Long memory and seasonality in US Consumer Price Inflation: an empirical investigation at varying levels of aggregation.

Marius Ooms

December 31, 1996

Econometric Institute, Erasmus University Rotterdam
P.O. Box. 1738, NL- 3000 DR Rotterdam
email: ooms@few.eur.nl
URL: http://www.eur.nl/few/ei/faculty/ooms
preliminary and incomplete
For Presentation at the Econometric Society North American Winter Meeting
New Orleans, 4 Jan 1997

Abstract

We examine the persistence of US consumer price inflation at different levels of aggregation using models allowing for fractional integration. We use 72 series of monthly data for the period 1959.1-1995.12. We use a flexible seasonal model which allows for different fractional orders of integration at different seasonal frequencies. We find that the order of integration, and therefore the persistence, is highest in the more aggregated series. We find the seasonal extension which allows for long memory at seasonal frequencies to be relevant at different levels of aggregation: for commodities less food and energy, for apparel and its sub-items, and for e.g. fish and seafood. Some inflation series seem to be antipersistent, like the item fresh vegetables. Although inflation in energy commodities is very volatile, it is not persistent compared with other categories like food, and service and utility items.

Contents:

1 Introduction

We examine the persistence of US consumer price inflation at different levels of aggregation using models allowing for fractional integration. The usefulness of the concept of fractional integration for flexible modeling of extreme persistence (slow decay in autocorrelation function) and corresponding low frequency spectral behavior is widely established, especially for post war inflation series, see e.g. Hassler and Wolters (1995). Baillie (1996) wrote an extensive survey on the history and use of long memory models in econometrics. Robinson (1994) gave a technical overview of the econometric literature. In an earlier paper Ooms(1995)we gave an overview of seasonal extensions of long memory models. The literature is still growing at a considerable pace.

The extra flexibility one obtains by allowing for fractional integration can have important consequences for inference about the (long-run) mean of the inflation process. The variance of the sample mean decreases very slowly with increasing sample size if the process is fractionally integrated. This leads to a large uncertainty about the "population" mean, see Hosking (1996) . A fractional order of integration also determines a "fractional" rate of growth of long-horizon out-of-sample forecast intervals for log prices. These might be more realistic than the familiar growth rates obtained for integer orders of integration, see Diebold and Lindner(1996).

A widely cited statistical explanation of long memory in inflation rates stems from Granger (1980), recently extended in Granger and Ding (1996), who showed that specific averaging of short memory processes can explain long memory in the aggregate. This argument to explain long memory has also been used in economic theory, see e.g. Haubrich and Lo (1989) .

We try to get insight in the relevance of the "aggregation explanation" by investigating 72 price indexes describing urban consumer price inflation as defined by the US Bureau of Labor Statistics. We use monthly seasonally unadjusted data for the period 1958-1995. We use univariate flexible seasonal ARFIMA models, described earlier in more detail in Ooms (1995), which allow for efficient simultaneous statistical analysis of fractional orders of integration at the zero frequency and at seasonal frequencies. We find many models with both a significant "short memory"-ARMA part and a significant (non)seasonal "long memory" part.

We find "long memory" in a large majority of the aggregated price indexes. We find higher orders of integration for higher levels of aggregation, which supports the aggregation argument. We estimate orders of integration at the zero frequency in the "infinite variance region" for inflation in CPI all item indexes, and for energy services. For the latter index the estimates also indicate long memory at seasonal frequencies. Inflation in e.g. energy commodity prices seems to be covariance stationary, whereas inflation in e.g. medical services does not.

The next section discusses empirical aspects of long memory and it introduces our model. In section 3 we consider the "aggregation explanation of long memory and we treat estimation and selection procedures. We investigate some aspects of the data in more detail in section 4. Section 5 discusses the estimation results. Section 6 concludes.

2 Flexible Seasonal Long Memory and the ARFIMA model

2.1 Notation and nomenclature

Consider the following discrete time model with a periodically varying mean for a series with 2 observations per year:

$$y_t = \beta_0 + \beta_\pi(-1)^t + x_t, \quad t = 1, 2, \ldots, T \tag{1}$$

where x_t is a zero mean covariance stationary process with *autocovariance function* $\gamma_k = E(x_t x_{t-k})$, $k = \ldots, -1, 0, 1, \ldots$, with *spectrum* $f_x(\omega)$, $-\pi < \omega \leq \pi$, and with *long run variances*

$$\sigma_0^2 = \lim_{T\to\infty} T \cdot Var[\widehat{\beta_{0T}}]$$

and

$$\sigma_\pi^2 = \lim_{T\to\infty} T \cdot Var[\widehat{\beta_{\pi T}}]$$

at frequencies 0 and π, respectively, where

$$\widehat{\beta_{0T}} = \sum_{t=1}^{T} x_t/T$$

is the (ordinary least squares) estimator of the mean of y_t, and where

$$\widehat{\beta_{\pi}} = \sum_{t=1}^{T} (-1)^t \, x_t/T$$

is the corresponding estimator of the (angular) frequency π component, i.e. the component with period 2. In the sequel we use subscripts and superscripts involving π to denote properties for specific seasonal frequencies.

Suppose x_t is *short range dependent* (*short memory).* Then it has three properties:

$$\lim_{n\to\infty} \sum_{k=0}^{n} |\gamma_k|$$

exists and is finite (absolute summability of covariances), $f_x(\omega)$ is bounded on $(-\pi,\pi]$ (finite spectrum) and σ_0^2 and σ_π^2 exist and are finite, so that $\sigma_0^2 = 2\pi f_x(0)$ and $\sigma_\pi^2 = 2\pi f_x(\pi)$.

