ECON 706 - Problem Set 2

André Victor D. Luduvice, Stefano Pietrosanti and Joseph Silverstein

March 18, 2015

Abstract

In this problem set we study the behavior of housing completions and starts in the frequency domain. To investigate the underlying cyclical behavior of housing starts and completions, we use parametric and nonparametric techniques to conduct detailed spectral analyses of the series in the univariate and multivariate cases.

1 Introduction

We conduct a frequency domain analysis, analogous to the time domain analysis we conducted in Problem Set 1. We again work with the Federal Reserve Bank of St. Louis data set, (FRED, 2014), which uses the results of the US Census.

The time series under study collects 45 years plus one month of monthly observation of housing starts and completions, from January 1970 to January 2015, for a total of 541 observations.

This article is organized into two sections: Univariate and Multivariate Analysis; these sections are then subdivided on the basis of the methods employed (lag-window, autoregressive, etc...), and methods are grouped with respect to whether they are parametric or non-parametric.

Methodology

The goal of spectral analysis is to understand the variability of a stochastic process over the frequencies rather than over the time domain. This means to look at the periodogram instead of investigating the autocorrelations. Thus changing perspective, we try to observe the Time $\times$ Level space as filled by many possible wave (sine-cosine) functions, each for a different frequency.

E.g., we can take a cosine function, and think to have a finite series of length T - this only for the present paragraph, for clarity sake. The $cos(u)$ function is defined over the $[0, 2\pi]$ space (actually, it is defined over $[-\pi, \pi]$, but being symmetric we can lightheartedly shift to the positive radiants). We observe the u unit over the circle in the following way: $u(\omega, t) = 2\pi * \frac{k}{T} * t$, where T is the total of our time observations; t is specific time point; k the number of cycles our wave completes in the time span we observe; hence, $\omega_k = \frac{2\pi * k}{T}$. We have thus restated the wave function on the Time-Value space, where the measure 2π over the circle is mapped (if possible) to the $t = \frac{T}{k}$ data point, and so on and so forth[1].

As we do this, we consider (loosely following Hamilton (1994), as in all other comments) data points as x_t, from the generic stationary time series $\{x_t\}_t$, as the result of the composition of these different sine and cosine waves at different frequencies.

$$x_t = \sum_k (a_k * \sin(\omega_k * t) + b_k * \cos(\omega_k * t)) \tag{1}$$

We can recognize something similar to a linear regression. By "similar" we mean that we can estimate this as an actual linear regression with perfect fit (why perfect fit is explained in the footnote n.1) considering a_k , b_k as coefficients. Yet, as long as we stay on theoretical ground, we must state that these coefficients are

[1]Hence moving to an infinite sample makes what above imprecise, yet not the intuition. With an infinite sample we can in principle observe an infinite number of cycles.

zero mean, with equal variance (σ_k^2), serially and mutually uncorrelated random variables. As in the regression framework, some k indexed explanatories will be more important and other less[2]. To measure this importance, we can ask ourselves the classical ANOVA question: "how much this component helps explaining the total variability?". In order to answer such question we can observe that[3]

$$\begin{aligned} E[x_t^2] = \gamma(0) = \sum_k (E[a_k^2] * \sin(\omega_k * t)^2 + E[b_k^2] * \cos(\omega_k * t)^2) = \\ = \sum_k \sigma_k^2 \end{aligned} \tag{2}$$

We also know that $\gamma(0)$ is just one instance of the autocovariance function, which we can recast completely on the frequency domain using the autocovariance generating function. For each frequency ω we have:

$$g_X = \sum_{t=-\infty}^{\infty} \gamma(t) z^t \tag{3}$$

evaluating on over the unit circle, hence $z = e^{-i\omega}$, and rescaling w.r.t. 2π we get

$$\begin{aligned} s_X(\omega) = \sum_{t=-\infty}^{\infty} \frac{1}{2\pi} \gamma(t) e^{-i\omega t} = \\ = \frac{1}{2\pi} \left[\gamma(0) + 2 \sum_{t=-\infty}^{\infty} \gamma(t) \cos(\omega t) \right] \end{aligned} \tag{4}$$

Where the last passage follows from the basic trigonometric identities and from symmetry of the involved functions around zero. This last function is the population spectrum, a continuous, real valued function of the continuous frequency

[2]When working with an actual sample of T observations, the maximum possible k is $\frac{T}{2}$. This since, with T observation, we can have at most T explanatories; each k brings with itself a sine and a cosine explanatory both, hence we can at most have T over a half ks, and perfect fit. Since we want to check for the contribution of all possible frequencies, we sum over $k \in \{0, \frac{T}{2}\}$, and we go for the perfect fit.

