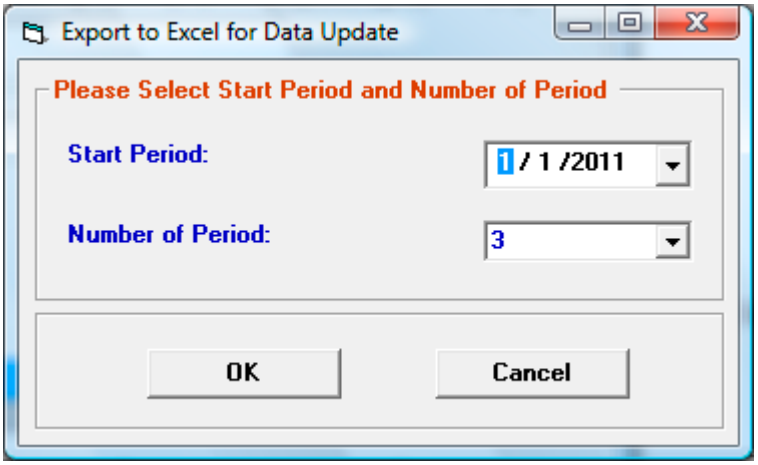
Click **Input Sheet** button at product screen or click **Utilities**, **Create Worksheet for Price Update** from toolbar.



Select outlets by ticking the small square boxes in the first column. Then click **Create Excel Sheet** button.

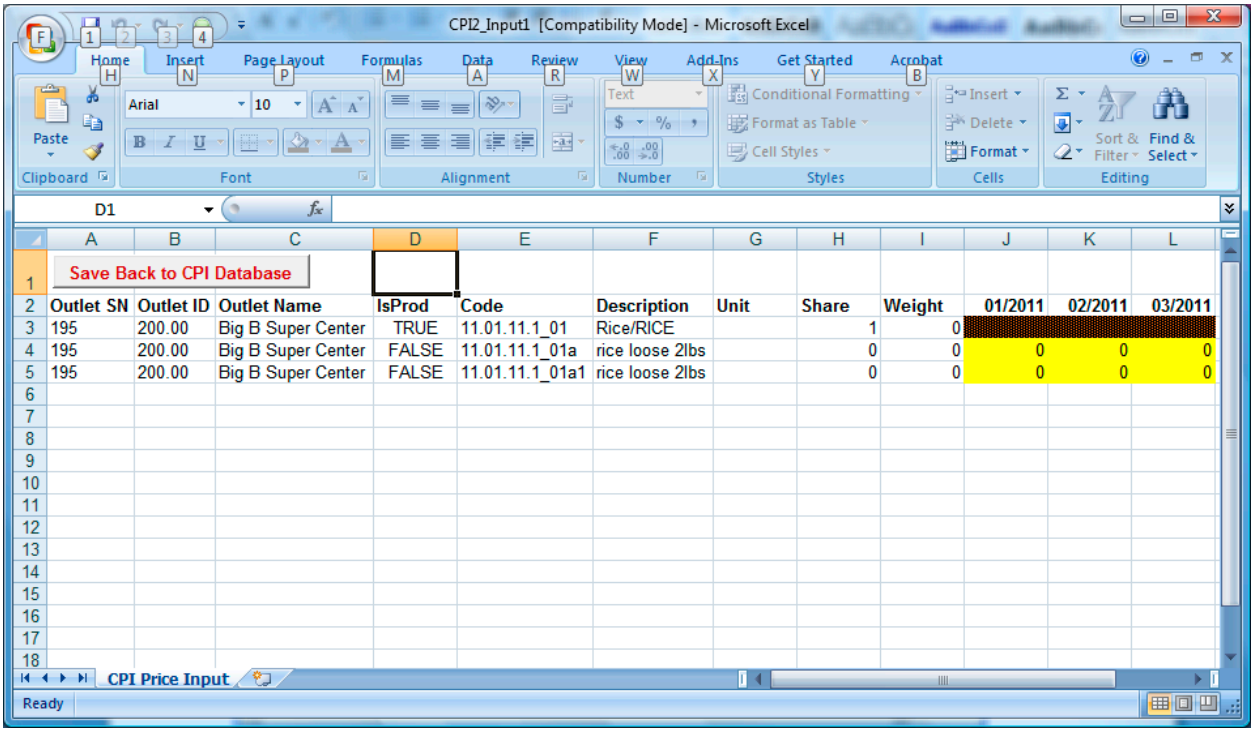


The user will see the following screen: which allows the user to specify the start period and the number of periods for the price data.



Click the **OK** button and System will generate the Excel sheet.

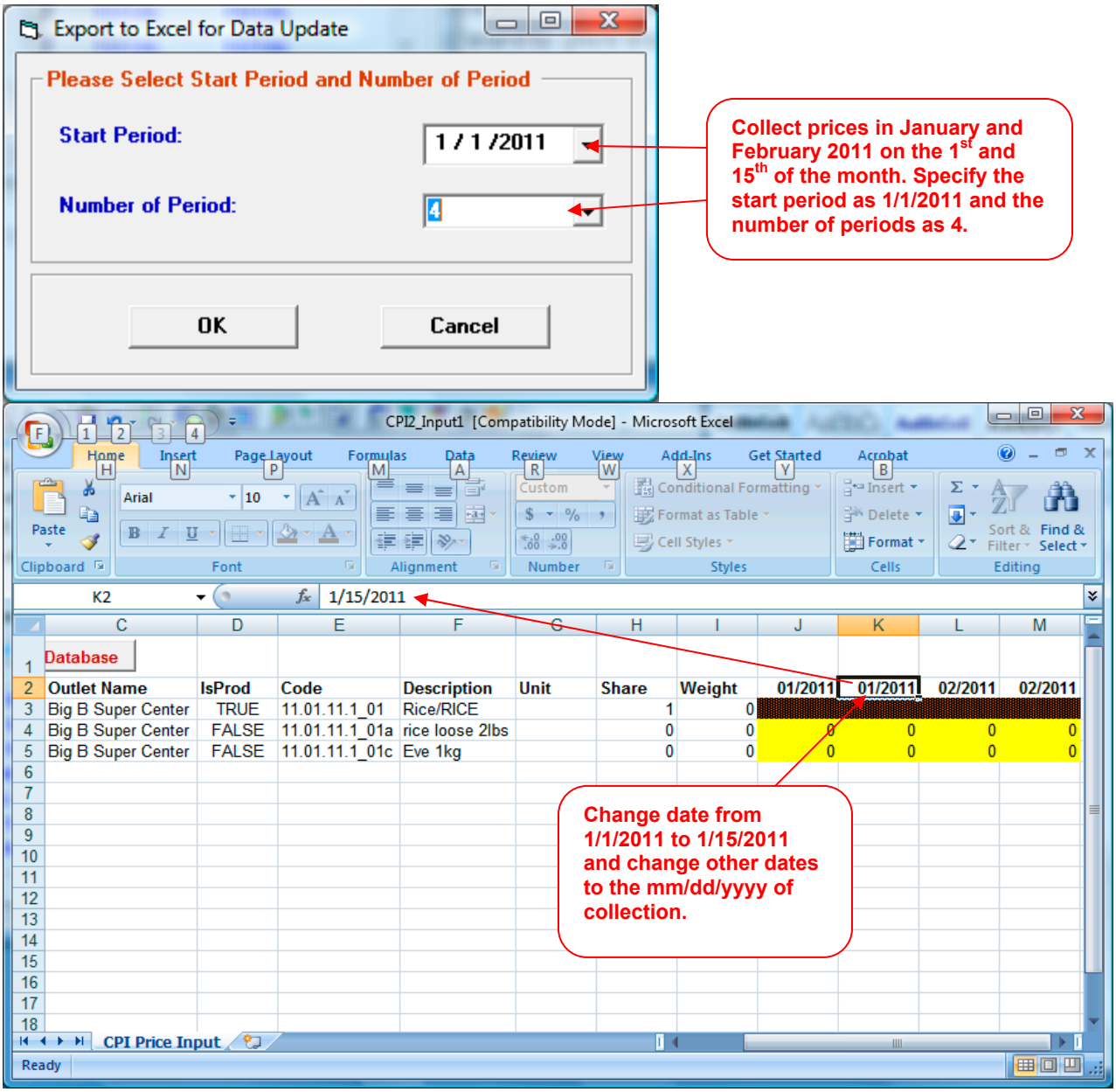
The user will see following Excel spreadsheet:



	Outlet SN	Outlet ID	Outlet Name	IsProd	Code	Description	Unit	Share	Weight	01/2011	02/2011	03/2011
3	195	200.00	Big B Super Center	TRUE	11.01.11.1_01	Rice/RICE		1	0			
4	195	200.00	Big B Super Center	FALSE	11.01.11.1_01a	rice loose 2lbs		0	0	0	0	0
5	195	200.00	Big B Super Center	FALSE	11.01.11.1_01a1	rice loose 2lbs		0	0	0	0	0

This worksheet is protected except for the yellow area, which is the area where the user can input price information. It is important to note that this is a pre-designed spreadsheet with a specific table format for price update purposes. Some cell information is for technical purposes and the table structure cannot be changed at the user's discretion. Any unsolicited change of the table structure will cause errors and data cannot be saved properly. But if it is absolutely necessary, the user can unprotect this worksheet by clicking **Tools, Protection, Protect Sheet** button In Excel 2003 or to **Review, Unprotect Sheet** in Excel 2007. The password is "123".

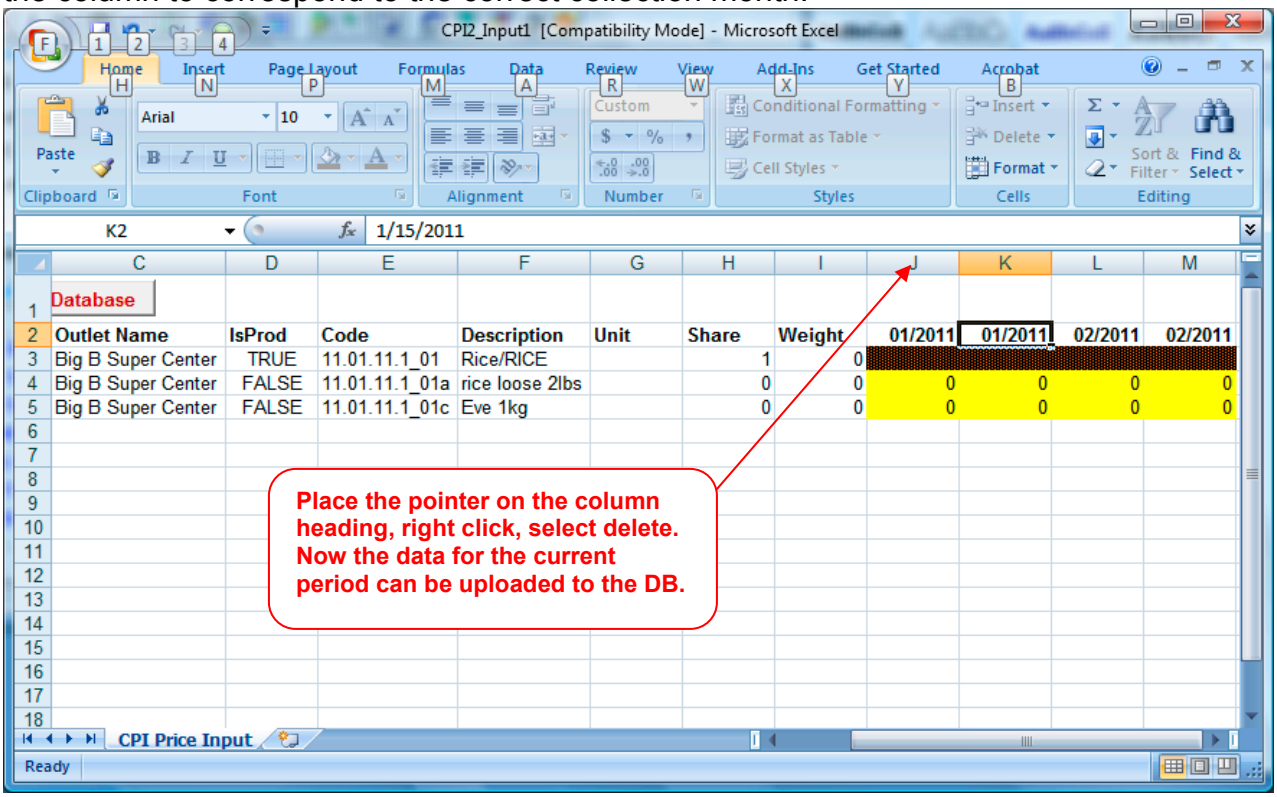
If the user enters more than one price observation, the system will take the average of these prices in the compilation process. The system can handle multiple price collections per month. The user needs to specify the price collection date in the Date column. For example, if prices are collected on the 1st and 15th, these dates have to be entered as 1/1/2011 and 1/15/2011. **NOTE**: **If Method 1** (Batch Upload from Excel) **is used for multiple collections during the month**, the user must change the dates on the column heads of the collection date. So if the user collects food prices twice per month and they want to print the Excel sheet for two price collection periods, they must specify the collection date in the date over the price column. See the following example.



When you generate the price input sheets, the prices stored in the DB are all printed, including the imputed prices that are stored. If you use the Excel sheets for multiple months, please note that each time you upload the sheet, it replaces the data in DB. For example, if your sheet has data for January after the most recent index compilation, it will contain January imputed prices. When you upload the February data, it will be stored in the DB. When you compile the February index, the system will impute values for the missing prices. If you use the same sheets in March (i.e., do not generate new sheets), all the data on the sheet will be uploaded to the DB and imputed values for February will be overlaid in the DB by the “0” price that appears on the update sheet. This causes a problem with the current month of March data, because when the system does the calculations for March, there will now be a missing value for February rather

than an imputed value to use in the calculation. The system will re-impute a value for February, but it will not use the March data in the calculation as it would if the imputed value for February were still there. If you continue to use the Excel sheet for more months, every time you upload the price data from that sheet, it will overlay the price data in the DB with the price data on the sheet. So all variety prices that are “0” on the sheet because you could not collect them, will go back into the DB and replace imputed prices.

The user can prevent this by deleting the columns in the sheet that have already been uploaded. The user can also add addition columns by copying the last column on the sheet and pasting it to adjacent columns. NOTE: you must change the date at the top of the column to correspond to the correct collection month.



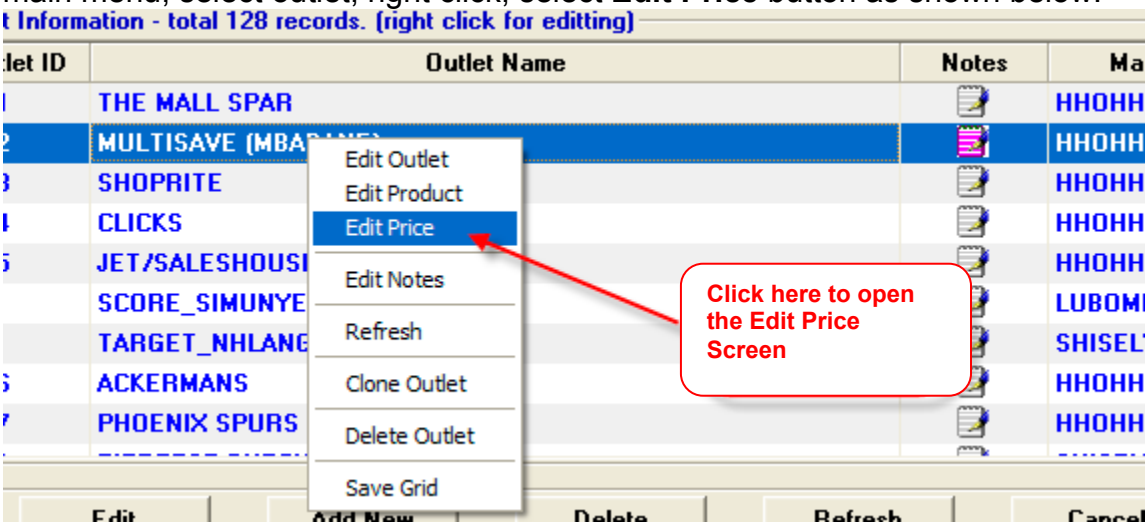
The other alternative is to generate new price input sheets for each month**.** If you do not there are implications for each additional month that you use the same price sheets. This update process (e.g., using the Excel price update sheet for 3 or more months) will not cause an issue unless you go back to re-run the previous months’ indexes. I am hoping that you are not going back and re-compiling back indexes each month. If you have been, you will probably notice differences as long as you use the same Excel update sheets for several months. Again, depending on the number of imputations that have to be made, the more different the results might be.

So to avoid over lay problem, either delete the columns for previous months or run the Excel price update sheets every month. Also, do not re-compile indexes

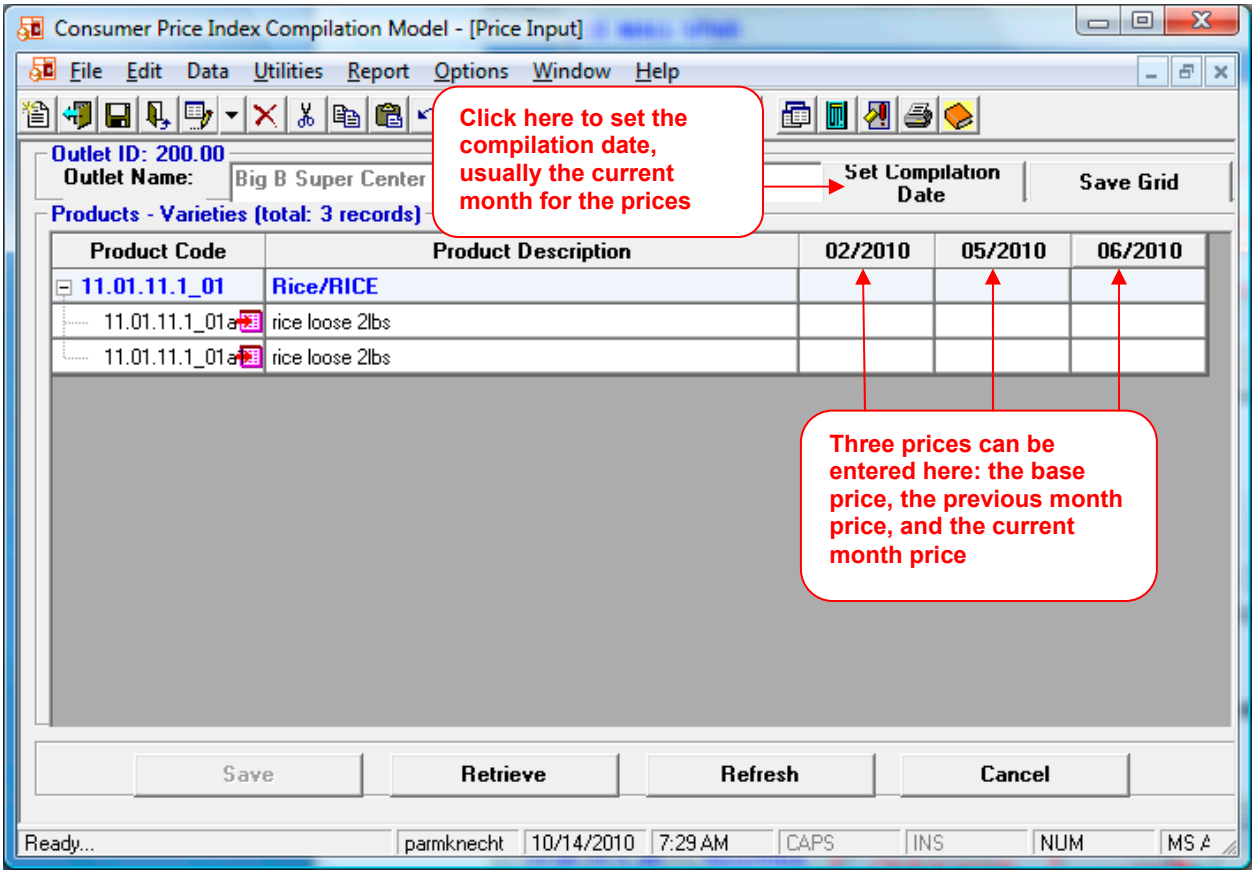
for past periods unless absolutely necessary to republish the indexes because of another error that requires a revised CPI to be published.

Method 2, Enter/Update Price through Windows Form

Click the **Price Input** button in the edit product screen, or click **Edit** from main menu, select outlet, right click, select **Edit Price** button as shown below:



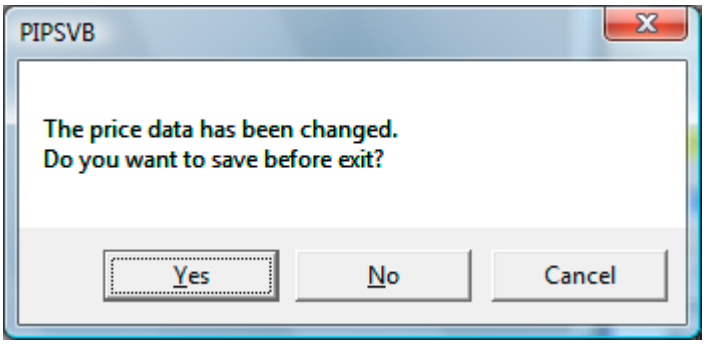
The Edit Price screen looks like:



Enter the price in the relevant cells. The user can enter/edit the price for 3 months (for the base period, preceding period, and current period). Note that once the cell is in editing mode the background color will turn yellow.

The user can retrieve historical price data by clicking the **Retrieve** button. To change the date, the user can click date cell.

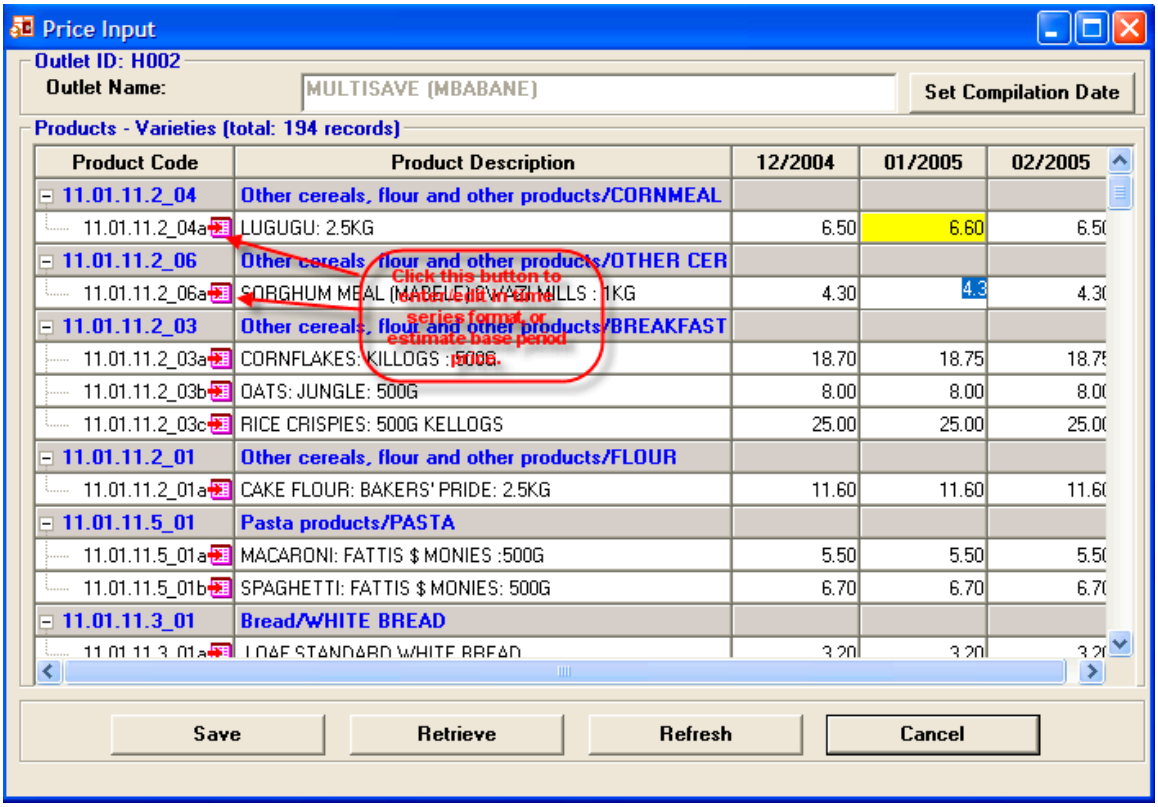
If the user exits the price edit form without saving the revised data, they will see following message:



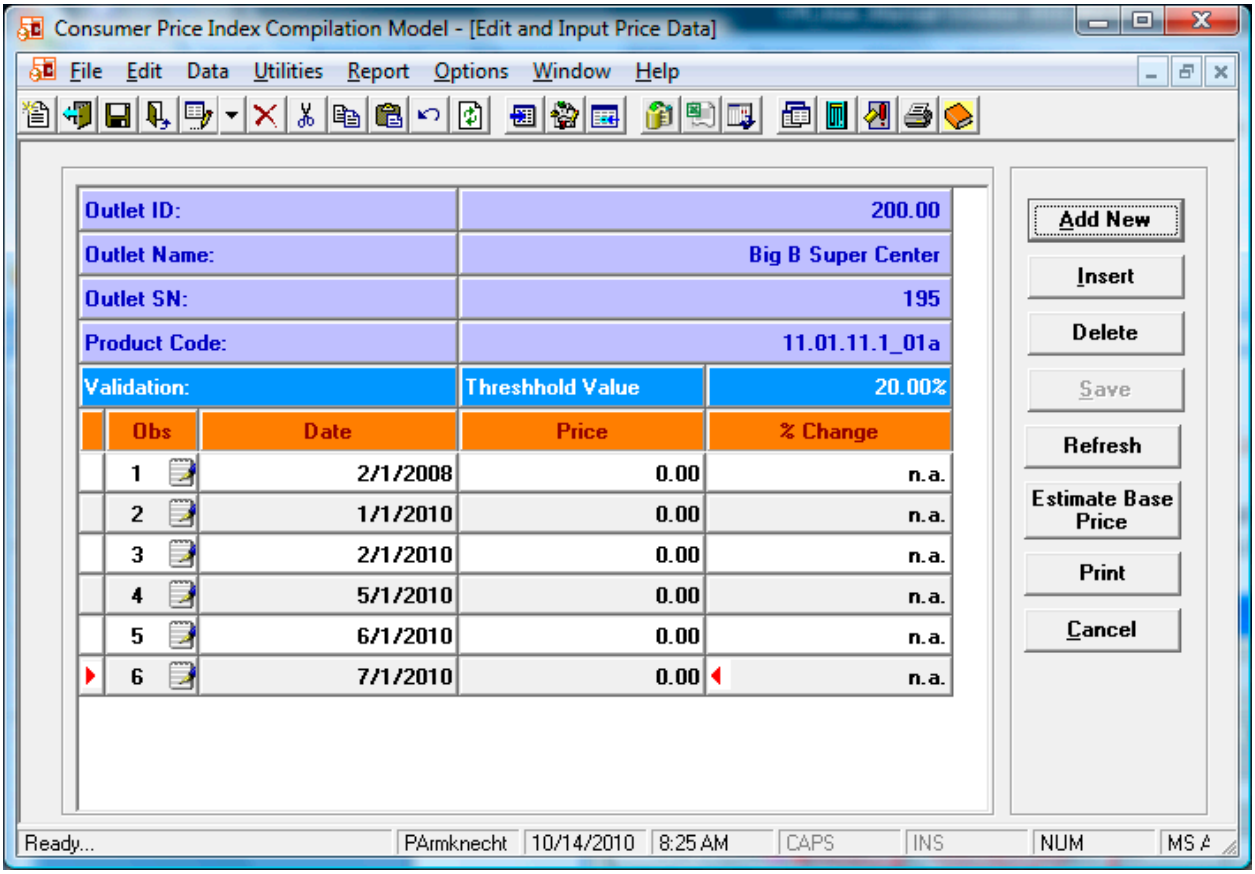
Click **Yes** to save the information, click **No** to ignore the changes, click **Cancel** to cancel the exit and continue to edit the price.

Method 3, Enter/Update Price in a Time Series Format

The user can enter/edit price data for one specific variety for multiple periods in time series format. Click the right side of cell that contains the variety code.



Click the red button, the user will be give following screen for price entry:



At this screen, the user can (a) add an observation at the end of time series; (b) can also insert one between two observations; (c) remove an observation; (d) estimate base period price using another series as a proxy. At this screen, the user can also define a threshold value, in percentage terms. If the percentage increase exceeds this threshold value, the font color of the percentage change will appear in red. Click icon to add a note to a price observation, and indicates that there is note for that observation.

Please note that if the price value is in red, this indicates that this value is an estimated value from an imputation.

If the user enters more than one price observation, the system will take the average of these prices in the compilation process. The system can handle multiple price collections per month. The user needs to specify the price collection date in the Date column. For example, if prices are collected on the 1st and 15th, these dates have to be entered as 1/1/2011 and /15/2011

Edit and Input Price Data

Outlet ID:	H002
Outlet Name:	MULTISAVE (MBABANE)
Outlet SN:	2
Product Code:	11.01.11.2_06a
Validation:	Threshhold Value 20.00%

The price exceeds the threshold value of 20%.

