A nonparametric ACD model

Antonio Cosma[*] Fausto Galli[§]

December 2017

Abstract

We propose a fully nonparametric approach to the analysis of the Autocorrelated Conditional Duration (ACD) process applied to durations between financial events. We use a recursive algorithm to estimate the nonparametric specification. In a Monte Carlo experiment, we analyse its forecasting performance and compare it with a correctly and a mis-specified parametric estimator. On a real dataset, the nonparametric estimator seems to mildly overperform in terms of predictive power. The nonparametric analysis can also provide guidance on the choice between alternative parametric specifications. In particular, once intraday seasonality is directly modelled in the conditional duration function, the nonparametric approach provides insights into the time-varying nature of the dynamics in the model that the standard procedures of deseasonalization may lead one to overlook.

Keywords: nonparametric, ACD, trade durations, local-linear, intraday seasonality.
JEL Classification: C14, C58, G17

[*]CREA, University of Luxembourg.

[§]DISES and CELPE, University of Salerno. Corresponding author. Address: Dipartimento di Scienze Economiche e Statistiche (DISES), Campus di Fisciano, Università degli Studi di Salerno, Via Giovanni Paolo II 132, 84084 Fisciano (SA), Italy. Email: fgalli@unisa.it, Phone: +39 089 96 8154.

1 Introduction

Waiting times between particular financial events, such as trades, quote updates or volume accumulation, are an important object of analysis in the econometrics of financial market microstructure. The statistical inspection of the durations between these events reveals the presence of a series of stylized facts (for instance, clustering and overdispersion) that are rather common features in financial data. For instance, they can be compared with the clustering and fat tails displayed by the time-varying conditional variance of financial returns. The traditional econometric approach to duration analysis therefore needs to be extended to be able to fit and reproduce these peculiarities.

To this end, the autoregressive conditional duration (ACD) model, originally introduced by Engle and Russell (1998), combines elements from the ARCH literature and of duration analysis. The main structure of this model is composed of a random variable (the so-called baseline duration), the distribution of which follows a law characterized by a positive support (such as an exponential or a Weibull), multiplied by a deterministic conditional duration, which in the seminal specification was a linear function of lagged values of the observations and of the conditional duration itself.

A rich family of parametric extensions followed this first specification of the ACD model. These contributions develop along two main lines: the functional form of the time-varying conditional duration mean and the distribution of the innovations of the conditional duration. Among the first line of extensions (which abound in the literature), consider the log-ACD proposed by Bauwens and Giot (2000), where the conditional mean function takes an exponential form, the asymmetric ACD, by Bauwens and Giot (2003), characterized by the presence of a threshold in the conditional mean, and the Box-Cox transformation proposed by Fernandes and Grammig (2006). Other interesting extensions are those of Ghysels et al. (2004) and Bauwens and Veredas (2004) who introduce an element of randomness in the conditional mean, which in the previous specifications was deterministically modelled. For a review of several of these model variants, see Pacurar (2008) or Hautsch (2011).

The second line of extensions addresses the modelling of the distribution of the conditional duration, suggesting laws characterized by higher degrees of parametrization and generality. Among the most commonly adopted densities are the Weibull, the gamma, the lognormal, the Burr (encompassing the Weibull), the generalized gamma (encompassing the Weibull and the Gamma) and the generalized F (encompassing the Burr). Combinations of distributions have also been advanced (see De Luca and Zuccolotto (2006) or Luca and Gallo (2008)).

In the ACD literature, this variety of parametric specifications for the conditional mean and distribution has only partially been matched by attempts to provide semiparametric expressions for the conditional mean, which have the advantage of being robust to misspecification. A number of semiparametric point process specifications have been proposed; see, for instance, Saart et al. (2015), Brownlees et al. (2012) or Gerhard and Hautsch (2007).

The aim of this work is to introduce an even more general, fully nonparametric form of the ACD family model, where the conditional mean is expressed as a generic function of the lagged observation and of its own past and is nonparametrically estimated. Bauwens et al. (2004) note that more complex specifications of the distribution of the innovations do not seem to provide substantial improvements in the goodness of fit, and thus we believe that a more flexible definition of the conditional mean function could provide improvements in fitting the data.

The main difficulty of estimating ACD models in a fully nonparametric way resides in the unobservability of one or some regressors. To overcome this difficulty, various solutions have been proposed in the literature on GARCH models, which share many commonalities with ACD models. Hafner (1998), proposes replacing the unobservable regressor with a function of several lagged values of the observations. We can understand this approach as an approximation of a GARCH(p,q) by an ARCH(∞) model. This approach is straightforward to implement, but because of the large number of regressors, it is computationally heavy and suffers severely from the curse of dimensionality. Another interesting solution comes from Franke and Muller (2002) and Franke et al. (2004), who employ a deconvolution kernel estimator, which relies heavily on the normality of the innovations (meaning that this approach would hardly be extendable to an ACD framework) and has a rather slow rate of convergence. The iterative scheme proposed by Bühlmann and McNeil (2002) is an approach that naturally adapts to ACD models. Under a central, and albeit rather restrictive, contraction hypothesis, the authors of the latter work show that the estimator is consistent and has a rate of convergence equal to that of a usual bivariate nonparametric regression technique. We thus expect this estimator to perform better than the deconvolution kernel.

The advantages of the Bühlmann and McNeil (2002) approach are, in principle, rather significant. As do all nonparametric methods, it only imposes very mild assumptions on the function to be estimated, almost entirely eliminating the risk of incorrect specification. However, its main cost is that the exact role played by an independent variable in the model cannot be summarized in a single vector of parameters, and this limits the scope for inference. The main finding of this paper is that the nonparametric ACD estimator improves, albeit not dramatically, on the forecasting precision of the linear specification. More interestingly, close inspection of the conditional mean surfaces that we obtain from the nonparametric analysis can provide valuable information on the possibly nonlinear structure of the conditional mean function and on how to take into account the seasonality in the data introduced by the time-of-the-day variable.

The outline of this work is as follows: Section 2 introduces the nonparametric estimator for financial duration. In Section 3, a Monte Carlo experiment compares the performance of the nonparametric estimator and of the maximum likelihood parametric estimator. In Section 4, we estimate on a financial dataset that is commonly used in the ACD literature and perform some forecast accuracy comparisons. Section 4 also presents an evaluation of the shock impact curve calculated on the basis of a nonparametric estimation and the results of joint estimation of the conditional duration and of the seasonality effects. Section 5 concludes.

2 The Theoretical framework

2.1 The Model

We introduce in this section the ACD model in the form that can be usually found in the literature and then rewrite it in a way that allows us to estimate it nonparametrically. Let $\{X_t\}$ be a nonnegative stationary process adapted to the filtration $\{\mathcal{F}_t,\, t \in \mathbb{Z}\}$, with $\mathcal{F}_t = \sigma(\{X_s; s \leq t\})$, such that

$$X_t = \psi_t \epsilon_t,$$

$$\psi_t = f(X_{t-1}, \ldots, X_{t-p}, \psi_{t-1}, \ldots, \psi_{t-q}), \quad (1)$$

where $p, q \geq 0$ and $\{\epsilon_t\}$ is an *iid* nonnegative process with mean 1 and a finite second moment. We assume $f(\cdot)$ to be a strictly positive function. Since $f(\cdot)$ is $\mathcal{F}_{t-1}$-measurable, we have that $E(X_t \mid \mathcal{F}_{t-1}) = \psi_t$, i.e., ψ_t is the time-varying conditional mean of the process. We focus on the case in which $p = q = 1$, this restriction being widely justified by empirical works. Several parameterizations of (1) have been introduced, the first being the linear specification:

$$\psi_t = \omega + \alpha \psi_{t-1} + \beta X_{t-1}. \quad (2)$$

More complex specifications followed (2), allowing for nonlinearity in the response of the conditional mean to the realizations of X_t or in the autoregressive part. In our setup, $f(\cdot)$ is allowed to be any function of the past realizations X_{t-1} and of the lagged conditional mean ψ_{t-1}. Moreover, parametric specifications of the ACD family often make use of highly parameterized functions for the distributions of the innovations ϵ_t, while here we only require the mean of the ϵ_t to be one and the variance to be finite. We expect our estimation to outperform parametric models in the case in which the 'real' f shows some accentuated nonlinearity as in the *threshold* models:

$$\psi_t = h(X_{t-1}) + \sum_i \beta_i \mathbb{I}_{[X_{t-1} \in B_i]} \psi_{t-1},$$

where B_i are disjoint subsets of $\mathbb{R}_+$ and $h(x)$ is again a strictly positive function.

