Heterogeneous seasonal patterns in agricultural data and evolving splines

Cáceres-Hernández, José Juan (jcaceres@ull.es)
Martín-Rodríguez, Gloria (gmartinr@ull.es)
Universidad de La Laguna

Poster paper prepared for presentation at the International Association of Agricultural Economists Conference, Gold Coast, Australia, August 12-18, 2006

Abstract

In this paper an appropriate model of the seasonal pattern in high frequency agricultural data is proposed that takes the specific nature of such a pattern into account. The methodological proposal is based on evolving splines that are shown to be a tool capable of modelling seasonal variations in which either the period or the magnitude of the seasonal fluctuations do not remain the same over time. The seasonal pattern in each year or agricultural campaign is modelled in such a way that the seasonal effect at each season is a function of the seasonal effects corresponding to some fixed seasons that act as reference points. The spline function is enforced to satisfy several conditions that provide some regularity in the adjusted seasonal fluctuation; on the other hand, the main source of changes in the adjusted seasonal pattern is obtained by assuming that the values of the seasonal effects at the fixed reference seasons do not remain the same year by year. If the length of the period in which the seasonal fluctuation is completed does not change, the proposed specification is flexible enough to test the hypothesis that the seasonal pattern in several consecutive years is fixed by using simple statistical procedures. This proposal is applied to capture the movements in a weekly tomato export series and the analysis is carried out inside the frame delimited by the structural approach to time series.

JEL Subject Codes: C22, Q17.

1. Research problem

Agricultural time series recorded at intervals shorter than a year are increasingly used in economic literature (Miller and Hayenga, 2001; Sorensen, 2002; Sanjuán and Dawson, 2003; Hill *et al*., 2004; Jarvis and Vera-Toscano, 2004; Rucker *et al*., 2005; Richards and Patterson, 2005). And, bearing in mind that better knowledge of seasonal variability is valuable for agricultural decision making at the farm, marketing or policy level, additional research on agricultural economics applied to monthly, weekly, daily or hourly series or even to data recorded at shorter intervals is usually needed.

These seasonal series are often characterised by the presence of observations corresponding to specific time periods in a year that are not always observed at the same period in other years. Such an irregular distribution of observations throughout the year is usually clear in high frequency series. For example, the horticultural products are not exactly harvested during the same weeks year by year. Therefore, the seasonal behaviour in this kind of series may be far from standard and can hardly be coped by means of conventional models. Note that the length of the period in which the seasonal fluctuation is considered to be completed does not usually remain the same over time and alternative more flexible models should be used. In this sense, the specification of changing deterministic components could be appropriate. Fixed splines are another interesting tool to capture this type of changes in the seasonal pattern (Martín and Cáceres, 2005). But, leaving aside anomalous movements, the changes in the behaviour of many agricultural series are gradual and smooth enough to be captured by means of intermediate models between deterministic and stochastic formulations. From this point of view, in this paper evolving splines are proposed as an appropriate method of dealing with a changing seasonal pattern.

2. Methods

In a time series model formulated as

$$y_t = \mu_t + \gamma_t + \varepsilon_t,\ t = 1,...,T, \tag{1}$$

where μ_t and γ_t are the trend or level component and the seasonal component, and ε_t is the irregular component, modelling the seasonal pattern by means of a set of regressors defining a spline function could be interesting.

When the seasonal pattern is fixed, $\gamma_t = \gamma_w$ if the observation at time t corresponds to the season w, $w = 1,...,s$; then, this component can be modelled by a periodic cubic spline. That is,

$$\gamma_w = g(w) + \xi_w, \tag{2}$$

where ξ_w is a residual term and $g(w)$ is a third degree piecewise polynomial function,

$$g_i(w) = g_{i,0} + g_{i,1}w + g_{i,2}w^2 + g_{i,3}w^3,\ w_{i-1} \le w \le w_i,\ i = 1,...,k-1, \tag{3.a}$$

