

Seasonal heteroskedasticity in Census Bureau construction series

Thomas M. Trimbur

Statistical Research Division

U.S. Census Bureau

Washington DC 20233-9100

Abstract

Seasonal heteroskedasticity exists in a number of monthly time series from major statistical agencies. Accounting for such variation in calendar month effects can be important in estimating seasonal and trend movements. In the context of seasonal adjustment, the standard procedure uses nonparametric (X-11) filters of different lengths in the signal extraction routine of X-12-ARIMA. This serves as a simple, pragmatic procedure that is, however, limited in its ability to adapt to different datasets. In this paper I extend the model-based methodology introduced recently by Proietti (2004) and Bell (2004). I discuss different forms of the seasonal specific model, showing examples of estimation and analysis of trend and seasonal components. A statistical test for seasonal heteroskedasticity is presented and applied to a number of Census Bureau series on housing starts and building permits. It is shown how seasonal noise can be separated from nonsystematic noise and included in the seasonal adjustment of a time series.

KEYWORDS: seasonal heteroskedasticity, time series, trends, unobserved components.

JEL classification: C22, C51, C82

1. Introduction

Seasonal heteroskedasticity exists in a number of monthly time series from major statistical agencies. Often the data are nonstationary and, since the estimation of seasonal and trend movements are naturally intertwined, both component estimates can be improved by using an econometric model that is specifically designed to handle variation in calendar month effects. In this paper I concentrate on Census Bureau construction series, though the discussion of methodology is relevant for any seasonally heteroskedastic economic variable.*

The X12-ARIMA seasonal adjustment program (see Findley et. al, 1998) provides a simple way to account for seasonal heteroskedasticity on a coarse level by using X-11 seasonal filters of different

lengths. This procedure is, however, limited in its ability to adapt to the diversity of seasonal patterns found in real datasets. Furthermore, the nonparametric framework gives no clear way to test for the presence of seasonal heteroskedasticity or to evaluate the plausibility of the heteroskedastic effects as a description of the dynamics of the time series.

This paper concentrates on statistical analysis of seasonal heteroskedasticity. The framework is provided by the seasonal specific models introduced recently by Proietti (2004) and by the alternative model-based approach developed by Bell (2004). The major aims are first, to develop methodological improvements and extensions, and second, to show how the models can be used to gain a better understanding of seasonal heteroskedastic movements.

The paper proceeds as follows. Section 2 reviews the seasonal specific levels model implemented in Proietti (2004). The goal is to clarify and explain the structure of the model, which differs from most standard time series models. A modified definition of the trend is introduced to preserve consistency with the standard interpretation of seasonality. In Section 3, I estimate the model and components for a highly heteroskedastic dataset to illustrate the application of the framework. Results for estimated trend and seasonal components are compared with the seasonal adjustment output from X-12-ARIMA.

Section 4 extends the framework by introducing a statistical test for the presence of seasonal heteroskedasticity. Finite sample effects and dependence on model parameters are accounted for by using Monte Carlo simulation to estimate the correct critical values for each series. I also consider the alternative seasonal specific model that is suggested by the modeling strategy used in Bell (2004); this gives an important benchmark for comparison that allows one to check the robustness of the test.

Also, by comparing the results of model estimation, it becomes clear how the different strategies are reflected in the extracted components and this leads to suggestions for how heteroskedasticity may be handled in seasonal adjustment. In particular, Section 5 shows how the seasonal noise may be extracted from a series to produce a fully adjusted series that could be useful in some applications. Section 6 summarizes the main issues and concludes.

***Disclaimer.** This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical or methodological issues are those of the author and not necessarily those of the U.S. Census Bureau.

2. Seasonal specific levels model

The rationale for the modeling approach used in Proietti (2004) is to handle seasonality directly by setting up separate processes for the different calendar months. This means that the series of successive January observations is assumed to follow the basic specification with certain parameter values. The same equation also applies to the February series but with distinct disturbances and possibly different parameter values, and similarly for the other months. Once the model parameters are determined, the division of the dynamics into long-term, or trend movements, and seasonal movements can be addressed. The assumed covariance structure of the disturbances across the season-specific equations determines the form of the heteroskedastic property.