We refer to processes which lack one of these properties to be *long memory* or *long range dependent.* Well known non-stationary long memory processes are the simple random walk process x_{0t} defined by

$$(1-L)\, x_{0t} = u_t, \ t = 2, 3, ...,$$

and the seasonal random walk process x_{2t} defined by

$$(1-L^2) x_{2t} = u_t, \ t = 3, 4, ...,$$

where L is the lag operator: $L^k x_t = x_{t-k}$, $k = \dots, -1, 0, 1, \dots$ and u_t is a white noise process with finite variance σ_u^2. It is immediately clear that x_{0t} and x_{2t} are long memory. The long run variance of x_{1t} does not exist at frequency zero:

$$T \cdot Var[\sum_{t=1}^{T} x_{0t}/T] = T \cdot Var[\sum_{t=1}^{T} \sum_{s=1}^{t} u_s/T] = O\left(T^2\right), T \to \infty \tag{2}$$

but

$$T \cdot Var[\sum_{t=1}^{T} (-1)^t x_{0t}/T] = T \cdot Var[\sum_{t=1}^{T} (-1)^t \sum_{s=1}^{t} u_s/T] \equiv T \cdot Var[(1-L^2)^{-1} u_t/T] = \sigma_u^2/2, \tag{3}$$

so that its long run variance does exist at frequency π. x_{2t} does not have a finite long run variance at either frequency 0 or π. These "integer unit root" processes can be made short memory using ordinary difference filters. x_{1t} needs first differencing to become short memory at all frequencies. One needs to apply the seasonal difference operator, $\left(1-L^2\right) = (1+L)(1-L)$ to x_{2t} to render it short memory. Applying only first differencing to x_{2t} one obtains a process which we denote by $x_{\pi t}$:

$$(1+L) x_{\pi t} = u_t, \, t = 2, 3, \dots,$$

which does have a long run variance at frequency zero, but does not at frequency π. The long memory in $x_{\pi t}$ is removed by taking yearly sums (or averages). Suppose now that we observe $x_{\pi t}$ only once every year, say at the even observations t':

$$x_{\pi t'} = \sum_{s'=1}^{t'} v_{s'}, \ v_{s'} = (u_{2s} - u_{2s-1}), \ t' = 1, 2, \dots, \ s = 1, 2, \dots$$

The process $x_{\pi t'}$ is again a random walk process, because $v_{t'}$ is white noise. The infinite long run variance of $x_{\pi t}$ at frequency π is shifted to an infinite long run variance of $x_{\pi t'}$ at frequency 0. This example of aliasing shows that long-memory at a seasonal frequency depends crucially on the observation frequency. A process with long memory at one of the seasonal frequencies which is observed at yearly intervals appears to behave as an "ordinary" zero frequency long memory process. Let t denote time at intervals of six months, so $t^{'}$ denotes time at yearly intervals. Let $x_{\pi t}$ denote changes in log prices. Suppose $x_{\pi t}$ is long memory at frequency π only. In this case we see that $(1+L)\,x_{\pi t}$, i.e. the inflation over a year, is short memory, whether observed at each t or only at each $t^{'}$, but $x_{\pi t'}$, i.e. the inflation of every last six-month period in each year is long memory at the zero frequency. These results continue to be relevant in more general cases considered below.

In the econometric literature the process x_{0t} with one "AR unit root" is called $\mathrm{I}_0(1)$: integrated of order one at frequency zero, whereas $x_{\pi t}$ is $\mathrm{I}_\pi(1)$ and x_{2t} is both $\mathrm{I}_0(1)$ and $\mathrm{I}_\pi(1)$. The (short memory) white noise process u_t which results after appropriate differencing is called I(0) at all frequencies, see Hylleberg et al. (1990). Suppose we difference u_t yet again. So we get an "MA unit root" $x_t = (1-L)\,u_t$. This process is covariance stationary, so by inspection of the spectrum $f_x(\omega) = (2\pi)^{-1}\,\sigma_u^2|(1-e^{i\omega})|^2$ we see that this so-called "overdifferenced" or "anti-persistent" process has zero long run variance at frequency 0. This process is called integrated of order minus 1 at frequency zero: $I_0(-1)$.

2.2 Fractional ARIMA models

In this paper we use generalizations of well known models for fractional integration that fill the gaps between I(0) and I(1) processes. The simple fractionally integrated process x_t is defined by

$$(1-L)^d\,x_t = u_t, t = 1, 2, \ldots, \tag{4}$$

where $(1-L)^d = 1 - dL - (1/2)d(1-d)L^2 - ....$ denotes the fractional differencing operator defined by an ordinary binomial expansion. x_t is covariance stationary for $d < 0.5$. Its autocovariance function dies out at a polynomial rate:

$$\gamma_k = O(k^{2d-1}),\ k \to \infty, \tag{5}$$

see e.g. Hosking (1981, Theorem 2). This entails an infinite spectrum and an infinite long run variance at frequency zero if d is positive:

$$f_x(\omega) \sim \omega^{-2d}, \omega \to 0 \tag{6}$$

and

$$Var[\sum_{t=1}^{T} x_t/T] = O\left(T^{2d-1}\right), T \to \infty, \tag{7}$$

while $f_x(\omega) = 0$ and $\sigma_0^2 = 0$ if $d < 0$. The process is still mean reverting, for $0.5 < d < 1$, i.e. the impulse response function dies out eventually, but then x_t is no longer finite variance.

These three basic are shared by a wider range of long memory processes. It is relevant here to note some empirical (finite-sample) properties of these processes, which can help to judge whether this type of model is appropriate for a particular data set.

Suppose $x_t$is mean stationary. How can we see whether a realization of this process is long memory with $0 < d < 1$? It is of primary importance to have a long realization, otherwise there will be too little information in the sample to discover anything. First one should look at a time series plot of the data, of course. This plot should display long "nonperiodic cycles" as described by Mandelbrot (1972). These long cycles lead to the slow convergence of the sample mean to the "population" mean for positive d noted in (7). Postwar US inflation has this pattern with a few persistent high inflation periods and several long low inflation periods, see appendix B, figures 1-3, where we plot 72 inflation series.

Second, one should look at the pattern of decline in the sample autocorrelations at high lags. See Newbold and Agiakloglou (1993) and Hassler (1994b) for recent expositions on this topic. The decline is slow. One has to bear in mind that ordinary sample autocorrelations, i.e., the correlations in deviation from the sample mean, grossly understate their population counterparts for $d$close to 0.5. The sample autocorrelation functions for $0.5 < d < 1$, are qualitatively similar to those for I(1) processes and decline approximately linearly. In order to tell the difference with I(1) processes one should also look at the differenced process. This has order of integration $d^* = d - 1 < 0$, so it has an inverse autocovariance function $\gamma^{(i)}(d^*)$ which equals the autocovariance function $\gamma(-d^*)$ of process x_t in (4) with $d = -d^*$. Consequently the population inverse autocorrelation dies off slowly for $d^* < 0$, whereas it cuts off immediately for $d^* = 0$. Chang and Dickey (1994) provided approximations for the sample inverse autocorrelation function for $d^* = -1$.

Third, one should take look at log of the sample spectrum or periodogram of the series. For $d < 0$, the *log* periodogram diverges to minus infinity near frequency zero, whereas it diverges to plus infinity for $d > 0$. There is a large literature on the behavior of the log-periodogram for long memory processes near frequency zero, see e.g. Robinson (1995).