$$g_k(w) = g_{k,0} + g_{k,1}w + g_{k,2}w^2 + g_{k,3}w^3,\ w_{k-1} \le w \le s, \tag{3.b}$$

where w_0 is the first season. Koopman (1992) and Harvey *et al.* (1997), based on Poirier (1976), propose to formulate the previous spline as the linear function

$$g(w) = \gamma_0^+ X_{0,w} + ... + \gamma_{k-1}^+ X_{k-1,w}, \tag{4}$$

where $X_{0,w},..., X_{k-1,w}$ are appropriate regressors defined as functions of the break points $\{w_i\}_{i=0,...,k-1}$ and $\{\gamma_0^+,...,\gamma_{k-1}^+\}$ are the observed values of the seasonal pattern at the previous break points. However, in this paper the seasonal pattern is proposed to be specified as $\gamma_w = g(w) + \xi_w$, where the spline $g(w)$ is expressed as

$$g(w) = \gamma_0^* X_{0,w} + ... + \gamma_{k-1}^* X_{k-1,w}, \tag{5}$$

where $X_{0,w},\ldots,X_{k-1,w}$ are the regressors previously defined and $\gamma_0^*,\ldots,\gamma_{k-1}^*$ are free parameters which must be estimated. In this way, the seasonal pattern can be incorporated into the model in Equation (1) as a function of such regressors.

The previous specification is flexible enough to capture a non-fixed seasonal pattern. The period under study could be divided in sub-periods of s time units (seasons). Suppose that there are m sub-periods. For the sub-period c, $c=1,\ldots,m$, appropriate regressors $X_{0,w}^c,\ldots,X_{k_c-1,w}^c$ could be defined as functions of the break points $\{w_i^c\}_{i=0,\ldots,k_c-1}$. Although the break points would be assumed to be the same for different sub-periods, in such a way that regresssors $X_{0,w}^c,\ldots,X_{k_c-1,w}^c$ are also the same, changes in the magnitude of seasonal variations are able to be captured by defining different parameters $\gamma_0^{c*},\ldots,\gamma_{k_c-1}^{c*}$ for each sub-period. Furthermore, when the period s in which the seasonal variation is completed does not remain the same over time, the length of the sub-period c can be defined as s_c, $c=1,\ldots,m$. That is to say, the seasonal pattern could be formulated as $\gamma_t = g(t)+\xi_t$, where the spline $g(t)$ is expressed as

$$g(t)=\sum_{c=1}^{m}\left[\gamma_0^{c*}X_{0,t}^c+\ldots+\gamma_{k_c-1}^{c*}X_{k_c-1,t}^c\right]D_{c,t}^c\,, \tag{6}$$

where $D_{c,t}^c=\begin{cases}1, & t\in sub-period\ c\\ 0, & in\ other\ case\end{cases}$, $c=1,\ldots,m$, and $X_{i,t}^c=X_{i,w}^c$, $i=0,\ldots,k_c-1$, if the observation at time t corresponds to the season w, $w=1,\ldots,s_c$. When the length s_c and the break points w_i^c are the same for all sub-periods, then $X_{i,w}^c=X_{i,w}$, $i=0,\ldots,k-1$, but the seasonal variations are able to evolve over time. When $\gamma_{i,w}^{c*}=\gamma_{i,w}^*$, $i=0,\ldots,k-1$, the seasonal pattern is fixed.

The critical point is the selection of the number and position of knots. In this sense, the decision has been adopted to select the combination of locations that minimises the residual sum of squares when the following regression model

$$\gamma_t^1 = \sum_{c=1}^{m} \left[\gamma_0^{c*} X_{0,t}^c + \ldots + \gamma_{k_c-1}^{c*} X_{k_c-1,t}^c \right] D_{c,t}^c + \xi_t , \qquad (7)$$

is estimated, γ_t^1 being a previous seasonal component approximation. For the chosen locations, the regressors $X_{0,t}^c D_{c,t}^c, \ldots, X_{k_c-1,t}^c D_{c,t}^c$, $c = 1, \ldots, m$, can be incorporated into the time series model in Equation (1) as exogenous variables. A structural time series model (Harvey, 1989) is usually an appropriate specification to model the instabilities in the components of the series. Seasonal pattern can be similar in several sub-periods. So, some F tests can be applied to check this assumption and simplify the model (7).