To estimate f, we rewrite (1) in the additive form:

$$X_t = f(X_{t-1}, \psi_{t-1}) + \eta_t, \tag{3}$$

$$\eta_t = f(X_{t-1}, \psi_{t-1})(\epsilon_t - 1).$$

The process $\{\eta_t\}$ is a white noise, as $E(\eta_t) = E(\eta_t|\mathcal{F}_{t-1}) = 0$ and $E(\eta_s\eta_t) = E[E(\eta_s\eta_t|\mathcal{F}_{t-1})] = 0$ for $s < t$. The conditional variance of X_t is $\mathrm{Var}(X_t|\mathcal{F}_{t-1}) = f^2(X_{t-1}, \psi_{t-1})(E(\epsilon_t^2) - 1)$. Thus, formally, $f(X_{t-1}, \psi_{t-1})$ could be estimated by regressing X_t on $f(X_{t-1}, \psi_{t-1})$. In practice, the ψ_t are unobserved variables. To overcome this problem, we adapt the recursive algorithm suggested by Bühlmann and McNeil (2002).

2.2 The estimation procedure

The algorithm is built as follows. Let $\{x_t; 1 \leqslant t \leqslant n\}$ be realizations[1] of the process (1), with $p = q = 1$. The steps of the algorithm are indexed by j.

Step 1. Choose the starting values for the vector of the n conditional means. Index these values by 0: $\{\psi_{t,0}\}$. Set $j = 1$.

Step 2. Regress nonparametrically $\{x_t; 2 \leqslant t \leqslant n\}$ on $\{x_{t-1}; 2 \leqslant t \leqslant n\}$ and on the conditional means computed in the previous step: $\{\psi_{t-1,j-1}; 2 \leqslant t \leqslant n\}$, to obtain an estimate $\hat{f}_j$ of f.

[1]We differentiate the population random variables X_t from the actual realizations x_t.

Step 3. Compute $\{\hat{\psi}_{t,j} = \hat{f}_j(x_{t-1}, \hat{\psi}_{t-1,j-1}); 2 \leqslant t \leqslant n\}$; remember to choose some sensible value for $\hat{\psi}_{1,j}$, which cannot be computed recursively.

Step 4. Increment j, and return to step two to run a new regression using the $\{\psi_t\}$ computed in Step 3.

Averaging the estimates of the last steps of the algorithm, when it becomes more stable, further improves the estimation procedure.

We refer to Bühlmann and McNeil (2002) for a justification and theoretical discussion of the algorithm. We state here from the main theorem of that paper, which provides the convergence rates of the estimates delivered by the algorithm. We first need some notation. Henceforth, $\|Y\|$ denotes the $\mathcal{L}_2$ norm of Y: $\|Y\|^2 = \mathbb{E}(Y^2)$. Let

$$\tilde{f}_{t,j}(x, \psi) = E(X_t | X_{t-1} = x, \hat{\psi}_{t-1,j-1} = \psi),$$

$$\tilde{\psi}_{t,j} = \tilde{f}_{t,j}(x_{t-1}, \hat{\psi}_{t-1,j-1});$$

That is, $\tilde{\psi}_{t,j}$ is the true conditional expectation of X_t given the value of $\hat{\psi}_{t-1,j-1}$ estimated at the previous step of the algorithm. Thus, the quantity

$$\Delta_{t;j,n} \equiv \tilde{\psi}_{t,j} - \hat{\psi}_{t,j}, \; j = 1, 2, \ldots, t = j+2, \ldots, n,$$

gives us the estimation error introduced at the j-th step solely due to the estimation of f. In nonparametric language, $\|\Delta\|$ is the stochastic component of the risk of the estimator $\hat{\psi}_{t,j}$ of $E(X_t | X_{t-1}, \psi_{t-1,j-1})$.

Theorem 1 (Theorems 1 and 2 in Bühlmann and McNeil (2002)) *Assume that*

1. $\sup_{x \in \mathbb{R}} |f(x, \psi) - f(x, \varphi)| \leqslant D|\psi - \varphi|$ *for some* $0 < D < 1$, $\forall \psi, \; \varphi \in \mathbb{R}_+$.

2. $E|\psi_t|^2 \leq C_1, \quad E|\psi_{t,0}|^2 \leq C_2, \quad \max_{2 \leq t \leq n} E|\hat{\psi}_{t,0}|^2 \leq C_3, \quad C_{1,2,3} < \infty,$
$\|\psi_j - \psi_{j,0}\| < \infty, \quad \|\hat{\psi}_{j,0} - \psi_{j,0}\| < \infty \;\; \forall j.$

3. $E(\{\tilde{\psi}_{t,j} - \psi_{t,j}\}^2) \leqslant G^2 E(\{\hat{\psi}_{t-1,j-1} - \psi_{t-1,j-1}\}^2)$ *for some* $0 < G < 1$, *for* $t = j+2, j+3, \ldots$ *and* $j = 1, 2, \ldots$

4. $\Delta_n \doteq \sup_{j \geqslant 1} \max_{j+2 \leq t \leq n} \|\Delta_{t;j,n}\| \to 0$, *as* $n \to \infty$ *for* $j = 1, 2, \ldots, t = j+2, \ldots, n$.

Then, if $\{X_t\}_{t\in\mathbb{N}}$ *is as in (1) with* $p = q = 1$ *and choosing* $m_n = C\{-\log\Delta_n\}$,

$$\max_{m_n+2\leqslant t\leqslant n} \|\hat{\psi}_{t,m_n} - \psi_t\| = O(\Delta_n), \text{ as } n \to \infty.$$

The theorem tells us that if all the assumptions hold, then the upper bound on the quadratic risk of the estimates of the $\{\psi_t\}$ is of the same order as Δ_n, that is, the error of a one-step nonparametric regression to estimate $\psi_{t,j}$ from (x_{t-1}, ψ_{t-1}). That is, in a bivariate nonparametric regression with an appropriate choice of the kernel function and of the smoothing parameter and assuming, for instance, that $f(x, \psi)$ is twice continuously differentiable, the convergence rates are $O(n^{-1/3})$. The authors suggest as a practical rule $m_n \sim 3\log(n)$.

We briefly discuss the assumptions of the theorem. For further insights, refer to Bühlmann and McNeil (2002). First, let us write

$$\|\hat{\psi}_{t,j} - \psi_t\| \leq \|\hat{\psi}_{t,j} - \tilde{\psi}_{t,j}\| + \|\tilde{\psi}_{t,j} - \psi_{t,j}\| + \|\psi_{t,j} - \psi_t\|. \tag{4}$$

The first two components of the risk (4) are the usual quadratic risk of an estimator $\hat{\psi}_{t,j}$ of $\psi_{t,j}$. The additional component $\|\psi_{t,j} - \psi_t\|$ is included because we do not observe ψ_t. Assumption 1 controls this last part of the risk. If there were no estimation error in passing from one step of the algorithm to the next, Assumption 1 combined with the recursive form of the algorithm would be sufficient to ensure the convergence of $\psi_{t,m}$ to the true value ψ_t. Assumption 2 is technical and needed to give an upper bound to the estimation error of the first step of the algorithm. Assumption 3 is used to control the second component of (4). It can be written in the following way: $\|\tilde{\psi}_{t,j} - \psi_{t,j}\| = \|E(X_t|X_{t-1}, \hat{\psi}_{t-1,j-1}) - E(X_t|X_{t-1}, \psi_{t-1,j-1})\|$ so Assumption 3 is a contraction property of the conditional expectation with respect to $\|\hat{\psi}_{t-1,m-1} - \psi_{t-1,j-1}\|$. It is again a technical property that Bühlmann and McNeil are obliged to impose on the process to prove the consistency of the estimates delivered by the algorithm. Assumption 4 bounds the first term of (4). It gives an upper bound to the one-pass regression of X_t on X_{t-1} and $\psi_{t-1,j-1}$.

2.3 The practical implementation

In our application to simulated and real data, we use the following settings. For the initial values of the $\{\psi_t\}$ to use in the first step of the algorithm, we choose a vector of random draws from an exponential distribution with expectation equal to the unconditional mean of the data series $\{x_t\}$. Bühlmann and McNeil (2002) suggest using a parametric estimate, which is to be improved in the following steps of the algorithm. Since our goal is to compare parametric with nonparametric estimates, we believe that challenging the nonparametric procedure by providing dull initial values would make the competition fairer and the results more reliable. Moreover, we noticed that the algorithm yields essentially the same outcome in both cases, that is, when providing the random draws or the parametric estimate as starting values. We can say that the algorithm is quite insensitive to changes in the choice of the initial values, provided that these are sensible.