Especially when focussing on trends, one may initially consider the strategy of incorporating the heteroskedasticity into the season-specific level equations. Thus, it is assumed that the disturbances driving the level processes may contain an idiosyncratic part that varies across seasons. This gives the following specification :

$$\begin{aligned} y_t &= \mathbf{z}'_t \boldsymbol{\mu}_t + \varepsilon_t, \quad t = 1, \dots, T \\ \boldsymbol{\mu}_{t+1} &= \boldsymbol{\mu}_t + \boldsymbol{\iota} \beta_t + \boldsymbol{\iota} \eta_t + \boldsymbol{\eta}_t^*, \quad \text{var}(\eta_t) = \sigma_\eta^2 \\ \beta_{t+1} &= \beta_t + \zeta_t, \quad \zeta_t \sim NID(0, \sigma_\zeta^2) \end{aligned} \quad (1)$$

where $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$ and $\boldsymbol{\eta}_t^* \sim NID(\mathbf{0}, \mathbf{N}_\eta)$. The symbol $\boldsymbol{\iota}$ stands for an $s \times 1$ vector of ones, with s denoting the number of seasons in a year, that is $s = 12$ for monthly data. The vector $\boldsymbol{\mu}_t = (\mu_{1t}, \dots, \mu_{st})'$ contains the month-specific levels, and \mathbf{z}_t is an $s \times 1$ selection vector that has a one in the position $j = 1 + (t - 1) \bmod s$ and zeroes elsewhere. The j th trend process $\mu_{j,t}$ evolves each period by a slope β_t that is common to all s elements. The disturbance ζ_t that drives the time-varying slope is assumed independent of the other disturbances in the model.

The s elements arranged in the vector $\boldsymbol{\eta}_t^* = (\eta_{1t}^*, \dots, \eta_{st}^*)'$ are idiosyncratic level disturbances. The variance of η_{jt}^* depends on the season j , and we will assume throughout that \mathbf{N}_η is diagonal, which ensures identifiability of the model (1). The disturbance η_t is uncorrelated with all η_{jt}^* and represents the part of the level shocks shared by the different $\mu_{j,t}$. The irregular ε_t is uncorrelated with the other disturbances in the model and accounts for the non-systematic noise in y_t .

Model (1) will be called the seasonal specific levels model. This model formed the basis for the illustrations in Proietti (2004); in Section 4, I will report

test results based on its application to a set of construction time series from the Census Bureau. I will, however, use a different trend estimator throughout the paper. Although the rationale for Proietti's (2004) estimator is clear, it suffers from a major drawback of giving biased measures of the level and seasonal. As explained in Bell and Trimbur (2005), the appropriate measure of the trend is given by an equal weighted average of the seasonal specific levels, that is $\mu_t = \mathbf{w}' \boldsymbol{\mu}_t$ with $\mathbf{w} = \boldsymbol{\iota}/s$. Likewise, the implied seasonal γ_t is computed at each time t by subtracting off the trend from the month-specific level for that time, that is $\gamma_t = \mathbf{z}'_t (\mathbf{I}_s - \boldsymbol{\iota} \mathbf{w}') \boldsymbol{\mu}_t$.

In (1), the process that determines the rate of growth, β_t , and the common level disturbance, η_t , link the various μ_{jt} together in forming the long run dynamics of the series. The model is similar in form to a multivariate local linear trend model. If one were to set $\eta_t = 0$, (1) would tend to produce smoother estimated trends; with $\text{var}(\eta_t) > 0$, the estimated trend will usually appear more responsive to temporary changes in the series. For the current problem, it is desirable to use the general local linear trend form since it can provide a better fitting model that is more informative about the time-variation in the seasonal and trend components.

3. Model-based estimation of trend and seasonal components

In (1) the relative variation in the η_{jt}^* 's determines the heteroskedastic structure. In setting up the model, there are a large number of variance parameters for the month-specific level shocks. However, a parsimonious model can be constructed on the basis of prior experience.

This section illustrates the use of the model for trend estimation and makes a comparison with X-12-ARIMA results for seasonal adjustment. Parameter estimates and results of testing for heteroskedasticity are presented. I consider nine regional housing starts and building permits series in the analysis.

In setting up model (1), each calendar month is classified as having either a high or low variance of the seasonal specific disturbance. The groupings of the months are displayed in table 1. Typically, the winter months, especially January, tend to cluster in the set of high variability months, though the precise classification depends on the series. The groupings are based on the knowledge and previous experience of Census Bureau staff in modeling and seasonal adjustment. With the exception of single-unit Northeast housing starts, where the high variance months are suggested by the analysis in Bell (2004), the groupings in table 1 represent the official

basis for the published seasonally adjusted data.

Figure 1 indicates monthly totals for the number of building permits issued in the Midwest region from January 1992 to December 2003. After the removal of trading day effects using X-12-ARIMA, logarithms of the series were taken. The trough of the series occurs with regularity in January, but the extent of the winter decline varies considerably throughout the sample. Overall, the amplitude of the seasonal component seems to fall off after 1999.