Obs	Date	Price	% Change
1	1/1/1900	n.a.	n.a.
2	6/1/2004	2.40	n.a.
3	7/1/2004	4.50	87.50%
4	8/1/2004	4.40	-2.22%
5	9/1/2004	4.40	0.00%
6	12/1/2004	4.30	-2.27%
7	1/1/2005	4.30	0.00%
8	2/1/2005	4.30	0.00%
9	3/1/2005	4.30	0.00%

Add New
Insert
Delete
Save
Refresh
Estimate Base Price
Print
Cancel

Select one price observation, right click, and the user will see:

Edit and Input Price Data

Outlet ID:	H002
Outlet Name:	MULTISAVE (MBABANE)
Outlet SN:	2
Product Code:	11.01.11.2_06a
Validation:	Threshhold Value 20.00%

Obs	Date	Price	% Change
1	1/1/1900	n.a.	n.a.
2	6/1/2004	n.a.	n.a.
3	7/1/2004	n.a.	n.a.
4	8/1/2004	4.40	n.a.
5	9/1/2004	4.40	0.00%
6	12/1/2004	4.30	-2.27%
7	1/1/2005	4.30	0.00%
8	2/1/2005	4.30	0.00%
9	3/1/2005	4.30	0.00%

Insert
Delete
Flag Price
Unflag Price
Deflate Price

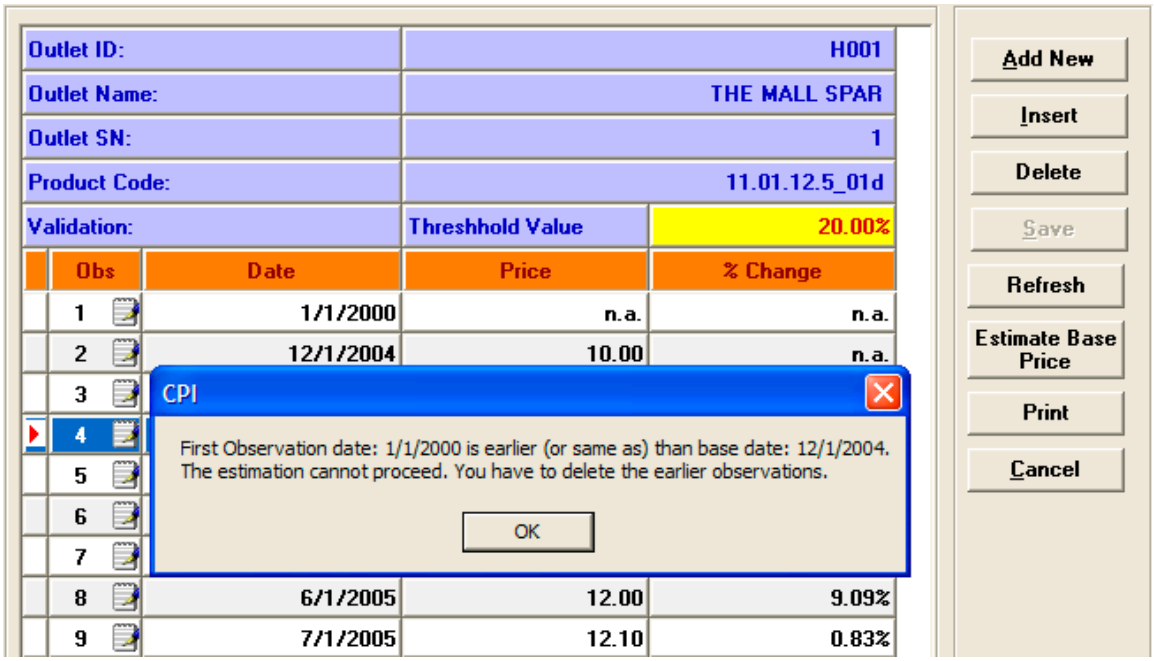
Add New
Insert
Delete
Save
Refresh
Estimate Base Price
Print
Cancel

The user should be able to **Insert, Delete, Flag Price, Unflag Price or Deflate Price.**

When the user substitutes a new variety for an old variety and there is a quality difference or the add a brand new variety to the price index compilation process, a base period price needs to be estimated if the user compiles with the long-term price relative estimation method.

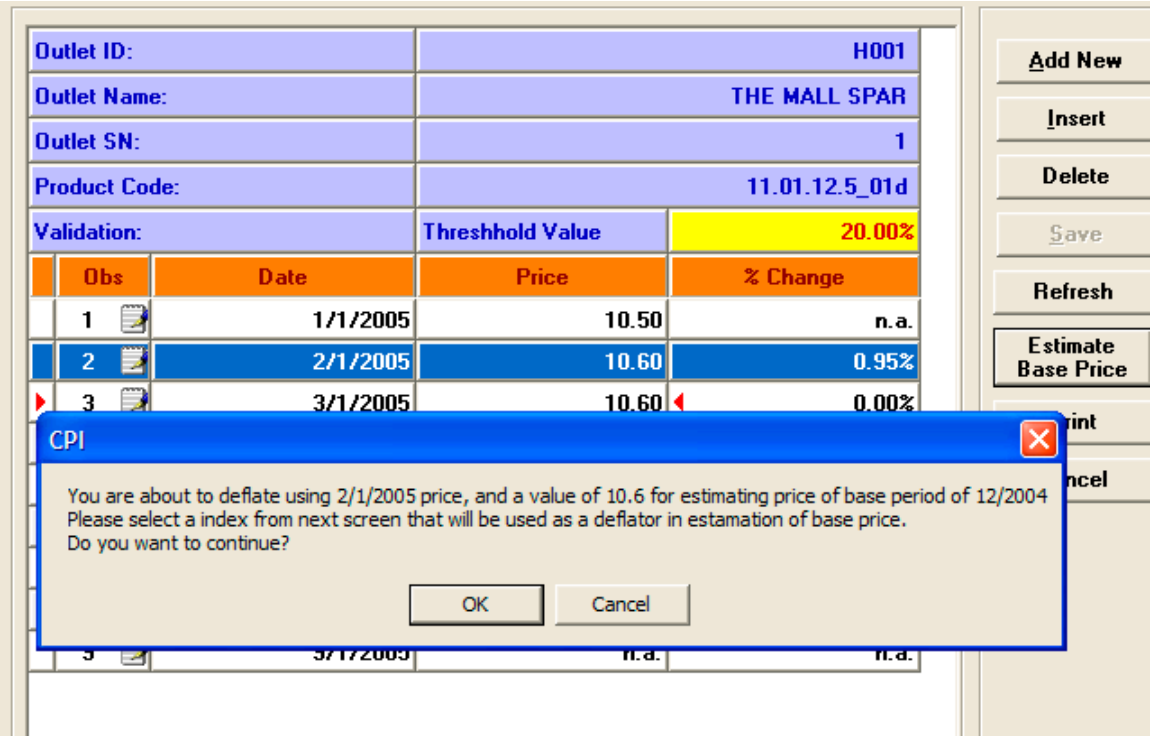
The user can click **Estimate Base Price** or **Deflate Price** button, to use another price index to estimate the base period price for a given price observation. (Note that you must first click on the line whose price you want to deflate and then right click to select the deflate option,)

If an observation exists prior to or at same period of that of base period, the user has to delete them before the estimation can be made.

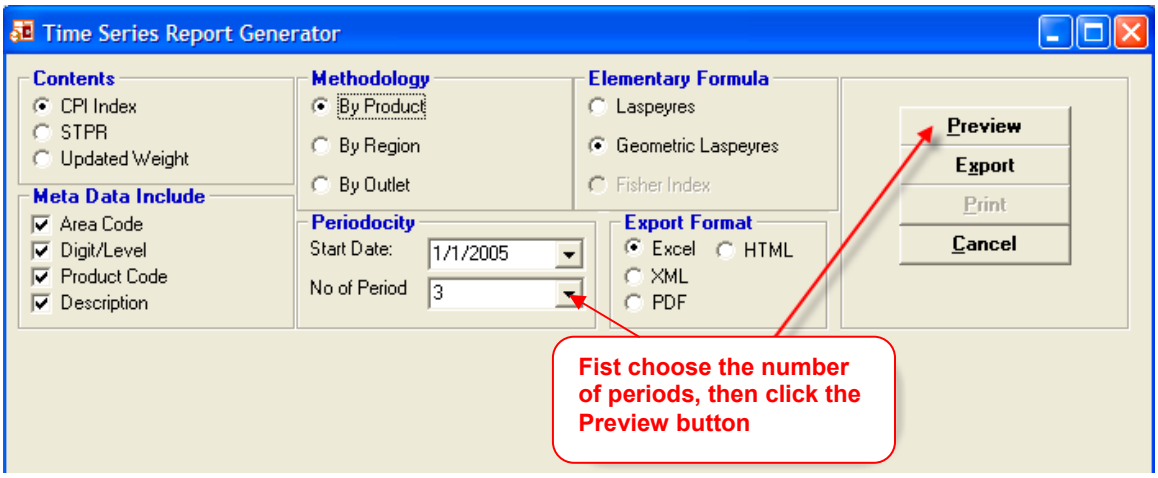


Delete the value by selecting the first column for the observation, right click and click the **delete** button as shown in the diagram below. (This operation does not work if you try to use the delete button on the right of the screen; you must use the “delete” option in the drop-down menu.)

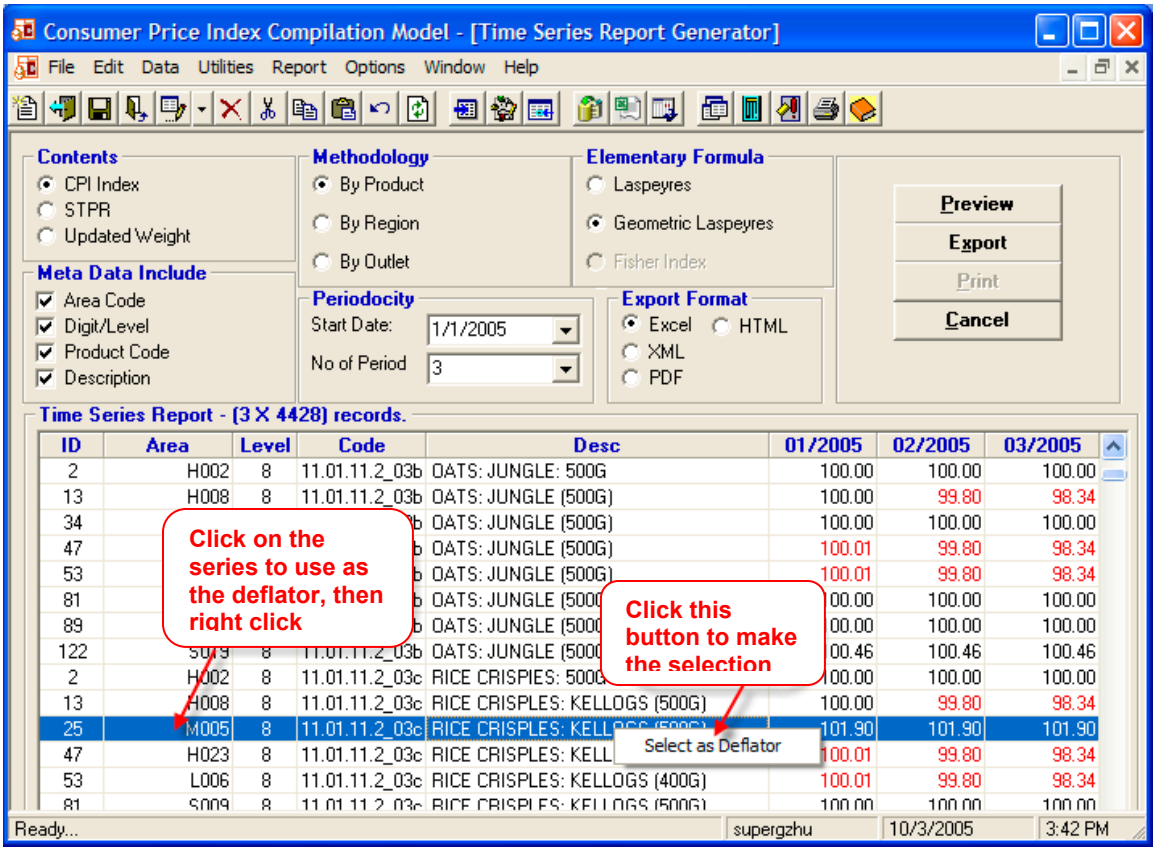
Then click the **Estimate Base Price** again.



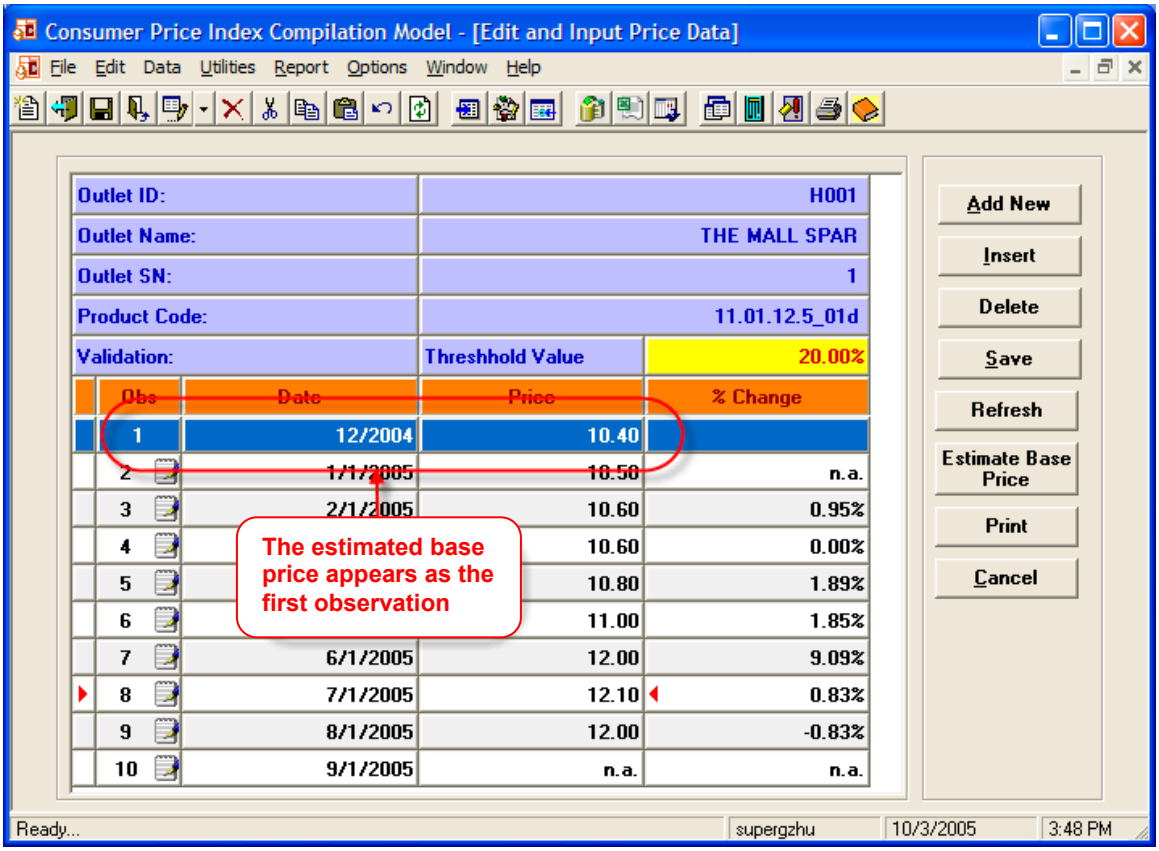
Click OK button to continue. At the next screen chose the number of periods to preview then click the **Preview** button.



Then select a time series to use as the **deflator** for estimating the base period price:



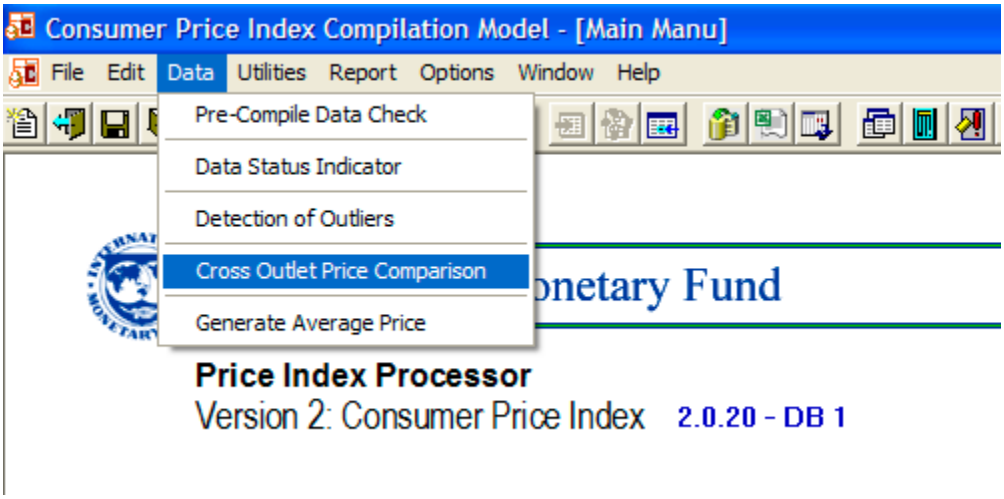
The System estimates the base period price as shown below:



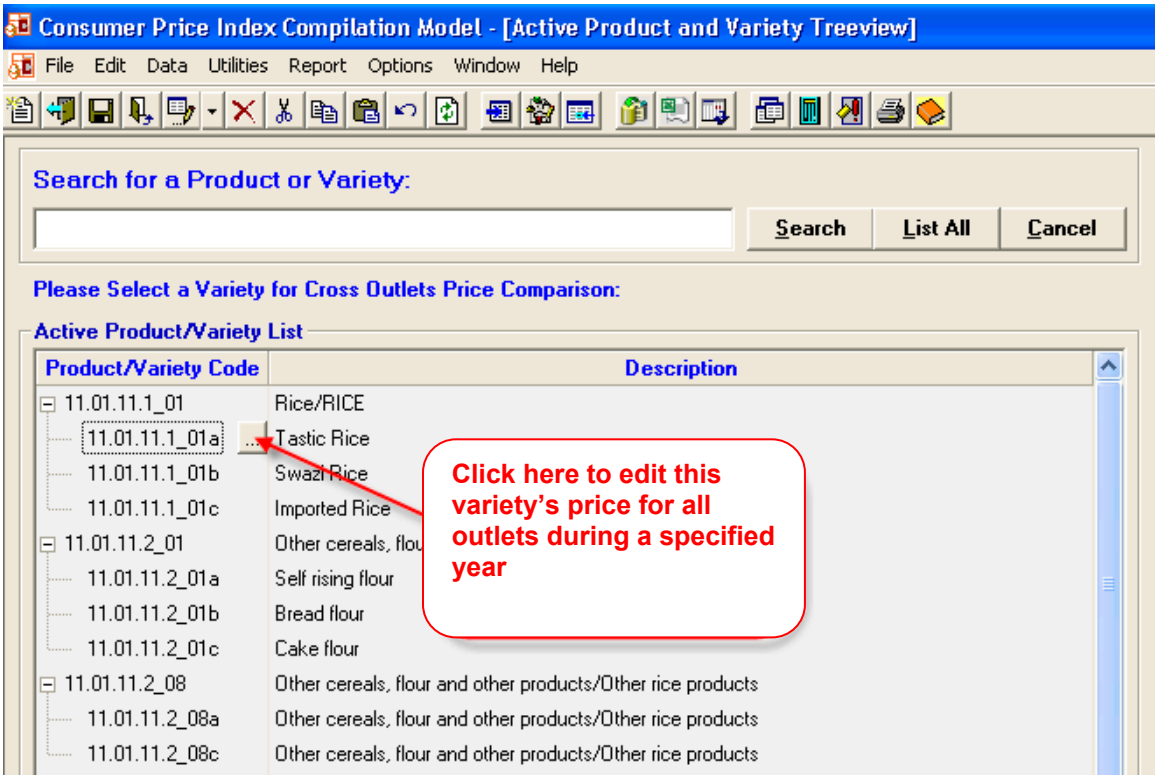
Click the **Save** button to save the price information.

Method 4, Enter/Update Price for Multiple Outlets and for Single Variety

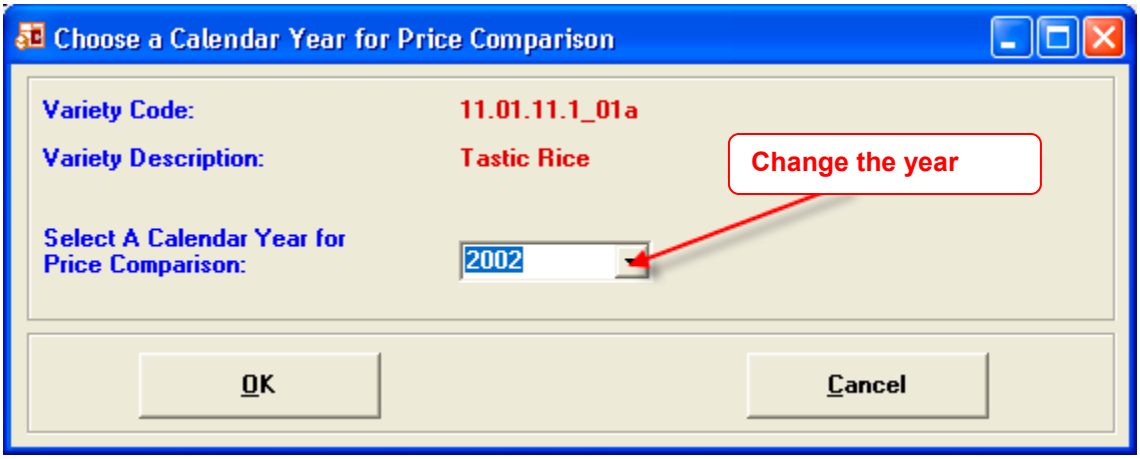
The user can also enter or edit the price for multiple outlets for a single variety. Click **Data**, **Cross Outlet Price Comparison** button as showed below:



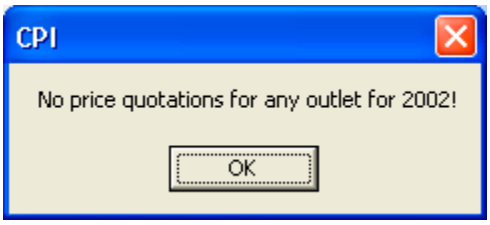
The user will see following screen:



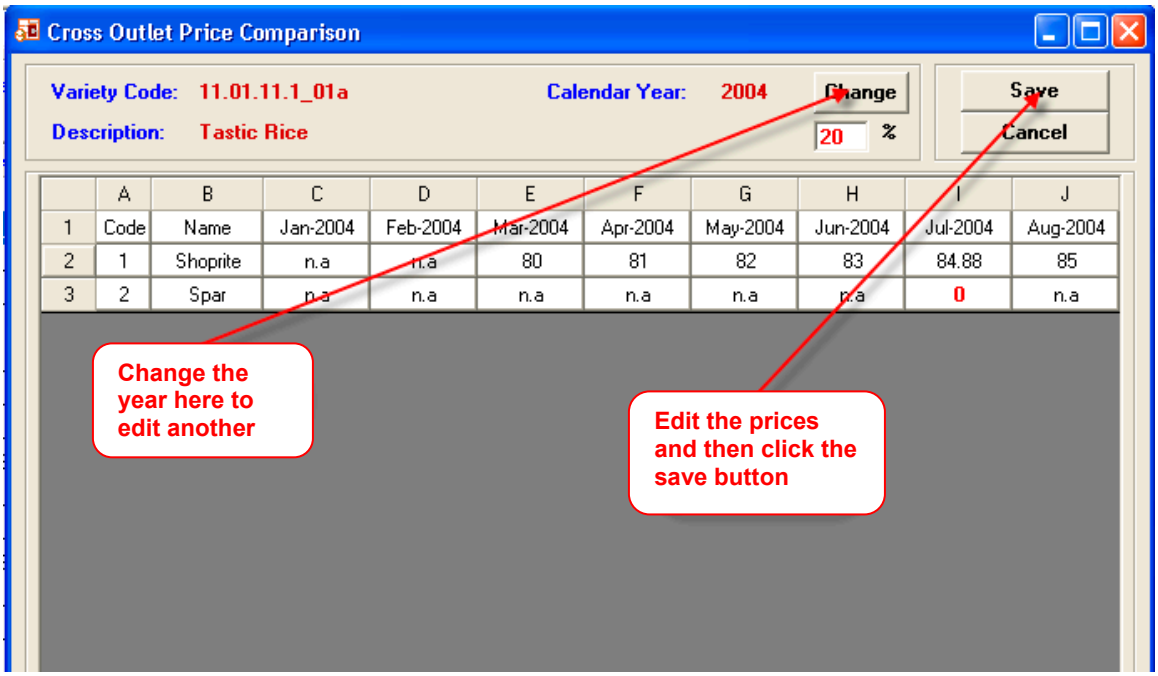
The user can click the grey button to edit the prices for variety 11.01.11.1_01a. The following screen appears to select a specific year:



Select year and then click the **OK** button. If no data is available, the user will see this message box:



Otherwise, the user will see:



	A	B	C	D	E	F	G	H	I	J
1	Code	Name	Jan-2004	Feb-2004	Mar-2004	Apr-2004	May-2004	Jun-2004	Jul-2004	Aug-2004
2	1	Shoprite	n.a	n.a	80	81	82	83	84.88	85
3	2	Spar	n.a	n.a	n.a	n.a	n.a	n.a	0	n.a

The user can change or input prices here and the click the **Save** button.

Treatment of Substitutions and Quality Adjustments

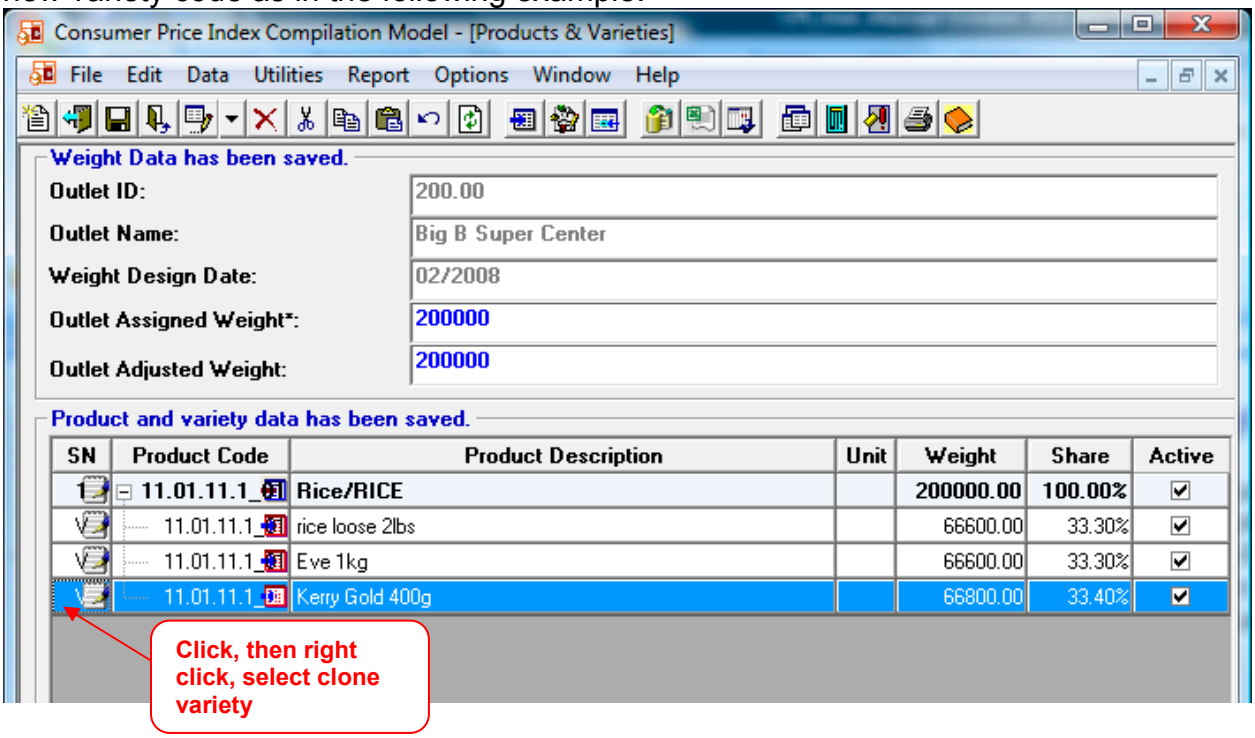
Over time, some varieties will become permanently unavailable and the outlet will no longer sell them. Often this fact is difficult for the enumerators to determine and they must contact the manager responsible for the product to make this determination. Once it is clear that the variety will no longer be available in that outlet, the enumerator will need to replace the it with a similar variety that consumer continue to purchase. The System is capable of handling this situation.

There are three ways to enter the price of a substitute:

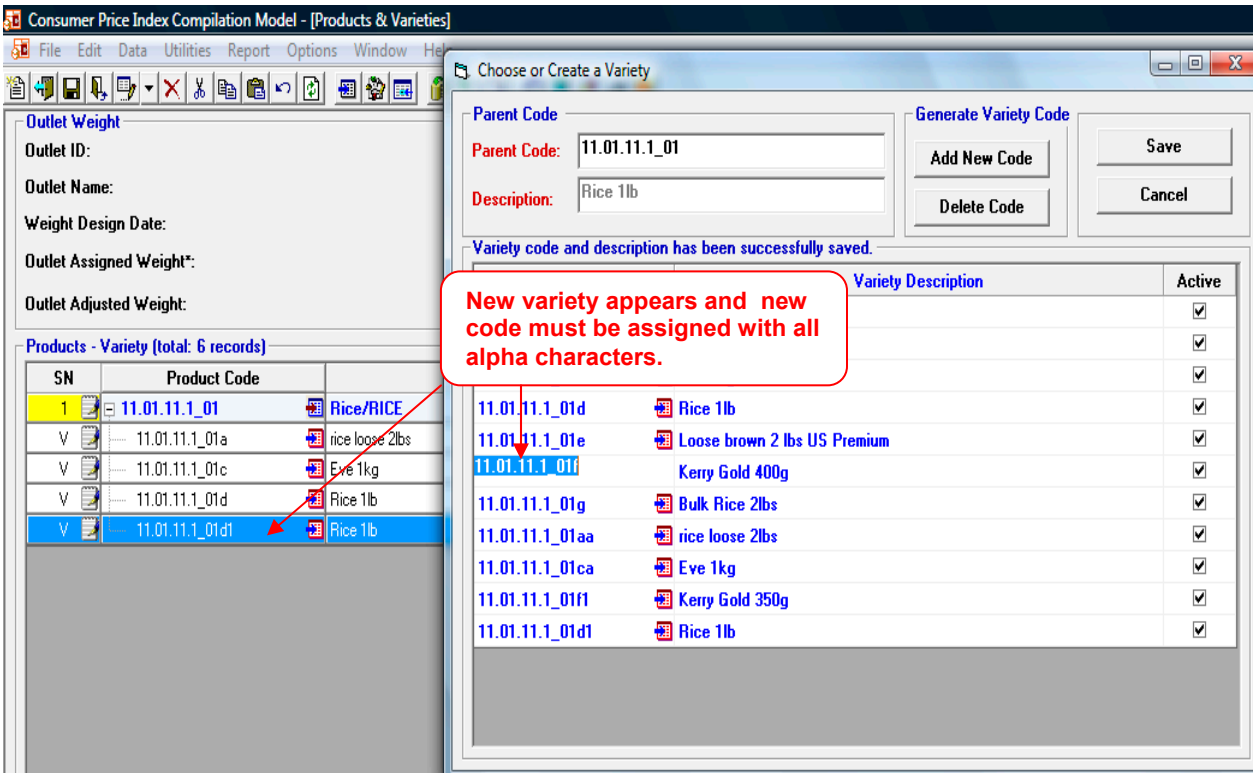
i) if the variety description is practically same as the previous one, the price of the replacement can simply be entered to continue the time-series;
ii) enter the replacement as a new variety and keep the price history of the old variety the same for the new variety; and
iii) enter the replacement as a new variety with no price history.