As far as the choice of the nonparametric technique is concerned, we use the locally weighted smoother (LOESS)[2], developed by Cleveland (1979). Hastie and Tibshirani (1990) provide a good introduction to this nonparametric technique. The main idea is to perform a local polynomial least squares fit in the neighbourhood of a point x_0. The design points entering the local regression are chosen as in the k-nearest neighbour method, and the value of the function at each design point is weighted with a tri-cube kernel. The degree of smoothing is determined by the percentage of the data points (also called the *span*) entering the local regression. Following the suggestion of Cleveland, we fit a local polynomial of order 1.

The reliance on nearest neighbours as an alternative to a symmetric, area-based (as in the case of standard kernel smoothing) criterion as a method of selecting of the neighbourhood of interest seems to be particularly useful given the particular features of our data. In our application, the predictors are the lagged durations X_{t-1} and the conditional means at the *j-th* step of the algorithm $\psi_{t-1,j}$. As Figure 1 shows, the predictor variables form a non-uniform random design in the $x\psi$ plane and are visibly more dense in the region next to the axes, drawing in the $x\psi$ plane a"falling star" pattern. We therefore need a method that is capable of adapting, in the neighbourhood of interest, to the local density of the predictors. Moreover, the bias of the local linear estimator does not depend on the marginal density of the predictors, thereby addressing the boundary in the domain of the regression function (both regressors are nonnegative). [Figure 1 about here] Following Hastie and Tibshirani (1990), we use a generalized cross validation (GCV) criterion to choose the span parameter. It can be proved that minimizing GCV is asymptotically equivalent to minimizing the mean square error of the regression. *[Figure 1 about here]* At each loop (and in the final averaged smoothing), we therefore use the span that minimizes

[2]We use the R implementation of LOESS.

the quantity

$$GCV = \frac{n - \sum_{t=1}^{n}(x_t - \hat{\psi}_{t,j})^2}{(n - trL)^2}, \tag{5}$$

where n is the sample size, $\hat{\psi}_{t,j}$ is the predictor of x_t corresponding to the loop j, and L is the smoother matrix, that is, the matrix that premultiplied to the vector of observed values $\{x_t\}$ yields the estimates. The quantity trL, the trace of the matrix L, plays a role analogous to the number of degrees of freedom in a standard linear regression.

Finally, the error appearing in Equation (3) is heteroskedastic, thus calling for a weighted fit. Obviously, the true weights would depend on the values ψ_t that we are estimating. We therefore replace them in each loop with the estimates of the conditional durations that were computed in the previous iteration.

3 Estimation of simulated processes

In this section, we assess the performance of the nonparametric specification via a comparison with the estimates of a linear ACD model on different simulated series. The first simulated series is characterized by an asymmetry in the conditional mean equation, which has the following form:

$$f(x_{t-1}, \psi_{t-1}) = 0.2 + 0.1x_{t-1} + (0.3\mathbb{I}_{[x_{t-1} \leqslant 0.5]} + 0.85\mathbb{I}_{[x_{t-1} > 0.5]})\psi_{t-1}, \tag{6}$$

and the conditional duration is Weibull distributed, with the scale parameter such that its mean is equal to one. The sizes of the generated samples are 1000, 5000 and 10000 observations. We simulate 50 series from model (6). The simulated series are estimated by ML with a linear ACD(1,1) specification and by the nonparametric smoother described in Section 2.2 with 10 basic iterations, and we perform a final smoothing based on the arithmetic mean of the last $K = 4$ iterations. The performance of the parametric and nonparametric estimators is compared by computing two widely used measures of estimation errors. The first one is the mean square error (MSE) based on a quadratic loss function:

$$MSE = \frac{1}{nM} \sum_{l=1}^{M} \sum_{i=1}^{n} (\hat{\psi_{il}} - \psi_{il})^2, \tag{7}$$

where $i = 1, \ldots, n$ denotes the $i-$th estimated conditional mean within the series, and $l = 1, \ldots, M = 50$ labels the 50 series simulated from Equation (6).
The second measure is the mean absolute error (MAE):

$$MAE = \frac{1}{nM} \sum_{l=1}^{M} \sum_{i=1}^{n} |\hat{\psi}_{il} - \psi_{il}|. \tag{8}$$

We perform the same type of analysis on series simulated from a standard ACD(1,1) model with no asymmetric component in the specification of the conditional mean equation. The functional form is of the conditional mean is

$$f(x_{t-1}, \psi_{t-1}) = 0.1 + 0.1x_{t-1} + 0.75\psi_{t-1}, \tag{9}$$

and the conditional distribution and the sample size are the same as in the first group of simulated series. The settings of the parametric and nonparametric estimators do not change from the first example. In particular, we estimate a parametric ACD(1,1) model, which is now correctly specified. *[Figures 2 and 3 about here]*

Figure 2 displays in a 200 data window an example of the evolution of the simulated ψ_t (hence the true dgp), and of those estimated parametrically and nonparametrically. Note that the parametric estimator seems to overreact and it yields too large estimates for a small number of points. Figure 3 shows the surfaces $f(x_{t-1}, \psi_{t-1})$ generated from the nonlinear model in Equation (6) in their simulated version and in that estimated nonparametrically. The abrupt change in the slope of $f = \hat{\psi}_t(x_{t-1}, \psi_{t-1})$ as a function of ψ_{t-1} for $x \leq 0.5$ and $x > 0.5$ is quite visible in the bottom part of the estimated surface (near the origin), where the data are very dense and the bandwidth is rather small. Farther from the origin, observations in the support become more sparse, and the result is somewhat more smoothed. In any case, it is clear that the slope increases as x increases. To complete the analysis on this group of simulations, we give in Table 1 an example of how the span parameter minimizing the generalized cross-validation criterion evolves with the steps of the algorithm. It is clear that the main loop converges quite early to a stable value. Since the value of the span depends on the distribution of the predictors, a stable value of this parameter indicates that there are no major changes in the distribution of the $\{\psi_t\}$, suggesting that the algorithm has converged. *[Tables 1 and 2 about here]*

Table 2 compares the performance of the nonparametric and parametric estimators in terms of MSE and MAE. In the case of the nonlinear model, both statistics show that the the nonparametric estimator outperforms the parametric estimator. Both MSE and MAE decrease at each loop and decline further after the final averaging. When instead the series is simulated starting from the linear ACD model, the parametric estimator is correctly specified and, unsurprisingly, obtains the lowest MSE and MAE. The nonparametric estimator selects a large span. This means that almost all data points enter the local linear regression, making the LOESS smoother behave more like a standard linear regression. We do not show the charts of the reconstructed surfaces in this case, as they appear to be rather uninformative and flat.

4 Estimation on a financial data set

In this section, we evaluate the performance of the nonparametric specification of the ACD model on a financial dataset. The estimated series consists in a set of trade, volume and price durations of the following stocks traded in 1997 on the New York Stock Exchange: Boeing, Disney and IBM.

4.1 Evidence on deseasonalized data

As noted in the seminal paper by Engle and Russell (1998), there is a strong intraday seasonality in tick-by-tick data, as durations have a tendency to be shorter on average at the beginning and close of trading sessions. It is therefore common to remove the seasonal pattern by means of a nonparametric regression of raw durations on the time of the day and to fit the adjusted data. In this round of estimations, we employ the simplified deseasonalization technique used both in Engle and Russell (1998) and Bauwens and Giot (2000). In a first step, one estimates the cyclical component. This is done by averaging the durations over 30-minute intervals for each day in our sample. The average value for each of the thirteen 30-minute bins (from 9h30 to 16h) is the value of the cyclical component at the mid point of each interval. We obtain the value of the cyclical component as a smooth function of the time-of-day variable by interpolating the 30 points using a cubic spline. Figures 4, 5 and 6 present the surfaces estimated nonparametrically with 10 loops and a final average of the last 4. *[Figures 4, 5 and 6 about here]* The visual analysis suggests some conclusions. First, some nonlinearity is present in almost all surfaces, although it never reaches the extreme features of the discontinuity as in the data simulated in the previous section. Second, for some datasets, notably Disney volume and, to a lesser degree, Boing price and IBM trade durations, the surface is almost linear. This is also supported by the very high values of the spans minimizing the generalized cross-validation (0.999, 0.995 and 0.981, respectively) criterion. In other datasets, the nonlinearities appear more marked. Third, in these cases, the real data-generating process in the conditional mean equation seems to place a low weight on the lagged observation X_{t-1}, and the dependence of ψ_t on ψ_{t-1} appears to diminish as X grows. This is a reasonable feature. Let us consider a regime-switching model, which is dependent on whether the market speeds up or slows down. When the market speeds up (short durations), we are more likely to observe bunching in the data, that is, there is a larger autocorrelation component in the conditional mean equation and hence a stronger dependence of ψ_t on ψ_{t-1}. When the market cools down, we observe less clustering in the duration data, and the conditional duration in the conditional mean is weaker.