3. Results

This section is concerned with the series of weekly *Las Palmas de Gran Canaria* tomato exports (measured in 6 kg boxes) from 1980/1981 to 2003/2004 harvests[1] (Figure 1). According to data identification purpose, each harvest is considered to start in week 27 of a year and conclude in week 26 of the following year.

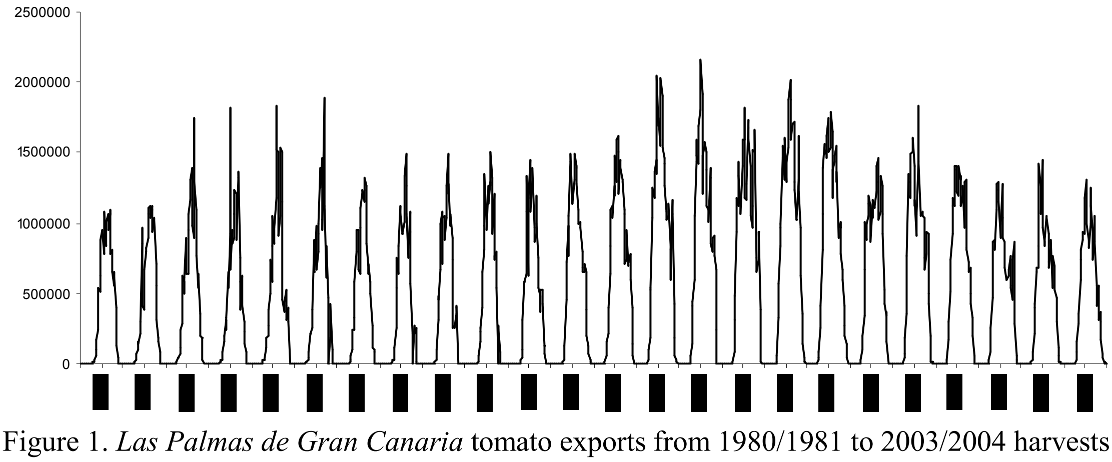


Figure 1. *Las Palmas de Gran Canaria* tomato exports from 1980/1981 to 2003/2004 harvests.

[1] Export statistics have been obtained from weekly data published by the provincial exporter association of *Santa Cruz de Tenerife* (*ACETO*) in its export season reports. In those weeks where this source did not register any data, a zero value has been assigned.

As regard the seasonal pattern, a harvest by harvest rising movement is observed that begins in October and finishes in January or February, followed by another downward movement that continues until May or June. However, from the 1991/1992 harvest, the harvests, often finished in early May, continued until June. Because exports are almost or exactly zero for some weeks in each year of the sample and the non-export period is longer until the 90/91 harvest, a model with a fixed seasonal period throughout the sample fails. In the first period, there are regular exports from week 42 of the year to week 19 of the following year. In the other one, the export activity could be considered to start in week 41 and conclude in week 22. So, if only observations corresponding to these weeks are considered, a new series is obtained and it will be referred to as $\{y_t\}_{t=1,\ldots,772}$, hereafter. It is appropriate to specify a model for the new series capable of capturing a seasonal pattern in which the period is 30 until the 90/91 harvest and other one in which the period is 34 from the 91/92 harvest. In this paper, the proposal for coping with these two seasonal patterns consists of using an evolving spline function.

The adequate specification of the spline, according to the previous section, requires obtaining a previous approximation of the seasonal component in each of the two periods. The stochastic formulation of the seasonal component requires a fixed seasonal period. Then, for each period, the results of estimating a basic structural model by maximum likelihood indicate that, in the first period, the level component is fixed and the seasonal component is stochastic, whereas, in the second period, the character of both components is changed.