Fourth one can view the sample variance time function of the process x_t, i.e. the growth of the sample variance of partial sums $\widehat{V_k^0} = \widehat{Var}[\sum_{t=1}^{t+k} x_t]/\widehat{Var}[x_t]$ as a function of k. The population variance time function

$$V_k^0 = O\left(k^{1+2d}\right),\ k \to \infty \tag{8}$$

for $-0.5 < d < 0.5$, see e.g. Diebold and Lindner (1996). Note that $\lim_{k\to\infty} V_k^0/k = \sigma_0^2/\gamma_0$ is finite for x_t with with $d \leq 0$. The use of these "variance ratios" has been popular in empirical macroeconomics, see e.g. Cochrane (1988), who applied it to the first differences of supposedly I(1) processes. In his case the partial sums for the differenced process corresponded to "long differences" for the integrated process, so they had a direct interpretation. Furthermore one can look at robust modifications of these functions like rescaled adjusted range functions, see e.g. McLeod and Hipel (1978). These functions are useful if the existence of the variance of x_t is in doubt because of fat (hyperbolic) tails in the distribution of the innovation process u_t. Lo (1991) developed a semi-parametric (MLR-)test for $d = 0$ based on the rescaled range.

Finally one can take a standard statistical approach by developing an appropriate parametric statistical model for the error process u_t, estimating and testing the model. That is our approach. Initially we confine ourselves to modeling the first and second order moments of x_t.

Some of the inflation series do not only have a large power spectrum near the zero frequency, but also near the seasonal frequencies. This "seasonal infinite spectrum" could also be of the long memory type. Suppose again for expository reasons that we only have two observation per year.

A straightforward seasonal extension of the fractionally integrated model is then given by

$$(1-L)^{d_0}(1+L)^{d_\pi} x_t = u_t, t = 1, 2, \ldots \tag{9}$$

where one allows for different orders of integration at frequencies 0 and π, in line with the econometric "seasonal unit root" literature, see e.g. Hylleberg et al. (1990).

Consider first the "pure frequency π model" with $d_0 = 0$. This model is covariance stationary and invertible for $-0.5 < d_\pi < 0.5$. Assume this condition holds. Then its periodically transformed autocovariance function dies off hyperbolically:

$$(-1)^k \gamma_k = O\left(k^{2d_\pi - 1}\right), k \to \infty$$

the spectrum behaves hyperbolically near π :

$$f_x(\omega) \sim |\omega - \pi|^{-2d_\pi}, |\omega - \pi| \to 0$$

and the periodic variance time function

$$\widehat{V_k^\pi} = \widehat{Var}[\sum_{t=1}^{t+k}(-1)^t x_t]/\widehat{Var}[x_t] = O\left(k^{1+2d_\pi}\right), k \to \infty,$$

which entails that $T \cdot Var[\widehat{\beta_{\pi T}}] = O\left(T^{2d_\pi}\right), T \to \infty$, see (1), so that this x_t is long memory at frequency π for $d_\pi > 0$. We use the notation of Proietti (1996) , who proposed seasonal analogues of the variance time functions, for frequencies $\pi/2$ and π.

Ray (1991,1993) was one of the first to apply a similar *flexible* fractional seasonal ARIMA model to economic time series. She used the flexible filter $\prod_{i=1}^r (1 + (-1)^{g_i} L^{s_i})^{d_i}$, $g_i = 0, 1$, $s_i = 1, 2, 3, \ldots$. Porter-Hudak (1990) successfully applied a *restricted* version to US monetary aggregates. She used the fractional seasonal difference operator $(1-L^s)^d$, which obtains here by putting $d_0 = d_\pi$ and $s = 2$. Other recent studies using a "flexible seasonal long memory" approach are Hassler (1994) , and Ooms (1994), who examined log-periodogram regression. Giraitis and Leipus (1995) extended earlier models proposed by Gray et al. (1989) and Anděl (1986) and Anděl and Gómez, (1988) to

$$(1-L)^{d_0} \prod_{i=1}^{r-1} \left(1 - 2\cos\lambda_i L + L^2\right)^{d_{\lambda_i}} (1+L)^{d_\pi} x_t = u_t \ , t = 1, 2, \ldots \tag{10}$$

to allow for different orders of integration d_{λ_i} at more than one unknown frequency λ_i, $0 < \lambda_i < \pi$. The operators $\left(1 - 2\cos\lambda_i L + L^2\right)^{d_{\lambda_i}} = (1 - e^{i\lambda_i}L)^{d_{\lambda_i}} \left(1 - e^{-i\lambda_i}L\right)^{d_{\lambda_i}}$ are analogously defined by a convolution of standard binomial expansions, now involving powers of $e^{i\lambda_i}L$. Frequency π can be seen as a special case: $(1+L)^{d_\pi} = (1-e^{i\pi}L)^{d_\pi} = 1 - d_\pi(-L) - (1/2)d_\pi(1-d_\pi)(-L)^2 + ...$ The exceptional definitions for the orders of integration for $\lambda_i = 0$ and $\lambda_i = \pi$ lead to analogous regions for covariance stationarity and invertibility at all λ_i: $-0.5 < d_{\lambda_i} < 0.5$, $i = 0, ..., r$.

Our flexible seasonal AutoRegressive Fractionally Integrated Moving Average (ARFIMA) model is obtained by extending u_t with a short memory, covariance stationary, invertible and identified ARMA(p, q) structure:

$$(1 + \phi_1 L + \ldots + \phi_p L^p)\, u_t = (1 + \theta_1 L + \ldots + \theta_q L^q)\, \epsilon_t, \tag{11}$$

where $\Phi(z) = (1 + \phi_1 z + \ldots + \phi_p z^p) = 0$ and $\Theta(z) = (1 + \theta_1 L + \ldots + \theta_q L^q) = 0$ have no common roots or roots inside the unit circle, and where ϵ_t is white noise process with finite variance σ^2_ϵ. We assume ϵ_t to be Gaussian. In the flexible seasonal ARFIMA model, the λ_i are fixed a priori at $2\pi i/S$, $i = 0, \ldots [S]/2$, where S is the number of observations per year and [.] denotes the entier function. $S = 12$ in our empirical work below. To avoid double indices we use the short hand notation d_i for d_{λ_i} from now on, so

$$d_0 = d_0,\ d_1 = d_{2\pi/12}, d_2 = d_{4\pi/12}, \ldots, d_6 = d_{12\pi/12} = d_\pi.$$

The spectrum, the AR(∞) and the MA(∞) expansion of this model are easily derived and computed. The autocovariance function is not. Viano et al. (1995) extended the model to complex-valued and non-polynomial filters. They also derived an expression for the autocovariances of the model (10) – (11), which they called "extended fractional ARIMA process", but their expression does not allow stable numerical computation of the autocovariance function. Therefore one cannot draw exact pseudo-samples with "flexible seasonal long memory" for Monte Carlo analysis as one can for simple ARFIMA models, where one can use forms of the Levinson-Durbin-Whittle algorithm as described in e.g. Hosking (1984). Except for some simple special cases like $x_t = \left(1 + L^2\right)^{d_3} u_t$, one has to resort to drawings from finite order AR or ARMA models, but these are either explosive or integrated of integer order by construction. Exact (Gaussian) Maximum Likelihood estimation is not possible either, since the exact likelihood requires an exact expression for the autocovariance function.