We now proceed to compare the forecasting performance of the nonparametric and parametric estimators with an in-sample forecasting experiment. Using the result of the estimation, we use the conditional duration mean for time t estimated both parametrically and nonparametrically as a forecast for the duration observed at time $t + 1$. We display the averages of the squared (MSE) and absolute (MAE) forecast errors in Table 3, along with the percentage gains obtained by nonparametric estimation. We also test for the significance of differences in forecast accuracy by performing a Diebold and Mariano (2002) test with an absolute loss function for MAE and a squared one fore MSE. In the table, percentage gains between parentheses are not significant at the 5% level. *[Table 3 about here]*

The nonparametric estimator seems to yield an improvement in the one-step-ahead forecast accuracy in almost all cases, although only a few of them appear to be significant. Overall, the differences tend to be small in magnitude, but whenever there is a statistically significant difference, it is in favour of the nonparametric forecast.

4.2 Empirical application: evaluation of the shock impact curve

Engle and Russell (1998) note that the ACD model has the tendency to overpredict after very long or very short durations. This would make a model with a concave shock impact function (that in the ACD is linear) better suited as a forecasting tool. The desirability of this feature has been explicitly acknowledged in the subsequent literature, and the Box-Cox transformation-based ACD family of specifications proposed by Fernandes and Grammig (2006) indeed shows concavity in the shape of the curve. The model proposed in this paper does not have an *a priori* form for the shock impact curve because, depending on the resulting estimated surface, the response of the expected conditional duration to a shock in the baseline duration can vary. As an experiment, we estimate our model with the same data (quote durations for the IBM stock) used in Fernandes and Grammig (2006) and compute the resulting shock impact curve by fixing ψ_{i-1} at 1 and letting ϵ_{t-1} vary to evaluate its impact on the value of the expected conditional duration ψ_t.

Figure 7 displays the curve resulting from the nonparametric estimation along with that resulting from the estimation of a parametric ACD model. The result seems to confirm the hypothesis of Engle and Russell (1998). The nonparametric estimator in fact seems to benefit from its greater flexibility and to produce a slightly concave response curve. We can also notice that the concavity resulting from our estimator seems less pronounced than that observed in the estimations of the modes proposed by Fernandes and Grammig (2006), at least on the basis of a simple visual evaluation. *[Figure 7 about here]*

4.3 Inclusion of the time of the day as a covariate

Maximum likelihood estimation of linear ACD models on deseasonalized data can be seen as a *de facto* semiparametric two-step procedure, where a first nonparametric deseasonalization is followed by the fully parametric estimation of the actual ACD model proper.[3] Yet, this standard practice does not take into account a possible time-dependence of the ACD parameters. The risk of this approach is the possibility of missing some of the information contained in the data and therefore providing suboptimal fit and forecasts if the seasonality-affected observations are the actual object of the analysis.

In this subsection, we exploit the flexibility of the nonparametric ACD estimator and include in the formula for the conditional duration the time of the day as an explanatory variable. In this setup, Equation (3) becomes

$$D_t = f(D_{t-1}, \psi_{t-1}, \tau_t) + \eta_t, \tag{10}$$

with

$$\eta_t = f(D_{t-1}, \psi_{t-1}, \tau_t)(\epsilon_t - 1),$$

where τ_t is the time of the day corresponding to the t-th observation, expressed in seconds from the beginning of the trading session, and D_t is the t-th un-desesonalized duration.

The estimation of the nonparametric ACD as described in Section 2 does not require

[3]The literature offers some exceptions to this practice: Rodríguez-Poo et al. (2008), Brownlees and Gallo (2011) and Bortoluzzo et al. (2010).

major changes in the recursive procedure, as τ_t is fully observable and can be treated as an exogenous covariate. The practical implementation is also straightforward since the LOESS function in R can accommodate a set of three explanatory variables instead of the two used with deseasonalized data. To make the forecasts comparable, we multiply the parametric ACD one-step-ahead predictions by the same values of the estimated time-of-the-day effect used to deseasonalize the raw data before estimation. *[Table 4 about here]*

Table 4 reports a comparison between forecast absolute and quadratic errors of the parametric and nonparametric estimators. As in the case of deseasonalized data, many differences in forecasting accuracy do not seem significant at the 5% level for price and volume durations. By comparing Tables 3 and 4, we can see that the performance of the nonparametric estimator improves for the figures related to the trade duration. In the case of IBM trade durations, the percentage decreases in both MSE and MAE are statistically significant, according to the Diebold-Mariano test. Although the absolute differences remain small, the general picture is in favour of better performance by the nonparametric estimator. As in the deseasonalized dataset, whenever the forecasting difference is statistically significant at the 5% level, it is the nonparametric estimator that is more accurate.

A visual account of the evolution of the nonparametrically estimated surface is provided by Figures 8, 9 and 10, which display contour plots of the estimated surface of Boing price, volume and trade durations computed at 30-minute intervals from 9:30 am to 4:00 pm. The shape of the estimated conditional duration clearly varies during the day. Moreover, the contour lines seemingly tend to shift gradually from one time period to the following period, suggesting a clear pattern of time dependency of the parameters of the model. Analogous time patterns of the estimated surface are present for all other durations of all stocks. This evidence clearly suggests that the standard practice of separating the estimations of the seasonality component and the conditional duration function risks missing the opportunity to exploit significant information present in the data. Even in a fully parametric specification, it may be therefore beneficial to include some form of interaction between time of the day and the other model parameters. *[Figures 8, 9 and 10 about here]*

5 Conclusion

The nonparametric specification of the ACD model encompasses most of the parametric forms thus far introduced to study high-frequency transaction data, the only exception being constituted by models with two stochastic components, such as the SCD. The model can easily be estimated by standard nonparametric techniques, although a recursive approach is necessary to address the fact that some regressors are not directly observable. The simulated examples show that in the presence of asymmetry in the specification of the conditional mean equation, the nonparametric estimator easily outperforms the symmetric parametric estimator. An estimation on a financial data set shows a marginally better performance of the nonparametric model in terms of forecasting power. When we include in the model time-of-the-day seasonality and estimate it jointly with the conditional duration surface, the gain in forecast accuracy from using the nonparametric estimator marginally improves for some stocks. Nevertheless, although not providing a specification test for parametric models, the nonparametric analysis can be useful as a benchmark in choosing the right parametric specification. The graphical study of the dependence of the conditional mean on its lags can provide valuable information on which type of parametric specification to choose. Including the time-of-the-day variable in the nonparametric analysis can provide valuable information on which deseasonalization procedure to use or suggest a possible time variation of the parameters of the parametric specification.

Finally, we discuss what could be a further use of this estimation strategy in empirical analysis. We believe that it could be beneficial to include in the regression of market microstructure variables, such as volume, prices, bid-ask spread or, when available, dummies for the arrival of news in the market. These variables have often been used in ACD estimations, but their impact on the frequency of trading is not always clear, and they could be easily the subject of a nonparametric or, eventually, a semiparametric analysis. We leave this development for further research.

References

Bauwens, L. and Giot, P. (2000). The logarithmic ACD model: an application to the bid-ask quote process of three nyse stocks. *Annales d'Economie et de Statistique*, 60(Special Issue, "Financial Market Microstructure"):117–149.

Bauwens, L. and Giot, P. (2003). Asymmetric ACD models: introducing price information in acd models with a two state transition model. *Empirical Economics*, 28(4):709–731.

Bauwens, L., Grammig, J., Giot, P., and Veredas, D. (2004). A comparison of financial duration models with density forecasts. *International Journal of Forecasting*, 20:589–609.