In the first case, the user can simply add a note to the variety metadata (click the icon on the variety line) stating that a replacement was made on certain date and the only change was a minor detail such as the color or some other characteristic that does not affect the price.

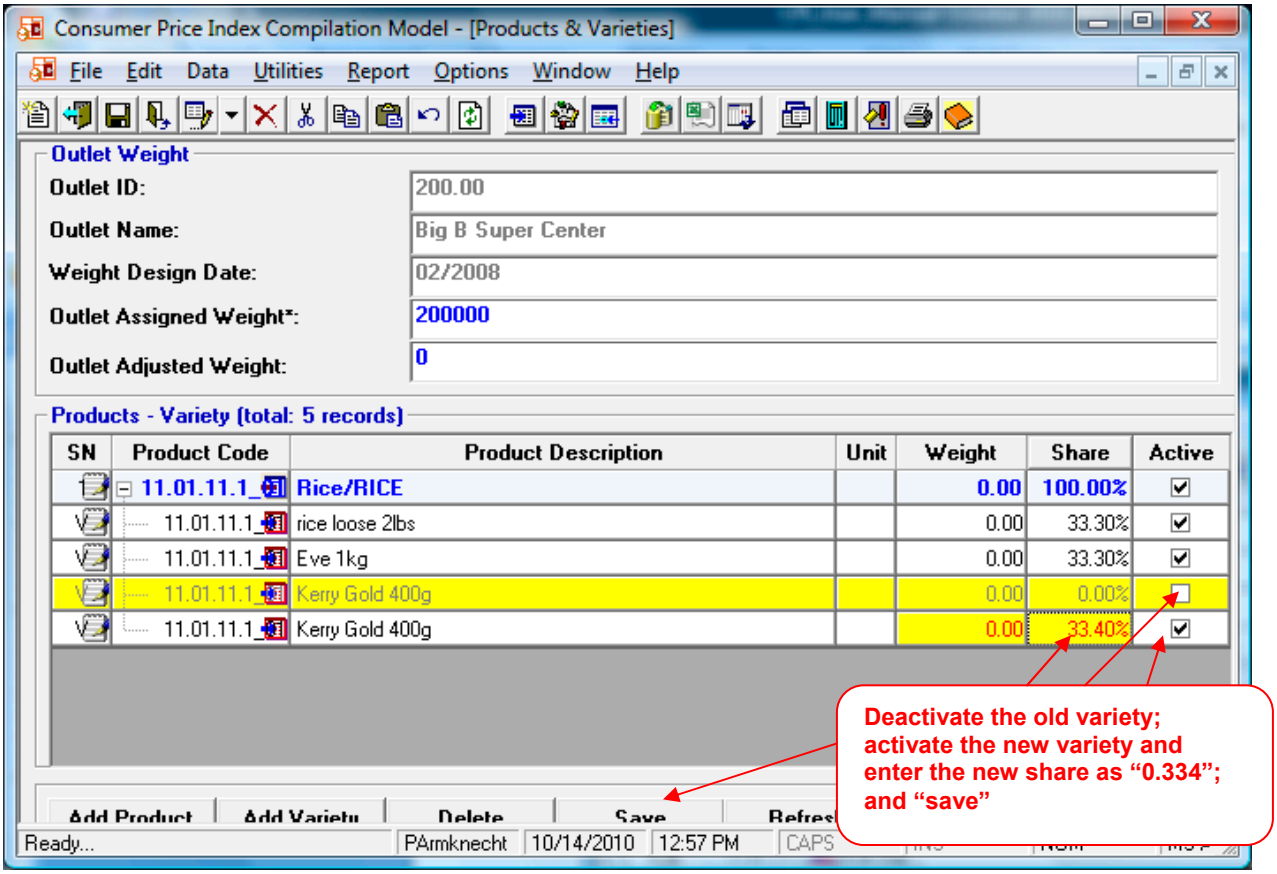
In the second case, the user can clone the old variety to create a new variety, with a new variety code as in the following example.



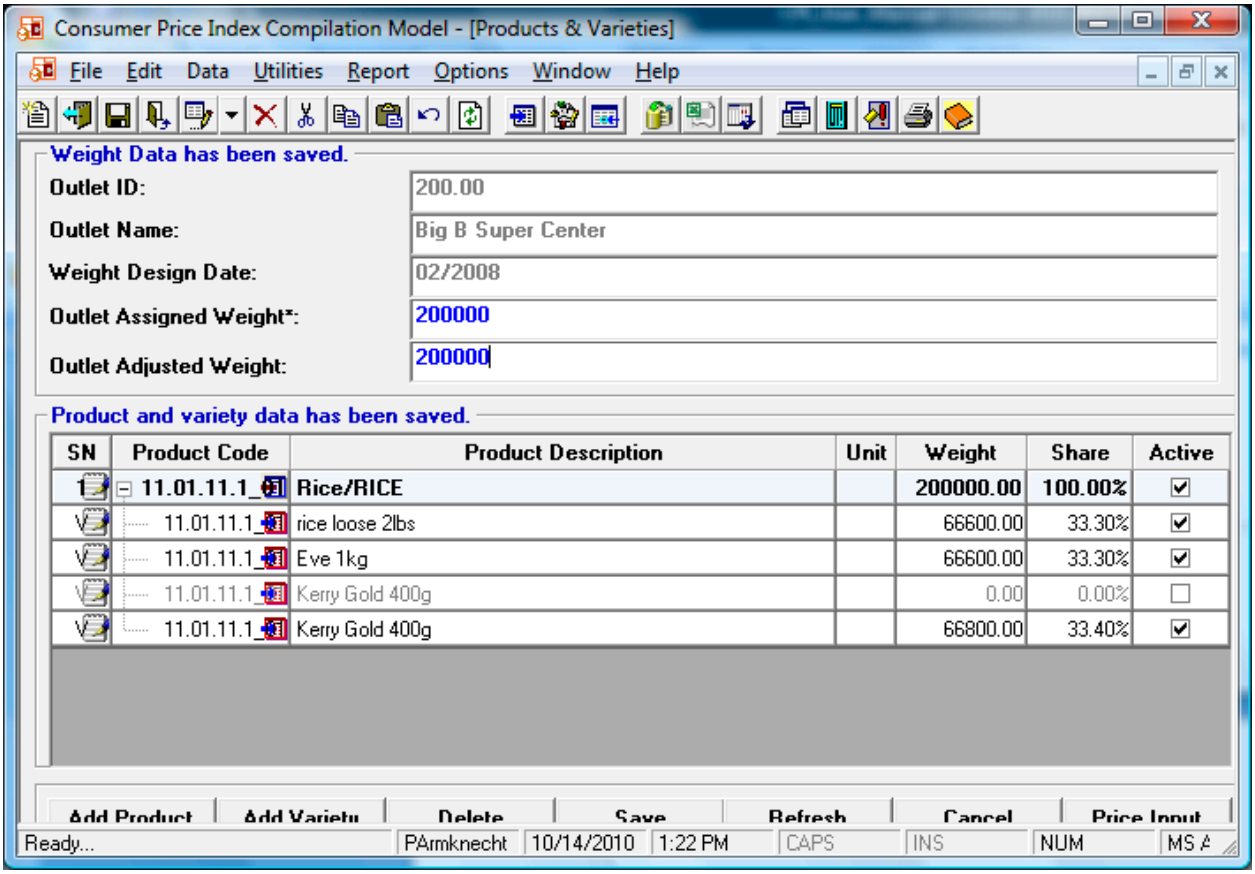
Open the edit Product screen. Click in the left corner of the variety to clone, and then right click to open the drop-down menu. Select “clone variety”. A new variety appears with the same variety description. The variety code has a “1” added as the last digit. This must be changed to an alphabetic character, e.g. “a” in the variety code field and in the variety code list. The user should add metadata in the notes for this new variety to describe the reasons for its creation.



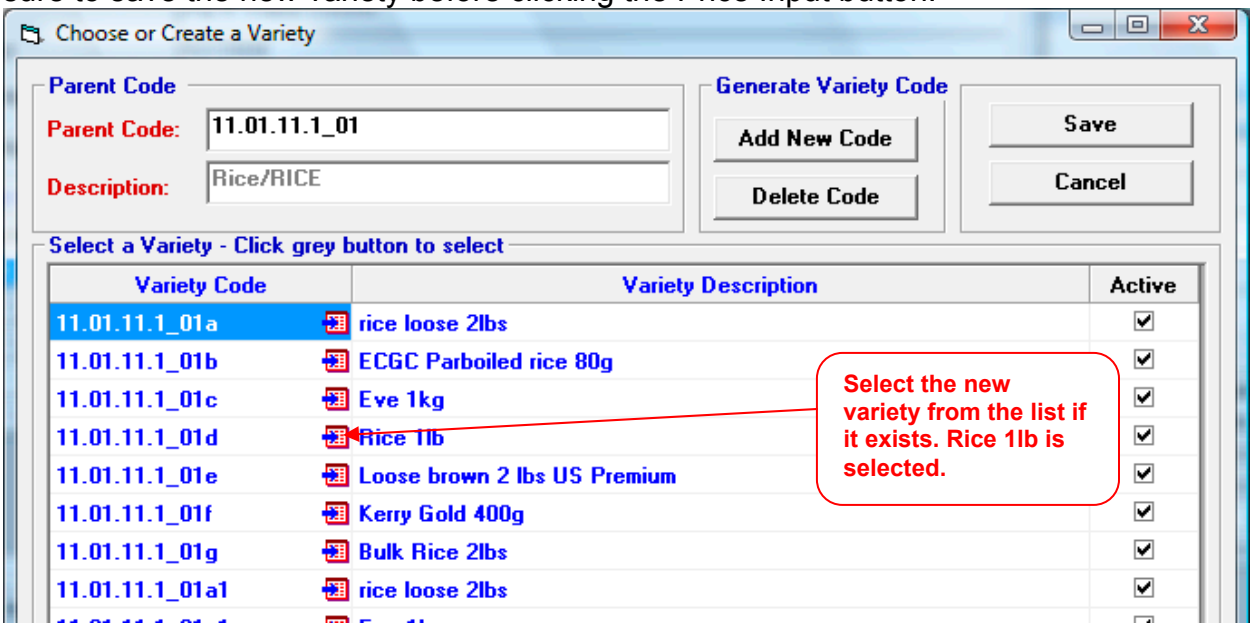
To use the new variety, the user should deactivate the old variety by clicking the "Active" box in the old variety's line, and activate the new variety by clicking the "Active" box on the new variety's line. When the old variety is not active, the share and weights become "0". The user must set the new variety weight to the old variety's weight. Enter 0.334 into the "Share" field to set the share as 33.4 % (Note that the system does not work if you try to enter 33.4 or .334; you must enter 0.334.)



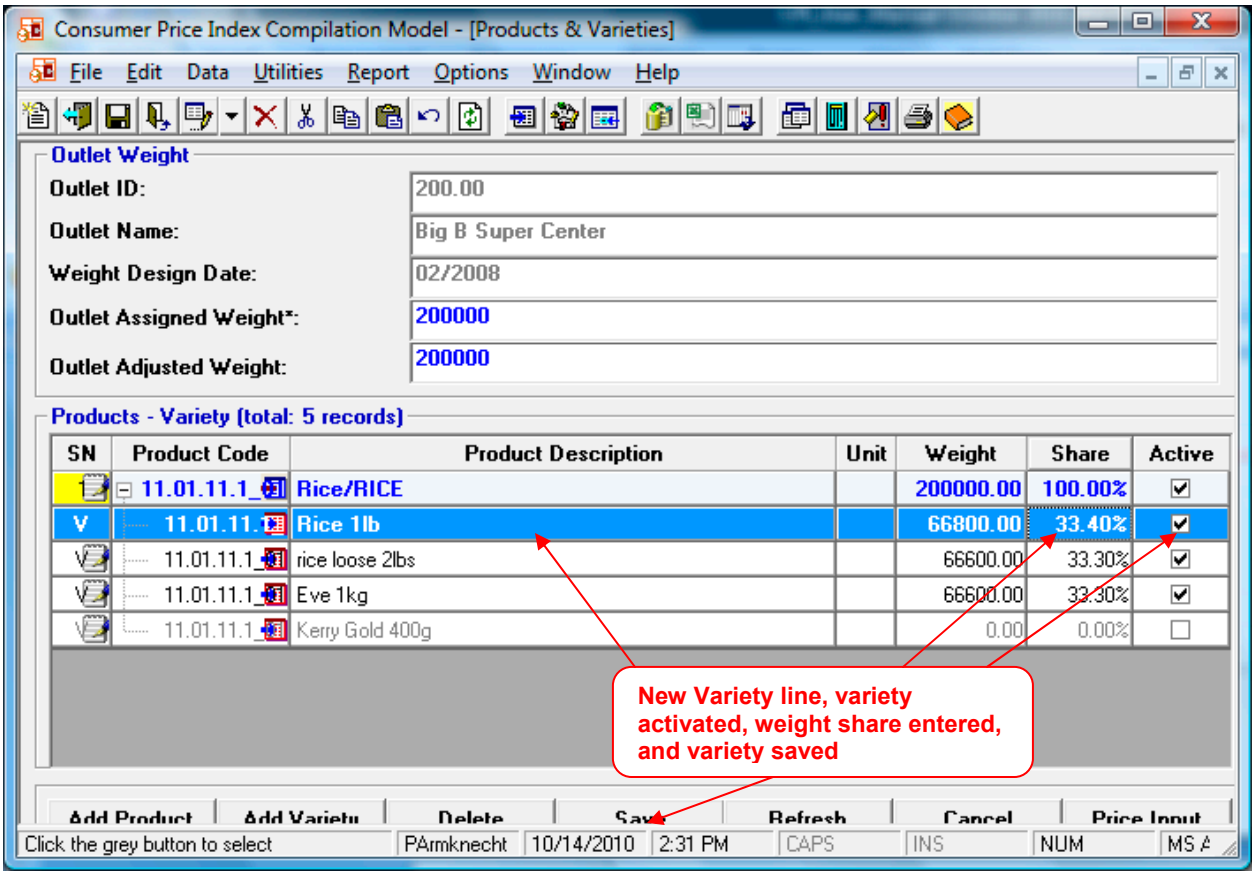
Click Save. The weights for the new variety will be set and the old variety will be inactive, which shows as grey text, as shown in the following screen:



For the third case, the user enters a new variety directly (no cloning involved). Open the Product screen, click on the left column product number, and add the new variety. The following variety list screen appears. The new variety may have the same description as an existing variety in the list or the user may have to enter a new variety description. Be sure to save the new variety before clicking the Price Input button.



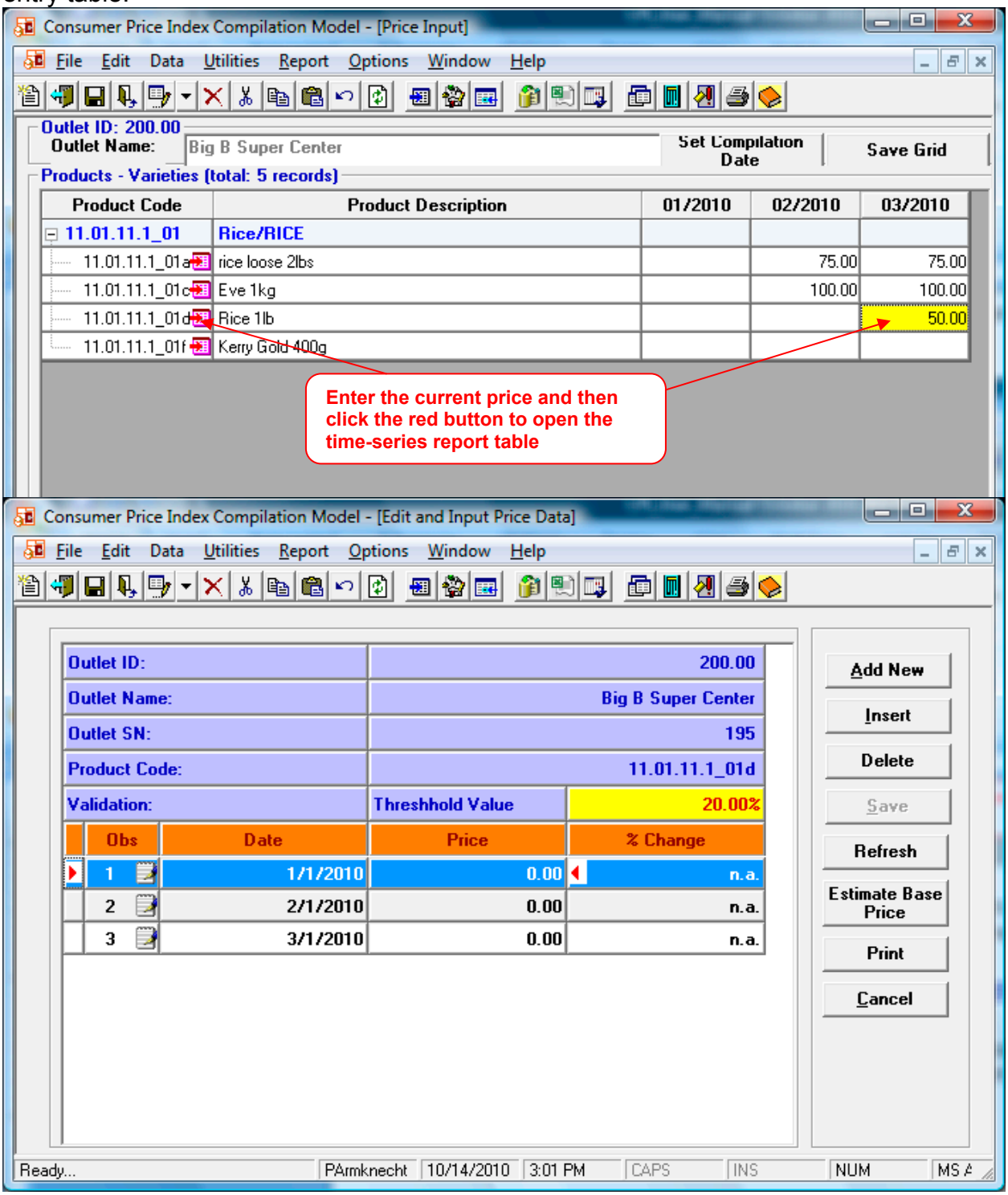
In this example, the replacement variety is 1 lb of rice that the user found already exists on the list. After clicking the red button to the left of “Rice 1lb”, the description appears in the Product screen below the product line. The user must deactivate the old variety (Kerry Gold 400g) and enter the share for the new variety using that for the old variety (0.334).



After saving, the user opens the Price Input screen. Enter the current price, and if the previous price was available, enter that price also. If using the recommended two-stage estimation method (Two-Stage box selected for compilation), no further action is needed. If using the long-term relative method, the user should proceed to estimate the base price externally or have the system estimate a base price.

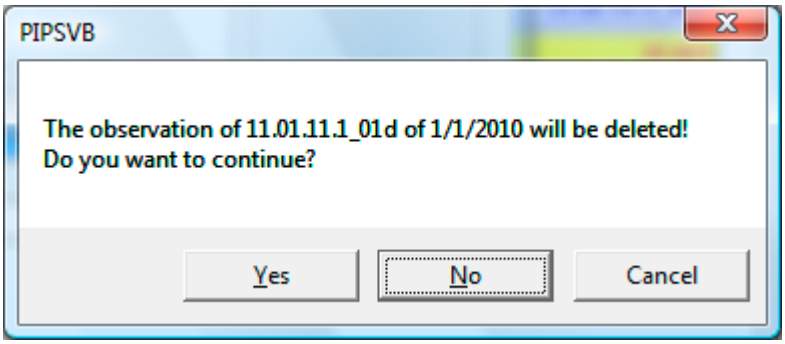
The user may also estimate the base price externally by taking the ratio of the new variety’s previous price to old variety’s price in the same period. This provides a quality adjustment (QA) ratio. Assume the price of Rice 1lb. in this outlet was 50 in the previous period and the price of Kerry Gold 400g. was 40. The QA ration is 50 (new variety’s price) divided by 40 (old variety’s price) to yield a value of 1.20. The QA ratio is multiplied by the base period price of the old variety to raise it to the level of the new quality. If the Kerry Gold’s base price was 37.0, then the estimated base price for Rice 1lb. in the this outlet is 44.40 (i.e., 1.20 x 37.0 = 44.40). The user will enter this price directly into the system as the price in 1/1/2010 as the externally derive price.

To have the system estimate the base price, proceed as follows: After entering the price for the new variety, click on the red button next to the description to open time-series entry table.

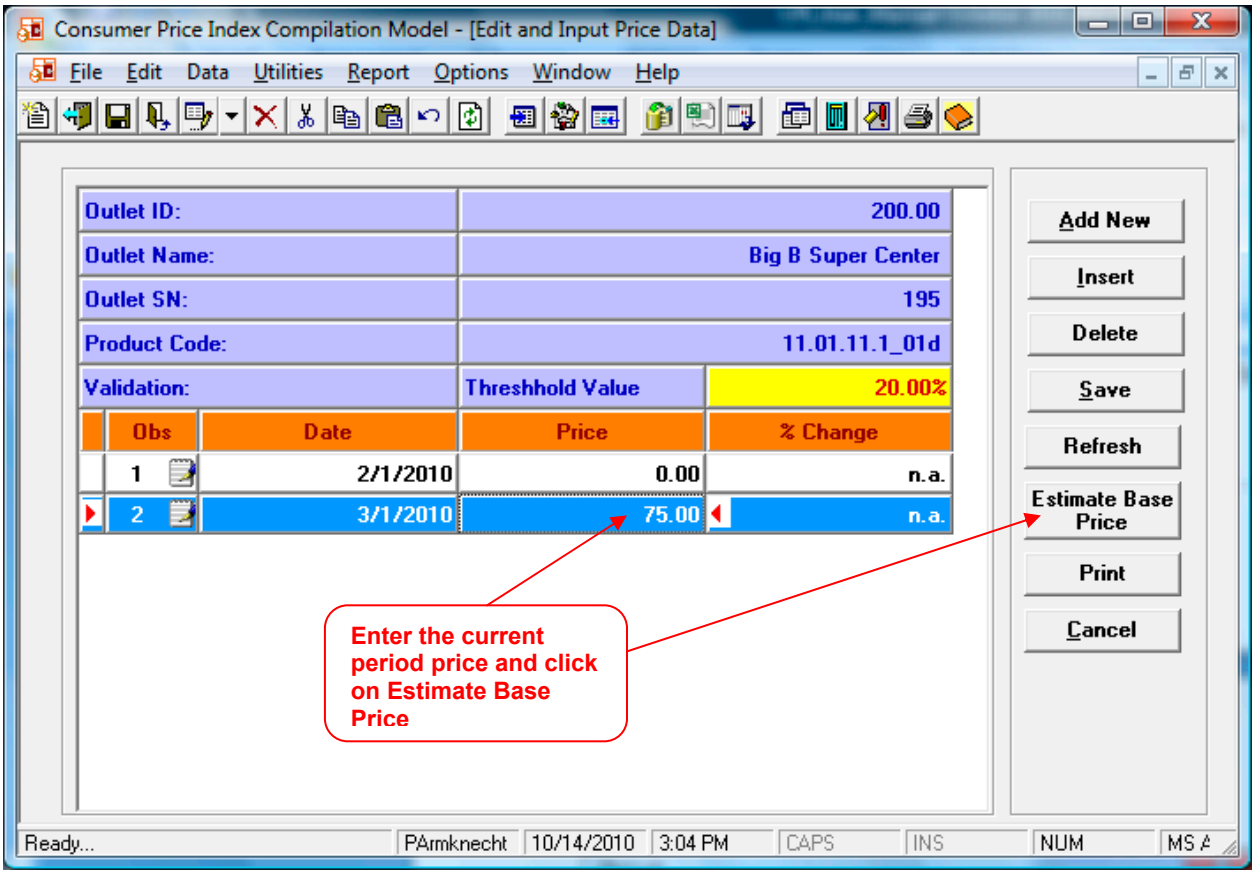


In this table, the base price month (1/1/2010) appears already. Before the system will estimate the base price, the user has to delete this line. Click in the left most box and

then right click to select delete. After selecting delete, the following box appears; choose Yes

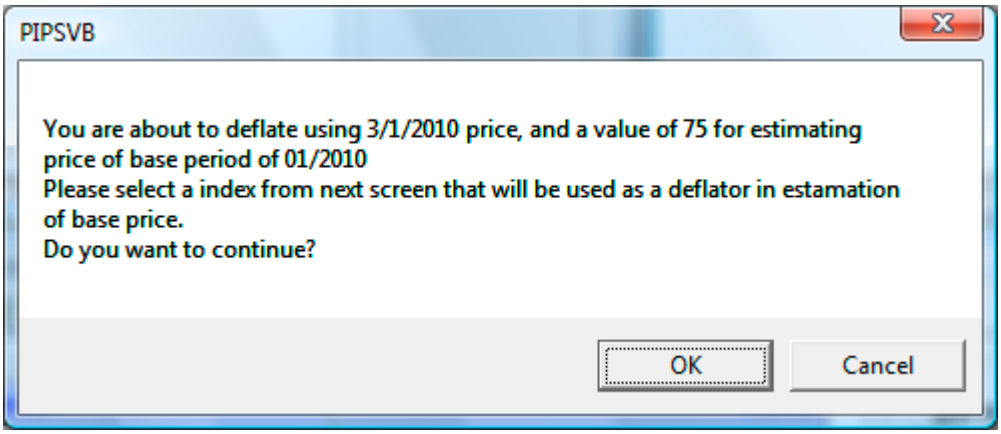


Enter the price for the current month (3/1/3010) and the previous month, if available. In this example, it is not available. Keep the current month line as the selection and then click the "Estimate Base Price" button (or right click and select deflate price).

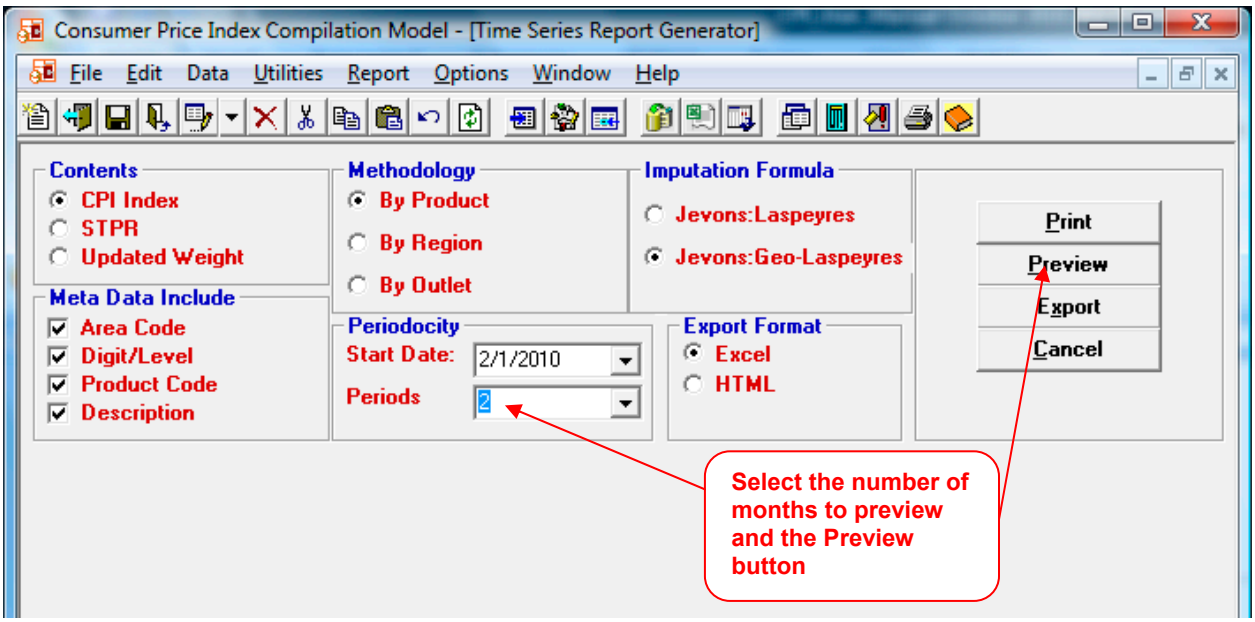


The following box appears and asks if you want deflate the value for the current month to estimate the price in the base price month. Click OK. (Note that the user can only complete this step after an index for the current month is compiled. The user would have to do a preliminary compile and save the resulting indexes to the database. After completing the base price estimate, the user would re-compile the indexes and save

them to the database again. Otherwise, the user would need the previous period price for this calculation along with the previous period index.)



The following screen appears and the user selects the number of periods to preview and the Preview button. The System will retrieve aggregate indexes for the periods so that the user may select the index to use as a deflator. The user should select the index for the product to which this variety is belongs.



The system provides a long list of indexes for selection. Scroll down the list to find an aggregate Product code for this variety. The list provides indexes for individual varieties as well as products in outlets. The user should use an aggregate index for which the ID column has a “0” value. Other values relate to outlet codes in the field. Once the user had identified the index (in the example, the index for Rice), click on the line, right click, and then click on the box saying “select deflator?”

Consumer Price Index Compilation Model - [Time Series Report Generator]

File Edit Data Utilities Report Options Window Help

Contents: CPI Index / STPR / Updated Weight

Meta Data Include: Area Code / Digit/Level / Product Code / Description

Methodology: By Product / By Region / By Outlet

Periodocity: Start Date: 2/1/2010; Periods: 2

Imputation Formula: Jevons:Laspeyres / Jevons:Geo-Laspeyres

Export Format: Excel / HTML

Print / Preview / Export / Cancel

Select the index series to use as the deflator, right click, and then click in the box "Use as Deflator"

Time Series Report - (2 X 4920) records.