2.3 Inference

Robinson (1994) and Ray (1991) discussed inference in models where the "long memory frequencies" λ_i are known a priori. A fully parametric approach using an asymptotically efficient estimator leads to standard Gaussian limit distributions for the parameter estimates and therefore to standard asymptotic Gaussian inference about the orders of integration, provided one does not test hypotheses on or outside the boundary of the stationarity region directly. The standard (seasonal) unit root hypotheses $d_i = 1$ and short memory hypotheses $d_i = 0$ as studied in Hylleberg et al. (1995) can of course be shifted into the stationary region by appropriate a priori differencing. One can also apply a *fractional* filter to x_t a priori, e.g. by using its AR(∞) form. The necessary presample values are then set equal to their (seasonal) sample means, see e.g. Beran (1995). Sowell (1987) showed the latter approach to be equivalent to the subsequent application of a Fourier transform to x_t, application of the fractional filter in the Frequency domain, and an inverse Fourier transform.

Note further that popular (augmented) Dickey-Fuller-type OLS-based tests for (seasonal) unit roots at a frequency λ_i (test of the model with $d_i = 1$) do not have the correct size if there is long memory remaining in the residual error process at frequency λ_i. See e.g. Chan and Terrin (1995), who studied the asymptotic behavior of these tests allowing for zero frequency long memory in the residual u_t. So OLS based unit root tests are difficult to apply in the context of (seasonal) long memory, to say the least. Jeganathan (1996) extended results of Chan and Terrin to tests for (seasonal fractional) (co)-integration in a GLS setting (accommodating for frequency zero long memory errors), but his results are not operationalized as yet.

On the other hand, tests for the I(0) hypothesis against I(1), like the Nyblom (1989) test for constancy of β_0 in (1) against random walk variation in β_0 can be useful. The KPSS test, Kwiatkowski et al. (1992), is probably the best known test for $\mathrm{I}_0(0)$ against $I_0(1)$ in econometrics. This test has asymptotic power against all $\mathrm{I}_0(d)$ hypotheses with $0 < |d| < 0.5$, see Lee and Schmidt (1995). These tests for short memory often use statistics like the variances of partial sums, as in (8). Canova and Hansen (1995) developed a flexible seasonal analogue of Nyblom's test for stability of e.g. β_π in (1). Although these tests have good asymptotic power properties, it is difficult to control their size if u_t is not white noise. Semi-parametric corrections which are needed to make them insensitive to short memory components in u_t under the null do not work well in finite samples, see e.g. Robinson (1994a, p. 1426), who stated that "a parametric approach might in any case be preferred when n is not very large", where n is the number of observations .

3 Aggregation, Estimation and Model Selection

There are several explanations of long memory in aggregate economic series. Most of this research on the "sources" of stationary long memory in econometrics is only theoretical. We confine ourselves to the analysis of aggregation of independent processes. Time aggregation of zero frequency long memory processes is a well studied subject. One of the essential properties of these processes is "asymptotic self-similarity", see e.g. Cox (1991). The basic asymptotic properties (5), (6) and (7) do not change if we study (long) averages

$$x_{t'} = \sum_{t=t'l+1}^{(t'+1)l} x_t/l,\ l = 1, 2, \ldots$$

instead of the original process x_t. In case there is long memory at other frequencies, one has to modify these results of course, see section 2.1.

3.1 Aggregation

One can introduce long memory by mixing an infinite number of short memory processes. Granger (1980) considered the aggregate of infinite number of independent short memory AR(1) components $w_{jt}, (1 - \alpha_j L)\, w_{it} = \epsilon_{it}$, where each α_j is a drawing from a beta-like distribution with support $(0, 1)$, parameters $u, v > 0$ and density:

$$\frac{2}{B(u,v)} \alpha^{2u-1} \left(1 - \alpha^2\right)^{v-1} I_{(0,1)}(\alpha)$$

with indicator function $I_{(0,b)}(\alpha)$. Long memory in the aggregate arises for all $0 < v < \infty$: The aggregate is an $I(1 - v/2)$ process according to Granger. Note that the distribution for α does not converge to zero as $\alpha \to 1$, so that "near unit root" processes can be drawn. For $v < 1$ there is even a pole at 1 and this leads to an infinite variance for the aggregate. In empirical examples it is interesting to see which underlying processes have near unit roots, and which have not. Granger and Ding (1996, section 4) discussed some extensions involving nonlinear transformations of the difference operator and aggregation of mixtures of $\mathrm{I}_0(d_j)$ processes for a range of distributions for d_j, The sum of a large number of $\mathrm{I}_0(d_j)$processes with d_j drawn from a density of the type $c(1-d)^{-1/2}$ with support $[0, D]$ was shown have an $I(D)$ component. This corresponds to aggregation results for a finite number of independent $I(d)$ processes, where the process with the highest order of integration "dominates" the spectrum near zero.

Cox (1991) introduced a different way to mix short memory processes in order to get a long memory aggregate. His approach boils down to aggregating a weighted mix of (continuous) time deformations for an infinite number of independent short memory processes. This time deformation entails that the large mass of the aggregate evolves much more slowly than the original underlying processes, with long memory features as a result.

These approaches work with a heterogeneous sample of underlying independent processes. It is not our purpose to identify these processes, but a disaggregated analysis can show whether more aggregated series display stronger long memory and it can show, at a given level of aggregation, which "section" of the underlying processes "causes" the long memory components. We can only achieve this if our series are long enough, so that there is enough information in the observed long run autocovariances. Our approach is univariate. At this stage we do not model relationships between the different series.