[3]We assume that x_t is mean zero.

variable. Given that the Fourier transformation can be inverted, from the population spectrum we can recover frequencies as follows:

$$\int_{-\pi}^{\pi} s_x(\omega) e^{-i\omega t} d\omega = \gamma(t) \tag{5}$$

and the fact that the autocovariance of our stationary process is the integral of the spectrum over the frequencies for $t = 0$ follows. We can appreciate how this works perfectly as long as we have infinitely many autocovariances. In a sample, with $T-1$ proper (not at lag 0) autocovariances, we will need to be careful. What follows is the tale of this carefulness.

2 Univariate Analysis

Here we are only interested in the study of the individual cyclical behavior of our series. A first and rough approach is to simply plot the raw Periodgram of our series. Though this estimator is unbiased, it is ridden with problems: first, a very large confidence interval that does not shrink with the increase in the number of data points considered (no consistency). We can think about this in the sine/cosine regression framework we stated in the introduction: as soon as we add new data to our sample, we can investigate new frequencies, hence we must add explanatories to our perfect fit regressions. The result is that - as the sample increases - we do not collect more information about the same parameters, but we add parameters instead.

The second problem is that, when observing a cluster of peaks in the sample raw spectrum, we are not totally sure about the "importance ranking" of the frequencies which correspond to these peaks. The main source of such a problem is the fact that the shift to frequency domain from time domain is a shift to a continuous space from a discrete space. As long as we have infinitely many discrete data points this is fine, and can be performed through Fourier transformation[4]. If we have finitely many discrete data points, though, we must take into account a lot of imprecision. This imprecision is due to the fact that, of the whole frequency line, we will only be able to consider the finitely many points spanned by t: $\frac{2\pi k}{T}t$, $t \in \{0, ..., T\}$. Formally, the raw estimate is:

[4]As it is made clear in the Methodology section, where we see that the Fourier transform takes the lags of the autcovariances from $-\infty$ to $+\infty$.

$$\hat{s}_x(\omega) = \sum_{t=-T+1}^{T-1} \frac{1}{2\pi}\gamma(t)e^{-i\omega t} = \\ = \frac{1}{2\pi}\left[\gamma(0) + 2\sum_{t=1}^{T-1}\gamma(t)\cos(\omega t)\right] \tag{6}$$

The underlying difficulty becomes clear as soon as we think of the theoretical Periodgram, where we account for contributions to the variance by *ranges* of ωs. This contribution is described by the area encompassed by the spectral density curve. If we try to estimate the said contribution with finitely many data points for each frequency *range*, we are estimating areas through points.

The result is pretty spiky and irregular (high - and irreducible - variance). Since this estimation is unbiased, though, it is a useful instrument to identify important ranges of frequencies, so to provide a "reality check" for our more sophisticated, consistent, yet often biased, estimation methods.

Before showing the plots, a brief preliminary account about how we think about it. On the ordinate axis, we show the values of the estimated spectral densities, which - as for every density function - are arbitrary, such that the only relevant information conveyed is the relative magnitude. We look for peaks, for relatively high-weight (density) frequencies, since we are interested in the frequencies which are the most important for variance decomposition. On the abscissae axis, we instead plot the "frequencies".

Yet here comes the tricky point, justifying our usage of quotes the last time we introduced the word frequencies. Since we employed the R's "ts" command to exploit the time series packages, it is as if we have been telling R to read frequencies over time, and not over the circle: an instance of the problem we mentioned before: for each frequency, we only have finitely many data points in the t spanned world.

Furthermore, also the number of cycles our waves can perform, k, is constrained. Given that we are dealing with monthly observations, T is the total

number of monthly points constituting our series. Hence, the maximal number of cycles our explanatory wave functions can accomplish, over the $T = 541$ periods we take into account, is $\frac{540}{2}$[5]. This is a bimonthly frequency: the fastest wave we can consider completes six cycles each year.