ID	Area	ev	Code	Desc	02/2	
0		0	0	All Products	101.99	103.30
0		1	1	GROSS DOMESTIC PRODUCT	101.99	103.30
0		2	11	INDIVIDUAL CONSUMPTION EXPENDI	101.99	103.30
0		3	11.01	FOOD AND NON-ALCOHOLIC BEVERA(	104.18	104.72
0		4	11.01.1	FOOD	104.02	104.64
0		5	11.01.11	Bread and cereals	101.82	100.68
0		6	11.01.11.1	Rice	99.05	93.77
0		6	11.01.11.2	Other cereals, flour and other products	102.47	102.78
0		6	11.01.11.3	Bread	100.16	100.11

Ready... | PArmknecht | 10/14/2010 | 3:28 PM | CAPS | INS | NUM | MS A

The System returns the following screen. Click "Save" and note that the base price appears in red because it is an imputed price.

Consumer Price Index Compilation Model - [Edit and Input Price Data]

File Edit Data Utilities Report Options Window Help

Outlet ID:	200.00	
Outlet Name:	Big B Super Center	
Outlet SN:	195	
Product Code:	11.01.11.1_01d	
Validation:	Threshhold Value	20.00%

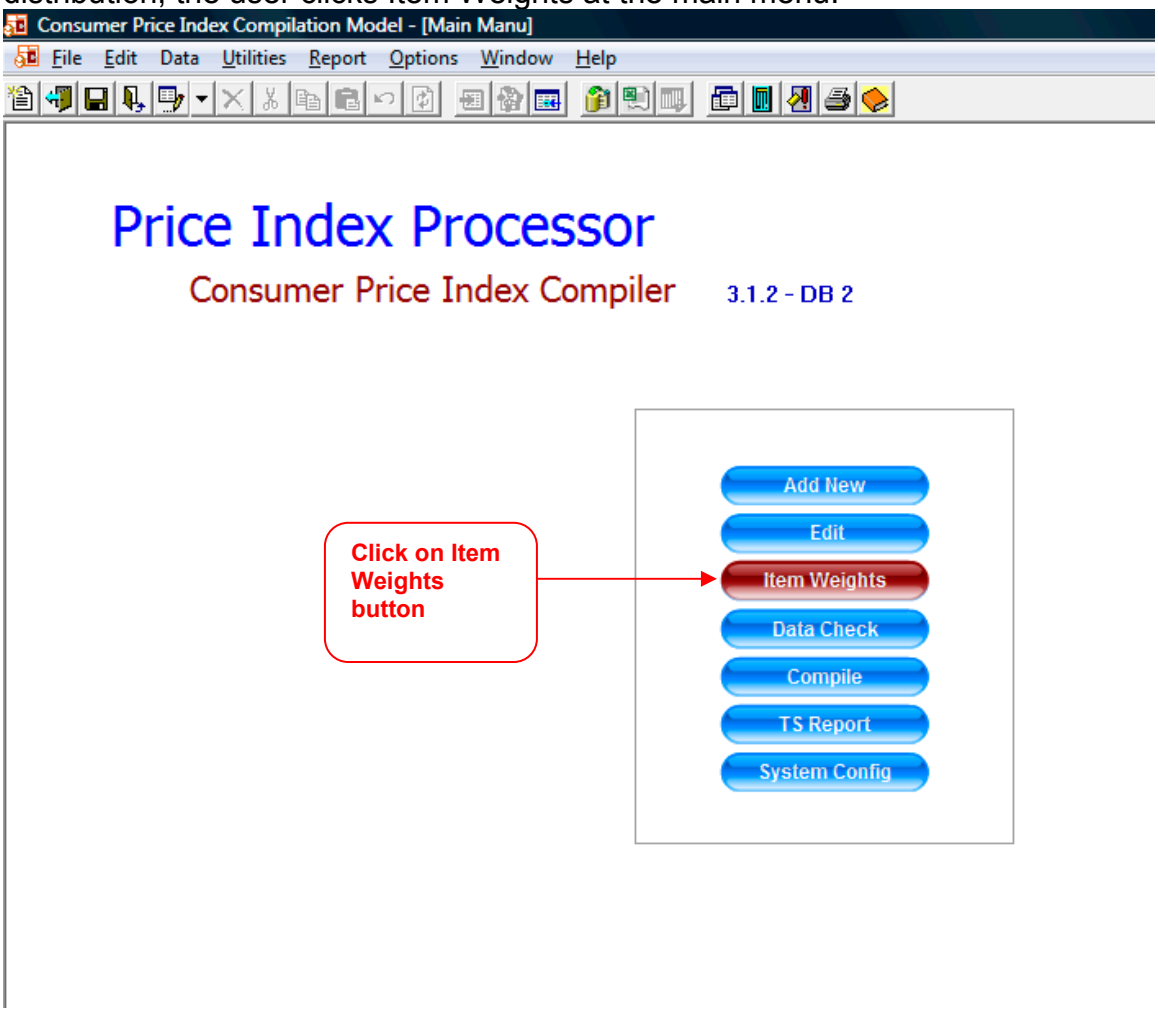
Obs	Date	Price	% Change
1	01/2010	79.99	
2	2/1/2010	0.00	n.a.
3	3/1/2010	75.00	n.a.

Add New / Insert / Delete / Save / Refresh / Estimate Base Price / Print

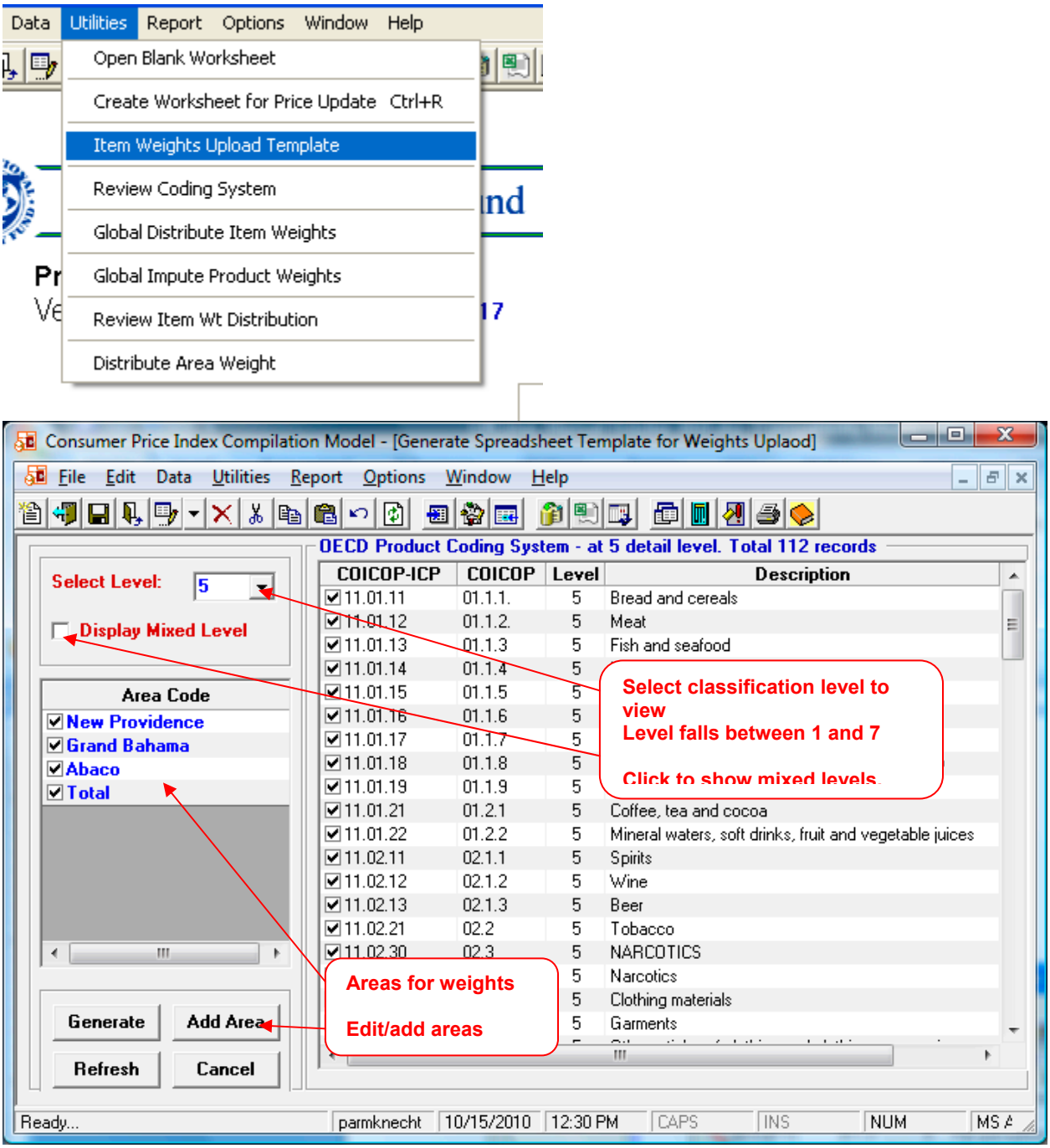
A final example may prove helpful to users for making quality adjustments when a replacement variety arrives. In the current period, the user enters the new variety for the product and its description but does not activate it. They keep the old variety active and allow the System to impute a price. In the following pricing period, the user activates the new variety, deactivates the old variety, and sets the new variety's share weight equal to that of the old variety. In this case, the new variety has both the current period and previous period price available so that the System can use them to compile the index by the two-stage (short-term price relative) estimation method. Alternatively, the user can have the System estimate the base period price using the previous period price for the new variety and the procedure for estimating the base price as explained above.

Item Weights Design and Distribution

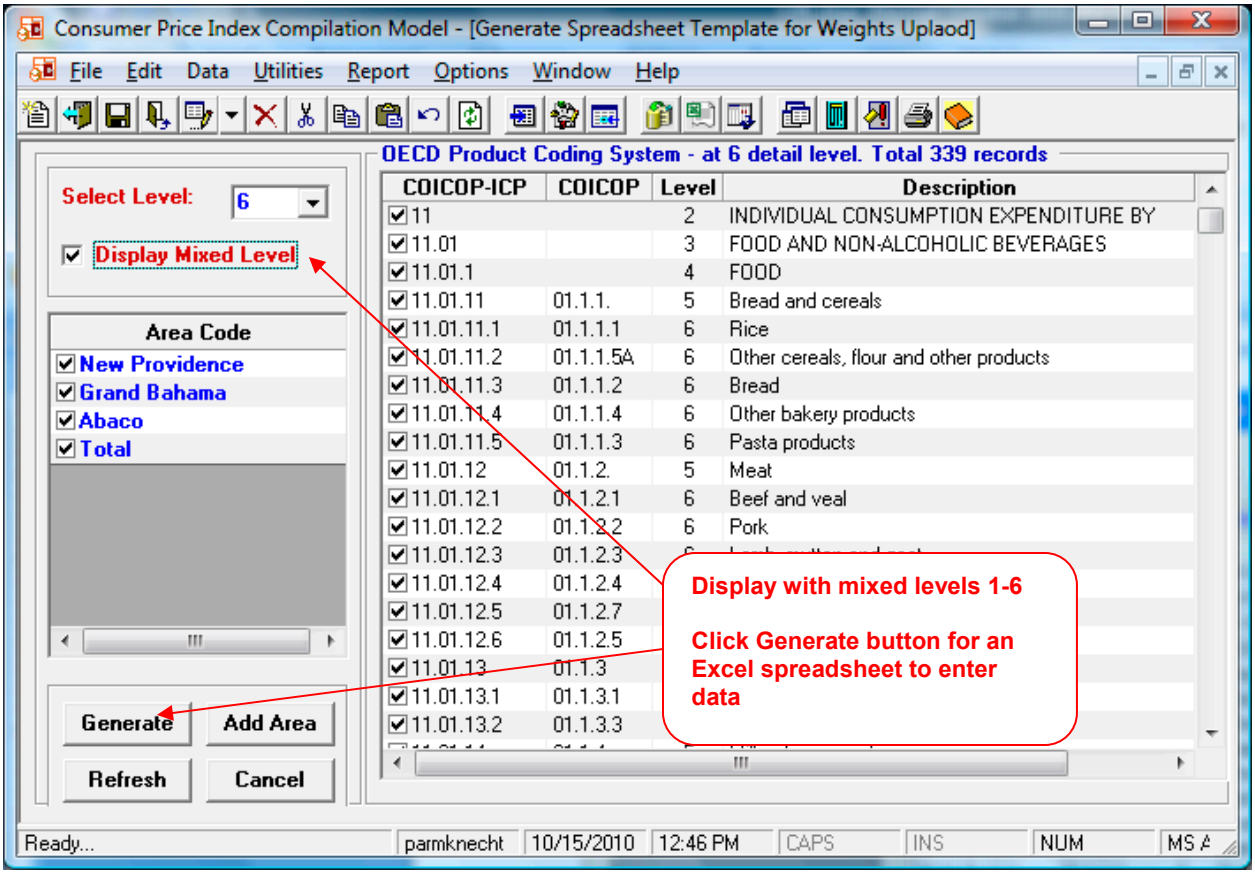
Item weights generally come from a household expenditure survey where expenditure estimates are available by area and by each individual item. To start the item weights distribution, the user clicks Item Weights at the main menu:



Click **Item Weights** button in the main menu or at click Utilities, Item Weights Upload Template button on toolbar:

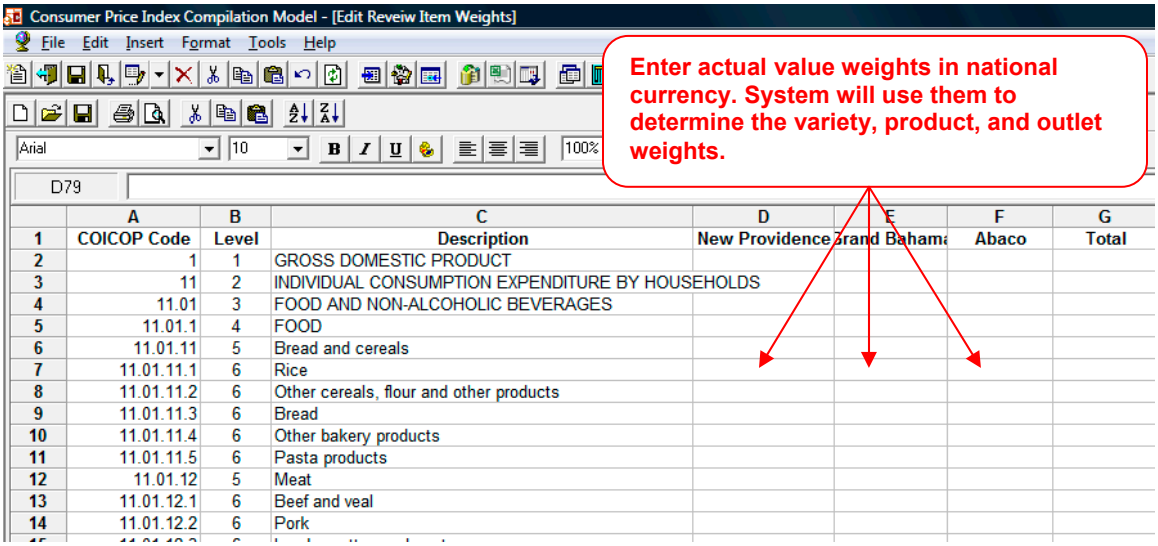


The user should decide the weights level from 1 to 7 according to the specific weight survey data. The user can also select item details at mixed level. The following screen shows that the user weight data is at mixed level:



It is important to note that the user should not select an item from level 5 and 6 at same time. For example, the user should not select "11.01.11 - Bread and Cereals" (level 5 item) and item 11.01.11.3 – Bread (level 6 item) simultaneously. By doing so, only the weight data at more detail level will be distributed to the price quotations underneath.

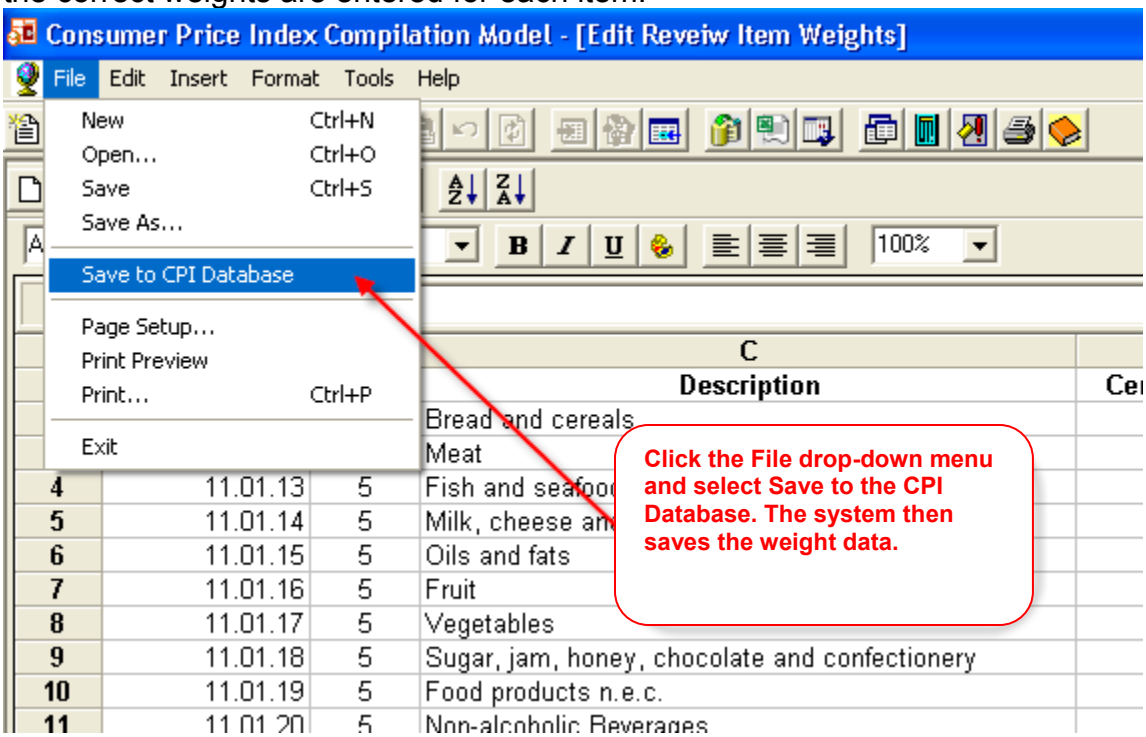
The user clicks the **Generate** button, and inputs the weight data on the following screen:



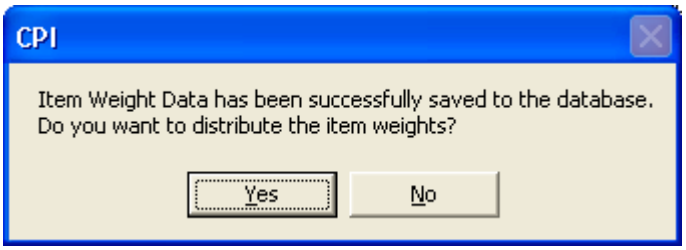
Please note that the user should input the unprocessed weight data in national currency denomination (raw weight survey data), even though correctly normalized data are also acceptable.

To correctly normalize the weight data, sum the columns to get regional totals first; then sum the regional totals to get the national total. Divide each cell by national total to get a fraction of each item in a specific weight area as a percentage of the national total—though the CPI System prefers to have expenditure weight data without normalization.

The user then should save the weight data to the CPI database by clicking **File**, **Save to CPI Database** as shown in the screen below. (The System saves the weight data to the default CPI database; use the system configuration to change the default database.) Note that this data will overlay any data already saved in the database so be sure that the correct weights are entered for each item.



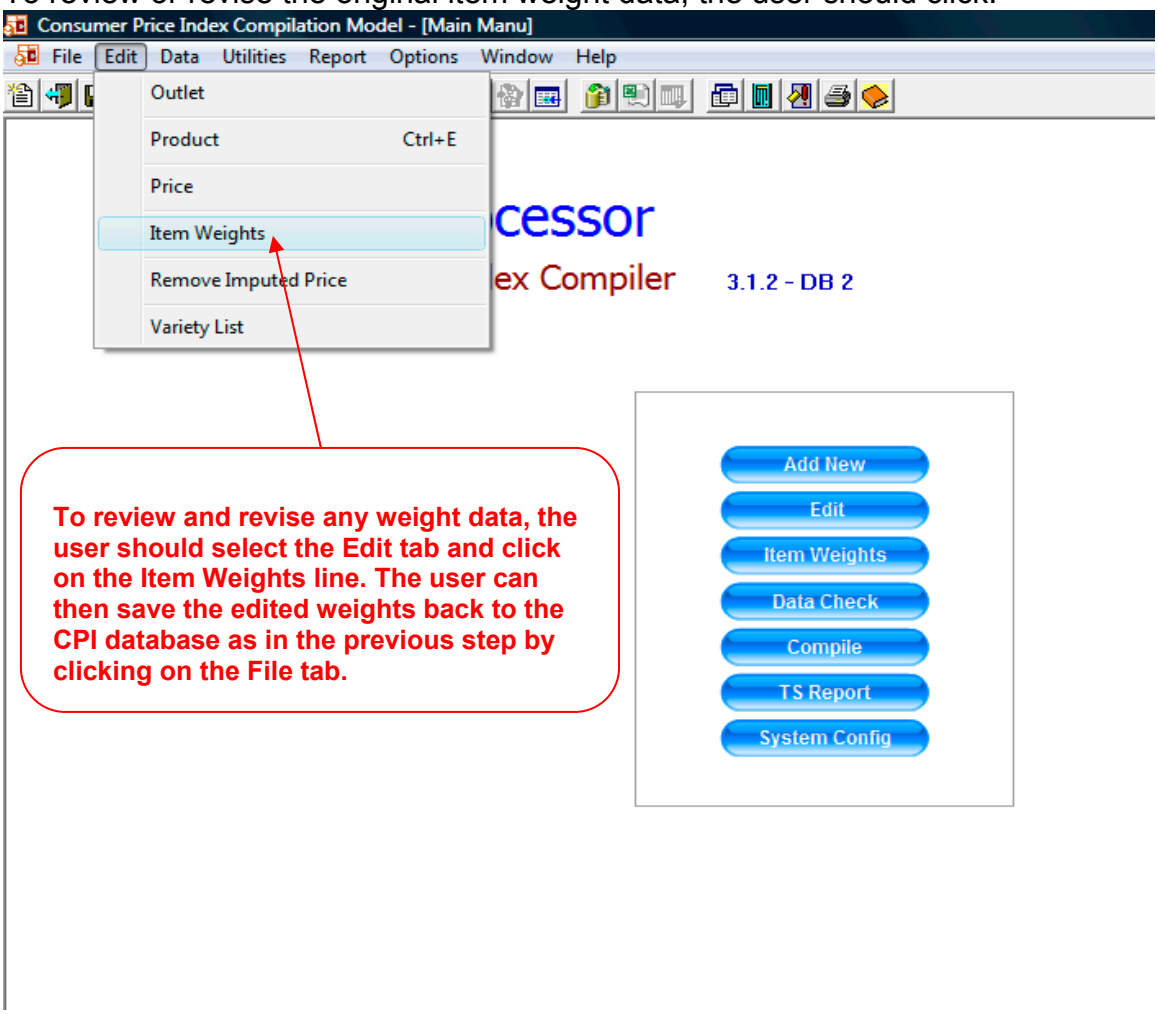
After the user saves the weight data, the following message box displays:



Click the Yes button to have the weights distributed to the price quotations under this item.

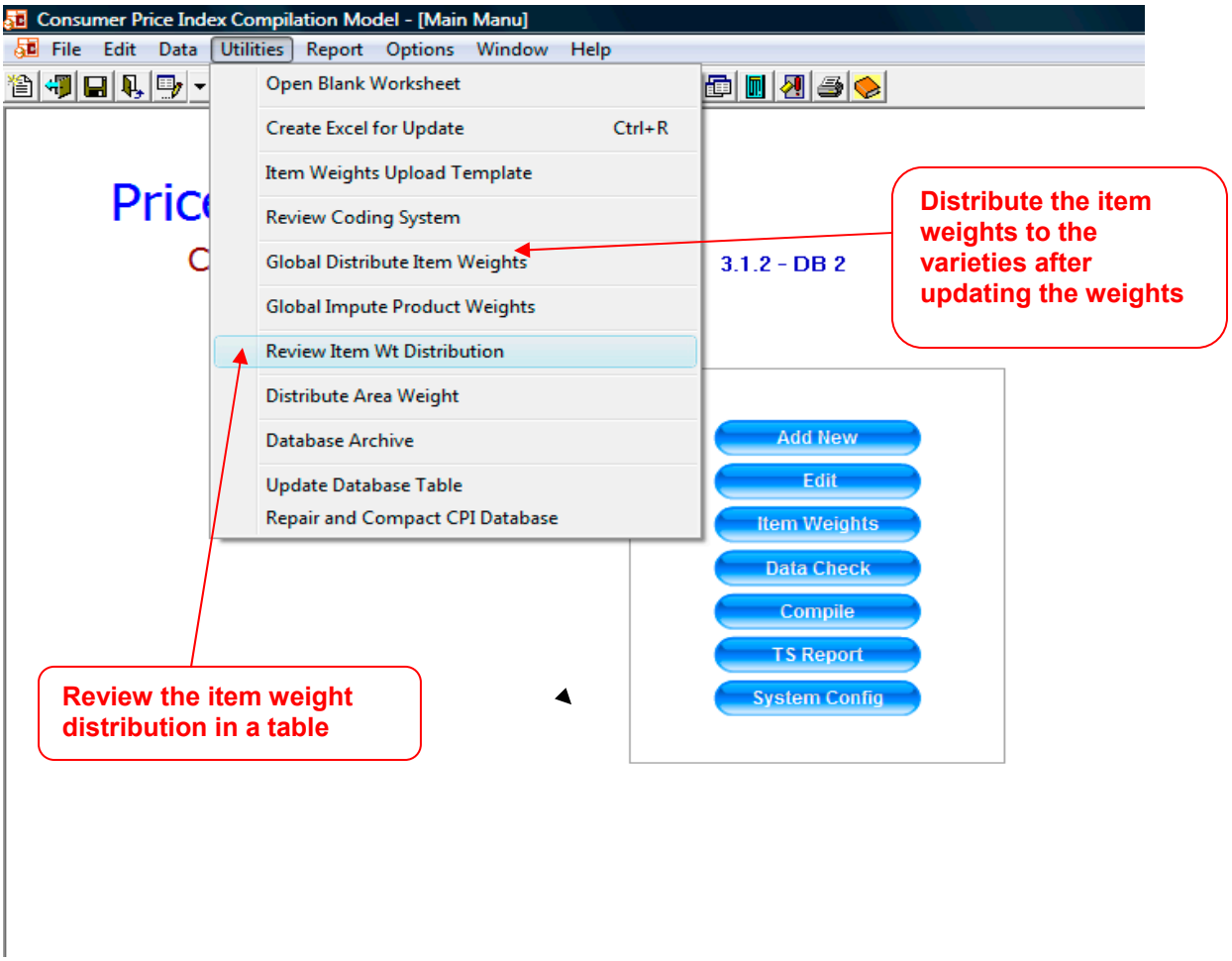
The message box indicates whether the System has distributed the item weights successfully.

To review or revise the original item weight data, the user should click:

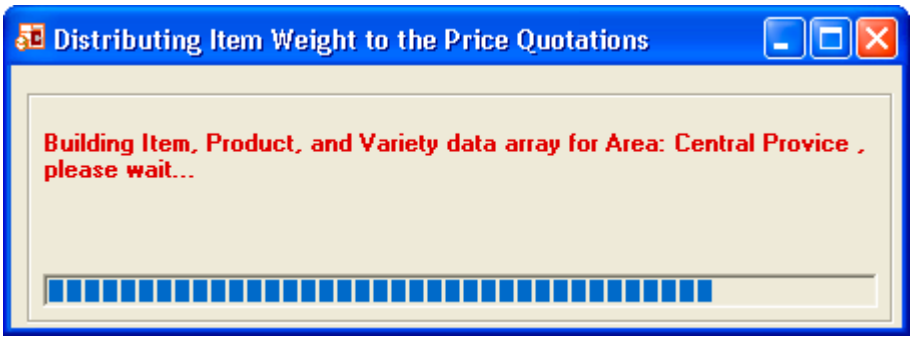


The CPI System can retrieve the weight data from database and display the data in spreadsheet format. The user can revise the item data if necessary, and click Save to CPI Database button.

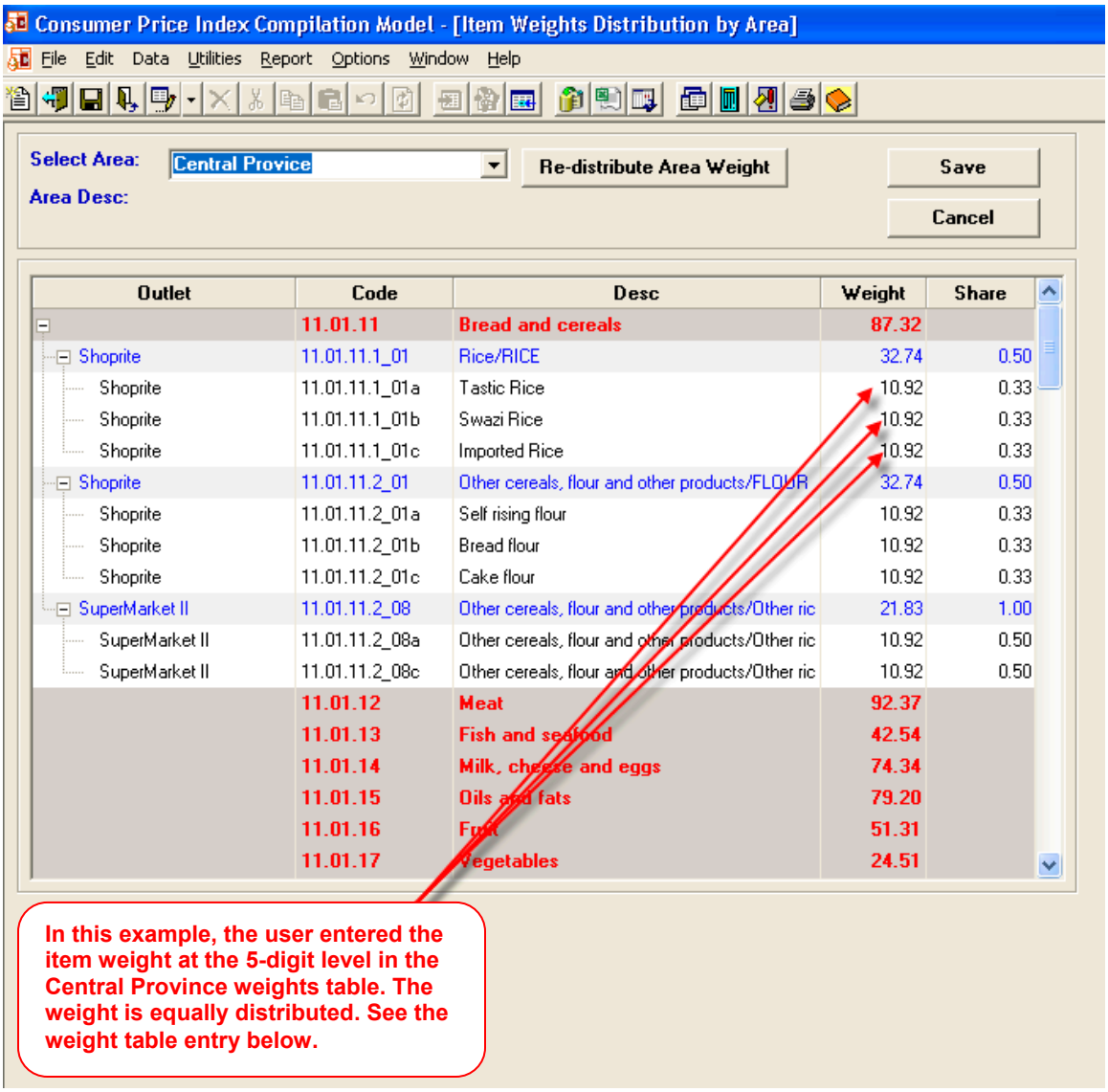
An area item weight will be equally distributed to every price quotation under this item if the user clicks **Utilities**, **Global Distribute Item Weights** button.