3.2 Estimation

There are three parametric approaches to estimation of ARFIMA models, exact maximum likelihood, operationalized by Sowell (1990), approximate frequency domain maximum likelihood, best known as Whittle estimation, proposed by i.a. Boes et al. (1989) and nonlinear least squares, see e.g. Beran (1995). For full years of data the Whittle estimator is easily shown to be independent of (estimates for) seasonal means, as explained in Ooms and Hassler (1996). This makes it computationally attractive, since the (seasonal) means of (seasonal) long memory processes are hard to estimate. The slow convergence for the OLS estimator for the means is clear from (7). The efficient (infeasible) GLS estimator, is only a touch more precise, see Samarov and Taqqu (1988) and Dahlhaus (1995). A feasible GLS type estimator would involve simultaneous estimation of 12 extra parameters in our case.

There are several asymptotically equivalent approximate frequency domain maximum likelihood estimators. Robinson (1994b) provided a detailed overview. We use the following estimator for $\psi = (d_0, d_1, \ldots, d_r, \phi_1, \phi_2, \ldots, \phi_p, \theta_1, \ldots \theta_q)$ given a data vector $x = (x_1, \ldots, x_T)$ of size T, where T is an integer multiple of the number of observations per year, S

$$\begin{aligned}
\max_{\psi} \ln L_x(\psi, x) &= -\frac{1}{2}\left\{\frac{1}{T}\sum_{j\in J} \ln g_x(\omega_j, \psi) + \widehat{\sigma}_\epsilon^2(\psi, x)\right\} \qquad (12)\\
\widehat{\sigma}_\epsilon^2(\psi, x) &= \frac{2\pi}{T}\sum_{j\in J} I_T(\omega_j)/g_x(\omega_j, \psi)\\
I_T(\omega_j, x) &= \frac{1}{2\pi T}\left|\sum_{t=1}^{T} x_t e^{it\omega_j}\right|^2\\
g_x(\omega_j, \psi) &= \left(\frac{2\pi}{\sigma_\epsilon^2}\right) f_x(\omega_j, \psi), \omega_j = 2\pi j/T\\
J &= \{1, ..., [T/2]-1\}/\{T/S, 2T/S, \ldots, [S/2-1]T/S\}
\end{aligned}$$

where $\widehat{\sigma}_\epsilon^2(\psi, x)$ denotes the residual variance estimator, $I_T(\omega_j, x)$ denotes the sample spectrum, or periodogram, at frequency ω_j, and $f_x(\omega_j, \psi)$ denotes the model spectrum for x at frequency ω_j. We exclude the frequencies where $\ln f_x(\omega_j, \psi)$ does not exist for all $-0.5 < d_i < 0.5$, $i = 0, \ldots, r$, $r = [S/2] = 6$. J denotes the index set of the remaining frequencies. Our estimator is therefore insensitive to changes in (seasonal) means for x_t: adding a fixed seasonal pattern to the data x_t should not change this approximate log-likelihood. Starting values for the optimization are derived as follows. We obtain values for $d_0, d_1, \ldots, d_r$, by log periodogram regression over the frequencies ω_j in the neighborhood of the corresponding long memory frequencies λ_i.

$$\ln I_T(\omega_j, x) = c_i + d_i \ln g_{ix}(\omega_j, d_i) + \eta_j, j \in J_i, i = 0, 1, \ldots r.$$

where $g_{ix}(\omega_j, d_i)$ is the spectral density of the "pure" fractionally integrated process at frequency $\lambda_i : g_{ix}(\omega_j, d_i) = g_x(\omega_j, 0, \ldots, 0, d_i, 0 \ldots)$. $J_0 = \{1, 2, \ldots m\}$, , $J_i = \{iT/S - m, \ldots iT/S\}/\{iT/S\}M$, $J_6 = \{T/2 - m, \ldots T/2 - 1\}$. We test whether d_i differs significantly from zero at significance level α, using standard regression inference, where we plug in the asymptotic variance for the error term $\sigma_\eta^2 = \pi^2/6$. In simulations we showed that this inference can be quite reliable for some pure FI models, Ooms (1995). We put the starting value for d_i equal to zero if the test does not reject. Otherwise we use the regression estimate as a starting value where we truncate it to lie in the stationary region.

We use quasi-Newton methods as described Press et al. (1986, section 10.7) Subsequently we filter x_t in the frequency domain using (10). Then we apply the least squares procedure as described in Lutkepohl (1993, section 7.3.4) to obtain starting values for the AR and MA parameters. We use quasi-newton methods as described Press et al. (1986, section 10.7) for optimization. We use a numerical Hessian of (12)to evaluate the covariance matrix for the parameter estimates. This method delivers sensible estimates in many cases, although familiar problems like lack of empirical identification in overparameterized models occur, where "non-automatic" procedures have to be followed.

3.3 Model Selection and Diagnostics

We have monthly data, so the model allows for 7 different long memory parameters in the long memory part (11), one for each frequency $\lambda_i = 2\pi i/12 i = 0, 1, \ldots, 6$. An ARMA(12,2) model for the short memory part (11) with the restrictions $\phi_3 = \ldots \phi_{11} = 0$ proved to fit adequately for large the majority of the series, see Tables 2-3 in appendix D below. We do not impose the stationarity restrictions on the short memory or long memory parameters, but check them after estimation. Note that estimates in the nonstationary region are not unlikely under the null of covariance stationarity with d_i close to 0.5. We use integer differencing at each frequency to check results. We apply a frequency domain test for the white noise hypothesis for the residual term $\widehat{\epsilon}_t$, due to Beran (1992) to check misspecification. We modify this test by leaving out the seasonal frequencies. This test can also be written in the time domain as a function of the sum of all estimable squared residual autocorrelation, so it is a kind of Box-Pierce test with a high number of lags. It may not be very powerful in the detection of misspecification of the short memory part, therefore we use a high significance level. We extend the model if the test rejects at a level of 20%. In some cases the model turns out to be overspecified, leading to near common roots in the AR-part and the MA-part, see examples in Table 3 in appendix D below. In that case we restrict the model and test it against the general model using likelihood ratio tests. It is known that estimated spectra for ARMA-models with near common roots near the unit circle can adapt to near-poles in the spectrum, typical for long memory processes, see Harvey(1989). This is one of the empirical identification problems that can occur. We compute the moduli and frequencies of the estimated roots of the estimated AR and MA process to monitor this problem. This selection process only uses the periodogram at standard frequencies as "sufficient" statistics.. This vector of periodogram ordinates has a one-to-one relation with the sample covariances of the data. Although our estimator does not depend on the overall seasonal means, it is not robust to large short run or moderate longer run changes in the seasonal means within sample.