Consequently, setting

```
freq=12
```

in the "ts" command, the Periodgram of any time series in our analysis will show frequency 6 as maximum abscissa. This means that the maximum number of cycles per year a wave component can complete in the environment of our analysis is 6. This said, the plots follow:

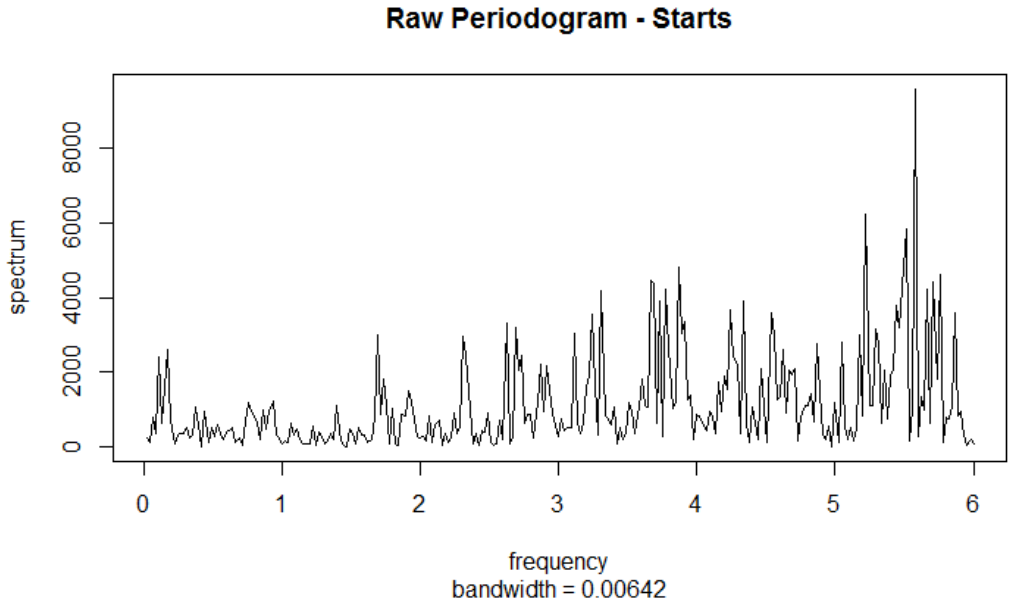

Figure 1

As we can see, our series spectra are mostly coherent in shape, showing:

[5]Since we cannot show "half-data ponts", we solve the problem removing one observation, as in Hamilton (1994).

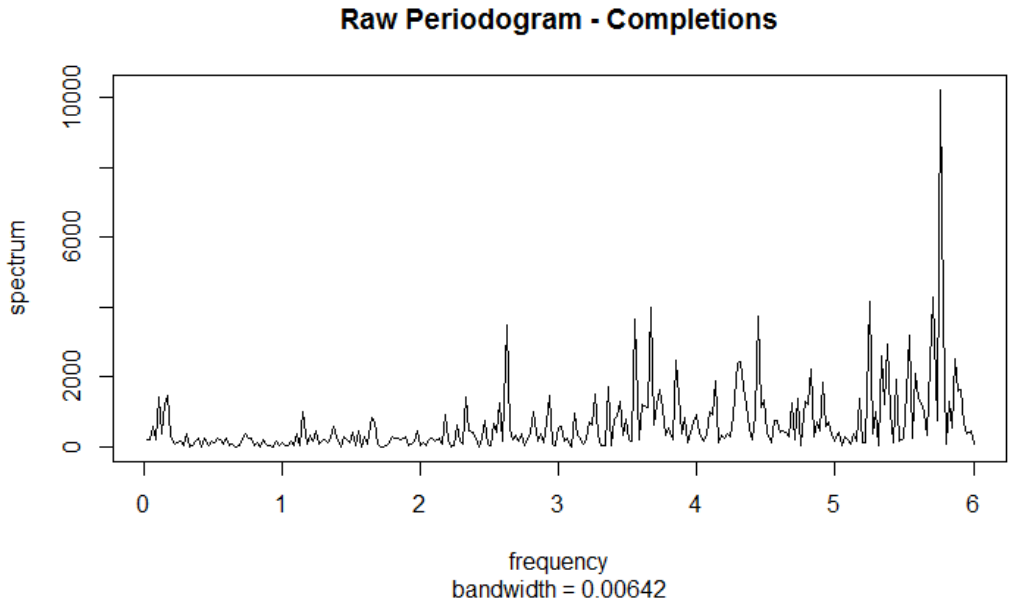

Figure 2

1. A low frequency component, which almost does not cycle, and hence has frequency concentrated near zero.

2. A group of spikes, located within the bi-quarterly and quarterly zone.

3. A very relevant (approximately) bi-monthly component, from which the most of the variability seems to be caused.

Now we procede improving our estimates through Non-Parametric and Parametric methods.