It was opted for obtaining other two approaches of the seasonal pattern in each period. First, a model is estimated with a three-segment linear spline capturing the trend component[2]. Then the residual term of this regression model is a rough approximation of seasonal variations, $\{\gamma_t^{1a}\}_{t=1,\ldots,772}$. Second, moving averages with period 30 until the last observation of 90/91 harvest and moving averages with period 34 since the first observation of 91/92 harvest are calculated. Then the difference between $\{y_t\}_{t=1,\ldots,772}$ series and moving average series is another approximation of seasonal variations, $\{\gamma_t^{1b}\}_{t=16,\ldots,755}$. These approximations suggest that the structural model approximation does not capture all the seasonal behaviour. Perhaps, as a response to the iterative estimating procedure, the variance of the trend component is high enough to capture some seasonal variations. In this sense, it is opted for using the last two approximations (Figure 2) to specify an evolving spline as it was indicated in methodological section.

Period I: 80/81 to 90/91 | Period II: 91/92 to 03/04

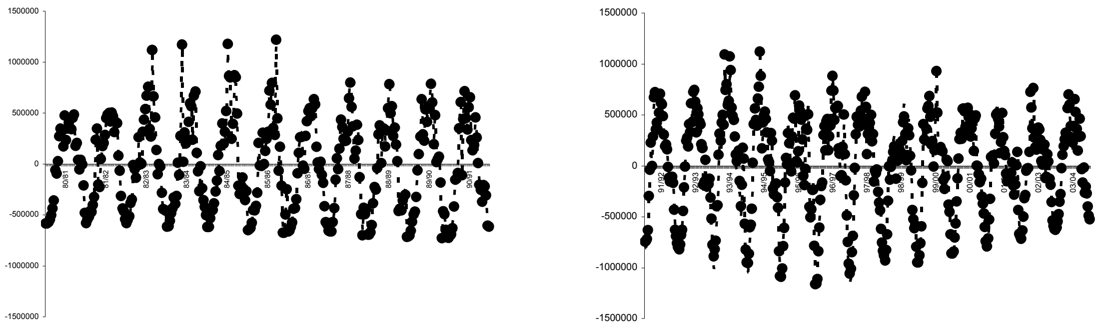


Figure 2. Seasonal components: $\{\gamma_t^{1a}\}_{t=1,\ldots,772}$ (•••) and $\{\gamma_t^{1b}\}_{t=16,\ldots,755}$ (– – –)

That is to say, the seasonal pattern could be formulated as $\gamma_t = g(t) + \xi_t$, where the spline $g(t)$ is expressed as

$$g(t) = \sum_{c=1}^{24} \left[\gamma_0^{c*} X_{0,t}^{c} + \ldots + \gamma_{k_c-1}^{c*} X_{k_c-1,t}^{c}\right] D_{c,t}^{c}\ , \qquad (8)$$

---

[2] The break points divide the sample under study in three periods: 80/81-90/91, 91/92-95/96 and 96/97-03/04. Note that these three periods differing by the long-term movement can be distinguished in Figure 1. The new trade situation of the Canary Islands with regard to the *EU* since July 1991 (reference prices were substituted by supply prices) and the full integration into the *EU* since January 1st 1993 (abolition of reference/supply prices) brought about a significant export boost. The general growth in exports in this second period was interrupted in 1996, coinciding with the introduction of a trade agreement between the *EU* and Morocco.

where $D^t_{c,t} = \begin{cases} 1, t \in harvest\ c \\ 0, in\ \ other\ \ case \end{cases}$, $c = 1,...,24$, and $X^c_{i,t} = X^c_{i,w}$, $i = 0,...,k_c - 1$, if the observation at time t corresponds to the season w, $w = 1,...,s_c$, where $s_c = \begin{cases} 30, c = 1,...,11 \\ 34, c = 12,...,24 \end{cases}$.