Series	Calendar months
Permits, MW, Tot	Jan-Mar, Dec
Permits, MW, 1U	(Mar, Aug, Sep)
Permits, NE, Tot	(Jul, Aug)
Permits, NE, 1U	Jan, Feb, Dec
Permits, S, Tot	(May)
Permits, W, 1U	Jan, Dec
Starts, MW, 1U	Jan, Feb
Starts, NE, 1U	Jan, Feb
Starts, S, 1U	Jan, Nov, Dec

Table 1 : Grouping of months used in heteroskedastic models for Census Bureau construction data. In the Series column, ‘Permits’ = Building Permits, ‘Starts’ = Housing Starts. ‘MW’ = Midwest region, ‘NE’ = Northeast, ‘S’ = South, ‘W’ = West as W. ‘Tot’ = Total number of structures, ‘1U’ = number of One-Unit structures. The second column shows the high variance months if this set contains six or fewer months; otherwise, the low variance months are shown enclosed by parentheses.

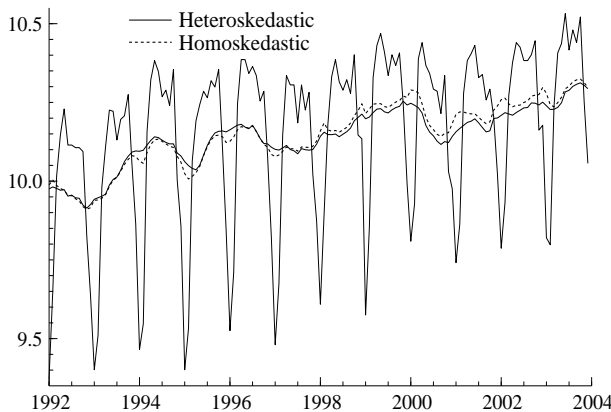


Figure 1: Estimated trends for total Midwest building permits (in logarithms), based on seasonal specific levels model for heteroskedastic and homoskedastic cases. The sample period is January 1992 to December 2003.

Model (1) is applied with the high and low idiosyncratic variances denoted by $\sigma_{\eta^*,h}^2$ and $\sigma_{\eta^*,l}^2$, respectively. Parameter estimation was carried out through maximum likelihood. In particular, the

likelihood function is evaluated for a given set of parameter values based on the prediction error decomposition, as described in Harvey (1989). The prediction errors are computed using the Kalman filter applied to the state space form of the model.

Maximum likelihood estimation was conducted using the BFGS algorithm to locate the optimum of the likelihood surface. This algorithm is an efficient quasi-Newton type method where the optimum is reached by an iterative updating procedure. Care must be taken in choosing starting values that yield the global maximum. The applications were performed by programs written in the Ox language of Doornik (1999) and relied on the Ssfpack library of routines for state space calculations described in Koopman et. al (1999). Parameter estimates are reported in Bell and Trimbur (2005).

For the series of Midwest total building permits, the variability in the seasonal-specific level rises by about 70% for January to March and December. Figure 1 shows the extracted trend. This series was computed by first estimating the vector of monthly levels, $\tilde{\boldsymbol{\mu}}_t = (\tilde{\mu}_{1t}, \dots, \tilde{\mu}_{st})'$, $t = 1, \dots, T$, with the Kalman smoother algorithm, and second, taking the product $\mathbf{w}'\tilde{\boldsymbol{\mu}}_t$. This gives the MMSE estimate of the level at each time t . Figure 1 also shows, for comparison, the trend produced under the homoskedastic restriction, that is $\text{var}(\eta_{jt}^*) = \sigma_{\eta^*}^2$ for all j . The trend computed from the signal extraction routine of X-12-ARIMA is typically close to the homoskedastic trend. Note that in each case, an additive decomposition is used for the logarithms of the original data, to ensure comparability of results.

In figure 1, the trend from the heteroskedastic model is relatively unaffected by the winter values. Though the general level matches that of the homoskedastic trend, the relative smoothness of the heteroskedastic estimates around the winter months marks a contrast. From 1994 to 1996, the severity of the January trough in the series induces temporary dips in the level for the homoskedastic model, but the heteroskedastic estimates are essentially robust.

Figure 2 shows the seasonal component by month plot for the heteroskedastic estimates. This plot presents the seasonal estimates specific to each calendar month. Thus, the lowest segment, which indicates successive January seasonals in each year from 1992 to 2003, gives a clear indication of the variability in the trough over the sample period. For comparison, the seasonal component from the official X-12-ARIMA seasonal adjustment (using the heterogeneous filters as noted above) is also shown. For the heteroskedastic model, the parametric estimates of the seasonal component capture an exten-

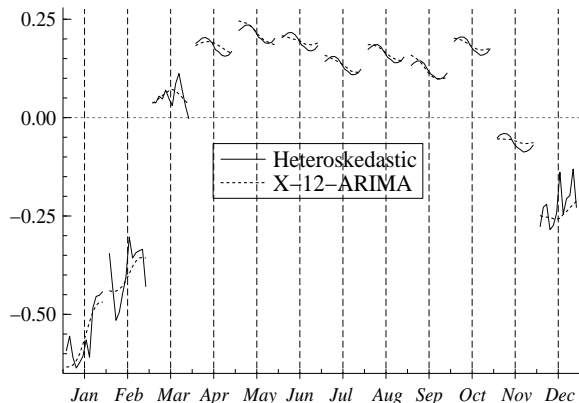


Figure 2: Seasonal component by month plots for total Midwest building permits (in logarithms), from heteroskedastic levels model and from an additive seasonal adjustment with X-12-ARIMA, sample 1992 to 2003.

sive range of movement in the highly variable calendar months. The evolution of the X-12 component is relatively smooth, for instance, the January trough gradually drifts upward over the sample period.