2.1 Non-parametric Estimation

Since the two main problems with the raw Periodgram are the improper concentration of spectral weight on single frequency points, and the inconsistent nature

of the raw Periodgram as an estimator, it would be nice to have a solution for both. Fortunately this solution is there, is unique, and addresses both problems. In plain words, only looking at the estimate we previously got (non-parametric), we will try to smooth them in a "clever" way, so to diminish the variance (redistribute the concentrated weight) and get back consistency doing the smoothing over fixed intervals, so that the increase in the number of data will give us again more informations about what we are estimating.

Formally, j being $\frac{k}{T}$, we use a $K(\omega_{j+m}, \omega_j)$ (the *kernel*) weighting function[6], and we compute the estimates of the Periodgram on the base of the raw estimates:

$$\hat{s}_X = \sum_{m=-h}^{h} K(\omega_{j+m}, \omega_j)\hat{s}_x(\omega_j) \tag{7}$$

where the h is the bandwidth, i.e. how many raw correlation data points we exploit to estimate each element of our final approximation of the Periodgram[7]. This mitigate our two main problems. On the one hand, the fixed nature of the bandwidth allows us to regain consistency: as we increase the number of data points, more of them will fall into the bandwidth, so we go back to a world in which more data means more information; on the other hand, redistributing the weights, we curb the wild behavior of the raw Periodgram, so to better grasp the contribution of *ranges* of frequencies to the variance. On the other end, this procedure is biased by nature: if we smooth a previously unbiased estimate, and then we take again the average of the smoothed values, the result will be different from the previous average (which was unbiased). Hence we have a bias-variance trade off

[6]In the sense that for each h, $\sum_{m=-h}^{h} K(\omega_{j+m}, \omega_j) = 1$.

[7]In the empirical analysis we define the bandwidth slightly differently: as $\frac{2h+1}{T}$, the total portion of disposable points we use for each estimate.

2.1.1 Spectral Window

Since we only have the $j = \frac{k}{T}$ frequencies at our disposal, in deciding the bandwidth we are deciding how many of these actually enter our smoothing procedure. This reasoning is at the base of the Spectral Window estimation method, in particular, we employ the modified Daniell's kernel.

This last is simply a centered moving average, where the weight of the last two elements has been decreased in favor of the central elements (we are doing this to avoid the leakage of variance from one frequency range to the other). To let the weight concentrate on the very central value - and decrease linearly outside - we can apply the kernel multiple times (convoluting the kernel). In such case, the bandwidth of the resulting kernel (the convolution of the kernels considered as a single kernel) will be $\frac{n*h+1}{T}$[8].

To do an actual example, let say we want the Spectral Window estimate of the Population Spectrum at the frequency ω_j, $\hat{s}_X\omega_j$, we have the raw Periodgram estimate $\hat{s}_x\omega_i$ for $i \in \left\{0, ..., \frac{T-1}{2}\right\}$; and we choose a bandwidth parameter equal to 1, applying it twice. This translates in the following mathematics:

$$\hat{s}_X(\omega_j)_1 = \frac{\hat{s}_x(\omega_{j-1}) + \hat{s}_x(\omega_j) + \hat{s}_x(\omega_{j+1})}{3}$$

then we apply the same kernel again, and we get (8)

$$\hat{s}_X(\omega_j)_2 = \frac{\hat{s}_x(\omega_{j-2}) + \hat{s}_x(\omega_{j+2})}{9} + \frac{2(\hat{s}_x(\omega_{j-1}) + \hat{s}_x(\omega_{j+1}))}{9} + \frac{\hat{s}_x(\omega_j)}{3}$$

The objective in this is to find the "best" way to smooth, hence the best bandwidth parameter. We do this *via* sheer comparison of the result of the application of different bandwidth parameters for both the applications of the kernel[9].

[8]Indeed, since in our final choice we opt for a two step convolution of the kernel, we have a bandwidth of $\frac{4*h+1}{T}$.

[9]We can apply, say, $h = 6$ the first time and $h = 3$ the second time. Actually, we choose for $8, 8$ for Start and $7, 7$ for completions.

To do this we cycle through the bandwidth parameters for both the passages of the kernel. The nested loop computes the estimate for each couple of bandwidth parameters, and plot it along with the parametric estimate (to be commented later) and the raw estimate.

I below show two extreme choices and the actual result for the Starts series(first the two extremes, then the result).