Bauwens, L. and Veredas, D. (2004). The stochastic conditional duration model: a latent factor model for the analysis of financial durations. *Journal of Econometrics*, 119(2):381–412.

Bortoluzzo, A. B., Morettin, P. A., and Toloi, C. M. (2010). Time-varying autoregressive conditional duration model. *Journal of Applied Statistics*, 37(5):847–864.

Brownlees, C. T., Cipollini, F., and Gallo, G. M. (2012). Multiplicative error models. *Handbook of Volatility Models and Their Applications. New Jersey: Wiley.*

Brownlees, C. T. and Gallo, G. M. (2011). Shrinkage estimation of semiparametric multiplicative error models. *International Journal of Forecasting*, 27(2):365–378.

Bühlmann, P. and McNeil, A. J. (2002). An algorithm for nonparametric GARCH modelling. *Computational Statistics and Data Analysis*, 40:665–683.

Cleveland, W. (1979). Robust locally weighted regresson and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836.

De Luca, G. and Zuccolotto, P. (2006). Regime-switching Pareto distributions for acd models. *Computational Statistics & Data Analysis*, 51(4):2179–2191.

Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1).

Engle, R. and Russell, J. R. (1998). Autoregressive conditional duration: a new approach for irregularly spaced transaction data. *Econometrica*, 66:1127–1162.

Fernandes, M. and Grammig, J. (2006). A family of autoregressive conditional duration models. *Journal of Econometrics*, 130(1):1–23.

Franke, J., Hardle, W., and Kreiss, J. (2004). Non parametric estimation in a stochastic volatility model. *Recent advances and trends in Nonparametric Statistics.*

Franke, J. H. and Muller, M. (2002). Nonparametric estimation of GARCH processes. *Applied Quantitative Finance*, pages 367–383.

Gerhard, F. and Hautsch, N. (2007). A dynamic semiparametric proportional hazard model. *Studies in Nonlinear Dynamics & Econometrics*, 11(2).

Ghysels, E., Gouriéroux, C., and Jasiak, J. (2004). Stochastic volatility durations. *Journal of Econometrics*, 119(2):413–433.

Hafner, C. M. (1998). Estimating high-frequency foreign exchange rate volatility with nonparametric arch models. *Journal of Statistical Planning and Inference*, 68(2):247–269.

Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*. Chapman and Hall.

Hautsch, N. (2011). *Econometrics of financial high-frequency data*. Springer.

Luca, G. D. and Gallo, G. M. (2008). Time-varying mixing weights in mixture autoregressive conditional duration models. *Econometric Reviews*, 28(1-3):102–120.

Pacurar, M. (2008). Autoregressive conditional duration (ACD) models in finance: A survey of the theoretical and empirical literature. *Journal of Economic Surveys*, 22(4):711–751.

Rodríguez-Poo, J. M., Veredas, D., and Espasa, A. (2008). Semiparametric estimation for financial durations. In *High frequency financial econometrics*, pages 225–251. Springer.

Saart, P. W., Gao, J., and Allen, D. E. (2015). Semiparametric autoregressive conditional duration model: Theory and practice. *Econometric Reviews*, 34(6-10):849–881.

Loop	Nonlinear	Linear
1	0.831	0.999
2	0.206	0.999
3	0.708	0.999
4	0.597	0.999
5	0.612	0.999
6	0.736	0.999
7	0.612	0.999
8	0.623	0.999
9	0.622	0.999
10	0.622	0.999
Average	0.621	0.999

Table 1: Evolution of the GCV-selected bandwidth in an estimation of a series of 5000 observations simulated from a nonlinear and linear ACD specification.

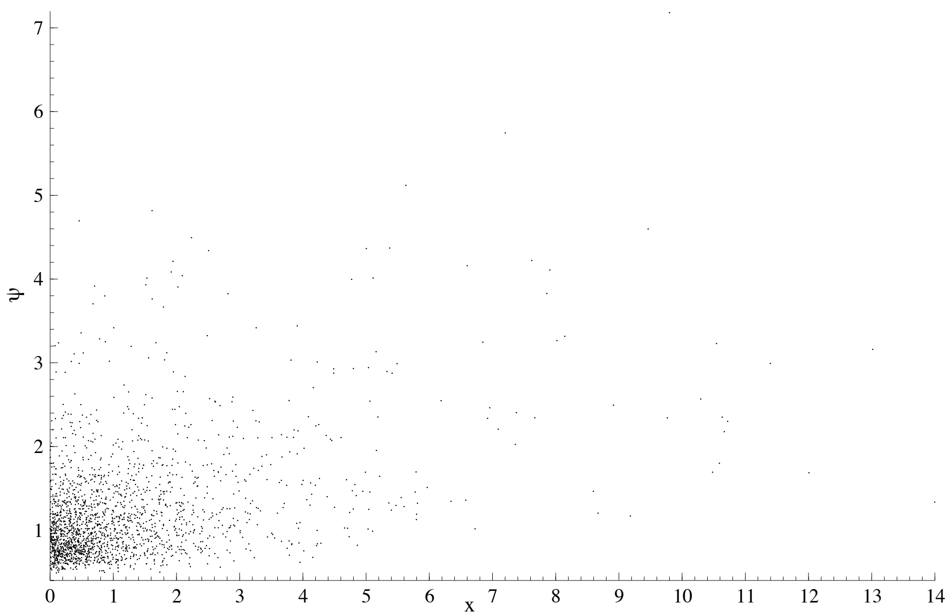


Figure 1: Scatterplot of a typical $x\psi$ domain

	1000 obs				5000 obs				10000 obs			
	Nonlinear		Linear		Nonlinear		Linear		Nonlinear		Linear	
Loop	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
1	0.060	0.120	0.0025	0.033	0.044	0.099	0.00054	0.015	0.042	0.098	0.00036	0.013
2	0.058	0.115	0.0028	0.036	0.037	0.088	0.00063	0.016	0.034	0.084	0.00039	0.013
3	0.057	0.114	0.0032	0.038	0.034	0.084	0.00072	0.017	0.030	0.078	0.00040	0.013
4	0.056	0.113	0.0036	0.040	0.032	0.082	0.00074	0.018	0.028	0.076	0.00042	0.014
5	0.056	0.113	0.0035	0.040	0.031	0.081	0.00074	0.017	0.027	0.074	0.00046	0.014
6	0.055	0.112	0.0035	0.040	0.031	0.081	0.00077	0.018	0.026	0.073	0.00045	0.014
7	0.057	0.113	0.0035	0.040	0.030	0.080	0.00078	0.018	0.026	0.073	0.00044	0.014
8	0.055	0.112	0.0036	0.040	0.030	0.080	0.00074	0.017	0.026	0.074	0.00044	0.014
9	0.056	0.113	0.0037	0.041	0.030	0.080	0.00077	0.018	0.026	0.073	0.00046	0.014
10	0.057	0.114	0.0038	0.040	0.030	0.080	0.00079	0.018	0.026	0.073	0.00049	0.014
avg	**0.054**	**0.112**	**0.0033**	**0.038**	**0.030**	**0.080**	**0.00074**	**0.017**	**0.026**	**0.072**	**0.00045**	**0.014**
par	**0.086**	**0.143**	**0.0012**	**0.025**	**0.077**	**0.138**	**0.00024**	**0.011**	**0.074**	**0.135**	**0.00016**	**0.009**

Table 2: Evolution of MSE and MAE for 50 series of 1000, 5000 and 10000 observations simulated from a nonlinear and linear simulated ACD specification.