The seasonal component by month plots for the homoskedastic form of the model and for seasonal adjustment using X-12-ARIMA with homogeneous seasonal filters may be found in Bell and Trimbur (2005); these indicate moderate improvement from altering the choice of X-11 filters in the seasonal adjustment. However, the use of parametric modeling gives a dimension of flexibility that is unavailable in the nonparametric X-12-ARIMA signal extraction. Based on continuous-valued parameters and components, the model-based filters can be better adapted to the characteristic dynamics of a given series.

Under the homoskedastic assumption, the variability across seasons is fixed; in model (1), this corresponds to setting $var(\eta_{jt}^*) = \sigma_{\eta^*}^2$ for all j . Figure 1 shows that in this case the model may be forced to account for a periodic increase in volatility through short-term adjustment of the trend. On the other hand, with the separation into high and low variance months, model (1) produces a robust trend, so that the extra variation in winter is absorbed by the seasonal component. Figure 2 shows how the model gives seasonal estimates that effectively concentrate on the heteroskedasticity.

4. A test for seasonal heteroskedasticity in a finite-length series

In general, it is useful to have statistical evidence for the existence of seasonal heteroskedasticity in any given dataset. This provides quantitative

empirical confirmation for cases such as figure 1, and a general test allows one to assess the practical situation where the heteroskedastic property might be less dominant. This section extends the analysis by introducing a test for seasonal heteroskedasticity and applying it to the series in table 1.

The homoskedastic model involves a single restriction relative to the heteroskedastic model with a pair of idiosyncratic variance parameters. Denoting the maximized log-likelihood for estimation of the restricted model as $\tilde{\mathcal{L}}^*$ and that for the unrestricted model as $\tilde{\mathcal{L}}$, the likelihood-ratio test statistic is constructed as $LR = -2(\tilde{\mathcal{L}}^* - \tilde{\mathcal{L}})$. Under the null hypothesis of homoskedasticity, the asymptotic distribution of LR is chi-squared with a single degree of freedom under regularity conditions. This gives a benchmark for comparison, but there are some important considerations that affect the sampling distribution of LR for the current problem.

The regularity conditions require that the values assumed by the unrestricted parameter set under the null hypothesis lie away from the boundary of the parameter space. The restriction of homoskedasticity is $\sigma_{\eta^*,h}^2 = \sigma_{\eta^*,l}^2$ in the seasonal specific levels model with two variance parameters. Below, I use an unrestricted alternative of $\sigma_{\eta^*,h}^2 \neq \sigma_{\eta^*,l}^2$; this ensures that, as long as the variances are sufficiently far from zero, the parameter values lie well within the boundary of the permissible space.

One may, however, prefer to rule out the possibility $\sigma_{\eta^*,h}^2 < \sigma_{\eta^*,l}^2$ by specifying the unrestricted alternative as $\sigma_{\eta^*,h}^2 \geq \sigma_{\eta^*,l}^2$. Thus, in setting up and implementing the model for the Census construction data, the assumption is that the sets of higher variability months have been correctly identified. The question would be whether, under the null $\sigma_{\eta^*}^2 = \sigma_{\eta^*,h}^2 = \sigma_{\eta^*,l}^2$, estimation of homoskedastic and heteroskedastic models would lead to a rejection of $\sigma_{\eta^*,h}^2 = \sigma_{\eta^*,l}^2$ in favor of $\sigma_{\eta^*,h}^2 > \sigma_{\eta^*,l}^2$. The constraint $\sigma_{\eta^*,h}^2 \geq \sigma_{\eta^*,l}^2$ can be imposed directly in estimating the unrestricted model. Alternatively, one could reparametrize the model (1) so that the variance differential $\sigma_{\eta^*,h}^2 - \sigma_{\eta^*,l}^2$ is reflected in an additional noise term that is present only in the months of high variability; using this approach, as in Bell (2004), implicitly guarantees the prior assumption $\sigma_{\eta^*,h}^2 \geq \sigma_{\eta^*,l}^2$ by virtue of the model specification. Under the null hypothesis, an extra variance parameter would be set to zero. In this case, the (asymptotic) null distribution of LR would be a mixture of discrete mass at zero (probability 1/2) and a continuous density. Instead, in what follows, I assume $\sigma_{\eta^*,h}^2 \neq \sigma_{\eta^*,l}^2$ for the unrestricted model; this approach is conservative and simplifies the analysis.