The spline is specified as a function of week w of the export period. That is to say, $w=1$ being the corresponding week of the year in which the export period is considered to start and $w=s$ being the last week of the following year in which the export period is considered to conclude. So, until 90/91 harvest, the length of the seasonal period is 30 in such a way that $w=1$ corresponds to week 42 of a year and $w=30$ corresponds to week 19 of the following year. From 91/92 harvest, the length of the seasonal period is 34 in such a way that $w=1$ corresponds to week 41 of a year and $w=34$ corresponds to week 22 of the following year.

To obtain a more parsimonious formulation, the break points are assumed to be the same for all harvests from 80/81 to 90/91. The same assumption is taken for all harvests from 91/92 to 03/04. Then $X^c_{i,w} = X^I_{i,w}$, $c = 1,...,11$, and $X^c_{i,w} = X^{II}_{i,w}$, $c = 12,...,24$, but the seasonal variations $\gamma^c_{i,w}$, $i = 0,...,k_c - 1$, could evolve over time. The resulting model is

$$\gamma_t = \sum_{c=1}^{11} \left[\gamma_0^{c*} X^I_{0,t} + ... + \gamma_{k-1}^{c*} X^I_{k_1-1,t} \right] D^c_{c,t} + \sum_{c=12}^{24} \left[\gamma_0^{c*} X^{II}_{0,t} + ... + \gamma_{k_2-1}^{c*} X^{II}_{k_2-1,t} \right] D^c_{c,t} + \xi_t . \qquad (9)$$

For each period, the decision has been adopted to select the number of knots by estimating the regression models

$$\gamma^1_w = \gamma^*_0 X^I_{0,w} + ... + \gamma^*_{k-1} X^I_{k-1,w} + \xi_w,\ w = 1,...,30, \qquad (10.a)$$

where γ^1_w corresponds to the averages values per week calculated from $\{\gamma^{1a}_t\}_{t=1,...,772}$ or $\{\gamma^{1b}_t\}_{t=16,...,755}$ series for the first period, and

$$\gamma^1_w = \gamma^*_0 X^{II}_{0,w} + ... + \gamma^*_{k-1} X^{II}_{k-1,w} + \xi_w,\ w = 1,...,34, \qquad (10.b)$$

where γ_w^1 corresponds to the averages values per week calculated from $\{\gamma_t^{1a}\}_{t=1,...,772}$ or $\{\gamma_t^{1b}\}_{t=16,...,755}$ series for the second period. The results of estimating previous models suggest a six-segment spline for both the first and the second period as an adequate specification. The combinations of locations that minimises the residual sum of squares when the regression models previously defined are estimated are the same using either of two approximations.

According to the results of estimating the model

$$\gamma_t^{1a} = \sum_{c=1}^{11} \left[\gamma_0^{c*} X_{0,t}^{I} + ... + \gamma_5^{c*} X_{5,t}^{I}\right] D_{c,t}^{c} + \sum_{c=12}^{24} \left[\gamma_0^{c*} X_{0,t}^{II} + ... + \gamma_5^{c*} X_{5,t}^{II}\right] D_{c,t}^{c} + \xi_t, \tag{11}$$

F tests to check the hypothesis that the seasonal pattern is the same in some harvests lead to conclude that the seasonal pattern is the same in 82/83 and 83/84 harvests; the same conclusion is obtained for 87/88 and 88/89 harvests. On the other hand, once the model

$$\gamma_t^{1b} = \sum_{c=1}^{11} \left[\gamma_0^{c*} X_{0,t}^{I} + ... + \gamma_5^{c*} X_{5,t}^{I}\right] D_{c,t}^{c} + \sum_{c=12}^{24} \left[\gamma_0^{c*} X_{0,t}^{II} + ... + \gamma_5^{c*} X_{5,t}^{II}\right] D_{c,t}^{c} + \xi_t, \tag{12}$$

is estimated, F tests lead to conclude that the seasonal pattern is the same between 82/83 and 83/84 harvests, 87/88, 88/89 and 89/90 harvests, and 91/92 and 92/93 harvests[3]. Given that two different specifications of the seasonal pattern are likely, a structural model can be formulated in which seasonal variations are capturing by means of the regressors $X_{0,t}^{I} D_{c,t}^{c},..., X_{5,t}^{I} D_{c,t}^{c}$, $c = 1,...,11$, and $X_{0,t}^{II} D_{c,t}^{c},..., X_{5,t}^{II} D_{c,t}^{c}$, $c = 12, ...,24$. That is to say, these regressors can be introduced into the structural model as exogenous variables; but, in order to avoid multicolinearity problems, the regressor $X_{5,t}^{II} D_{24,t}^{c}$ is dropped and the structural model can be formulated as