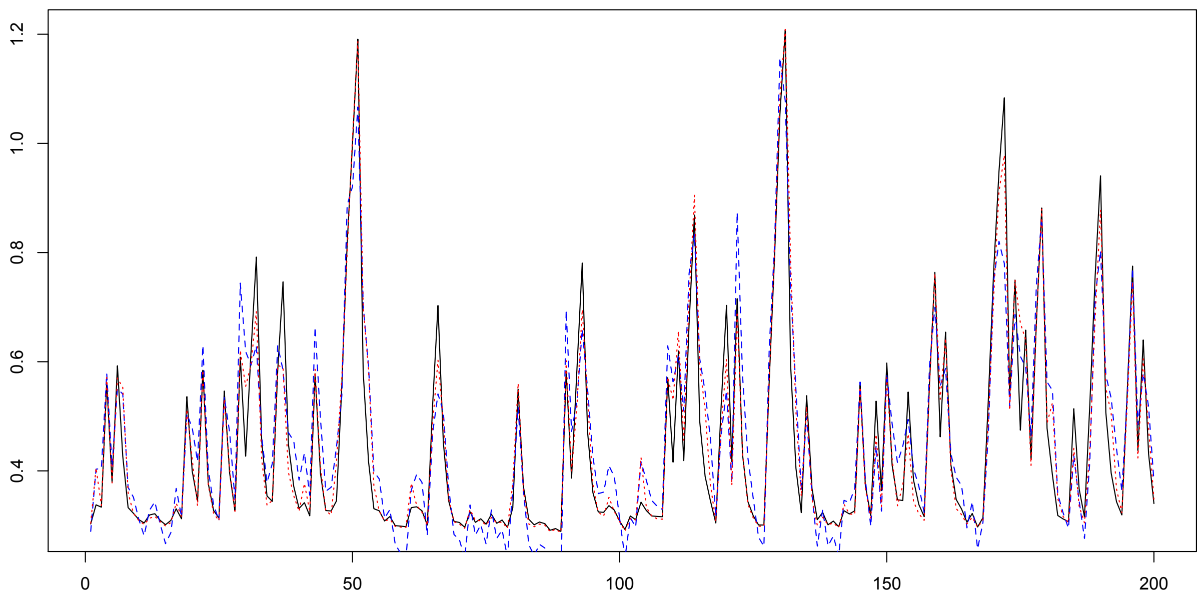


Figure 2: Nonlinear ACD, simulated conditional mean (black, solid), parametric estimate (blue, dashed) and nonparametric estimate (red, dotted).

Type	Obs	Stock	MSE Par	MSE Nonpar	MSE %	MAE Par	MAE Nonpar	MAE %
Trade	75435	**Boeing**	1.417	1.416	*(0.02)*	0.772	0.768	*0.41*
	61630	**Disney**	1.426	1.425	*(0.06)*	0.818	0.816	*(0.10)*
	136020	**Ibm**	1.240	1.240	*(0.02)*	0.742	0.741	*0.17*
	74170	**Exxon**	1.275	1.274	*(0.08)*	0.771	0.771	*(0.02)*
Price	8682	**Boeing**	1.750	1.738	*(0.58)*	0.821	0.806	*1.86*
	5985	**Disney**	1.362	1.360	*(0.11)*	0.765	0.761	*(0.50)*
	18878	**Ibm**	1.305	1.301	*(0.21)*	0.754	0.749	*(0.59)*
	12974	**Exxon**	2.243	2.107	*6.44*	0.849	0.831	*2.04*
Volume	4261	**Boeing**	0.415	0.409	*1.31*	0.483	0.478	*1.01*
	3450	**Disney**	0.355	0.356	*(-0.42)*	0.465	0.465	*(-0.09)*
	9684	**Ibm**	0.413	0.379	*(0.18)*	0.457	0.457	*(0.46)*
	5597	**Exxon**	0.413	0.412	*(0.29)*	0.488	0.485	*(0.56)*

Table 3: In-sample MSE, MAE and percentage gain of the nonparametric estimator on a set of trade, price and volume durations, the intraday seasonality of which was removed and the average of which was normalized to one. Percentage gains are in parentheses if the forecasts are not significantly different for the 5%-sized corresponding Diebold-Mariano test.

		Stats				MSE			MAE	
Type	Obs	Mean	Stdev	Stock	Par	Nonpar	%	Par	Nonpar	%
trade	75435	32	42	**Boeing**	1584	1581	*(0.22)*	25	25	*0.33*
	61630	39	50	**Disney**	2351	2351	*(0.01)*	32	32	*(0.12)*
	136020	18	22	**Ibm**	438	437	*0.22*	13	13	*0.42*
	74170	33	40	**Exxon**	1487	1485	*0.17*	25	23	*(0.06)*
price	8682	274	434	**Boeing**	160962	163029	*(-1.28)*	226	224	*(0.82)*
	5985	396	555	**Disney**	273336	263796	*3.49*	309	308	*0.44*
	18878	127	172	**Ibm**	24133	25915	*0.83*	98	97	*0.74*
	12974	184	307	**Exxon**	85822	86463	*(-0.75)*	158	161	*(-1.68)*
volume	4261	560	463	**Boeing**	147635	150132	*(-1.69)*	275	275	*(-0.22)*
	3450	690	499	**Disney**	190170	192158	*(-1.04)*	325	324	*(0.30)*
	9684	249	207	**Ibm**	28009	27879	*(0.46)*	116	115	*(0.66)*
	5597	428	343	**Exxon**	86696	87806	*(-1.28)*	212	211	*(0.34)*

Table 4: In-sample MSE, MAE and percentage gain of the nonparametric estimator on a set of trade, price and volume durations with time-of-the day entered as a regressor. Percentage gains are in parentheses if the forecasts are not significantly different for the 5%-sized corresponding Diebold-Mariano test.

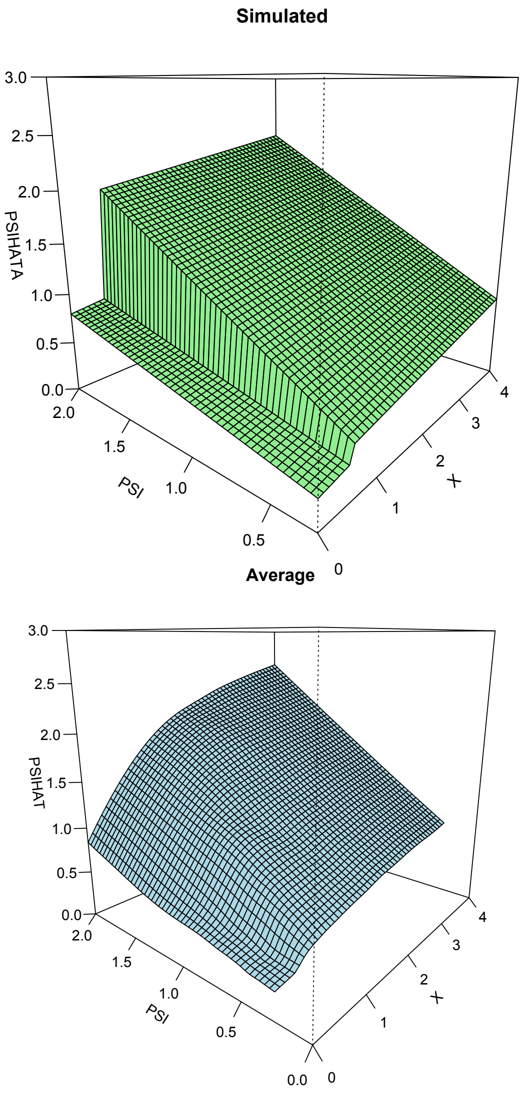


Figure 3: Nonlinear ACD, simulated and estimated surface (final average).