As the construction series are of modest length, the actual behavior of test statistics may deviate in important ways from asymptotic limits. These finite sample effects will in general depend on the model and parameter values for each case. For some of the construction series, As an example, when the common idiosyncratic variance $\sigma_{\eta^*}^2$ is close to zero for the estimated null model, which is the case for several of the construction indicators, then this parameter has a clear and specific effect on the behavior of the test for model (1).

To draw from the true sampling distribution of LR , the true parameter values for the homoskedastic model for each series are required. This is clearly impossible since the parameters are unknown, but the ML estimates may be used as substitutes. It will be assumed that this gives an acceptable approximation for the purpose of studying the sampling distribution of LR . Using the estimated homoskedastic model for each series as the data generating process, 5000 artificial series are generated. For each of the artificial series, re-estimation of homoskedastic and heteroskedastic models yields a likelihood ratio, which gives a draw for LR under the null.

For each of the construction series in table 1, I compiled an empirical density as a histogram for the 5000 draws for LR . Estimates of critical values for the test statistic may be formed by the appropriate percentiles. Figure 3 shows the resulting null distribution for the model fitted to the series of single-family Northeast housing starts; for comparison, the density function for a $\chi^2(1)$ is shown as a solid curve.

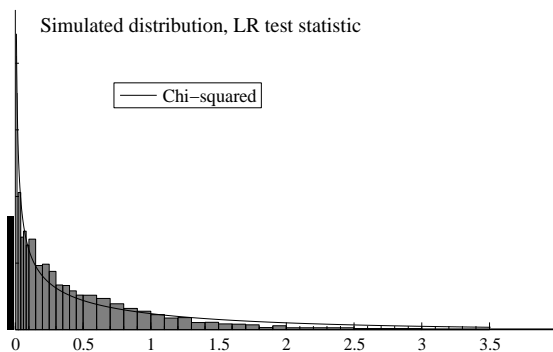


Figure 3: Distribution of likelihood ratio test statistic for simulations from homoskedastic model fitted to single-unit Northeast housing starts. The area of the solid black bar on the left gives the probability of zero.

In figure 3, the solid black bar on the bottom-left has area equal to the probability of obtaining a value of zero for LR . It is well-known that when

an unobserved components model has a variance parameter that is equal to or close to zero, the finite sample estimate of that variance will come out as exactly zero with positive probability; see Tanaka (1996), for instance. Apparently, a similar pile-up at zero occurs for the idiosyncratic variance parameters in the season-specific levels model (1), and further, the distribution of LR inherits this property, as I now explain.

Recall that, for each of the nine construction indicators in table 1, the ML estimates for the homoskedastic model are taken as the ‘true’ parameter values for the Monte Carlo simulations. That is, the estimated homoskedastic model is used as the data generating process. When the ‘true’ value of $\sigma_{\eta^*}^2$ for a given indicator is near zero, then for a moderate fraction of the 5000 artificial series generated on this basis, the maximum likelihood estimates of both $\sigma_{\eta^*,h}^2$ and $\sigma_{\eta^*,l}^2$ for the heteroskedastic model* will be zero. When this happens, the ML estimate of the homoskedastic variance $\sigma_{\eta^*}^2$ must also be zero for the simulated series; in this case, the maxima for homoskedastic and heteroskedastic models coincide, ensuring that $LR = 0$.

Series	LR_{η^*}	5%	Pr(0)
Permits, MW, Tot	19.6	2.37	0.025
Permits, MW, 1U	36.9	2.58	0.016
Permits, NE, Tot	32.4	2.16	0.063
Permits, NE, 1U	9.30	1.65	0.365
Permits, S, Tot	2.08	2.29	0.041
Permits, W, 1U	5.50	1.64	0.367
Starts, MW, 1U	9.72	2.28	0.036
Starts, NE, 1U	7.31	2.00	0.085
Starts, S, 1U	4.56	1.74	0.205

Table 2 : For each series, $LR = -2(\tilde{L}^* - \tilde{L})$ where \tilde{L}^* is the maximized log-likelihood under the null of homoskedasticity, and \tilde{L} is the maximized log-likelihood for the unrestricted heteroskedastic model. Estimated 5% critical values, shown in the ‘5%’ column, are based on 5000 Monte Carlo draws taking the estimated homoskedastic model as the true model.