[3] Before estimating model (12), a new series $\{\gamma_t^{1b}\}_{t=1,...,772}$ was obtained by substituting the unknown values at times $t = 1,...,15,756,772$ for the corresponding values of the series $\{\gamma_t^{1a}\}_{t=1,...,772}$.

$$y_t = \mu_t + \sum_{c=1}^{11} \left[\gamma_0^{c*} X_{0,t}^{I} + \ldots + \gamma_5^{c*} X_{5,t}^{I}\right] D_{c,t}^{c} + \sum_{c=12}^{23} \left[\gamma_0^{c*} X_{0,t}^{II} + \ldots + \gamma_5^{c*} X_{5,t}^{II}\right] D_{c,t}^{c} + \\ + \left[\gamma_0^{24*} X_{0,t}^{II} + \ldots + \gamma_4^{24*} X_{4,t}^{II}\right] D_{24,t}^{c} + \varepsilon_t \tag{13}$$

As regards the long term movements, a stochastic level component with stochastic slope has been formulated. The results of estimating such a structural model suggest to specify μ_t as a fixed term, that is to say, $\mu_t = \mu$. Another option is to formulate the level component as a three-segment linear spline defined by using the break points identified in this section. Note that the assumption $\mu_t = \mu$ does not imply a fixed level because spline regressors are capturing some long term movements. Furthermore, when the level is modelled by means of a linear spline the model is able to capture the long term movement throughout a harvest. So, this model is chosen. Again, F tests can be applied to check the hypothesis that the seasonal pattern is the same in some harvests. According to them, the conclusion is obtained that the seasonal pattern is the same in 82/83 and 83/84 harvests, and in 87/88 and 88/89 harvests.

Now, an appropriate structural model is

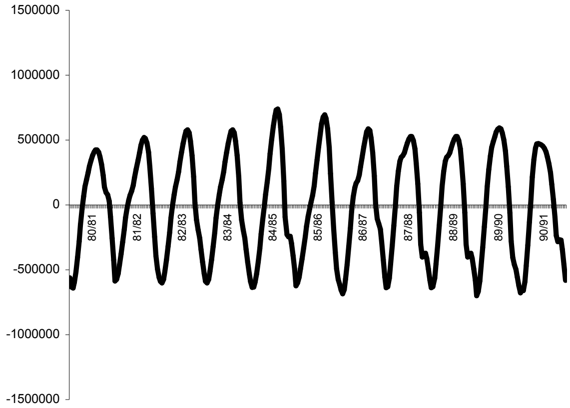
$$y_t = \mu_t + \sum_{c=1,2,5,6,7,10,11}^{11} \left[\gamma_0^{c*} X_{0,t}^{I} + \ldots + \gamma_5^{c*} X_{5,t}^{I}\right] D_{c,t}^{c} + \left[\gamma_0^{3,4*} X_{0,t}^{I} + \ldots + \gamma_5^{3,4*} X_{5,t}^{I}\right]\left(D_{3,t}^{c} + D_{4,t}^{c}\right) + \\ + \left[\gamma_0^{8,9*} X_{0,t}^{I} + \ldots + \gamma_5^{8,9*} X_{5,t}^{I}\right]\left(D_{8,t}^{c} + D_{9,t}^{c}\right) + \sum_{c=12}^{23} \left[\gamma_0^{c*} X_{0,t}^{II} + \ldots + \gamma_5^{c*} X_{5,t}^{II}\right] D_{c,t}^{c} + \\ + \left[\gamma_0^{24*} X_{0,t}^{II} + \ldots + \gamma_4^{24*} X_{4,t}^{II}\right] D_{24,t}^{c} + \sum_{m-\ell} \lambda_{m-\ell} I_{m-\ell} + \varepsilon_t \tag{14}$$

$I_{m-\ell}$ being an intervention variable capturing outliers which takes value 1 when the observation corresponding to the week m of the year ℓ and is equal to zero in other case. The final estimates of the seasonal variations are shown in Figure 3[4].