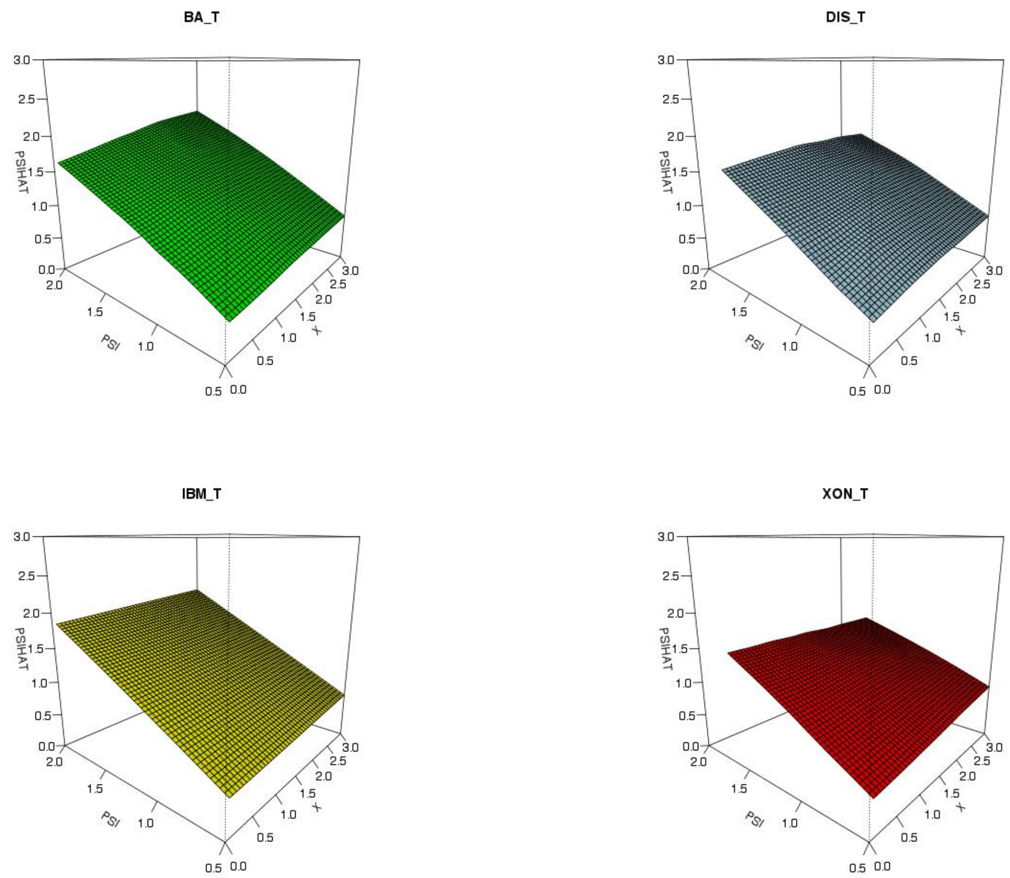


Figure 4: Estimated surfaces (final average) for trade durations of Boeing, Disney, IBM and Exxon stocks.

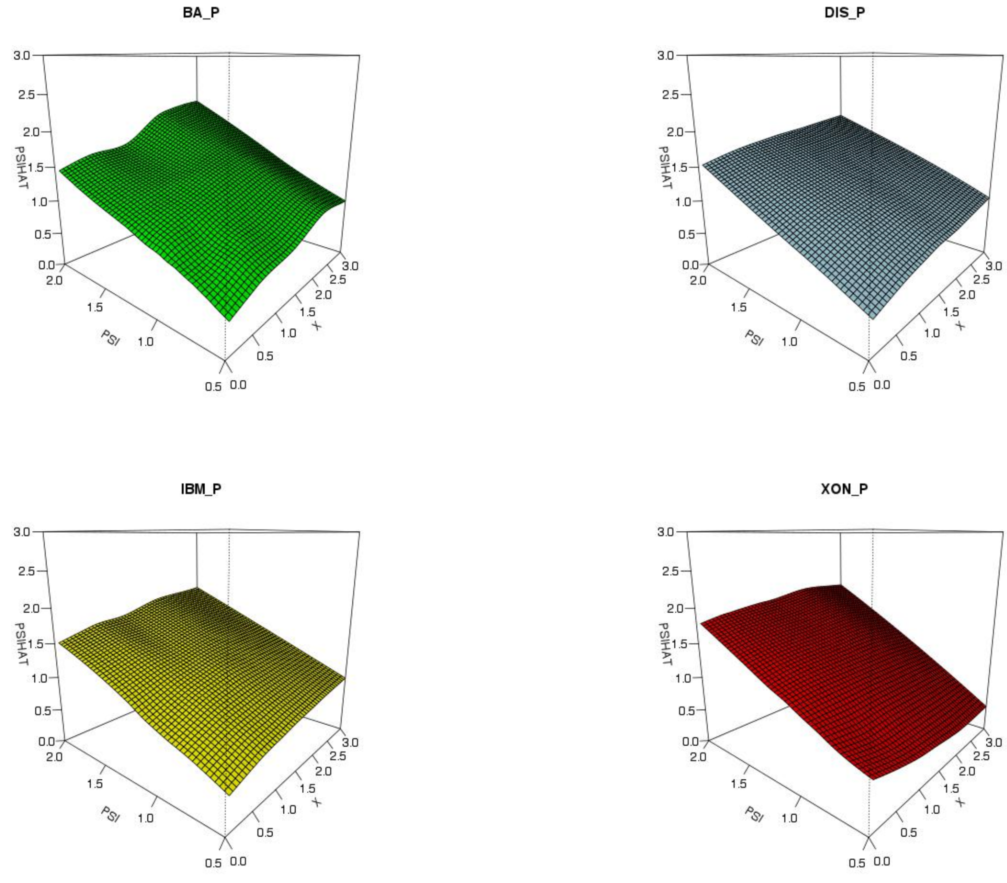


Figure 5: Estimated surfaces (final average) for price durations of Boeing, Disney, IBM and Exxon stocks.

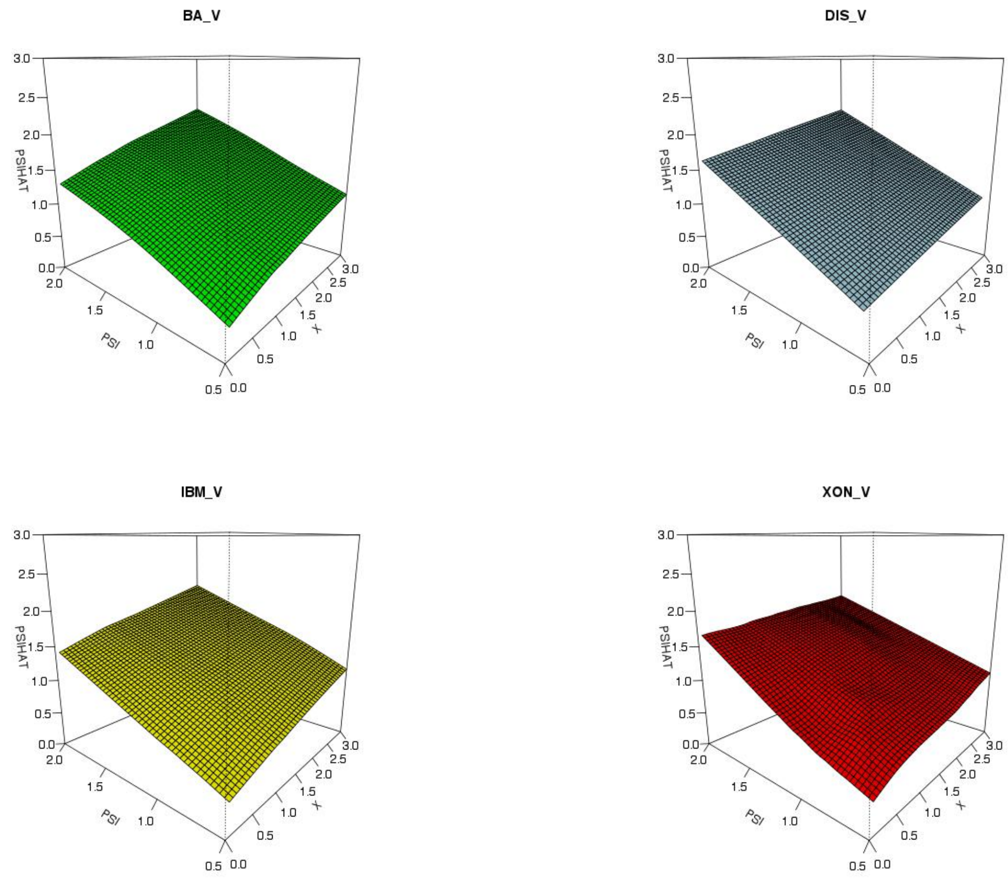
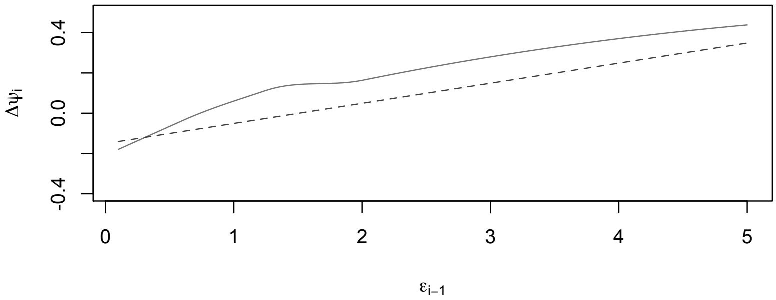


Figure 6: Estimated surfaces (final average) for trade durations of Boeing, Disney, IBM and Exxon stocks.