4 Data

We examine the US CPI index for "all items". The index is published every month. It is one of the main economic indicators in the US. It is used to keep a large number of payments in line with changing prices. Data are made publicly available by the US Bureau of Labor Statistics. The CPI index is an aggregate in a number of ways. We study the aggregation over the prices of different consumer goods and services. We consider national averages price for urban consumers. We do not look at regional or rural prices.

Since we want to study long memory and aggregation we need long series for the same period. There are 72 monthly series available which start at least as far back as 1958. The BLS codes and their explanations are listed in Table 1 in appendix A.

There are four levels of aggregation, roughly corresponding to present number of digits in the BLS item code. The series with codes starting with "sa0" are the *overall* ,"all item", aggregates. These series represent overall price movements of a large majority of the items of consumer expenditure. One important category of items is left out at a time. The index "sa0l1e" (all items less food and energy) determines the well known "core inflation". At a lower level there are two types of divisions. One is based on broad expenditure categories: food, housing, transport etc., the codes for these series start with "sa" followed by a single-digit or two-digit number. The other division is based on properties of the items: commodities or services, durables or nondurables, etc. Codes for these series start with "sa" followed by a letter. The third level of aggregation considers broad three digit categories like "cereals and bakery products", here the codes start with "sa" followed by more digits. The lowest level of aggregation concerns small categories like "pork". These prices are only available for food items. Their codes start with "se0" or "se1". At a similar level of aggregation but, much more important for the consumer budget, we have prices for energy services and utilities. Codes for these series start with "se2". A number suggesting the level of aggregation is presented in the last column of Table 1, but the resulting division is not meant to indicate clear "jumps" in the level of aggregation between classes.

Most important prices are covered for the whole period, except for the price of home ownership, an item which has a weight of up to 20 percent in the basket of goods and services which is used to determine the CPI. The available series starts in 1968, and a structural change in the construction occurred in 1983, when the "rental equivalence" approach replaced the "asset price" approach used earlier.

In Appendix B, Figures 1-3, we present time series plots for the inflation series, i.e. monthly changes in the natural logarithm of the price index. The data range from 1958.1 to 1995.12. There are 444 observations. We plot separate series for each month to identify persistent changes in seasonal pattern.

Overall persistence in the "all item" series is clear, with high inflation in the seventies and low inflation in the mid eighties. At the second level of aggregation one sees food price inflation to be less persistent. The same goes for energy price inflation, although volatility in these series is higher than in the "all item" series. The inflation series for utilities and public transportation and the series for Apparel and upkeep show remarkably persistent changes in their seasonal pattern.

Persistence in inflation seems to be even lower for the less aggregated food components like meat and dairy products. Persistent changes in the seasonal component are still observed for categories like "Fruits and vegetables" and "women and girls apparel". At the lowest level of aggregation for the "se*"-series we see intriguing permanent changes in the seasonal pattern of electricity prices after 1974. Many of the inflation series for separate meat products display exceptionally high inflation figures in the latter part of 1973. Energy showed exceptional upward price changes in the beginning of 1974 and exceptional downward movements in the beginning of 1986. These outliers could justify the use of robust methods in a later stage of our statistical analysis. For the moment we confine ourselves to standard estimation techniques.

Appendix C, figures 4-6, presents the sample autocorrelation function (ACFs), $\widehat{\gamma_k}/\widehat{\gamma_0}$, $k = 1, \ldots 240$, for the series (in deviation from their seasonal sample means). Remember that these statistics are the basis for our modeling. Again we connect the values 12 periods apart to elicit periodicities. Most aggregate series show a steady overall decline reconfirming the impression of persistence in the inflation series. Residential rent (se2101) shows a very regular linear decline up to lag 140, which may indicate $d_0 > 0.5$. The inflation series for apparel (sa3*) show a slow strongly periodic decline: the sample autocorrelation function decreases steadily from large positive values to negative values at lags 12, 24, 36 whereas the series at e.g. lags 4,16, 28,... starts with large negative values and shows at steady increase. Note that the ACF at lags 12, 24, 26, corresponds to the ACF of a series observed only once a year. This series will "collect" all the integration features at the frequencies $2\pi i/12$, $i = 0, 1, 2, 3, 4, 6$, as explained in section 2.1. Some of the correlation functions for agricultural inflation series do not indicate clear persistence at all, see e.g. eggs or bananas. The well known cyclical movement in some meat markets shows itself in the oscillating pattern of the autocorrelation function for bacon (se0401).

We do not present the sample autocorrelation function for the monthly changes in inflation here. The seasonal series still show persistence at lags $12, 24, 36$, for the reasons described above. Most other series display an SACF like an MA(1) process with a unit root: $\widehat{\gamma_1}/\widehat{\gamma_0} \approx -0.5$ and an irregular ACF around zero for higher lags.

5 Results

We present estimation results in the tables in appendix D. For use a common estimation period 1959.1-1995.12, 444 observations, for all models. This paper deals mainly with the effect of aggregation on long memory at the zero frequency, so the main focus is on the estimates for d_0, where we want to see whether a high level of aggregation corresponds to a high estimate for d_0. One has take empirical identification problems into account, so we present extra information for each model estimate. In order to avoid too much "bias by model selection" we present each model for all 72 series, when possible.

Tables 2 and 3 give the results for the first general model that we present here, a flexible seasonal ARFIMA(12,7,2) model where ϕ_3 to ϕ_{11} are in the AR part are restricted to zero. We present estimates of the integration parameters d_i in the first part. In the second part we present the largest inverse roots of the estimated AR- and MA polynomials of the short memory part. We give the norm and the frequency, which is denoted in cycles per year. So frequency 6 corresponds to six cycles per year, with a corresponding period of one sixth of a year, i.e. two months. We put an asterisks for inverse roots with norm larger than 0.95. Note again that simple (seasonal) unit root tests do not apply in the context of long memory errors, so this is not meant as a formal test. We also provide an R^2 in deviation from seasonal means, an approximate frequency domain likelihood, and a simple AIC criterion. This large model provides a reasonable fit in the frequency domain, according to the Beran test for no correlation in the residuals, presented in the column under "test". We see that the model is overspecified for some series, see e.g. SA0LE, which has near common roots in the AR and MA part close to the unit circle. The last column shows the standard error of the residual term, which is lower for the aggregated series. We see indeed that there are strong indications for long memory in most aggregated inflation series. We find that the order of integration, and therefore the persistence, is highest in the more aggregated series. We even find estimates for d_0 in the "infinite-variance mean-reversion" region $0.5 < d_0 < 1$, but standard errors are so large that one cannot reject finite variance hypotheses $d_0 < 0.5$ very comfortably. The evidence against short memory $d_0 \leq 0$ is quite strong for the most aggregated series, see the $t-$values for the test for $d_0 = 0$.