The estimated probability mass for the pile-up at zero is reported in table 2. For the series of single-unit Building Permits in the West and in the Northeast, the value of $\sigma_{\eta^*}^2$ for the simulations is zero (this

* Of course, even though for any null model the true values satisfy $\sigma_{\eta^*,h}^2 = \sigma_{\eta^*,l}^2$, in general, the estimates for any particular artificial series will differ from the true values and will not be equal to one another. The event where both $\sigma_{\eta^*,h}^2$ and $\sigma_{\eta^*,l}^2$ are estimated as zero guarantees the event $LR = 0$.

being the MLE from estimation of the homoskedastic model for these two series), and there is about a 37% chance that the log-likelihood ratio is exactly zero under the null. The minimum probability for $LR = 0$ is 1.6% for single-unit Midwest permits.

Generally, relative to a $\chi^2(1)$, the probability mass for LR is shifted toward the origin. In the example of figure 3, even after renormalizing the histogram to account for the pile-up at zero, the empirical sampling distribution lies above the chi-squared between 0 and 0.5, and it decays more rapidly than the chi-squared for $LR > 1$. The thinner tail has the effect of lowering the critical value.

Table 2 shows results for the likelihood ratio tests. In eight out of nine cases, we reject the null hypothesis at a 5% level of significance in favor of seasonal heteroskedasticity. The 5% critical values range from 1.64 for single-unit West Building Permits to 2.58 for single-unit Midwest Building Permits; all lie well below the $\chi^2(1)$ value of 3.84, implying that the usual likelihood-ratio test is conservative. In five out of nine cases, the test statistic is highly significant with $LR > 9$, which leads to a decisive rejection of the null. Even if the inflated critical value of 3.84 were used, one would reject the null for all but one of the series.

In (1), the heteroskedastic structure is built directly into the specification of the month-specific levels, and the additional volatility in winter is captured in the seasonal component obtained from the model. An alternative approach is to specify the irregular as a seasonally heteroskedastic component, as adopted in Bell (2004). In (1) the analogue is produced by shifting the heteroskedastic part of the model from the level processes to the irregular, which is then assigned a time-varying variance that changes with the season j . Thus, the homoskedastic form of model (1), where $\boldsymbol{\eta}_t^* \sim NID(\mathbf{0}, \sigma_{\eta^*}^2 \mathbf{I})$, is generalized to include a heteroskedastic irregular, that is $var(\varepsilon_t) = \sigma_{\varepsilon_j}^2$.

The random shocks ε_t are uncorrelated across seasons, and their variance depends only on the season index j . In this heteroskedastic irregular model, the pattern of heteroskedasticity is now determined by how $\sigma_{\varepsilon_j}^2$ varies according to calendar month. This leads to different model properties and a distinct decomposition into seasonal, trend, and noise.

In the heteroskedastic levels model, (1), a large shock in a high variance month j at time t carries over into future periods. If seasonal changes in variability are linked to temporary influences such as unusual weather in certain months, then the heteroskedastic irregular model may give a better de-

scription of the dynamics; the heteroskedastic shocks ε_t are stationary, each shock being a simple one-period effect. The variability in month j is directly expressed in the variance $\sigma_{\varepsilon_j}^2$. It is less straightforward to link the variance parameters in (1) to individual calendar months because the twelve month-specific processes are nonstationary, and the idiosyncratic shocks have persistent effects over time.

Table 3 shows results of likelihood-ratio tests for the heteroskedastic irregular model. Now there are two series where the null cannot be rejected at a 5% level of significance. For single-unit housing starts in the South, LR_{ε^*} is about one-third of LR_{η^*} ; this reduces the test statistic to below the 10% critical value. However, most of the basic conclusions are unchanged from table 2.

Series	LR_{ε^*}	5%
Permits, MW, Tot	19.1	4.18
Permits, MW, 1U	34.7	2.60
Permits, NE, Tot	36.0	4.33
Permits, NE, 1U	16.5	2.82
Permits, S, Tot	0.50	4.40
Permits, W, 1U	9.17	3.57
Starts, MW, 1U	9.56	4.43
Starts, NE, 1U	8.67	4.58
Starts, S, 1U	1.51	4.38
5% critical value, $\chi^2(1)$		3.84

Table 3 : For each series, $LR = -2(\tilde{L}^* - \tilde{L})$ where \tilde{L} is the maximized log-likelihood for the unrestricted model, that is with heteroskedastic irregular.

For seven of the regional housing starts and building permits series, the results tend to support the use of heterogeneous filters in seasonal adjustment. Relative to X-12-ARIMA, the greater adaptability of the parametric model-based approach is reflected in the estimated seasonal and trend filters, that reflect the properties of the series through the estimated model parameters. Though not investigated in this paper, the shape of the model-based kernel for estimating the seasonal could be directly compared to the nonparametric X-11 filters.