It is important to note that the user can review the weight distribution by clicking the **Review Item Wt Distribution** button shown in the screen above.



Wait until the system finishes the process. The user receives notice that the weights are distributed across the outlets.

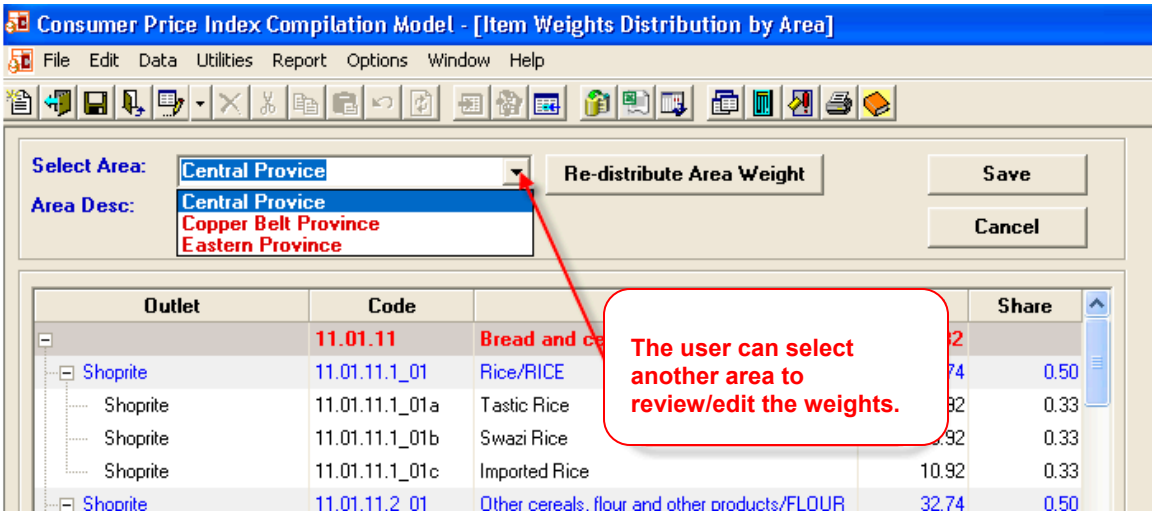


The item weight for Central Province (a weight area) for item 11.01.11 — Bread and cereals — is 87.32 which is equally distributed among the price quotations under it.

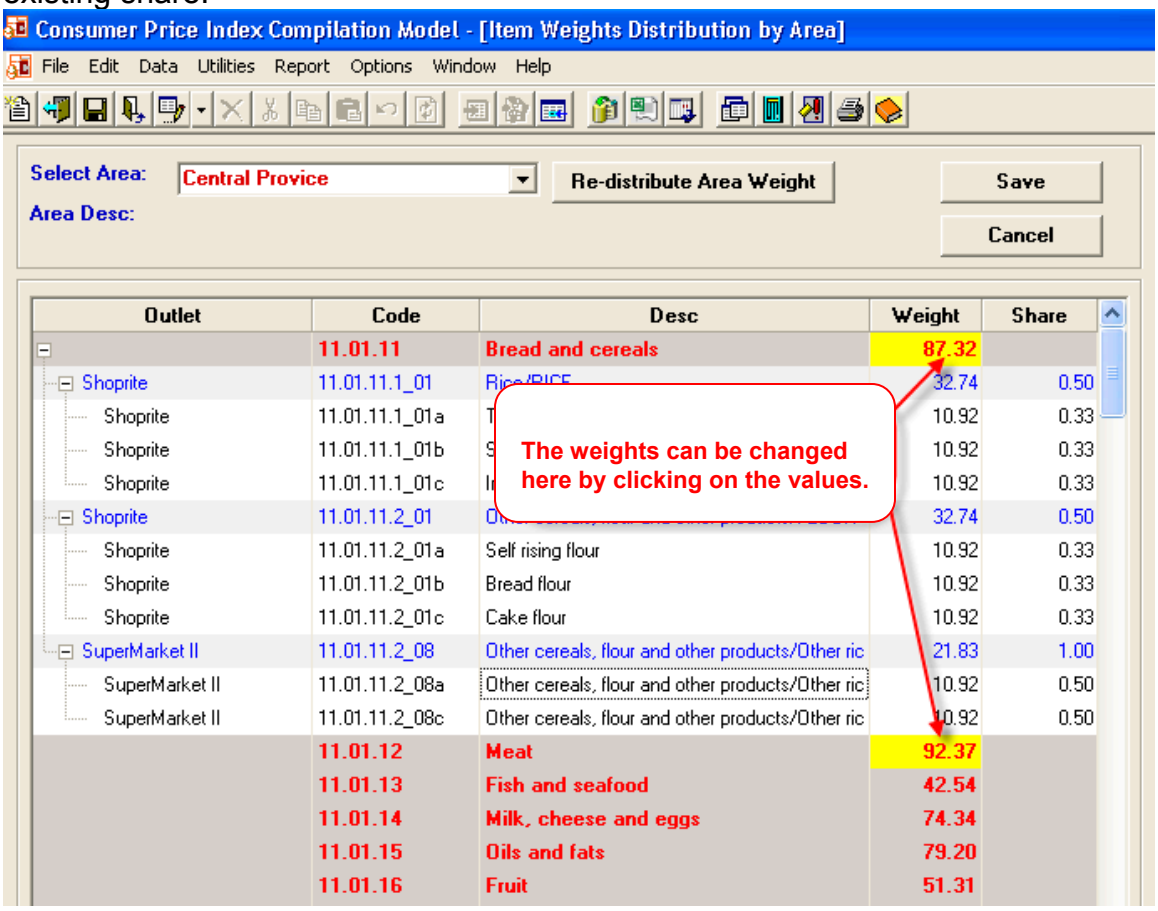
A	B	C	D	E
OECD Code	Level	Description	Central Provice	Copper Belt Pro
11.01.11	5	Bread and cereals	87.32	40.76
11.01.12	5	Meat	92.37	44.97
11.01.13	5	Fish and seafood	42.54	82.28
11.01.14	5	Milk, cheese and eggs	74.34	55.68
11.01.15	5	Oils and fats	79.2	49.97
11.01.16	5	Fruit	51.31	4.87
11.01.17	5	Vegetables	24.51	5.29

This weight is equally distributed.

To view another area's weight distribution, click:



The weight for the particular item for a specific area are editable, Suppose that the user wants to change the weight data for item 11.01.11 for Central Province, they can double click 87,32 and input a new number. This new figure will be redistributed based on the existing share.

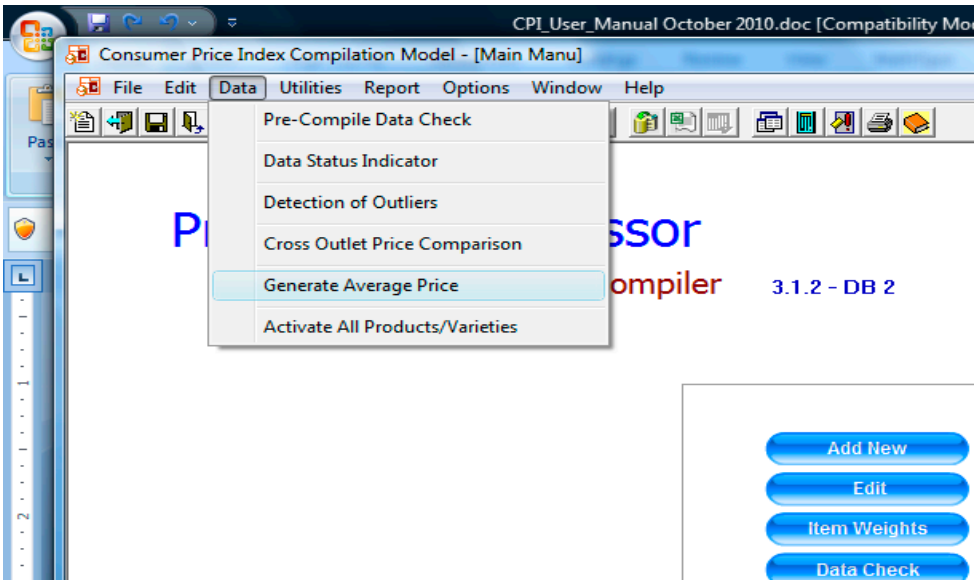


Rows in red indicate the items whose weights are available. If there are products or varieties below these entries, the view can expand via a tree structure.

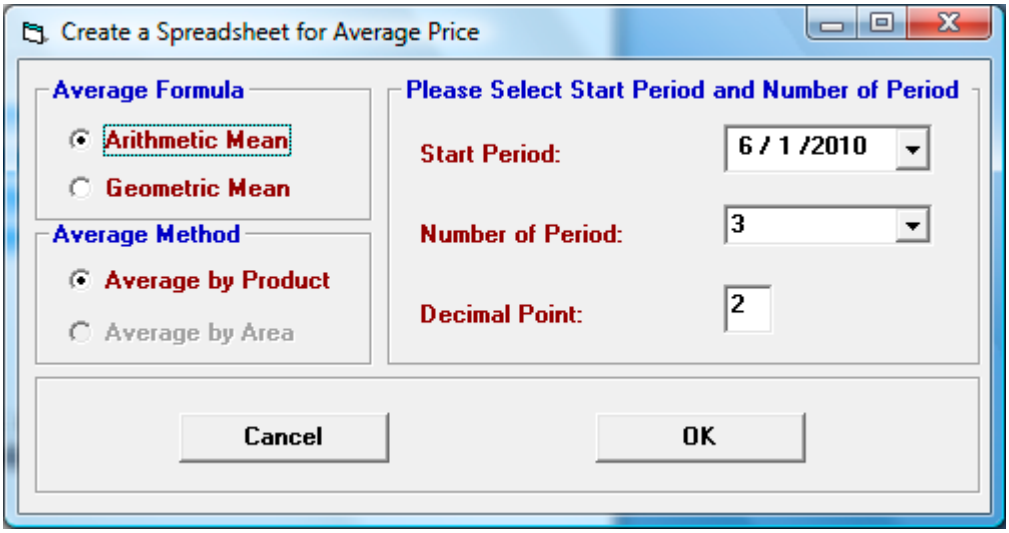
Once each variety obtains its weight from the item to which it belongs, the system can impute the product weights as well as outlet weights.

Generate Average Price for Varieties

The system can generate a report for average price for all the varieties across the outlets. The user can click **Data**, **Generate Average Price** button as showed below:



The user will see the following screen:



The user can select an arithmetic mean or geometric mean to use in the calculation. The user should also select **Start Period**, **Number of Period** and **Decimal Point** they would like to have in the report.

The following report shows the arithmetic average price for different varieties across the outlets in the country for 3 periods.

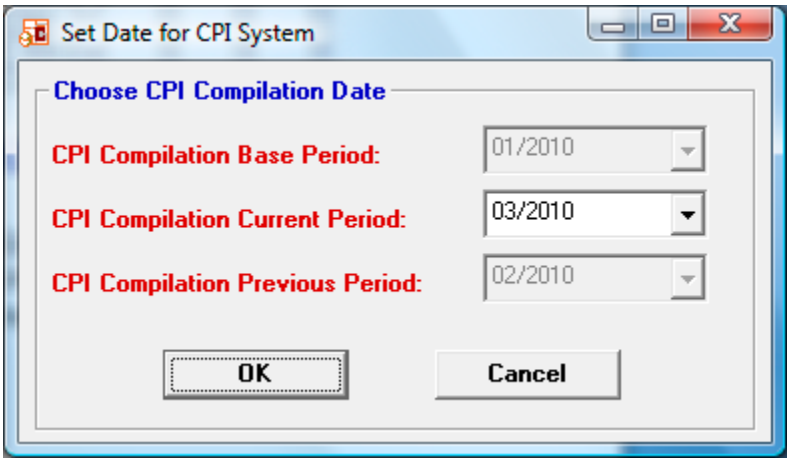
Consumer Price Index Compilation Model - [Edit Reveiw Item Weights]

File Edit Insert Format Tools Help

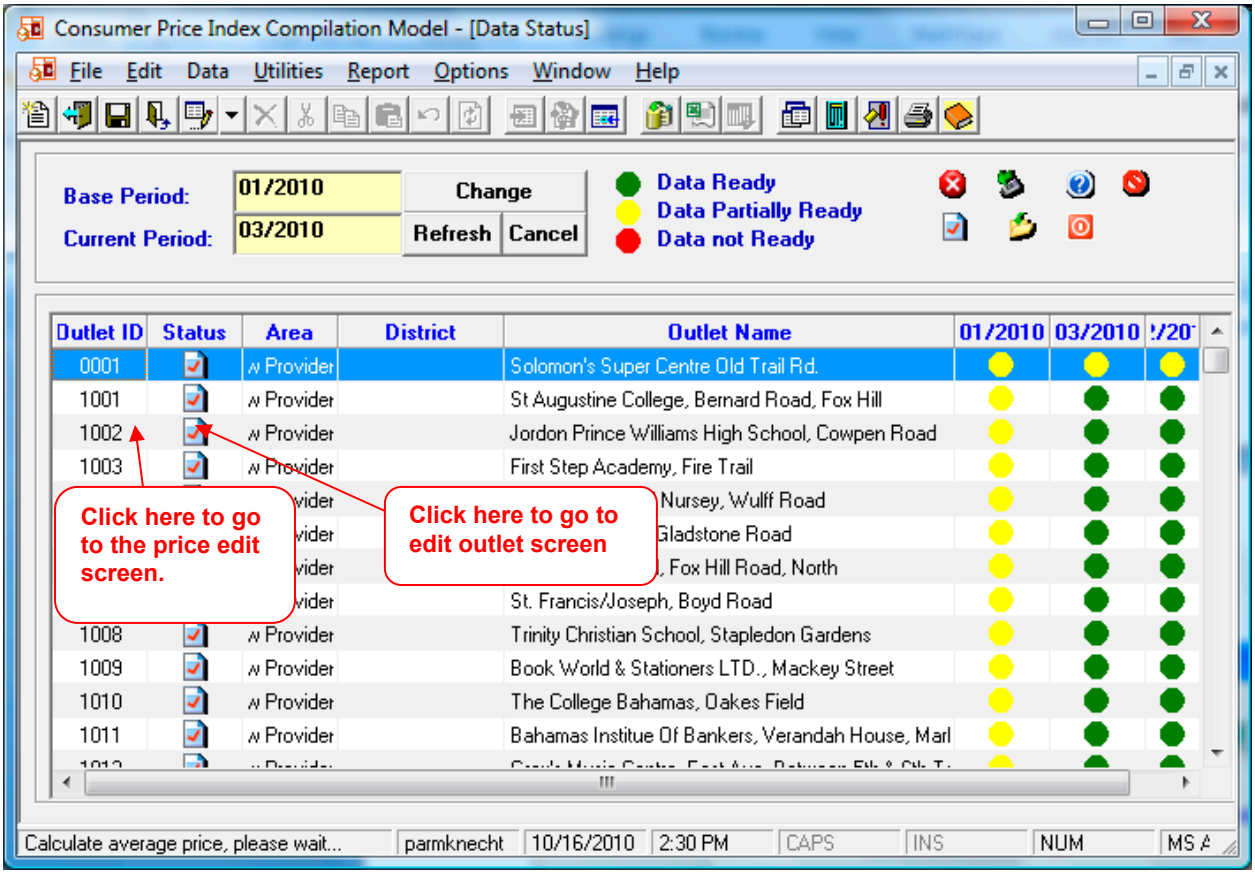
	A	B	C	D	E	F	G	H
1	Arithmatic Average of Price							
2	Outlet SN	Area Name	Region	Description		6/1/2010	7/1/2010	8/1/2010
3								
4	528	New Providence	New Providence	11.04.11.1_01a	YAMACRAW- unfurnish water included, mint green/ whit	1600	1600	1600
5	528	New Providence	New Providence	11.04.11.1_01aa	ENGLERSTON- 2 bed 1 bath water included, white/ gree	550	550	550
6	528	New Providence	New Providence	11.04.11.1_01ab	KENNEDY- 2 bed 1 bath water included, white/ hot pink,	500	500	500
7	528	New Providence	New Providence	11.04.11.1_01ac	KENNEDY- 1 bed 1 bath water & cable included, white/	550	550	550
8	528	New Providence	New Providence	11.04.11.1_01ad	GARDEN HILLS- 2 bed 1 bath, water & light included, pe	450	450	450
9	528	New Providence	New Providence	11.04.11.1_01ae	GARDEN HILLS- 2 bed 1 bath, pink/ white duplex, Door	450	450	450
10	528	New Providence	New Providence	11.04.11.1_01af	FORT CHARLOTTE- 2 bed 1 bath water included, apt. co	775	775	775
11	528	New Providence	New Providence	11.04.11.1_01ag	FORT CHARLOTTE- 2 bed 1 bath, door #1, white/ burgu	700	700	700
12	528	New Providence	New Providence	11.04.11.1_01ah	FORT CHARLOTTE- unfurnish 2 bed 1 bath, lime green/	400	400	400
					BAIN TOWN- All white 2 storey apt. (Davis Apt.), door			

Data Status Indicator and Pre-compile Data Check

Data Status Indicator is a graphical feature provided to help the compiler to examine the data availability and status before undertaking the index compilation. In the Data drop down menu shown above, select “Data Status Indicator.” The user has to determine the compilation period by selecting it in the following screen.



Click **OK** button, the user will see following screen:

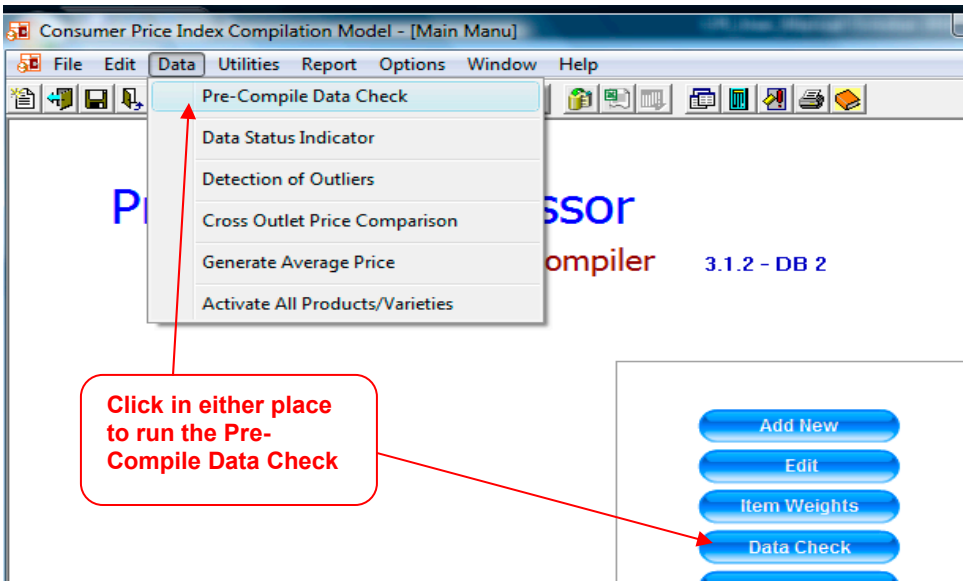


The user will have graphic views of the data status. Green dots indicate that data are available; yellow dots indicate that data are partially available and red dots indicate that data are not available. By clicking the grey button, which will appear when the user single clicks the **Outlet ID** column, the user will enter into price edit screen; from there the user can find the missing price data. If the user clicks the **Status** column, the user will be able to edit outlet information. This feature provides the user quick access to the data to resolve the missing data problem.

Different icons represent different outlet status. For example, represents **Initiated and reported** ("initiated" means that the outlet has entered the active sample and representative items have been selected for pricing); represents **Refuse to participate**; represents **Out of business**; represents **Could not locate**; represents Not get initiated; represents Resending for initiation and for **No relevant product**.

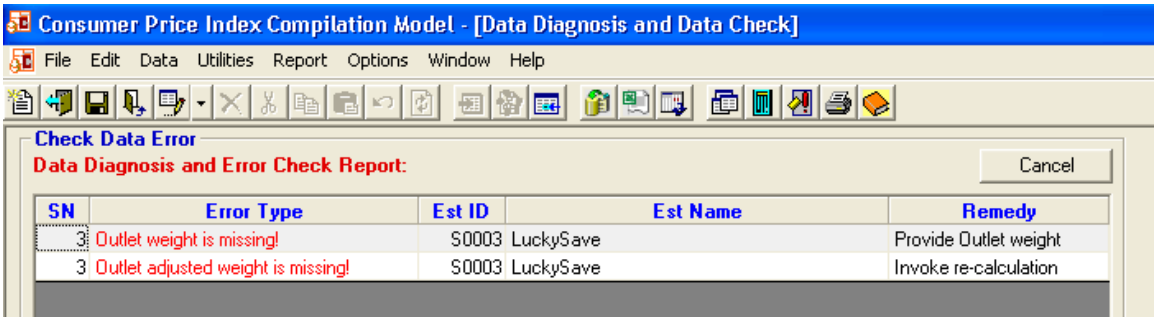
The user can change the base period and current period by clicking the **Change** button.

To examine whether the data is good for compilation, the user can click **Data, Pre-Compile Data Check** button.



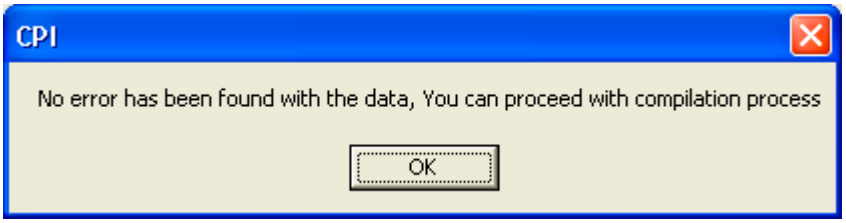
The system is able to identify the following data errors: (1) missing outlet ID; (2) missing weight area; (3) missing product and variety weight information; (4) missing outlet status; (5) missing outlet weights and (6) whether product and variety weights exceed 100%.

For example, the following screen shows that there are two errors for outlet S0003. They are interrelated: the outlet weight is missing and the adjusted weight is missing.

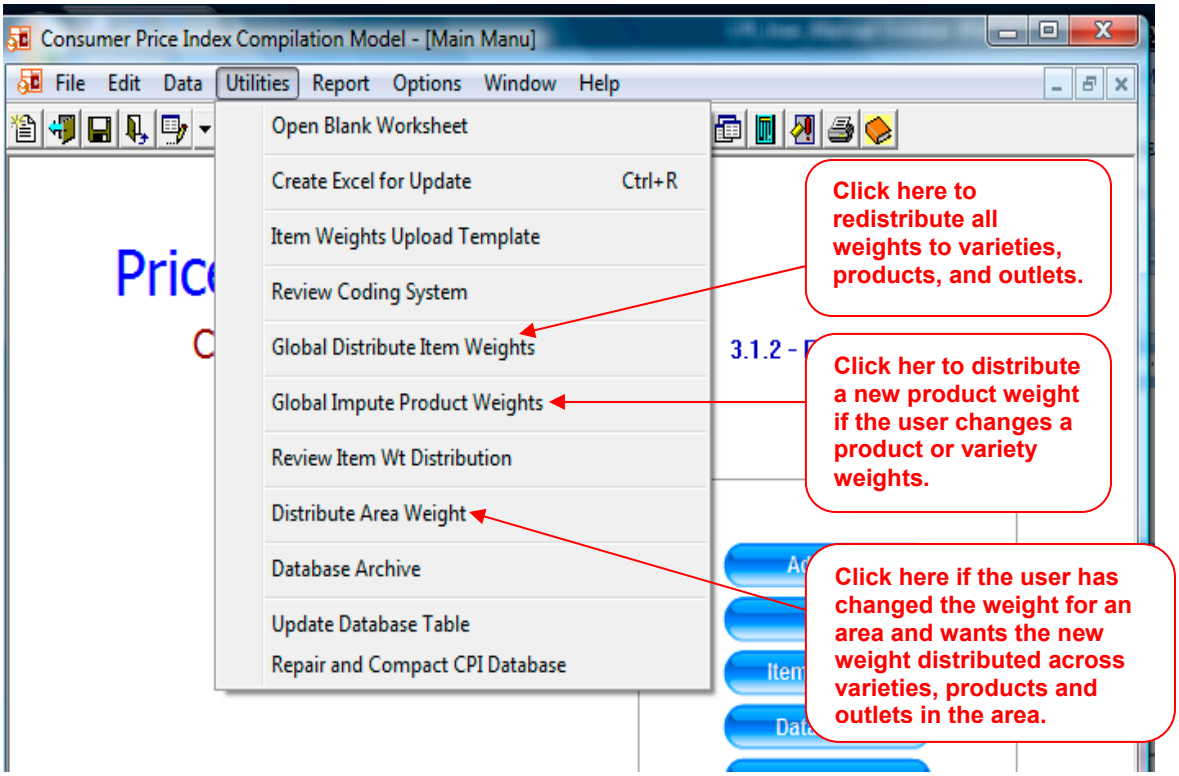


By double-clicking the red colored fonts, the program will bring to the screen where the error may occur and where the user can make corrections.

If no error is listed, the user will see the following message box:



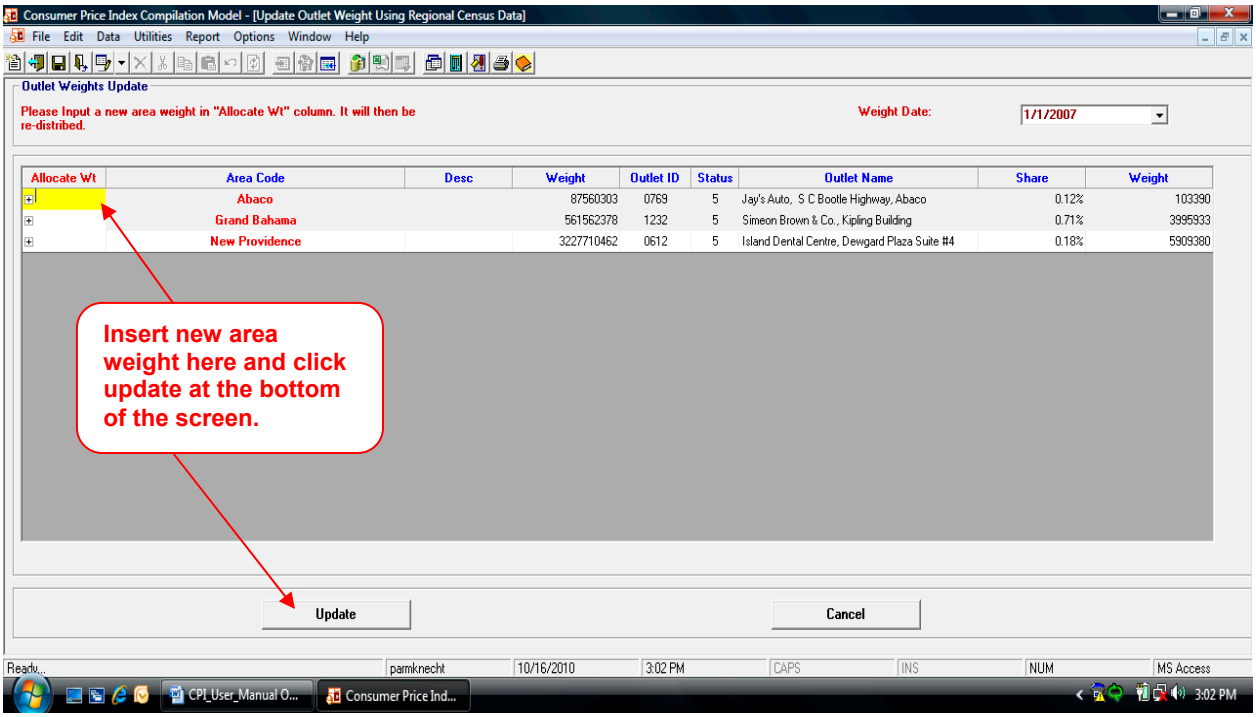
Recalculation of Weights



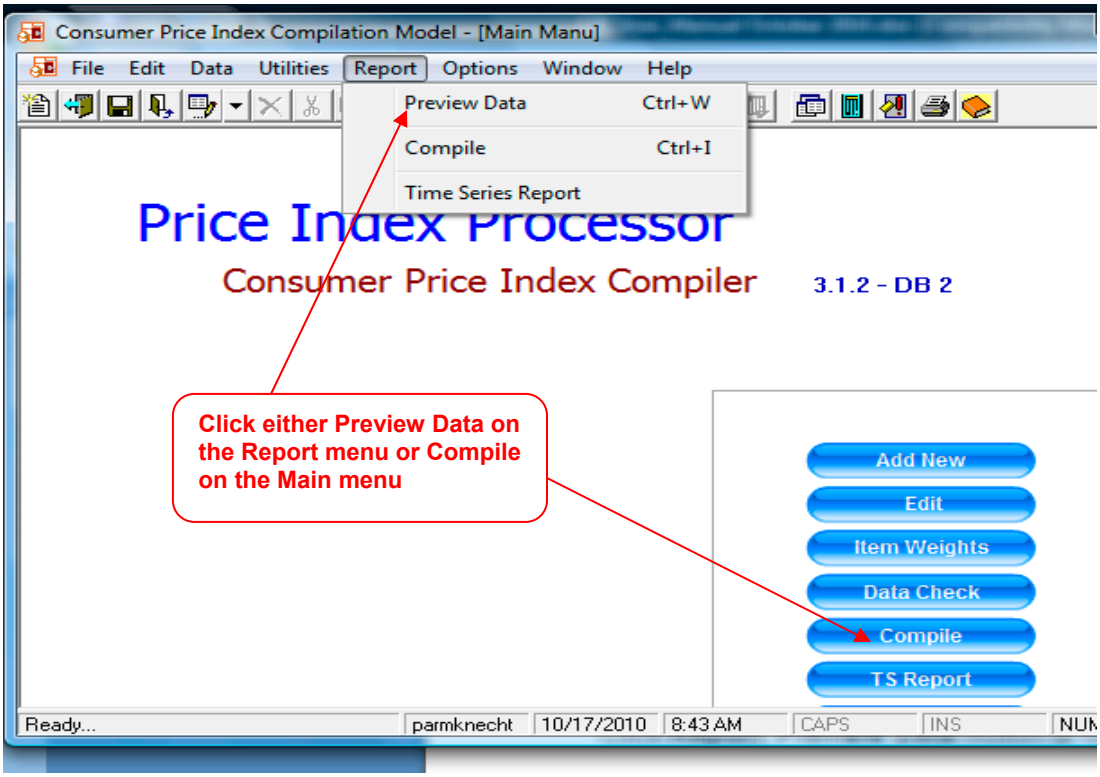
If item weights by area are revised, the user should click **Global Distribute Item Weights** to trigger recalculation process.