As we can see, the graph for bandwidth $\frac{4*3+1}{540} = 0.0241$ (lowest bandwidth considered for two passages) does not perform properly in its smoothing task. On the other hand, the $\frac{4*10+1}{540} = 0.0759$ (looking at the raw estimation) is clearly biased, since it is over-smoothed, which results in a possible misestimation of the autocovariance of the sample at lag 0. Finally, the $\frac{4*8+1}{540} = 0.0611$ accomplish the smoothing purpose without exceeding in it. So we settle on this choice.

The Completion final result follow, where the selection logic is the same.

2.1.2 Lag Window

The Lag Window differs from the Spectral Window with respect to the intuition and the type of kernel involved, yet the result is not this remarkably different - as we shall see.

First, the intuition. The main choice we had to perform with respect to the Spectral Window estimator was about over how many raw Periodgram point estimates we were going to smooth. Which is, over how many frequencies on the abscissae.

For the Lag Window, we approach the issue from the autocorrelation perspective, for each frequency. We use the Bartlett kernel, defined as

$$\kappa_j^* = \begin{cases} \text{if } t \in \{1, ..., q\} & : \ 1 - \frac{t}{q+1} \\ \text{otherwise} & : \ 0 \end{cases} \tag{9}$$

We apply it, in combination with the sample autocovariances, in the following way:

$$\hat{s}_X(\omega) = \frac{1}{2\pi}\left[\hat{\gamma}(0) + 2\sum_{t=1}^{q}\left(1 - \frac{t}{q+1}\right)\hat{\gamma}(t)\cos(\omega t)\right] \tag{10}$$

So, choosing q, we choose how many sample autocovariances to use in our estimation, as if there was some underlying, finite MA process involved. After a similar inspection procedure as the one employed for the Spectral window, we conclude that the best q equals 25 for the Starts series and 30 for the Completion series. Given the closeness in the selection procedure with respect to the previous section, we will just show the final result

Below the plots:

2.2 Parametric - Autoregressive Spectral Density Estimation

A possible way to solve our estimation problem, given the close relationship between spectral density and variance of a stochastic process, is to exploit its underlying autoregressive structure to recover the spectral density.

This procedure's underlying logic is[10]:

let assume we do know our generating process has a stationary, MA(∞) nature[11]. Then we know about the existence of an appropriate ψ(L) polynomial that describes this MA(∞) process, which means we can refer to the following autocovariance generating function:

[10] In the following I am using capital letters for the population, and non-labelled variables (as σ, the error's std deviation) for error related quantities.

[11] And we do: the time domain analysis showed how and ARIMA$(1,1,2)$ was a good approximation of both of the time series, so we have good evidence in favor of an MA(∞) of some sort generating the series of the *differences* in houses starts and completions.

$$g_X(z) = \sigma^2 \psi(z)\psi(z^-1) \tag{11}$$

We can then evaluate this expression on the unit circle, which means at $z = e^i\omega$, and then divide everything times the circumference of the unit circle (2π). We thus get:

$$s_X(\omega) = \frac{\sigma^2 \psi(e^{i\omega})\psi(e^{-i\omega})}{2\pi} \tag{12}$$

To have a practical example, say we have an AR(1), whose related filter is $\psi(L) = \frac{1}{1-\phi L}$, then, its population spectrum is:

$$\begin{aligned} s_X(\omega) = \frac{\sigma^2}{2\pi(1-\phi e^{i\omega})(1-\phi e^{-i\omega})} = \\ = \frac{\sigma^2}{2\pi(1+\phi^2-2\phi\cos(\omega))} \end{aligned} \tag{13}$$

where the last passage exploits Euler equation[12].

To do this in R, we exploit the spec.ar command, which fits an AR model to our sample with the help of the AIC, and then recovers the $\hat{\phi}$s, so to empirically compute the spectral density over the frequency domain[13]. What we are going to show in the following is the plot of the spectral density of the estimated model, obtained through the variance generating function.

[12]Feynman's beloved $e^{i\omega} = \cos(\omega) - i\sin(\omega)$

[13]Here we must recognize that we are using a command that only allows for AR, and not MA, specification. It turns out that this is the most convenient way to handle the estimation given our sample. Tough we are allowing for a plethoric AR specification, we do this in the spirit of Hamilton (1994, p.165): "Even if the model is incorrectly specified, if the autocovariances of the true process are reasonably close to those for an ARMA(p,q) specification, this procedure should provide a useful estimate of the population spectrum". For we have good evidence in favor of assuming an MA(∞) generating process - and not, say, a finite order MA - we count on that "reasonable closeness".