5. Adjustment for seasonal noise

In the seasonal component by month plot in figure 2, the estimates for the heteroskedastic model are considerably more variable than the nonparametric output of X-12-ARIMA. For the high variance months, the heteroskedastic component spans a wider range and undergoes larger short-term fluctuations than the X-12 seasonal. Given their relative noisiness, at least part of the extra variation in the

parametric estimates seems to be linked to seasonal increases in volatility.

The concept of seasonal heteroskedasticity loosely defines a changing degree of variability for different calendar months. The precise structure of such movements is left open, so it is natural to consider different modeling approaches to the problem.

Using the statistical models for heteroskedasticity, the additional variability linked to calendar month can be measured. Thus, for instance, in (1), the ratio of the idiosyncratic variances can be compared. However, since each of the month-specific processes μ_{jt} continues to evolve throughout the sample, it is unclear exactly how to localize the effect on the observation y_t of a shock to μ_{jt} in month j . For (1), in contrast, the measurement of the extra noise by calendar month is straightforward.

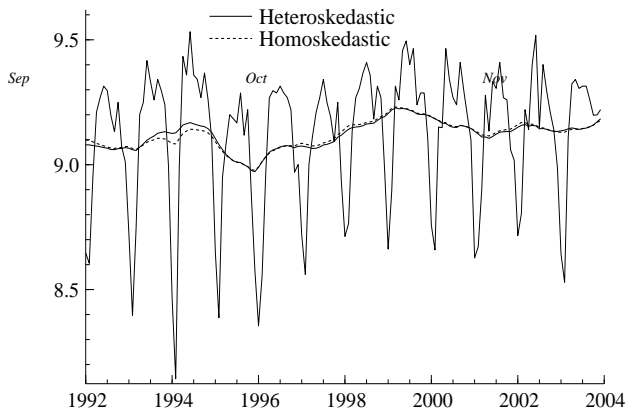


Figure 4: Estimated trends for single-unit Northeast housing starts (in logarithms). Sample period is January 1992 to December 2003.

Figure 4 shows the series of single-unit Northeast housing starts. The estimated trend based on the heteroskedastic irregular model is less responsive to winter volatility than is the homoskedastic trend. In the months surrounding early 1994, the difference is especially noticeable; with the heteroskedastic irregular, the level of the model-based estimates is sustained as it passes through the unusually steep decline arising from severe weather (reports indicate that January 1994 was a severely cold month). The homoskedastic model produces a trend that is influenced at several leads and lags and whose growth rate dips in the immediate vicinity of the trough.

Consider now a decomposition of the white noise variation $\varepsilon_t = \varepsilon_t^s + \varepsilon_t^n$, so that a distinction is made between seasonal and nonseasonal components of the irregular. Therefore, ε_t^s represents the excess volatility present in certain months, while the non-

systematic randomness present throughout the year is captured in ε_t^n . The series can be decomposed as

$$y_t = \mu_t + \gamma_t + \varepsilon_t^s + \varepsilon_t^n = \mu_t + s_t + \varepsilon_t^n \quad (2)$$

where $s_t = \gamma_t + \varepsilon_t^s$ is the total seasonal influence and ε_t^n is a ‘pure’, or constant variance, irregular.

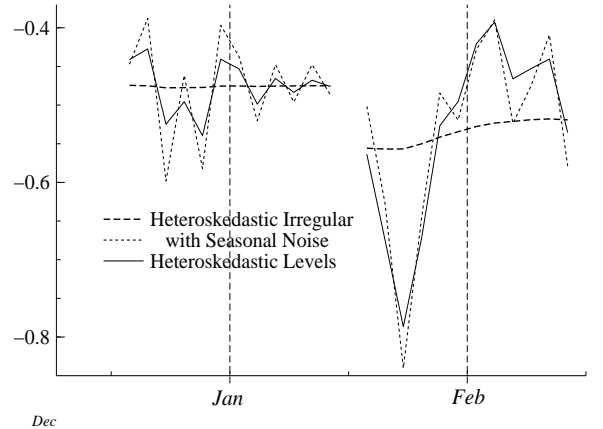


Figure 5: Seasonal components for single-unit Northeast housing starts, sample 1992 to 2003.

Figure 5 shows the smoothed seasonal $\tilde{\gamma}_t$ in January and February for the heteroskedastic irregular model; it is more or less flat. The rise in variability that punctuates the winter is being absorbed by the irregular $\tilde{\varepsilon}_t$. The MMSE estimate of the seasonal noise $\tilde{\varepsilon}_t^s$, conditional on estimated parameters and $\tilde{\varepsilon}_t$, is a simple signal extraction, and figure 5 shows the total seasonal component \tilde{s}_t resulting from the addition of $\tilde{\varepsilon}_t^s$ to $\tilde{\gamma}_t$. For comparison, the seasonal estimated from model (1) is shown as well.