It is important that the user can manually define the relative importance of a product or variety by assigning its weight. Clicking **Global Impute Product Weights** will trigger a recalculation based on outlet status and sample group and newly assigned weight. If excess weight is assigned to an outlet, this extra weight will be taken from other outlets. Similarly, if less weight is assigned to a particular outlet, excess weight will be redistributed to the other outlets.

Distribute Area Weights allows the user to undertake a weights reset or update. The user can reset an old area weight with a new one, which will be distributed based on the existing share of each outlet. Click **Utilities**, **Global Impute Product Weights** button, newly assigned outlet weights will be redistributed to the product and varieties below.

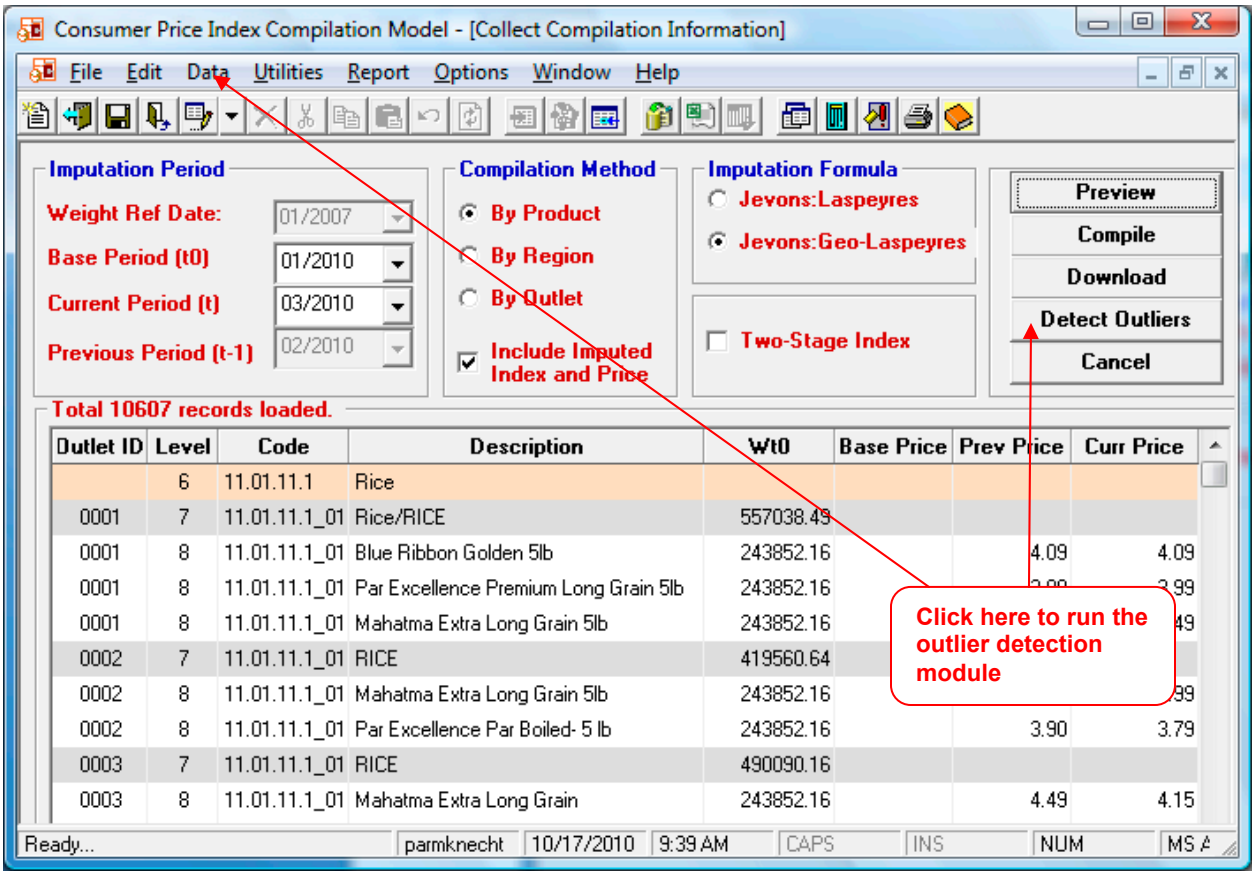


Pre-compilation Preview



Click **Report**, **Preview Data** button on the toolbar to preview the data before compilation or click **Compile** on the Main Menu to open the Preview/Compile screen.

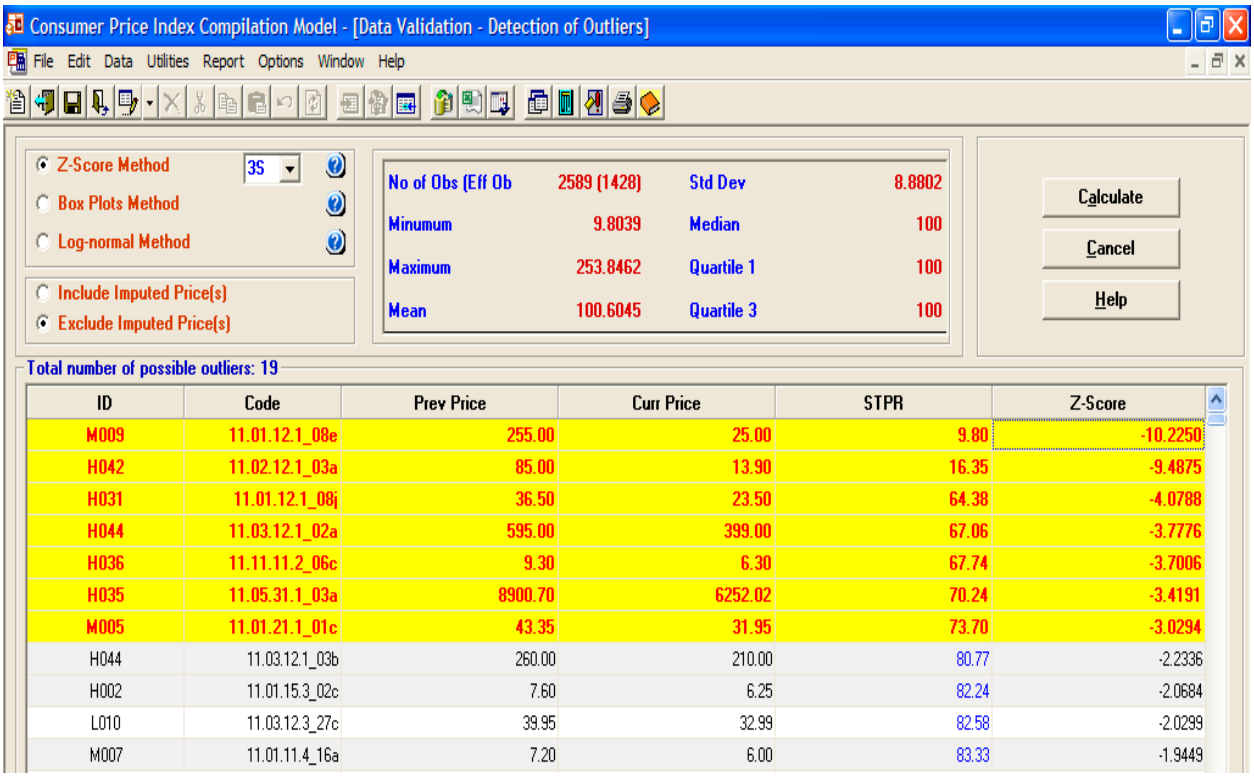
The user can select the Base Period and Current Period. Wt0 column indicates weights data followed by the base period price, previous period price, and current period price.



Detecting Outliers

To minimize data entry errors, the system has an independent Detect Outliers module available for checking outliers among the short-term price relatives (one-month price changes). For the methodology, please refer to Chapter 3, Detection of Outliers. Click **Data, Detection Outliers** button or from the **Preview** screen click the **Detect Outliers** button.

Wait for the system to load the preview data into the table. Click the Detect **Outliers** button**. Click the Z-Score** box in the fixed row to sort the information. The system will sort by the score (lowest to highest, click again for highest to lowest).

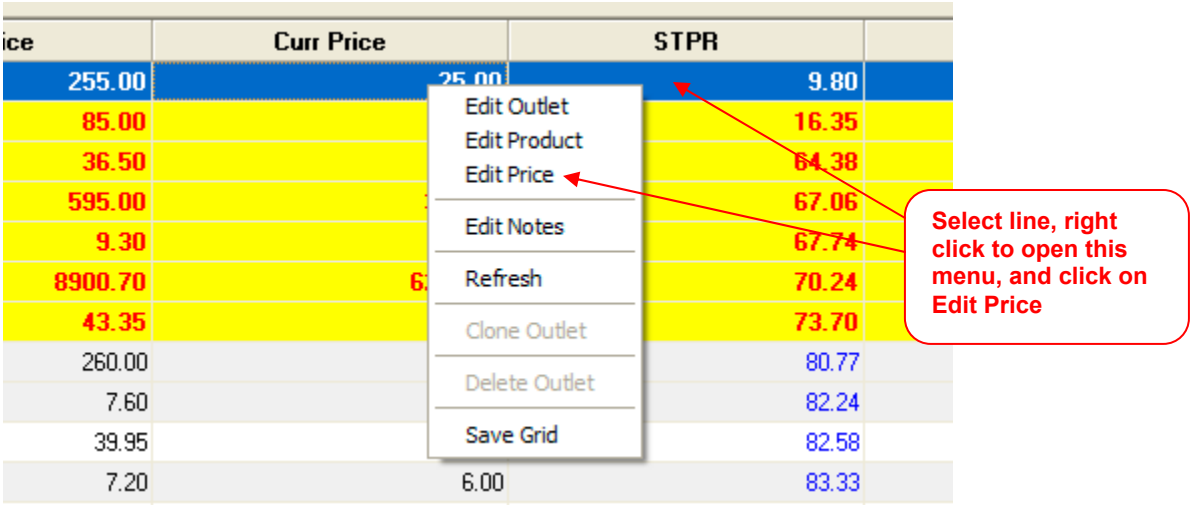


ID	Code	Prev Price	Curr Price	STPR	Z-Score
M009	11.01.12.1_08e	255.00	25.00	9.80	-10.2250
H042	11.02.12.1_03a	85.00	13.90	16.35	-9.4875
H031	11.01.12.1_08j	36.50	23.50	64.38	-4.0788
H044	11.03.12.1_02a	595.00	399.00	67.06	-3.7776
H036	11.11.11.2_06c	9.30	6.30	67.74	-3.7006
H035	11.05.31.1_03a	8900.70	6252.02	70.24	-3.4191
M005	11.01.21.1_01c	43.35	31.95	73.70	-3.0294
H044	11.03.12.1_03b	260.00	210.00	80.77	-2.2336
H002	11.01.15.3_02c	7.60	6.25	82.24	-2.0684
L010	11.03.12.3_27c	39.95	32.99	82.58	-2.0299
M007	11.01.11.4_16a	7.20	6.00	83.33	-1.9449

The default method is the Z-score test; the user can also select the Box Plots Method or the Log-normal Method. The system presents a summary statistical table with the information on short-term price relatives including: (1) Number of observations (and the number with reported prices) ; (2) Minimum value; (3) Maximum value; (4) Mean; (5) Standard deviation; (6) Median; (7) Quartile 1; (8) Quartile 3. The user can select whether to exclude imputed values from the calculations. The system shows outliers in red with yellow background as seen in the above screen capture.

The first 7 rows are detected to be outliers. Apparently, there is the very strong possibility of a data entry error in product 11.01.12.1_08e in which the previous price is 255.00 and current price is 25.00.

To fix the problem, select the record, right click to open the menu, and then click the Edit Price button. The system will open the price input screen and highlight the price observation in question. When finished editing, cancel the price input screen to return to the outliers for further review/edits.



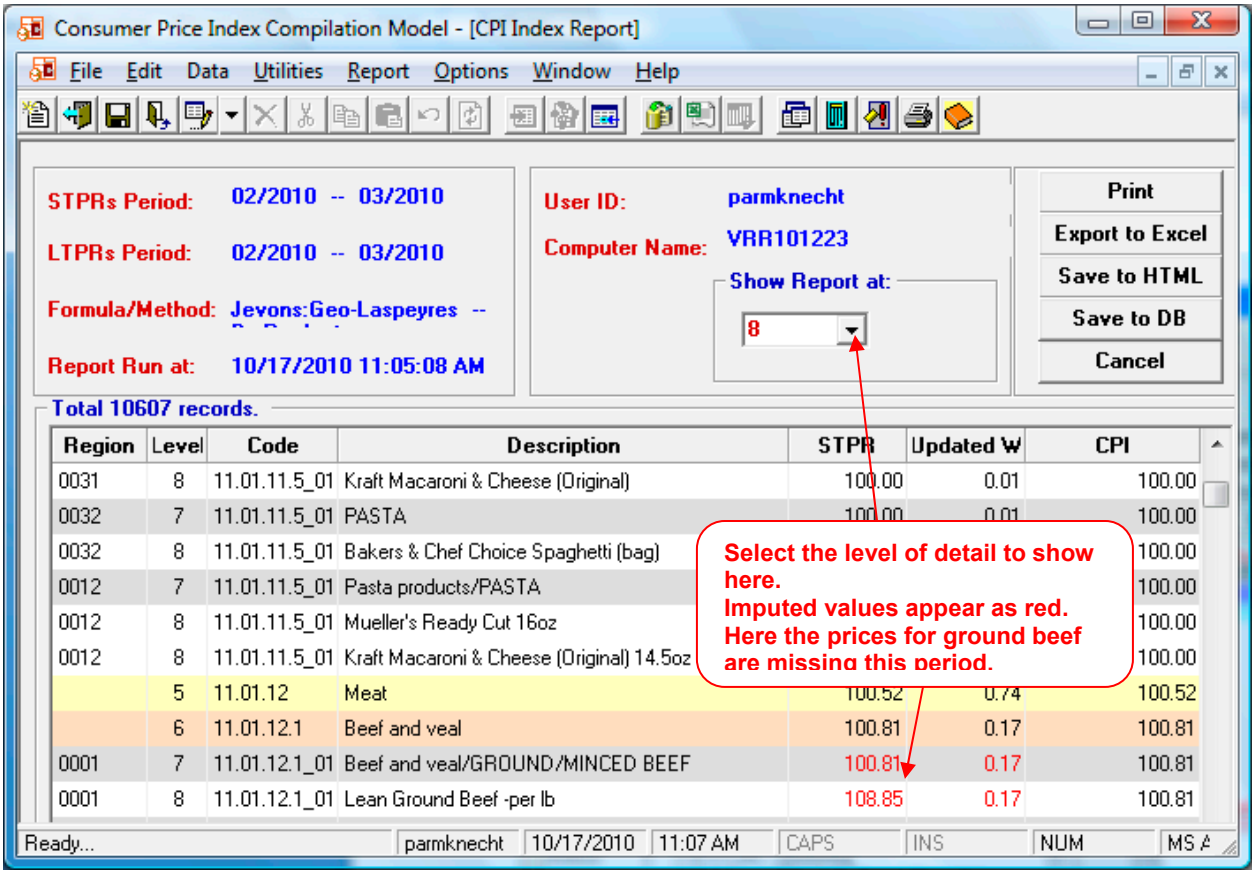
Index Compilation

The user defines the compilation method by choosing one of the options: **By Product**, **By Region** or **By Outlet**. At the elementary level where the observations are equally weighted the system always uses the geometric mean. The user can also choose either Laspeyres-type or Geometric Laspeyres-type formula in the compilation at the weighted level. (The actual index formula used—Laspeyres, Lowe, Young, Geometric Laspeyres, Cobb-Douglas—depends on the weight structure. See Chapter 3, Methodology.)

If the **Modified/Two-stage** is checked, the system uses the previous period Long-Term Price Relatives ($LTPR_{t-1}$), which is the price increase from period 0 to period t-1, in the calculation along with the Short-Term Price Relatives from the previous to current period (please refer the Chapter 3, Methodology for the details of the formula). Alternatively, the base period price, P_{t0}, and the current period price, P_t, will be used to calculate the current period indexes and the current and previous period price, P_{t-1}, for the calculating the short-term relatives. The system estimates the current period indexes for all levels—from the individual price observation to the aggregate index for the total CPI.

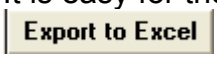
If there are missing prices or missing indexes, the system will impute the missing values as discussed in Chapter 3, Methodology. The system displays Imputation results in red.

When the user completes the review for the period and is satisfied that all data are correct, they need to save the data by clicking the **Save to DB** button. If the user forgets to save the compiled indexes to the database, the system may have trouble compiling the indexes for the next period and return a message that a critical error occurred during compilation.



The system generates three sets of results: (1) **STPR**, Short Term Price Relatives, which is the index of price changes from previous period to current period. (2) **CPI**, which is the Long Term Price Index, reflecting accumulated price increase since period t0. (3) **Updated W** is the updated cost weight, i.e. the original weight expressed as a share, updated by the long-term price change from t0.

The system's default view is from the most detailed level (8-digit variety) to the top-level aggregate CPI. The user can select an alternative level for the report by using the drop-down box in the center of the screen. For publication results, the user may only need the report displayed at the 5- or 6-dgit level. The user can then export the table.

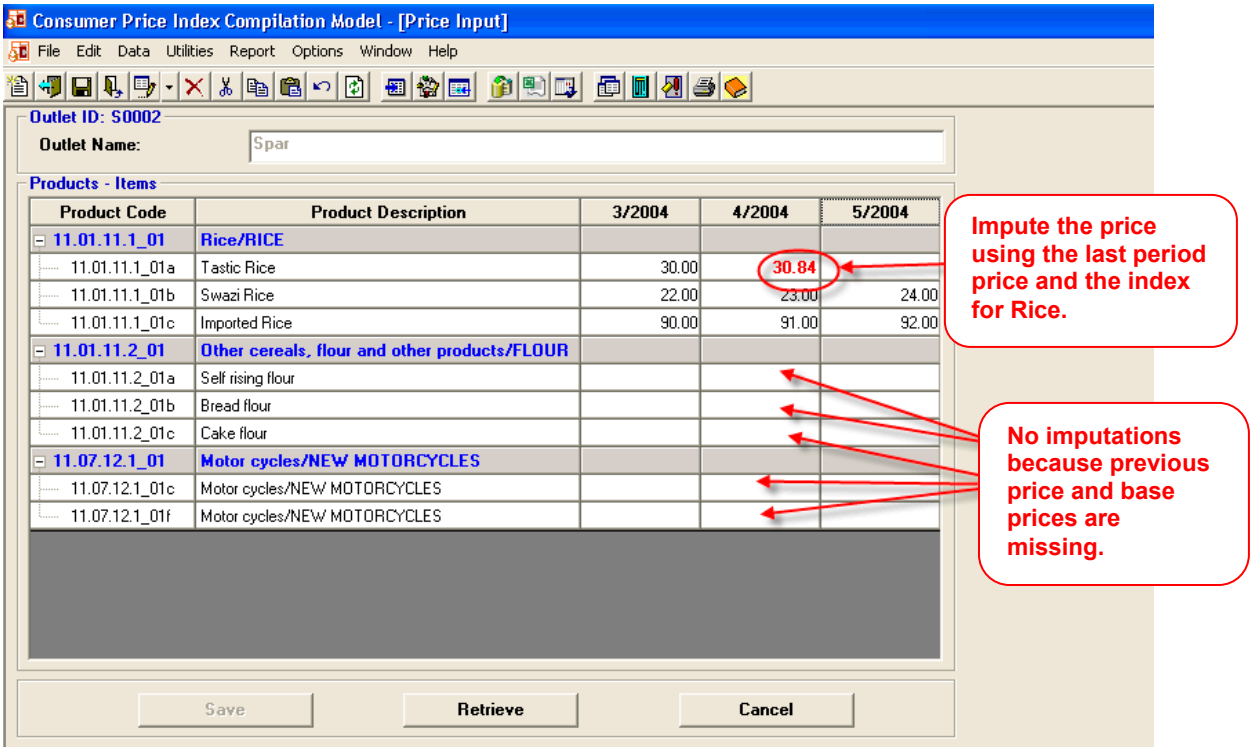
It is easy for the user to export the results to Excel. The user can click **Export to Excel**, 

button, and then an Excel icon will appear in the screen.

Click the link, the system transfers the report results to the Excel spreadsheet. From Excel the data can be stored in a historical format to use for developing publication tables.

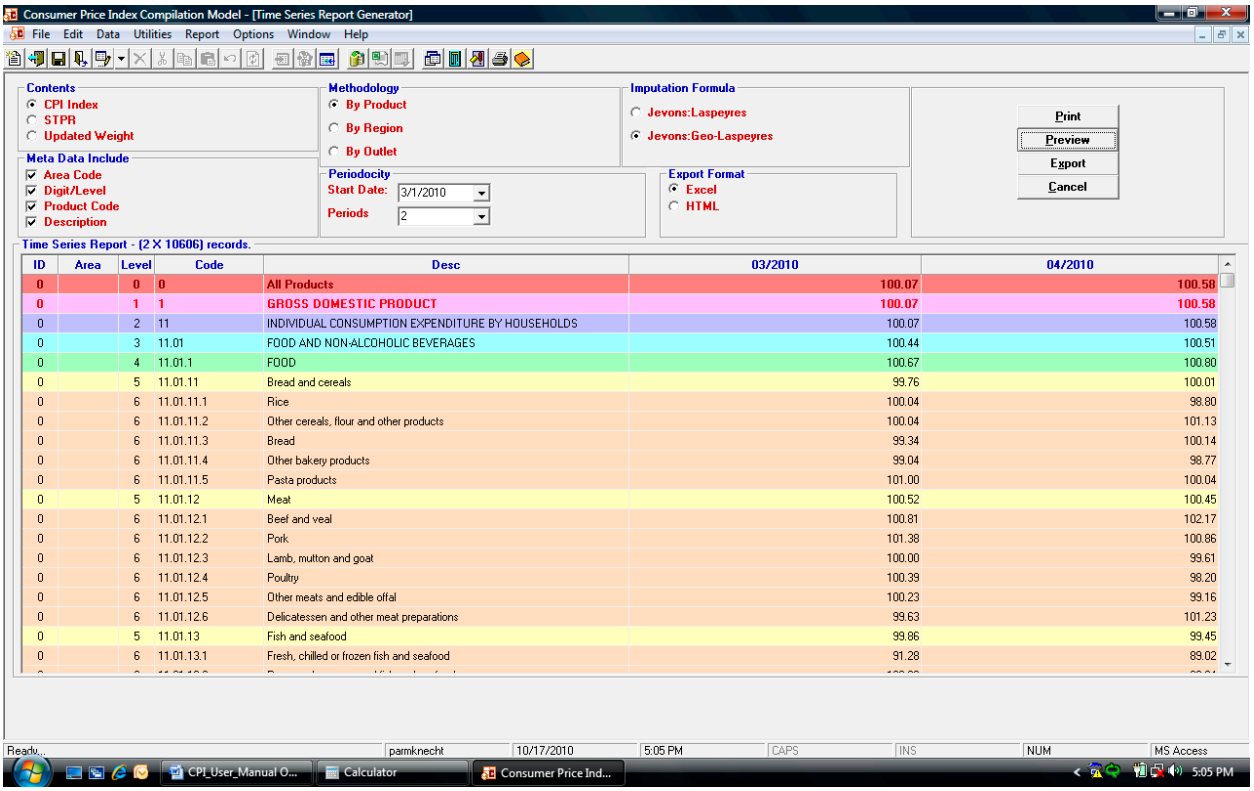
The system imputes missing price indices by taking the indices from their parent group. If parent group indices are missing, the system has to go one level up again, until it finds a calculated index. The system displays imputed indices in red. When the user saves the indices to the database, the system automatically imputes the missing prices and stores them in the database with a flag. The system displays missing prices on the price edit screen in red.

The system imputes current period missing prices by multiplying the last period price by the STPR of that variety. If the last period's price is missing, then the imputed price will take the product of the base period price and index for the higher product group. If both the previous and base period prices are missing, the system makes no imputations.



Time Series Report

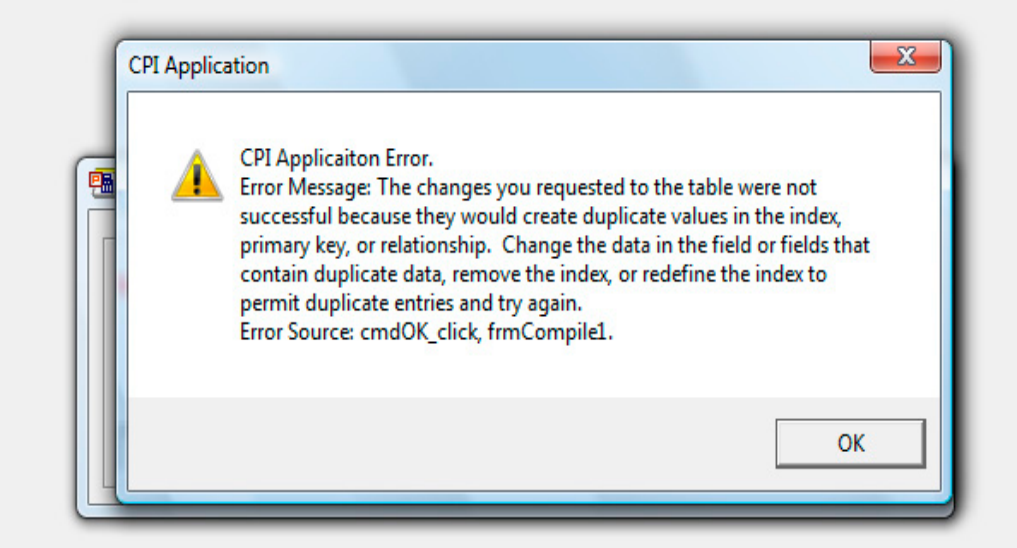
The system can generate stored indexes in time series format (one index for multiple periods, each column indicates a period). Time series presentations give the user one index at a time for the multiple periods. The user can select which, if any, metadata they would like to have appear in their report (area code, digit level, product code and description). The system can readily export the time series report to an Excel spreadsheet. The following screen shows the details of such a report.



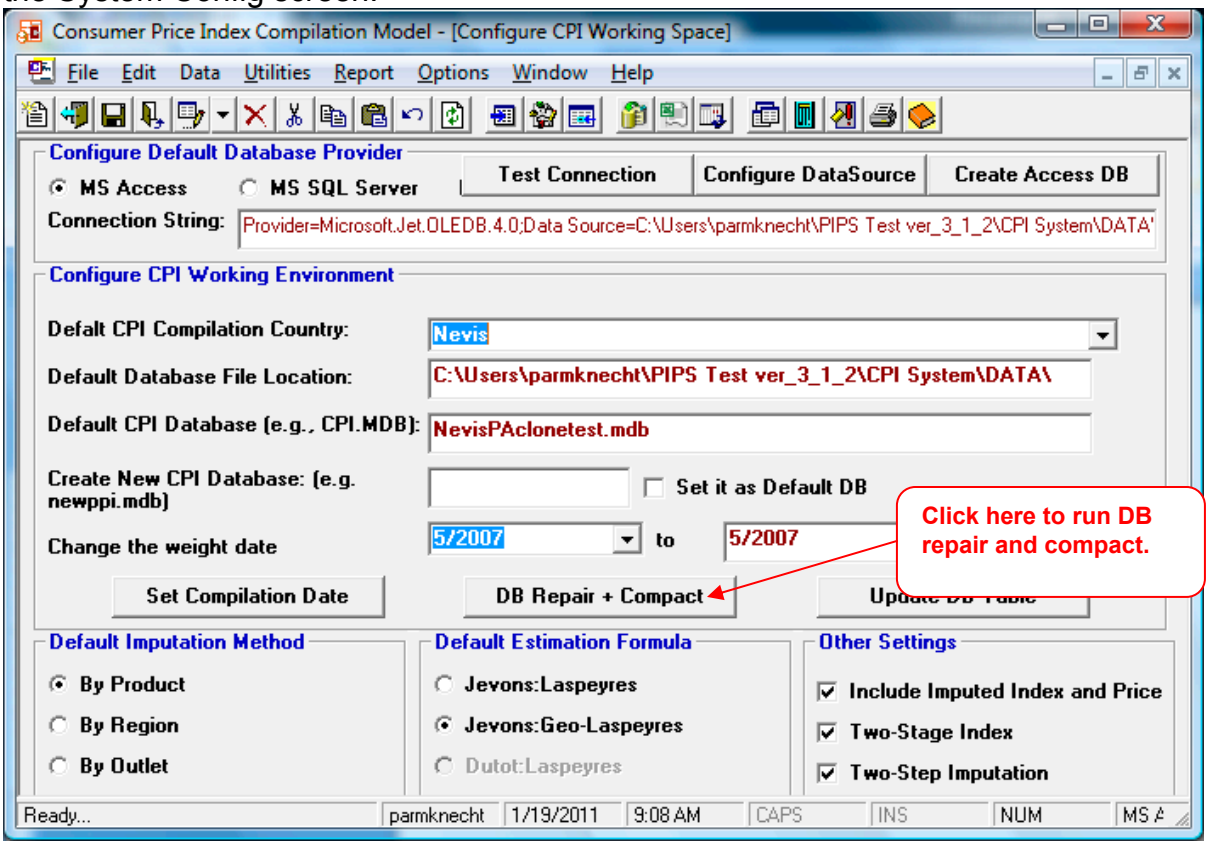
The system can produce up to 12-months of data for each item in the report. The user can save the data in Excel and use Excel to generate publication tables.