We find the seasonal extension which allows for long memory at seasonal frequencies to be relevant at different levels of aggregation: for commodities less food and energy, for apparel and its sub-items, SA3, SA3111, and for e.g. fish and seafood, SE07. Some inflation series seem to be antipersistent, like the item fresh vegetables, SE12. Although inflation in energy commodities is very volatile, (see the standard error to e.g. aggregate energy price, SE0E, which is five times larger than for "core inflation" SA0L1E). It is not persistent compared with other categories like food and service and utility items, see the estimates for d_0 for SACE and SA0E. Note that the series for energy services (SE26) shows a high integration parameters at all frequencies except frequency 6, which is captured in the AR part. The propagation of energy price changes through the service channel seems to be rather slow.

Tables 4 and 5 present results for the flexible seasonal ARFIMA(2,7,2) model. Many results are comparable with the results in the larger ARFIMA(12,7,2) model. We find fewer roots in the AR-part close to the unit circle, although the common root problem in the short memory part is still apparent for series like SE4. The overspecification for this series shows also in the failure of the optimization algorithm for the ARFIMA(12,7,2) model, where a lower local optimum was found than for the ARFIMA(2,7,2) model.

Tables 6 and 7 present results for the ordinary ARFIMA(12,1,2) model, which is rejected according to an LR test against the general model for many series, especially for the strongly seasonal series of course, like the "apparel series" SA3, SA31, ..., where totally different estimates for d_0 are obtained as compared with the flexible seasonal model. Our main concern here is to avoid gross misspecification which could lead to wrong results for the zero frequency analysis. The Beran test also indicates misspecification, but not nearly as consistently as the LR test.

In order to compare our results with more familiar models for seasonal time series we estimate flexible seasonal ARFIMA(13,7,13) models, again with the a priori restriction $\phi_3 = \ldots = \phi_{11} = 0$ and $\theta_3 = \ldots = \theta_{11} = 0$. This model allows for the seasonal ARIMA structure $(1-L)(1-L^{12})x_t = (1-\theta L)(1-\Theta L^1 2)\epsilon_t$. We present the results in tables 8 and 9. This extensive model produces unreasonably high estimates for d_0 for the most aggregated series series, but it seems still appropriate for some seasonal series like SA3111 and SA3112. It still shows a significant seasonal long memory part, despite the increased flexibility to capture seasonal movement in the ARMA part.

6 Conclusion

Our results indicate that aggregation may be an important factor in the explanation of long memory in consumer price inflation. Although our model is flexible in the capturing of the autocorrelation structure, it may be interesting to make it more robust to fat-tailed innovation processes and exogenous changes in (seasonal) means. These are topics for further research. Our explicit modeling of the autocorrelations should allow for a simple GLS testing against alternative models with a shift in (seasonal) means. This seems to be a relevant alternative for a number of series. Estimation under the alternative with changing seasonal means would be easier using an extension of a nonlinear least squares method as in Beran (1995).

References

Andel, J. (1986), "Long Memory Time Series," *Kybernetika (Academia, Prague)*, 22, 105–123.

Andel, J., and Gómez, M. (1988), "Two-dimensional Long Memory Models," *Kybernetika (Academia, Prague)*, 24, 1–16.

Baillie, R.T. (1996), "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics*, 73, 5–59.

Beran, J. (1992), "A Godness-of-fit Test for Time Series with Long Range Dependence," *Journal of the Royal Statistical Society, Series B*, 54, 749–760.

——— (1995), "Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models," *Journal of the Royal Statistical Society*, 57, 659–672.

Boes, D.C., Davis, R.A., and Gupta, S.N. (1989), "Parameter estimation in low order fractionally differenced ARMA processes," *Stochastic Hydrology and Hydraulics*, 3, 97–110.

Canova, F., and Hansen, B.E. (1995), "Are Seasonal Patterns Constant Over Time? A Test for Seasonal Stability," *Journal of Business and Economic Statistics*, 13, 237–252.

Chan, N.H., and Terrin, N. (1995), "Inference for Unstable Long-Memory Processes with Appliacations to Fractional Unit Root Regressions," *The Annals of Statistics*, 23, 1662–1683.

Chang, M.C., and Dickey, D.A. (1994), "Recognizing Overdifferenced Time Series," *Journal of Time Series Analysis*, 15, 1–18.

Cochrane, J.H. (1988), "How Big is the Random Walk in GNP?," *Journal of Political Economy*, 96, 893–919.

Cox, D.R. (1991), "Long-Range Dependence, Non-Linearity and Time Irreversibility," *Journal of Time Series Analysis*, 12, 329–335.

Dahlhaus, R. (1995), "Efficient Location and Regression Estimation for Long Range Dependent Regression Models," *The Annals of Statistics*, 23, 1029–1047.

Diebold, F.X., and Lindner, P. (1996), "Fractional integration and interval prediction," *Economics Letters*, 50, 305–313.

Granger, C.W.J. (1980), "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics*, 14, 227–238.

Granger, C.W.J., and Ding, Z. (1996), "Varieties of long memory models," *Journal of Econometrics*, 73, 61–77.

Gray, H.L., Zhang, N.-F., and Woodward, W.A. (1989), "On Generalized Fractional Processes," *Journal of Time Series Analysis*, 10, 233–257.

Harvey, A.C. (1989), *Forecasting, structural time series models and the Kalman Filter.* Cambridge University Press, Cambridge, UK.

Hassler, U. (1994a), "(Mis)specification of Long Memory in Seasonal Time Series," *Journal of Time Series Analysis*, 15, 19–30.

——— (1994b), "Sample Autocorrelations of Fractionally Integrated Noise," Discussion paper, Institute of Statistics and Econometrics, Freie Universität Berlin, Boltzmanstrasse 10, 14195 Berlin, Germany.

Hassler, U. and Wolters, J. (1995), “Long Memory in Inflation Rates: International Evidence,” *Journal of Business and Economic Statistics*, 13, 37–46.

Haubrich, J.G., and Lo, A.W. (1989), “The Source and Nature of Long-Term Memory in the Business Cycle,” Discussion paper, NBER Working Paper, No. 2951.

Hosking, J.R.M. (1981), “Fractional Differencing,” *Biometrika*, 68, 165–176.