The resemblance between \tilde{s}_t and the seasonal from (1) suggests the models show similar effectiveness in capturing the heteroskedastic variation. The heteroskedastic levels model directly gives a seasonal that absorbs a good deal of short-term variation but is less erratic than the total seasonal produced by the second-step addition of seasonal noise. Since in (1) the heteroskedasticity is embedded in the season-specific level processes, which together form a sort of multivariate trend model, this may help ensure some persistence in the implied seasonal.

The broader definition of seasonality may be used to produce an adjusted series that is subject mainly to nonseasonal uncertainty. Figure 6 shows the fully adjusted series in a plot by month format for January and February. This enables one to abstract from the seasonal volatility and focus on the nonseasonal variation such as the trend movements and the remaining noise $\tilde{\varepsilon}_t^n$, which is closer to constant variance white noise.

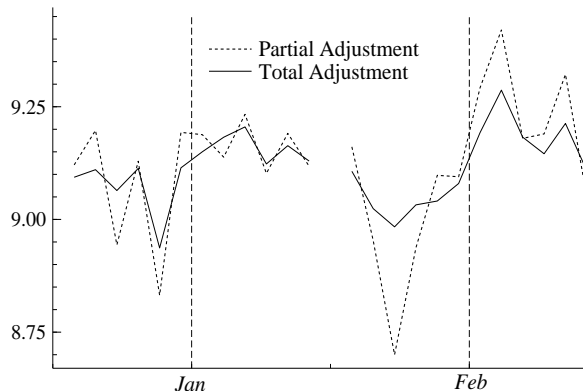


Figure 6: Seasonally adjusted series for single-unit Northeast housing starts, sample 1992 to 2003.

The additional noise recurring in winter is still influential for the partially adjusted series. In the absence of additional knowledge, observations such as the dip at the beginning of 1994 may lead to special concern on the part of policymakers. However, once the effect of the seasonal influence is accounted for, it is seen that, relative to the course of events in a typical calendar year, their severity is overstated.

The above examples show how the seasonal specific model with heteroskedastic disturbances can give a smoothed trend that, in the presence of heightened variability in certain months, successfully maintains its long-run course. This is possible since the extra noise, associated with seasonal increases in volatility, are effectively absorbed by a separate component. In the heteroskedastic irregular model, the noise is separated from the smoothed seasonal, and it is possible to extract the seasonal noise and study its effect directly. The advantage of (1) is that the direct estimate of the seasonal component already captures a good deal of the seasonal noise variation but preserves a moderate degree of smoothness.

6. Conclusions

This paper concentrates on models for seasonal time series that are set up as a collection of processes, one for each season. Unlike traditional structural, or unobserved components time series models, in models such as (1), the seasonal and trend components are enmeshed in a vector of season-specific level processes. The applications have illustrated how this model structure is convenient for handling the property of seasonal heteroskedasticity when there is specific knowledge about which months contribute to the increase in variability.

In using a likelihood ratio test of the property,

the heteroskedastic levels model gives similar conclusions to the heteroskedastic irregular model for a set of Census Bureau construction series. There are, however, some differences, and the finite sample distribution of LR under the null is better behaved for the heteroskedastic irregular model. The empirical analysis demonstrates the feasibility of testing series of limited length for seasonal heteroskedasticity.

References

- Bell, W., 2004. On RegComponent time series models and their applications. In A. Harvey et al. (eds), *State space and unobserved components models : Theory and applications*, Cambridge : Cambridge University Press.
- Bell, W., and Trimbur T., 2005. Seasonal heteroskedasticity in time series : modeling, estimation, and testing. Working paper.
- Doornik, J. A., 1999. *Ox: An Object-Oriented Matrix Programming Language* (London: Timberlake Consultants Ltd.).
- Findley, D., Monsell B., Bell W., Otto M., and Chen B.-C. 1998. New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program, *Journal of Business and Economic Statistics* 16 (2), 127-77.
- Harvey A.C., 1989. *Forecasting, structural time series models and the Kalman filter* (Cambridge: Cambridge University Press).
- Koopman, S. J., Shephard N., and Doornik J., 1999. Statistical algorithms for models in state space using SsfPack 2.2, *Econometrics Journal* 2, 113-66.
- Proietti, T. 2004. Seasonal specific structural Time Series, *Studies in Nonlinear Dynamics & Econometrics*, 8 (2), Article 16.
- Tanaka K., 1996. *Time Series Analysis* (New York: John Wiley & Sons, Inc.).
- Acknowledgements:** Thanks go to William Bell, David Findley, John Hisnanick, Tucker McElroy, and Brian Monsell for their comments. Tomasso Proietti provided Ox code, which I modified and extended to handle the applications in this paper. Kathleen McDonald-Johnson supplied the construction dataset and gave background information on the seasonally adjusted estimates published by the U.S. Census Bureau.