Database Maintenance and Repair

Sometimes the Access DB may get corrupted or slight out of sequence, particularly when the user makes a number substitutions and uses the clone variety feature. When this happens, you may get the following error message when compiling the index.



The user can resolve this problem by running the “DB repair and compact” feature on the System Config screen.



When it completes, it may say no errors were found or that errors were found and corrected. Click OK and you will get a message “do you want to continue to compact you database?” And you should respond yes.

You should periodically run the DB repair and compact feature to reduce the size of the DB to make it more manageable. This feature can reduce the size of the DB by 50-75%.

References

- Consumer Price Index Manual: Theory and Practice (2004), *International Labor Organization*. http://www.ilo.org/public/english/bureau/stat/guides/cpi/ and at http://www.imf.org/external/pubs/cat/longres.cfm?sk=17165.0

- COICOP, Detailed structure and explanatory notes, Statistics Division, UN. http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=5&Lg=1

- Practical Guide to Producing Consumer Price Indices (2009), UN Economic Commission for Europe. http://www.unece.org/stats/publications/Practical_Guide_to_Producing_CPI.pdf

Index

How to Use PIPS Access to/from SQL Database Utility Program

By PIPSUS Support Team
Dr. Chouikha
(achouikha@gmail.com)
October 10, 2010, Version 1.0

1. Introduction

PIPS (Price Index Processor Software) Data Transfer Utility Program (DTUP) is developed for transferring PIPS data from Microsoft Access to SQL server or vice verse. It works with the database created by the Price Index Processor – Consumer Price Index Compiler, version 3.0.0 - DB 2 or above. **The DTUP is designed to help the user to migrate or upgrade the database and moving data from one format to another**.

2. Background

The PIPS is a database application and CPI Compiler currently supports both Access and SQL/SQL Express databases. There are some distinct differences between Access and SQL server as the backend. (1) Access, which is part of Microsoft Office Suite, uses small desktop databases with a frontend graphic user interface and SQL uses client - server model; (2) SQL server is a more robust system and is able to handle larger volume data than Access. (3) SQL keeps the data on the server as the backend and the forms, reports in the frontend. As a result, PIPS with SQL Server works better in multiuser environment to keep all the records together in one place and in sync whereas the PIPS with Access database is typically portable and used in single (or less than 5) user(s) environment.

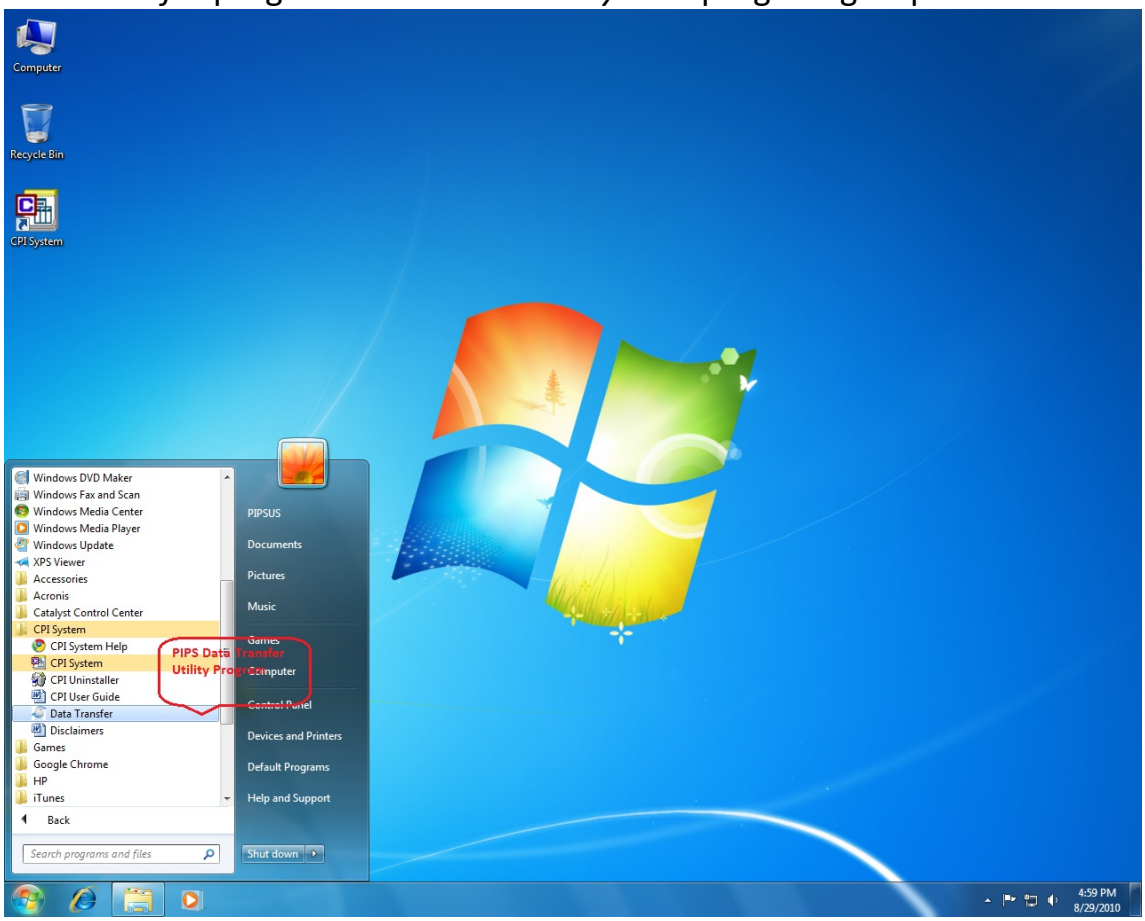
Because the way the relational database schema (i.e. tables, fields, relationships, views, indexes, procedures and functions, etc.) is designed

in the PIPS application, there are some technical barriers need to overcome before the data can be successfully copied from one format to another. For example, the identity field with incremental numbers should be carried over when being copied and predefined relationships among tables should be preserved to ensure the application will function in an identical way.

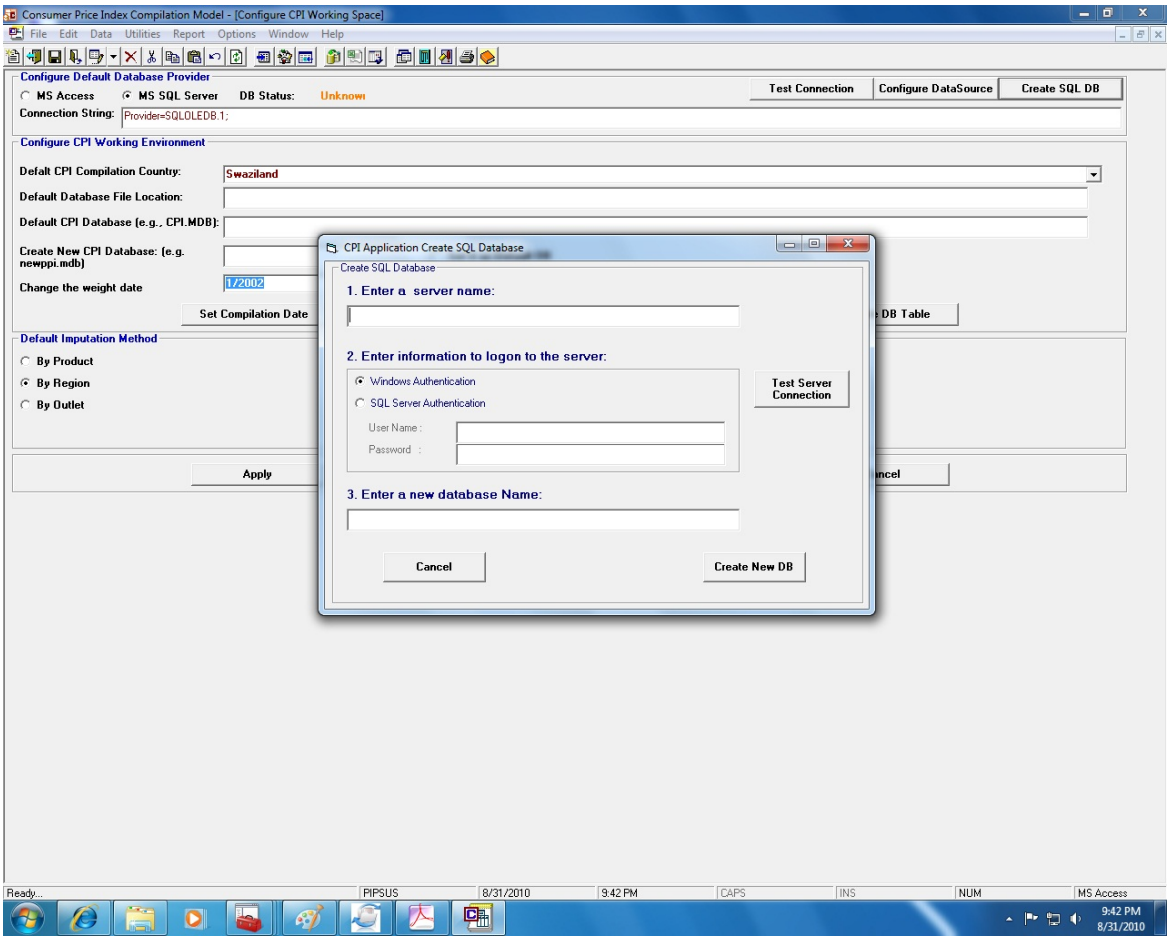
3. How PIPS DTUP works

When PIPS is installed, a shortcut of the "PIPS Data Transfer" icon will be added to the Desktop as well as in the "CPI System" program group, which is shown in following diagram

Data Transfer program is added to *CPI System* program group.



Before start the Data Transfer process, please make sure the PIPS database template is created. This can done using PIPS application, in the System Configuration module.



3.1 Technology Overview

There are several technologies available to achieve the objective of copying the database from one format to another. We can either push the data from Access or pull the data from SQL Server or vice verse.

Pushing the data - (1) Recordset to Recordset Method. Copy the data row by row and field by field from one recordset to the other. (2) Recordset to Insert query. Create an Insert statement for each row in the recordset. (3) Linked table insert Link to the SQL Server table and use an Insert statement to insert the data.

Pulling the data - (1) DTS package - Create a DTS package in SQL to pull the data, we can execute the DTS package either from the Access database or by calling a stored procedure on the SQL server which executes the DTS

package. (2) Linked Server, we can create a linked server in SQL which points at our Access DB, we can then call a stored procedure to insert the data from the linked server into your SQL table.

The method we are using in our utility software is by calling OPENROWSET function. This method is an alternative to accessing tables in a linked server and is a one-time, ad hoc method of connecting and accessing remote data using OLE DB. The OPENROWSET function can be referenced in the FROM clause of a query as though it is a table name. The OPENROWSET function can also be referenced as the target table of an INSERT, UPDATE, or DELETE statement, subject to the capabilities of the OLE DB provider. Although the query may return multiple result sets, OPENROWSET returns only the first one.

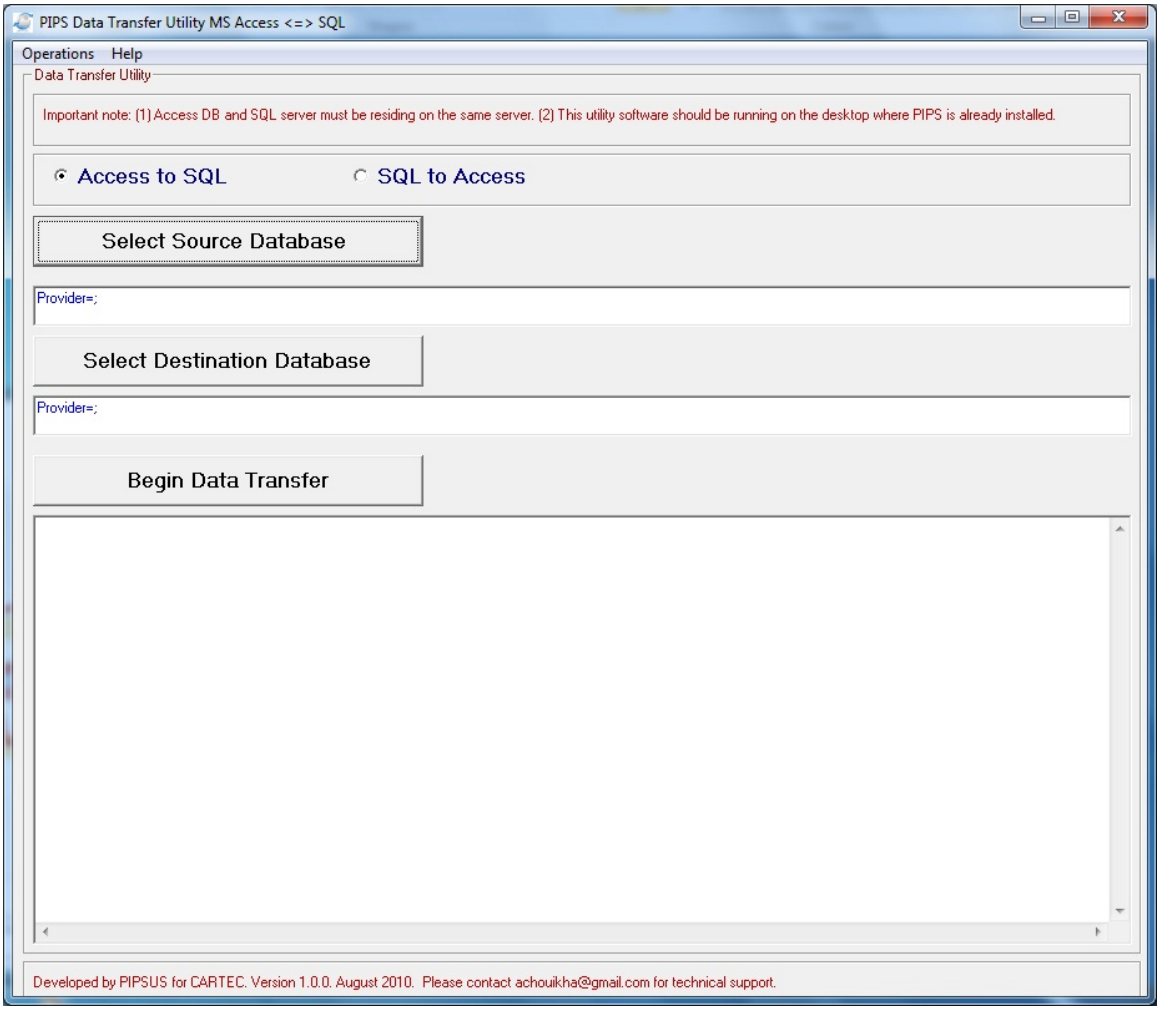
By using this approach, the Access DB that needs to be migrated and SQL server must be residing in the same physical server. It's important to note that is only required for data transfer utility NOT for the operation of index compiler software itself.

Apart from this requirement, PIPS application must be installed together with data transfer utility program.

Below is the sample SQL insert statement with an OPENROWSET function to an Access database.

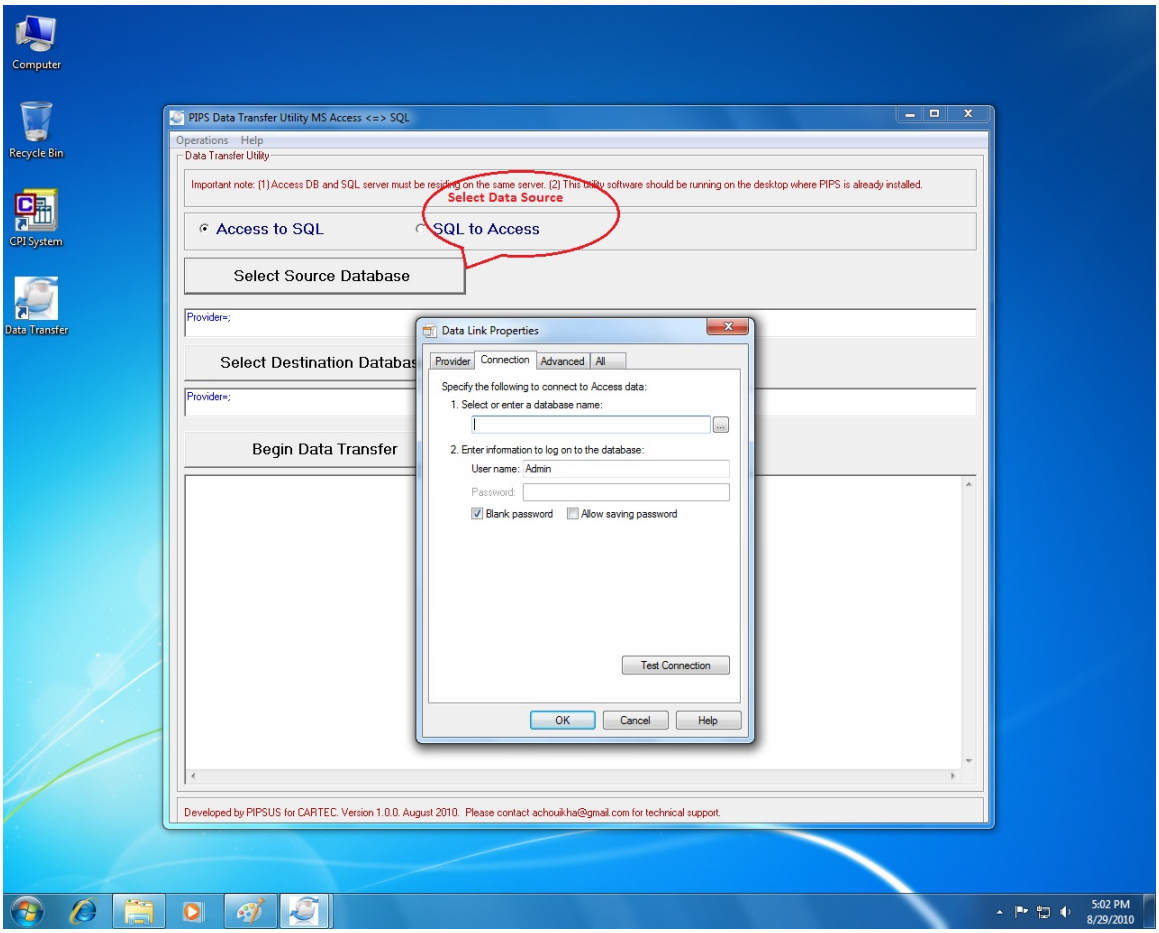
```
INSERT INTO
dbo.mytable (myfield)
SELECT
myfield
FROM OPENROWSET('Microsoft.Jet.OLEDB.4.0',
'c:\mypath\mydb.mdb';'admin';'mypwd', mytable)
GO
```

3.2 Double click the Data Transfer icon to start the program.

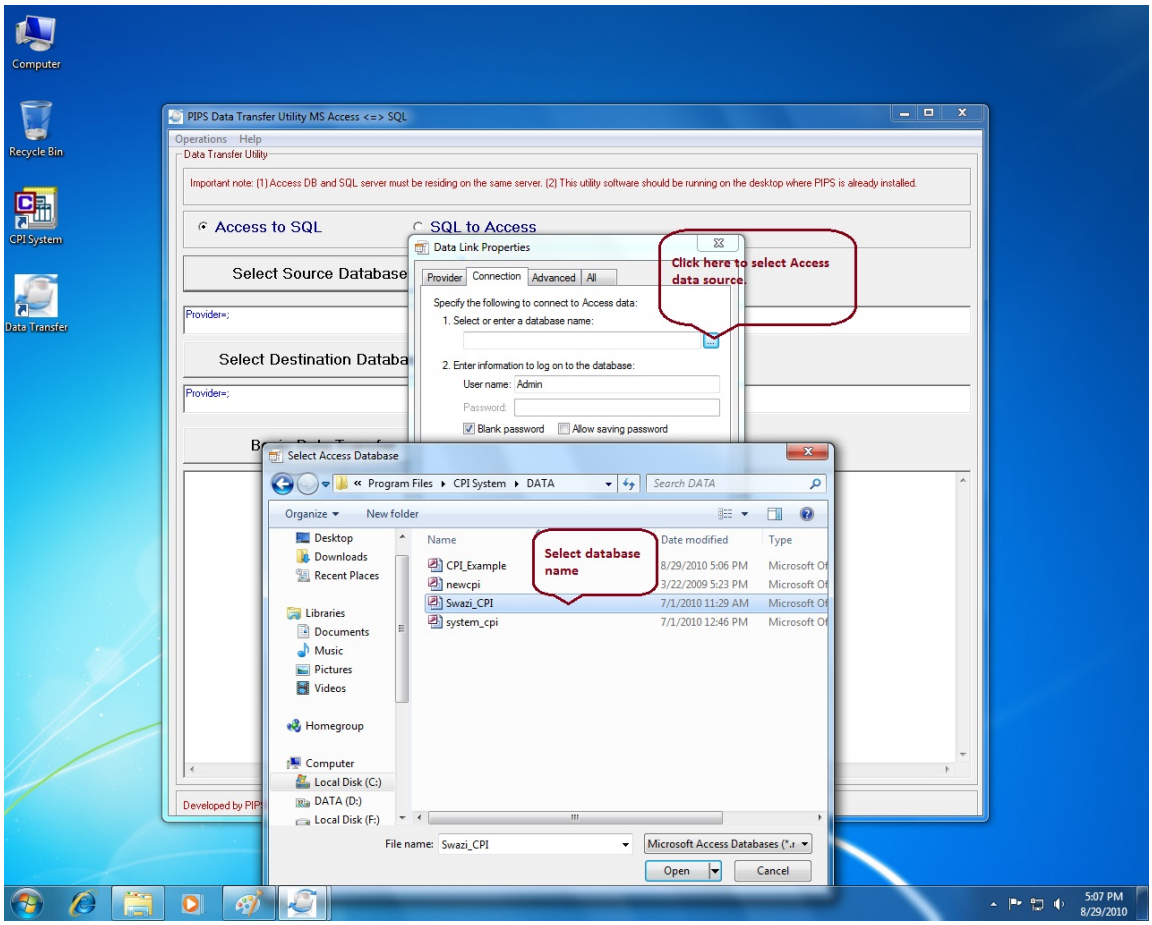


3.3 Select database transfer direction either from Access to SQL or SQL to Access. Then select Source Database.

Select Access data source

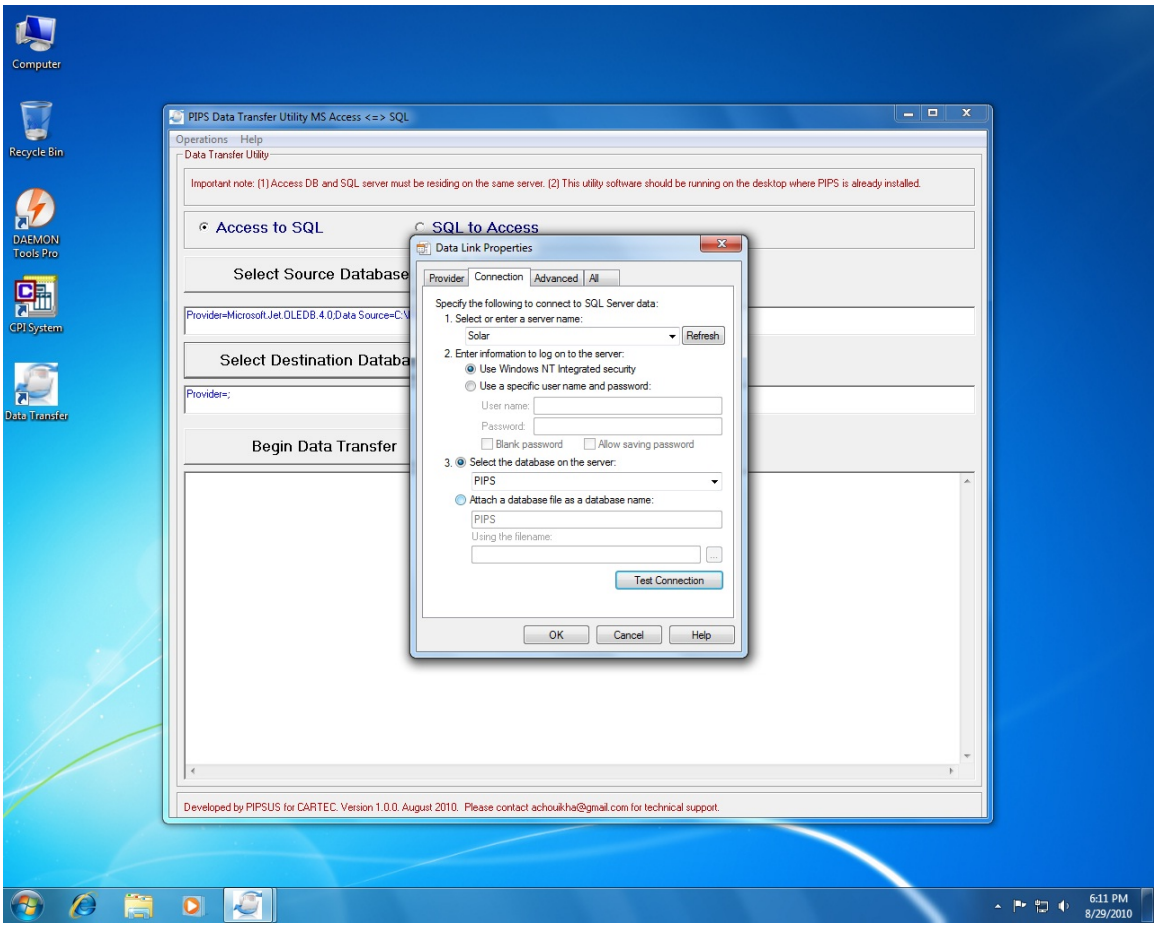


Select PIPS Access Database

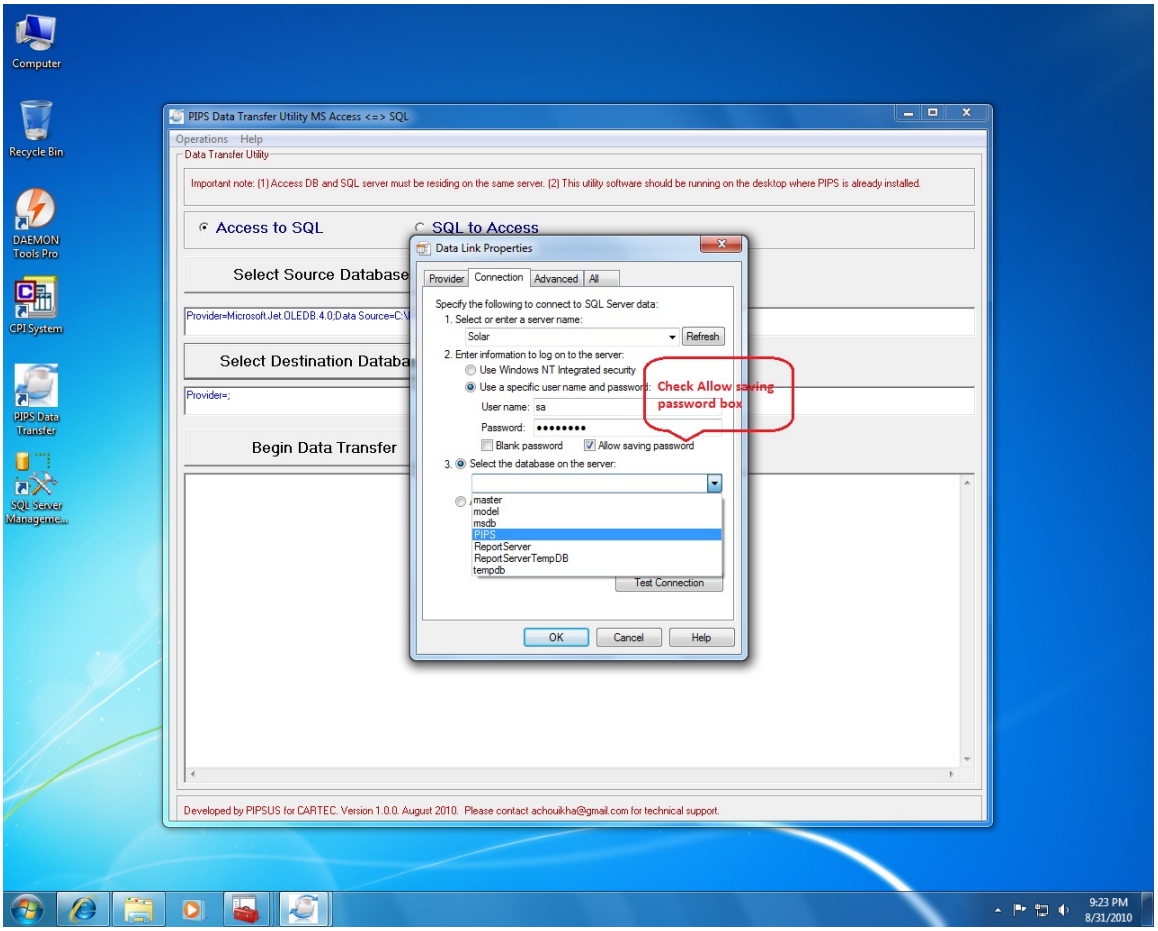


3.4 Select Destination SQL Data Source

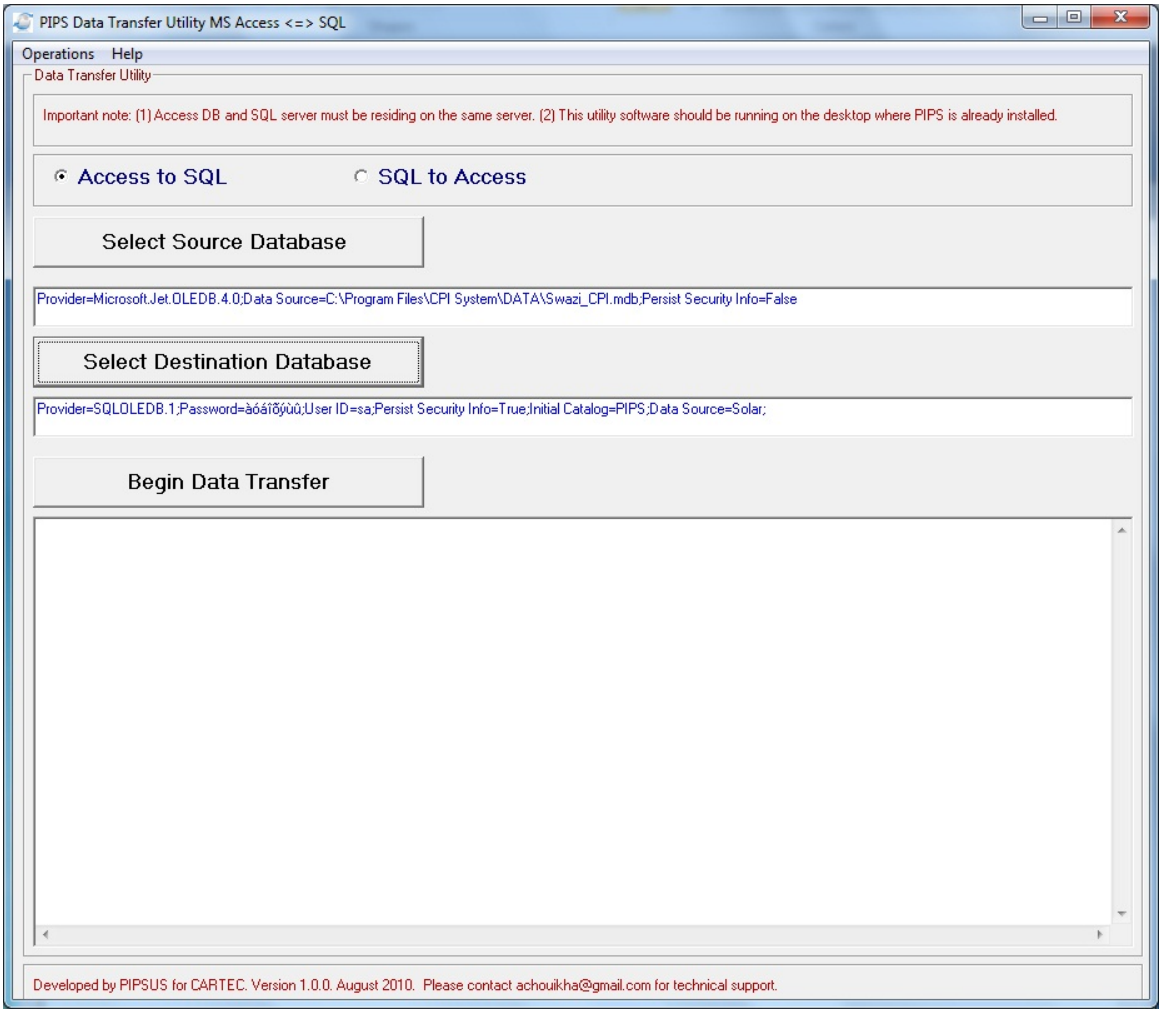
Please enter the correct credential information to access SQL server. The user must have sufficient rights to delete and insert the records in SQL Server. If you are encountering any issue here, you may need to seek assistance from your DBA.