---

[4] The estimates of the seasonal component obtained from the estimates of spline parameters are corrected in such a way that the seasonal variations sum up to zero over each harvest. Then, the estimates of the level component are also properly corrected so that the same variations are not captured simultaneously by trend and seasonal components.

Period I: 80/81 to 90/91



Period II: 91/92 to 03/04

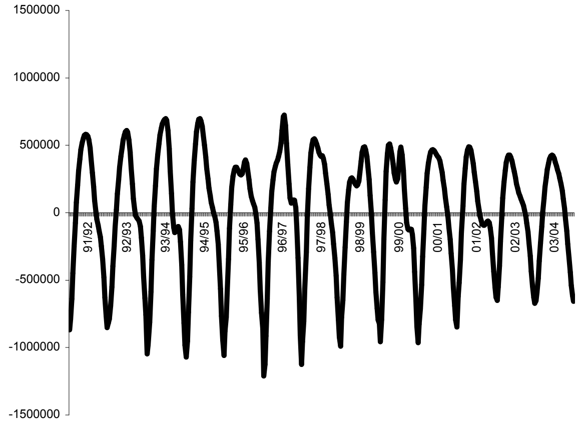


Figure 3. Seasonal component

References

Harvey, AC (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Harvey, AC, Koopman, SJ and Riani, M (1997). ‘The modelling and seasonal adjustment of weekly observations’, *Journal of Business and Economic Statistics*, 15, pp. 354-368.

Hill, HSJ, Mjelde, JW, Love, HA, Rubas, DJ, Fuller, SW, Rosenthal, W and Hammer, G (2004). ‘Implications of seasonal climate forecast on world wheat trade: a stochastic dynamic analysis’, *Canadian Journal of Agricultural Economics-Revue Canadienne d Agroeconomie*, 52, pp. 289-312.

Jarvis, L and Vera-Toscano, E (2004). ‘Seasonal adjustment in a market for femate agricultural workers’, *American Journal of Agricultural Economics*, 86, pp. 254-266.

Koopman, SJ (1992). ‘Diagnostic Checking and Intra-Daily Effects in Time Series Models’. Thesis Publishers, Tinbergen Institute Research Series, 27. Amsterdam.

Martín, G and Cáceres, JJ (2005). ‘Modelling weekly Canary tomato exports’, *Agricultural Economics*, 33, pp. 255-267.

Miller, DJ and Hayenga, ML (2001). ‘Price cycles and asymmetric price transmission in the US pork market’, *American Journal of Agricultural Economics*, 83, pp. 551-562.

Poirier, DJ (1976). *The Econometric of Structural Change with special emphasis on Spline Functions*, Amsterdam, North Holland.

Richards, TJ and Patterson, PM (2005). ‘Retail price fixity as a facilitating mechanism’, *American Journal of Agricultural Economics*, 87, pp. 85-102.

Rucker, RR, Thurman, WN and Yoder, JK (2005). ‘Estimating the structure of market reaction to news: information events and lumber future prices’, *American Journal of Agricultural Economics*, 87, pp. 482-500.

Sanjuán, AI and Dawson, PJ (2003). ‘Price transmission: BSE and structural breaks in the UK meat sector’, *European Review of Agricultural Economics*, 30, pp. 155-172.

Sorensen, C (2002). ‘Modeling seasonality in agricultural commodity futures’, *Journal of Futures Markets*, 22, pp. 393-426.