As we can see, under AR specification, both series seem to have at least five important component: an intercept that is not "cycling", a relatively slow component cycling over 12 months periods, an almost bi-quarterly component, an almost quarterly component and, finally, an almost bimonthly component from which the most of the data variation seems to generate.

3 Univariate Analysis - Non-seasonally-adjusted

We now analyse the non-seasonally adjusted completion series. The choice of this series is motivated by the fact that in the last problem set we have assessed that, among the two, it is the most well-behaved for time domain estimations. The following graphs plot the series for the whole sample and also for a short window of the years 2012-2014. In the latter, we can see clear the seasonal component within each year.

As noted previously, the completions seasonally-adjusted series is non-stationary and therefore we take a one-lag difference to correct it. For the sake of completeness, we have also tested via the Augmented Dickey-Fuller (ADF) test with and without drift for both the AIC and BIC criteria the non-seasonally-adjusted series. We could not reject the hypothesis of the non-stationarity of the series on a 10% significance level for all of them. Thus, we have also taken the first difference and then performed the same ADF tests obtaining a strong rejection of the null-hypothesis for all implemented. We omit the showing the output of the tests as it is straightforward to obtain them from the code. Hence, all our analysis is done over the first difference of the series in the same way of the previous section.

The plot below shows the raw periodogram of the series.

The first clear observation is that the series has many peaks spreading through the harmonics of a 12 months period. This shape follows closely what character-

izes the "Granger's Typical Spectral Shape of an Economic Variable" (Granger, 1966) with the difference that the peaks do not decrease on magnitude at higher frequencies.

The highest peak of the plot is on the bimonthly cycle showing its strong relative importance over the periodogram. This stands in accordance to all the analysis done with respect to the seasonally adjusted counterpart. The other two outstanding peaks are, respectively on magnitude order, the quarterly and yearly component. The peak on the quarterly component shows that the series carries an intense effect of an year's seasons. The peak on the number 1 of the x axis, i.e., the 12 months cycle component, shows the true seasonal component of the series in which house completions cycles repeat itself year after year. Comparing with the periodogram of its seasonally-adjusted counterpart, it is clear that all those harmonic components where eliminated in the seasonal filtering applied by the FRED.

3.1 Non-parametric Estimation

3.1.1 Spectral Window

We proceed with the spectral window estimation using the same methodology of the previous section. We use two passes of the Daniell's kernel obtaining a triangular weighting. This filter is specially appealing to the non-seasonally-adjusted series as the relevant frequency components are just the ones in which we have the peaks discussed before.

The graph below shows our chosen approximation where, after the repetition of many iterations of different bandwidth parameters, we select $h = 2$ yielding us a window bandwidth of $\frac{4*2+1}{540} = 0.016$. The selection criterion was, as before, inspection and comparison with the raw periodogram. Fortunately, despite the

rather narrow bandwidth, we preserve the positioning and magnitude order of all important peaks previously identified, as well as smooth lower values, and thus we escape the worry of the bias-variance trade-off.

3.1.2 Lag Window

The lag window estimation is, once again, conducted using the Bartlett lag window. We smooth the periodogram by choosing the truncation on lag q of the autocovariance function. After repeating the computational cycle truncating from $q = 20$ until $q = 110$ we observe that above the 70th lag our estimation begins to assume the desired behavior replicating the peaks of the periodogram. We then choose $q = 75$ in order to augmented precision while preserving parcimony of the number of lags.

3.2 Parametric - Autoregressive Spectral Density Estimation

Our parametric estimation with the R function *spec.ar* selects an AR(25) through the AIC criterion and yields the following spectral density:

The estimation preserves the peaks in the harmonic components of the 12 months period observed in the raw periodogoram. The bimonthly importance is well-represented as well as the other minor components at the 2,3 and 5 in the x axis. However, the magnitude of yearly cycle is heavily emphasized, exceeding the one of the quarterly cycle and achieving the same relevance of the bimonthly cycle. This is a unexpected result and goes in the contrary direction of all the intuition previously discussed.

Bibliography

FRED, Federal Reserve Bank of St. Louis from US Bureau of Census. 2014. "US. Bureau of the Census, New Privately-Owned Housing Units, Start and Completion."

Granger, C. W. J. 1966. "The Typical Spectral Shape of an Economic Variable." *Econometrica* 34 (1):150–161.

Hamilton, James Douglas. 1994. *Time series analysis*. Princeton University Press.