3.5 Check “Allow saving password” box if you are using SQL specific user name and password

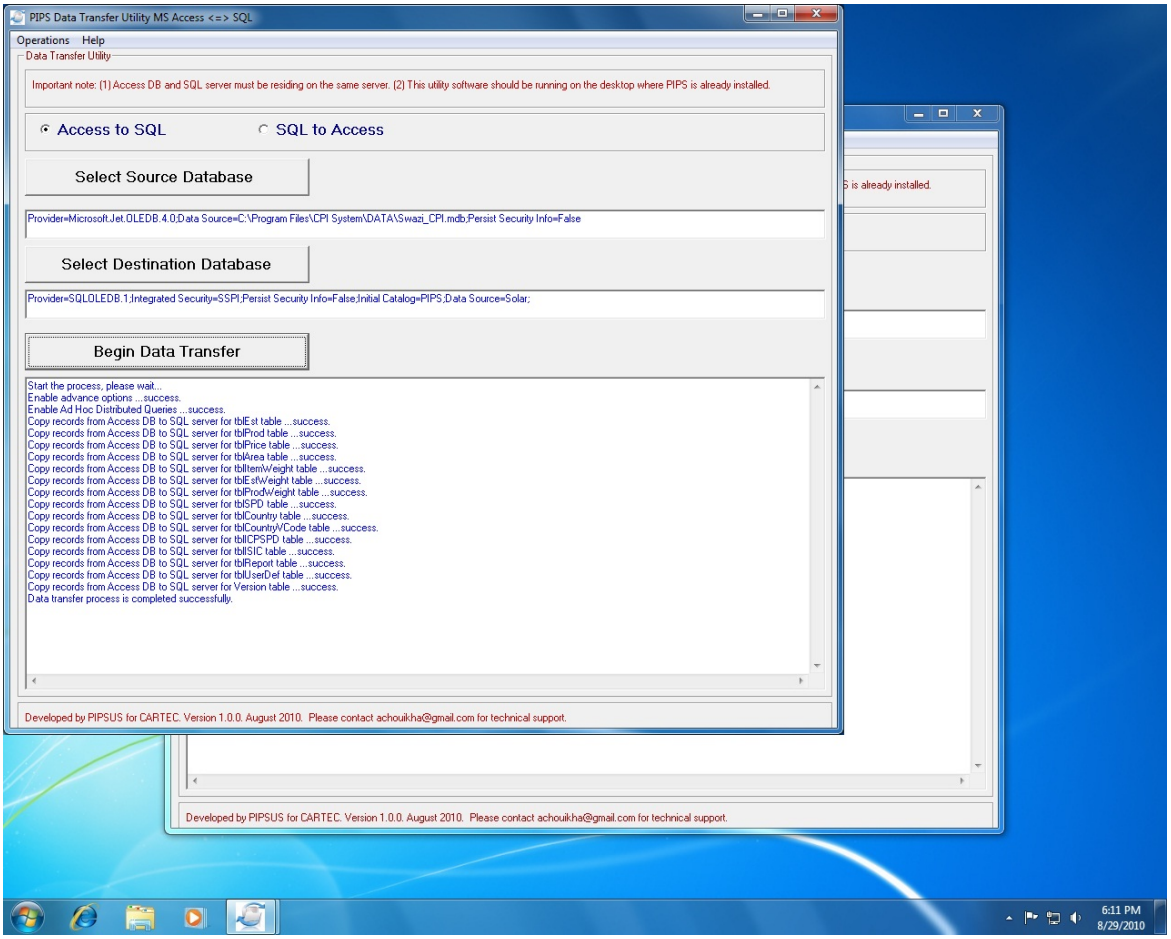


You must check the box of “Allow saving password” if a SQL password is entered. The password will be encrypted in the connection string.



3.6 Click Data Transfer Button

You should see the message if the data transfer is successful.



3.7 Click the Help – Readme to access this short documentation

4. Trouble Shooting

4.1 Database Connection Issues

The DTUP provides the progress report in the message window at the bottom of the Window form. Depending on the Access database size, it may take a while for the all operational steps to be completed.

Upon completion of the each specific step, the utility program will print a message in the message window.

The first two steps are building the connection between the source and destination databases.

“Start the process, please wait...
Enable advance options ...success.
Enable Ad Hoc Distributed Queries ...success.”

OPENROWSET function requires the remote connection to be established. Therefore, the DTUP has to first of all, configure the database instance to enable Ad Hoc Distributed Queries in the installed SQL Server database instance where the Ad Hoc query will run. This step is done through “sp_configure“ stored procedure to enable the ad hoc connections to remote data sources. The SQL query statement is:

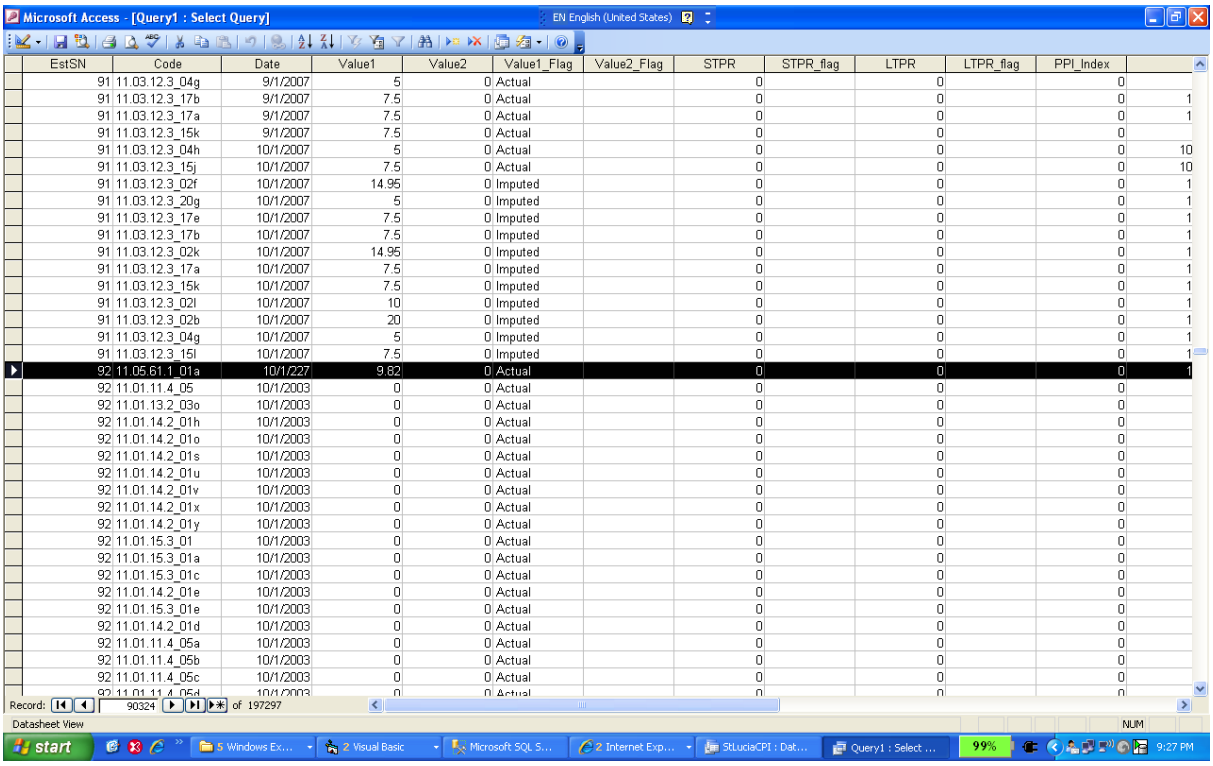
"sp_configure 'show advanced options', 1"
“reconfigure”
"sp_configure 'Ad Hoc Distributed Queries', 1"

If this step is not successful, please check your SQL database configuration and make sure that the Access database file is located in the same server as SQL server. Further technical information can be found by googling “How to enable the use of 'Ad Hoc Distributed Queries' by using sp_configure”.

4.2 Copy Data Error during the Data Transfer Process

It is possible some errors occur during the data transfer. The most common error is the data type of source does not compatible with the data type of defined in the destination database. Please note when SQL database is created using PIPS application, the database schema is created in the closest possible way to match that of Access DB. However, there might be some cases in which the data does not match the configuration at server side or the data is modified accidently altered in the original Access database leading to the failure the table data transfer. Below is an example of this type of error:

The record 90324 has a “bad” data in [Date] field and was mistakenly persisted.

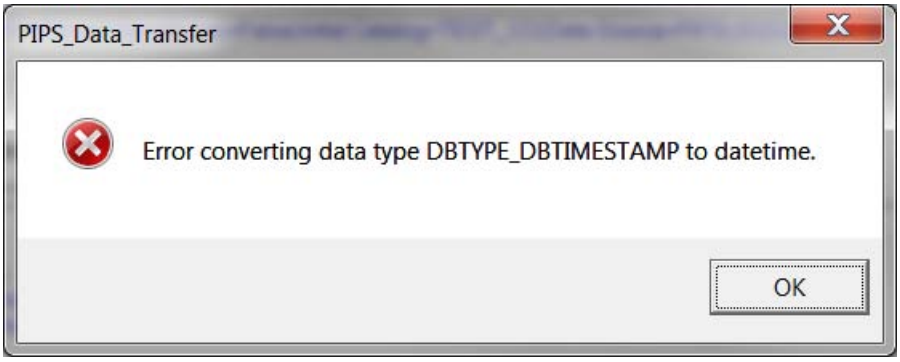


Microsoft Access - [Query1 : Select Query]

EstSN	Code	Date	Value1	Value2	Value1_Flag	Value2_Flag	STPR	STPR_flag	LTPR	LTPR_flag	PPI_Index	
91	11.03.12.3_04g	9/1/2007	5	0	Actual		0		0		0	
91	11.03.12.3_17b	9/1/2007	7.5	0	Actual		0		0		0	1
91	11.03.12.3_17a	9/1/2007	7.5	0	Actual		0		0		0	1
91	11.03.12.3_15k	9/1/2007	7.5	0	Actual		0		0		0	
91	11.03.12.3_04h	10/1/2007	5	0	Actual		0		0		0	10
91	11.03.12.3_15j	10/1/2007	7.5	0	Actual		0		0		0	10
91	11.03.12.3_02f	10/1/2007	14.95	0	Imputed		0		0		0	1
91	11.03.12.3_20g	10/1/2007	5	0	Imputed		0		0		0	1
91	11.03.12.3_17e	10/1/2007	7.5	0	Imputed		0		0		0	1
91	11.03.12.3_17b	10/1/2007	7.5	0	Imputed		0		0		0	1
91	11.03.12.3_02k	10/1/2007	14.95	0	Imputed		0		0		0	1
91	11.03.12.3_17a	10/1/2007	7.5	0	Imputed		0		0		0	1
91	11.03.12.3_15k	10/1/2007	7.5	0	Imputed		0		0		0	1
91	11.03.12.3_02l	10/1/2007	10	0	Imputed		0		0		0	1
91	11.03.12.3_02b	10/1/2007	20	0	Imputed		0		0		0	1
91	11.03.12.3_04g	10/1/2007	5	0	Imputed		0		0		0	1
91	11.03.12.3_15l	10/1/2007	7.5	0	Imputed		0		0		0	1
92	11.05.61.1_01a	10/1/227	9.82	0	Actual		0		0		0	1
92	11.01.11.4_05	10/1/2003	0	0	Actual		0		0		0	
92	11.01.13.2_03o	10/1/2003	0	0	Actual		0		0		0	
92	11.01.14.2_01h	10/1/2003	0	0	Actual		0		0		0	
92	11.01.14.2_01o	10/1/2003	0	0	Actual		0		0		0	
92	11.01.14.2_01s	10/1/2003	0	0	Actual		0		0		0	
92	11.01.14.2_01u	10/1/2003	0	0	Actual		0		0		0	
92	11.01.14.2_01v	10/1/2003	0	0	Actual		0		0		0	
92	11.01.14.2_01x	10/1/2003	0	0	Actual		0		0		0	
92	11.01.14.2_01y	10/1/2003	0	0	Actual		0		0		0	
92	11.01.15.3_01	10/1/2003	0	0	Actual		0		0		0	
92	11.01.15.3_01a	10/1/2003	0	0	Actual		0		0		0	
92	11.01.15.3_01c	10/1/2003	0	0	Actual		0		0		0	
92	11.01.14.2_01e	10/1/2003	0	0	Actual		0		0		0	
92	11.01.15.3_01e	10/1/2003	0	0	Actual		0		0		0	
92	11.01.14.2_01d	10/1/2003	0	0	Actual		0		0		0	
92	11.01.11.4_05a	10/1/2003	0	0	Actual		0		0		0	
92	11.01.11.4_05b	10/1/2003	0	0	Actual		0		0		0	
92	11.01.11.4_05c	10/1/2003	0	0	Actual		0		0		0	
92	11.01.11.4_05d	10/1/2003	0	0	Actual		0		0		0	

Record: 90324 of 197297

The DTUP can't continue the data transfer and thus produce the following error.



The issue can be resolved by correcting the data error.

4.3 A Summary of Main Difference between Access and SQL

Here is a list of data types in each environment, and how they are different. Some datatypes from SQL Server were left out (e.g. SQL_VARIANT, TABLE).

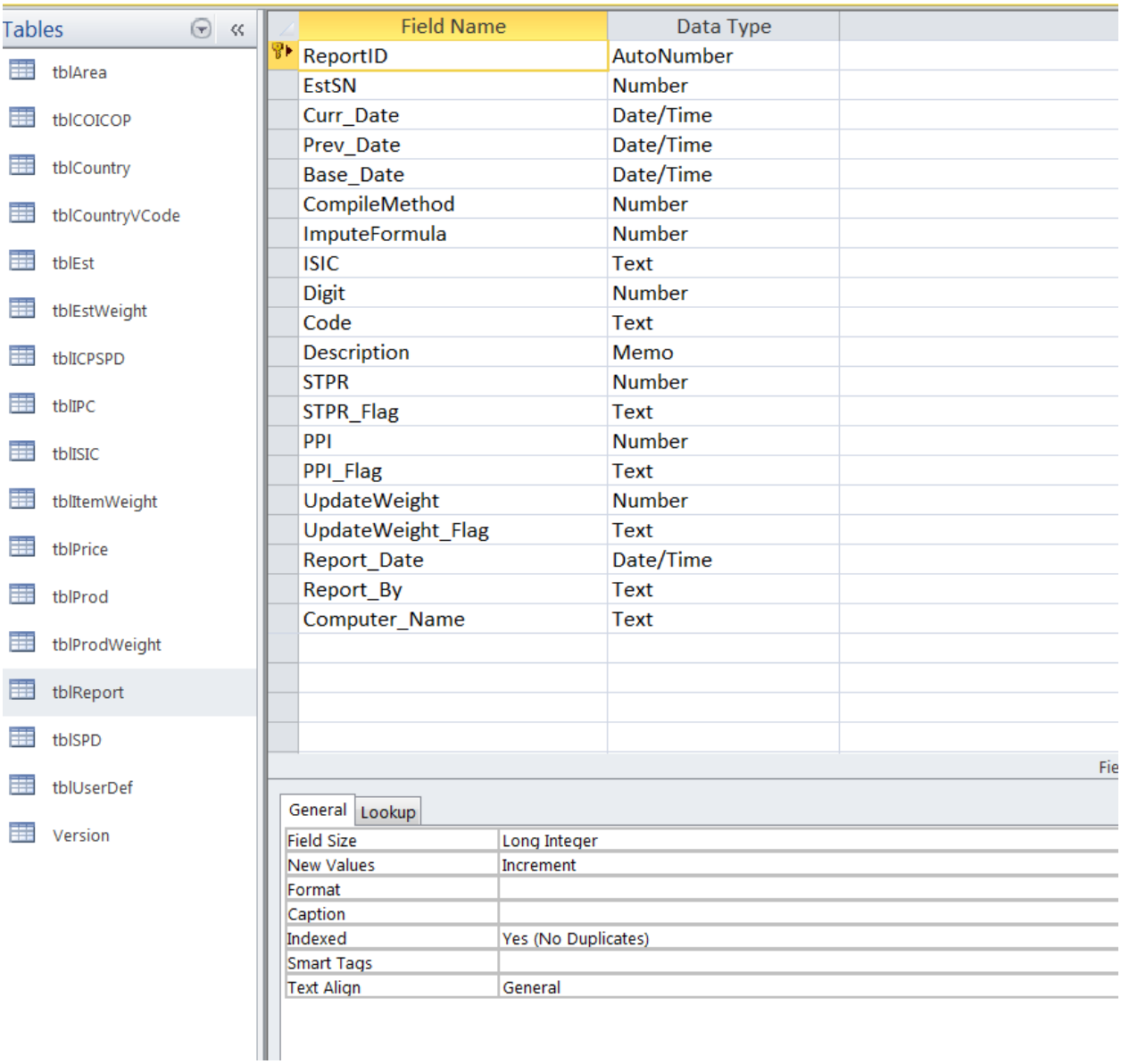
Access	SQL Server	SQL Server Definition
Yes/No	BIT	(Integer: 0 or 1)
Number (Byte)	TINYINTw	(Positive Integer 0 -> 255)
Number (Integer)	SMALLINT	(Signed Integer -32,768 -> 32,767)
Number (Long Integer)	INT	(Signed Integer -(2^31) -> (2^31)-1)
(no equivalent)	BIGINT	(Signed Integer -(2^63) -> (2^63)-1)
Number (Single)	REAL	(Floating precision -1.79E + 308 -> 1.79E + 308)
Number (Double)	FLOAT	(Floating precision -3.40E + 38 -> 3.40E + 38)
Currency	MONEY	(4 decimal places, -(2^63)/10000 -> ((2^63)-1)/10000)
Currency	SMALLMONEY	(4 decimal places, -214,748.3648 -> 214,748.3647)
Hyperlink	(no equivalent - use VARCHAR())	
Decimal	DECIMAL	(Fixed precision -10^38 + 1 -> 10^38 - 1)
Numeric	NUMERIC	(Fixed precision -10^38 + 1 -> 10^38 - 1)
Date/Time	DATETIME	(Date+Time 1753-01-01 -> 9999-12-31, accuracy of 3.33 ms)
Date/Time	SMALLDATETIME	(Date+Time 1900-01-01 -> 2079-06-06, accuracy of one minute)
Text(n)	CHAR(n)	(Fixed-length non-Unicode string to 8,000 characters)
Text(n)	NCHAR(n)	(Fixed-length Unicode string to 4,000 characters)
Text(n)	VARCHAR(n)	(Variable-length non-Unicode string to 8,000 characters)
Text(n)	NVARCHAR(n)	(Variable-length Unicode string to 4,000 characters)
Memo	TEXT	(Variable-length non-Unicode string to 2,147,483,647 characters)
Memo	NTEXT	(Variable-length Unicode string to 1,073,741,823 characters)

OLE Object	BINARY	(Fixed-length binary data up to 8,000 characters)
OLE Object	VARBINARY	(Variable-length binary data up to 8,000 characters)
OLE Object	IMAGE	(Variable-length binary data up to 2,147,483,647 characters)
Autonumber Autoincrement	IDENTITY	(any numeric data type, with IDENTITY property)

4.4 tblReport Issue

The tblReport table generated by the previous version of PIPS had a different schema as that of the current version. The correct the table design (for the most current version) of tblReport is shown below.

The table has ReportID field as identity field which does not allow null value. In the Identity column, no duplicated value is allowed (each integer is unique). Value in the column is auto-generated and with one increment in each row. The column also has got primary key on it which plays a pivotal role in the relational database. Without these settings, the software is not going to work properly.



4.5 Timeout Expires Error.

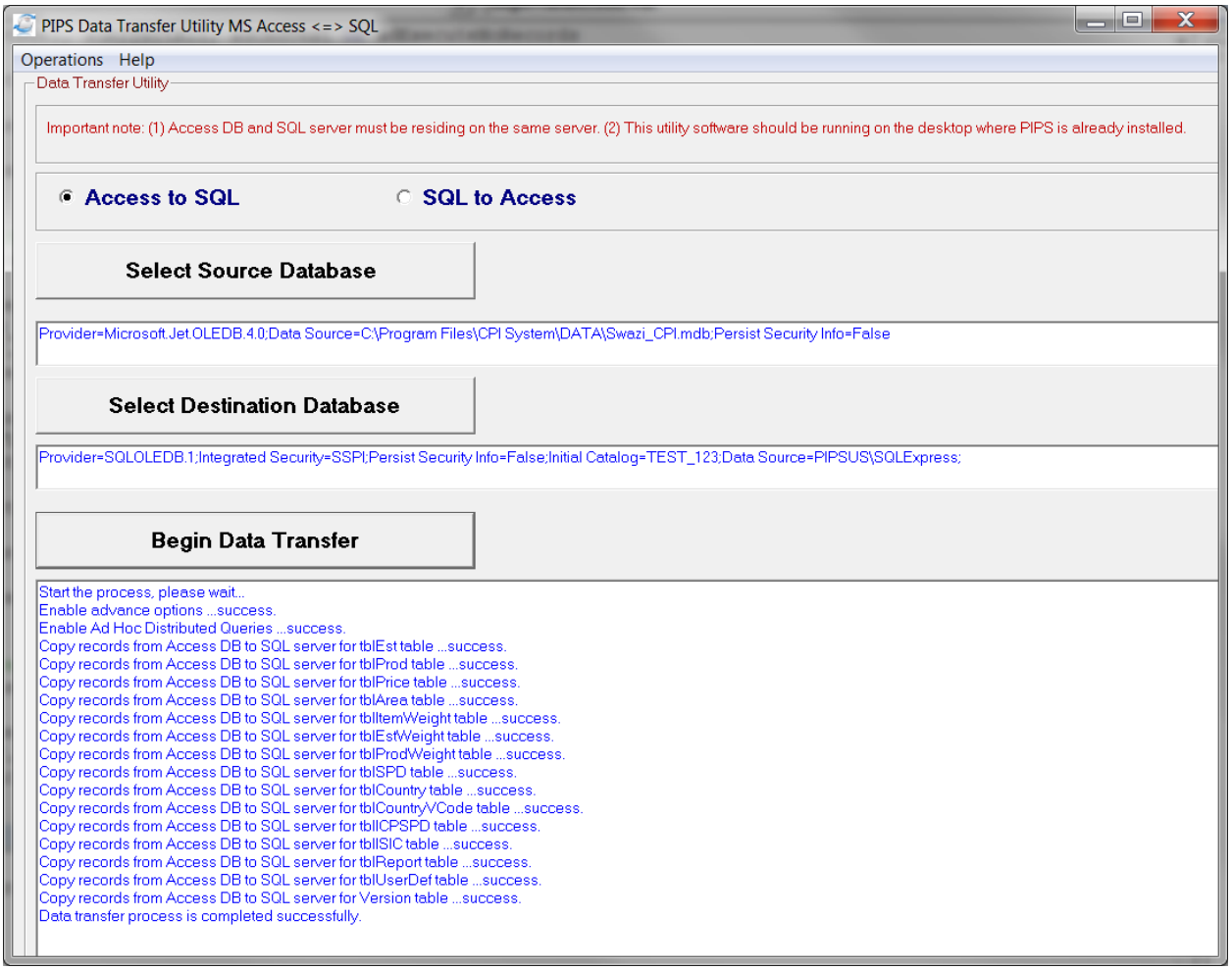
In the earlier trial version of DTUP, user will sometimes get “Timeout Expires” error. This has been resolved by extending the command time out. Please contact us if this error occurs again.

4.6 The Integrity of the Data during the Transfer process

The DTUP will delete the rows before insert the new data from the source database. In case of the insert transaction fails, the deletion of the previous data will be automatically rolled back. Only when both steps are completed successfully, then the transaction will be committed.

4.7 The Completion of the Data Transfer

There are altogether 17 tables in the PIPS database, and all of them need to be transferred before the whole process is finished successfully. At the end the process, user will see a message of “Data transfer process is completed successfully”.



Appendix II: Enable Remote Connections on SQL Server

Enable Remote Connections on SQL Server

By PIPSUS Support Team
Dr. Chouikha

achouikha@gmail.com
October 10, 2010, Version 1.0

Grant User Access

To allow access to users in a Windows domain

1. Open the SQL Server Management Tool.
2. Log on to the instance of SQL Server. This step presumes that you are the administrator of the computer and the instance of SQL Server.
3. In Object Explorer, expand the Security node.
4. Right-click Logins and click New Login....
5. Type the name of the user to allow by using the *domain\username* format.
6. Click Search and use the dialog to confirm that the user is a Windows authenticated user. Then close the New Login dialog.
7. In Object Explorer, expand the Databases node.
8. Expand the node of the database that you want to grant access to.
9. Expand the Security node of the database.
10. Right-click the Users node and click New User.
11. In the User name box, type in a name for the user.
12. In the Login name box, type the name of the user by using the domain\username format.
13. In the Role Members list box, select the role that you want to grant the user for the database. Common options are db_dataread and db_datawriter.

Enable a Port

This is a one-time procedure that you perform on the computer that hosts the instance of SQL Server to allow access to the computer through a specific port. The default port set by SQL Server is 1433. If you change the default, follow the directions below and change the port number as appropriate.

To enable port 1433 on Windows Vista

1. On the Start menu, click Control Panel.
2. Under Security, select Allow a program through Windows Firewall.
3. If prompted, click Continue. This presumes you are the administrator of the computer.
4. In the Windows Firewall Settings dialog, click Add Port.

5. In the Name box, type a name, such as SQL Server Port.
6. In the Port number box, type 1433. Use the default protocol TCP.
7. Click OK.

To enable port 1433 on Windows 7

8. On the Start menu, click Control Panel.
9. Click System and Security.
10. Click Windows Firewall.
11. Click Advanced Settings.
12. In the Windows Firewall with Advanced Security on Local Computer dialog, click Inbound Rules.
13. In the Actions panel, click New Rule….
14. In the New Inbound Rule Wizard, click Port, and then click Next.
15. In the Specific local ports: box type 1433. Ensure that the (default) TCP radio button is selected and then click Next.
16. Ensure that the (default) Allow the Connection radio button is selected and then click Next.
17. Clear or select the Domain, Private, and Public checkboxes as appropriate, and then click Next.
18. In the Name box, type an appropriate name, such as SQL Server Port, and then click Finish.

Set a Protocol

This is also a one-time procedure that you perform on the computer that hosts the instance of SQL Server to indicate which protocol to use when communicating with remote clients. This procedure uses the TCP/IP protocol.

To set the protocol

1. Open the **SQL Server Configuration Manager** application. This is found in the **Configuration Tools** folder of the Microsoft SQL Server 2008 folder.
2. Expand the **SQL Server Network Configuration Manager** node.
3. Expand the **SQL Server Network Configuration** node.
4. Click **Protocols for MSSQLSERVER**.
5. Right-click **TCP/IP** and click **Enable**.

   After setting the protocol you must restart the SQL Server service.

To restart the SQL Server service

1. In the **SQL Server Configuration Manager** application, click the **SQL Server Services** node.
2. Right-click **SQL Server (MSSQLSERVER)** and click **Restart**.