——— (1984), “Modeling persistence in hydrological time series using fractional differencing,” *Water Resources Research*, 20, 1898–1908.

——— (1996), “Asymptotic dsitributions of the sample mean, autocovariances, and autocorrelations of long-memory time series.,” *Journal of Econometrics*, 73, 261–284.

Hylleberg, S., Engle, R.F., Granger, C.W.J., and Yoo, B.S. (1990), “Seasonal Integration and Cointegration,” *Journal of Econometrics*, 44, 215–238.

Jeganathan, P. (1996), “On Asymptotic Inference in Cointegrated Time Series with Fractionally Integrated Errors,” Discussion paper, Department of Statistics, University of Michigan, Ann Arbor, MI.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., and Shin, Y. (1992), “Testing the null hypothesis of stationarity against the alternative of a unit root; How sure are we that economic time series have a unit root?,” *Journal of Econometrics*, 54, 159–178.

Lee, D., and Schmidt, P. (1995), “On the Power of the KPSS test of Stationarity Against Fractionally-Integrated Alternatives,” Discussion paper, Department of Economics, Michigan State University, East Lansing, MI, USA.

Mandelbrot, B.B. (1972), “Statistical Methodology for Nonperiodic Cycles: From the Covariance to R/S Analysis,” *Annals of Economic and Social Measurement*, 1/3, 259–290.

McLeod, A.I., and Hipel, K.W. (1978), “Preservation of the Rescaled Adjusted Range. 1. A Reassessment of the Hurst Phenomenon,” *Water Resources Research*, 14, 491–508.

Newbold, P., and Agiakloglou, C. (1993), “Bias in the sample autocorrelations of fractional noise,” *Biometrika*, 80, 698–702.

Nyblom, J. (1989), “Testing for the Constancy of Parameters over Time,” *Journal of the American Statistical Association*, 84, 223–230.

Ooms, M. (1994), *Empirical Vector Autoregressive Modeling*, vol. 407 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin, Germany.

——— (1995), “Flexible Seasonal Long Memory and Economic Time Series,” Discussion Paper 9515/A, Econometric Institute, Erasmus University Rotterdam.

Porter-Hudak, S. (1990), “An Application of the Seasonal Fractionally Differenced Model to the Monetary Aggregates,” *Journal of the American Statistical Association*, 85, 338–344.

Proietti, T. (1996), “Persistence of Shocks in Seasonal Processes,” *Journal of Applied Econometrics*, 11, 383–396.

Ray, B.K. (1991), “Fractionally Differenced ARMA Processes: Seasonality and Forecasting Issues,” Ph.D. thesis, Graduate School of Arts and Sciences, Columbia University, USA.

——— (1993), “Long-range forecasting of IBM product revenues using a seasonal fractionally differenced ARMA model,” *International Journal of Forecasting*, 9, 255–269.

Robinson, P. (1994), “Time Series with Strong Dependence,” in *Advances in Econometrics, sixth World Congress*, ed. by Sims, C.A., pp. 47–95. Cambridge University Press, Cambridge.

Robinson, P.M. (1994), "Efficient Tests of Nonstationary Hypotheses," *Journal of the Americal Statistical Association*, 89, 1420–1437.

——— (1995), "Log-Periodogram Regression of Time Series with Long Range Dependence," *Annals of Statistics*, 23, 1048–1072.

Samarov, A., and Taqqu, M.S. (1988), "On the Efficiency of the Sample Mean in Long-Memory Noise," *Journal of Time Series Analysis*, 9, 191–210.

Sowell, F. (1990), "The Fractional Unit Root Distribution," *Econometrica*, 58, 495–505.

Viano, M.C., Deniau, Cl., and Oppenheim, G. (1995), "Long-Range Dependence and Mixing for Discrete Time Fractional Processes," *Journal of Time Series Analysis*, 16, 323–338.

Appendix A, BLS Codes Inflation Series

Table 1: *US Monthly Consumer Price Indexes, Base Year 1982-1984*

BLS Code	Description	Aggregation
SA0	All items	0
SA0LE	All items less energy	0
SA0L1	All items less food	0
SA0L1E	All items less food and energy	0
SA0L2	All items less shelter	0
SA0L5	All items less medical care	0
SA11	Food	1
SA21	Shelter	1
SA4	Transportation	1
SA5	Medical care	1
SA3	Apparel and upkeep	1
SAC	Commodities	1
SACE	Energy commodities	1
SACL1	Commodities less food	1
SACL1E	Commodities less food & energy commodities	1
SAD	Durables	1
SAN	Nondurables	1
SANL1	Nondurables less food	1
SANL13	Nondurables less food and apparel	1
SAS	Services	1
SASLE	Services less energy services	1
SA0E	Energy	1
SAS22	Housekeeping & home maintenance services	2
SAS24	Utilities and public transportation	2
SAS4	Transportation services	2
SA22	Fuel and other utilities	2
SA111	Food at home	2
SE19	Food away from home	2
SA41	Private transportation	2
SA512	Medical care services	3
SA712	Personal care	3
SE2101	Rent residential	3
SA1111	Cereals and bakery products	3
SA11121	Meats, poultry, fish	3
SA111211	Meats	4
SA1113	Dairy products	4
SA1114	Fruits and vegetables	4
SA11141	Fresh fruits and vegetables	4
SA11142	Processed fruits and vegetables	4
SA31	Apparel commodities	4
SA311	Apparel commodities less footwear	4
SA3111	Mens and boys apparel	4
SA3112	Womens and girls apparel	4
SE11	Fresh fruits	4
SE12	Fresh vegetables	4
SE15	Sugar and sweets	4
SE17	Nonalcoholic beverages	4
SE25	Fuel oil and other household fuel components	4
SE2501	Fuel oil	4
SE26	Gas & electricity (energy services)	4
SE2601	Electricity	5
SE2602	Utility (piped) gas	5
SE0201	White bread	5
SE03	Beef and veal	5
SE0301	Ground beef excluding canned	5
SE0302	Chuck roast	5
SE0305	Round steak	5
SE04	Pork	5
SE0401	Bacon	5
SE0402	Pork chops	5
SE05	Other meats	5
SE06	Poultry	5
SE0601	Fresh whole chicken	5
SE07	Fish and seafood	5
SE08	Eggs	5
SE0901	Fresh whole milk	5
SE1101	Apples	5
SE1102	Bananas	5
SE1103	Oranges, including tangerines	5
SE1201	Potatoes	5
SE1202	Lettuce	5
SE1203	Tomatoes	5