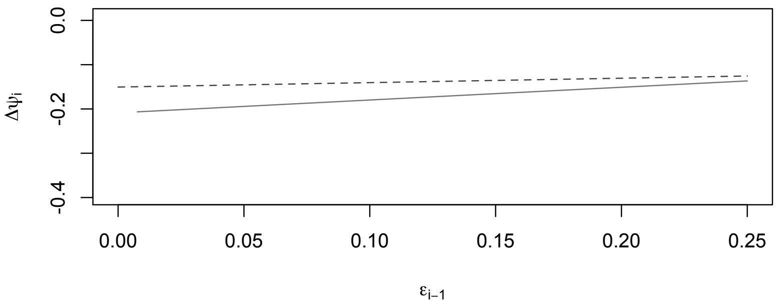


Figure 7: Empirical shock impact curves of a parametric (dashed) ACD estimation and a nonparametric (solid) estimation. Wider (above) and smaller (below) intervals for ϵ_t.

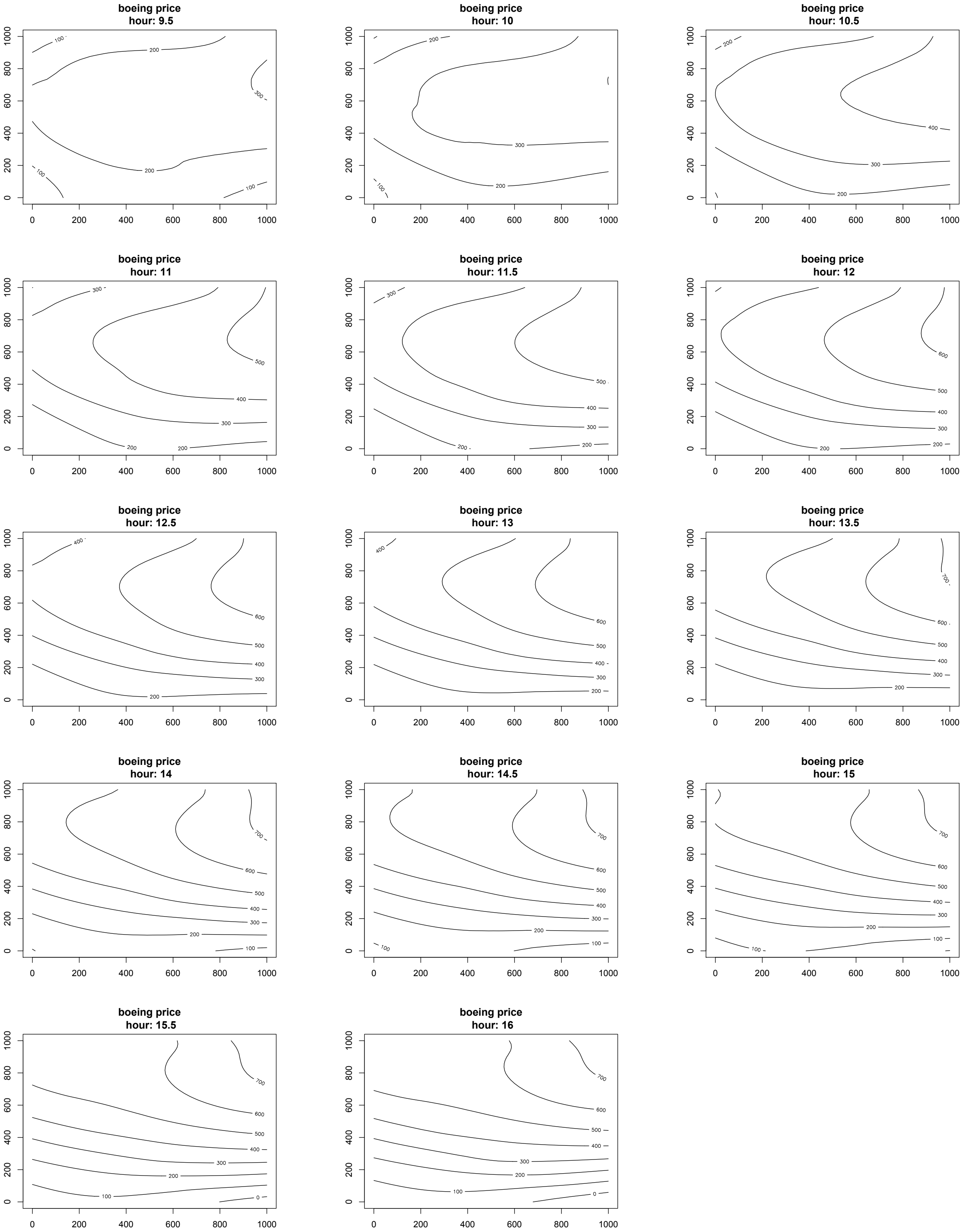


Figure 8: Contour plots of the estimated surface for Boeing price durations computed at 30-minute intervals.

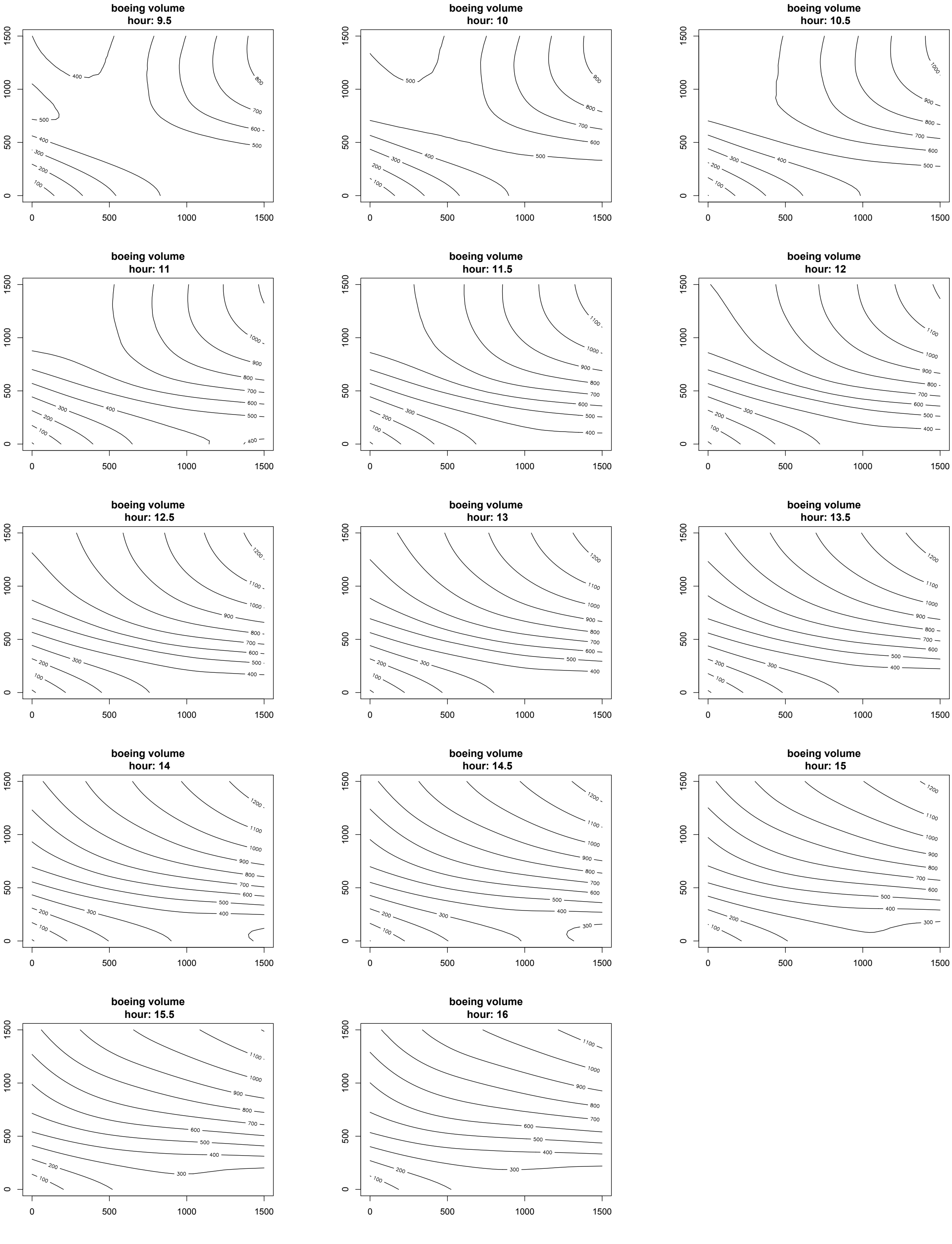


Figure 9: Contour plots of the estimated surface for Boeing volume durations computed at 30-minute intervals.

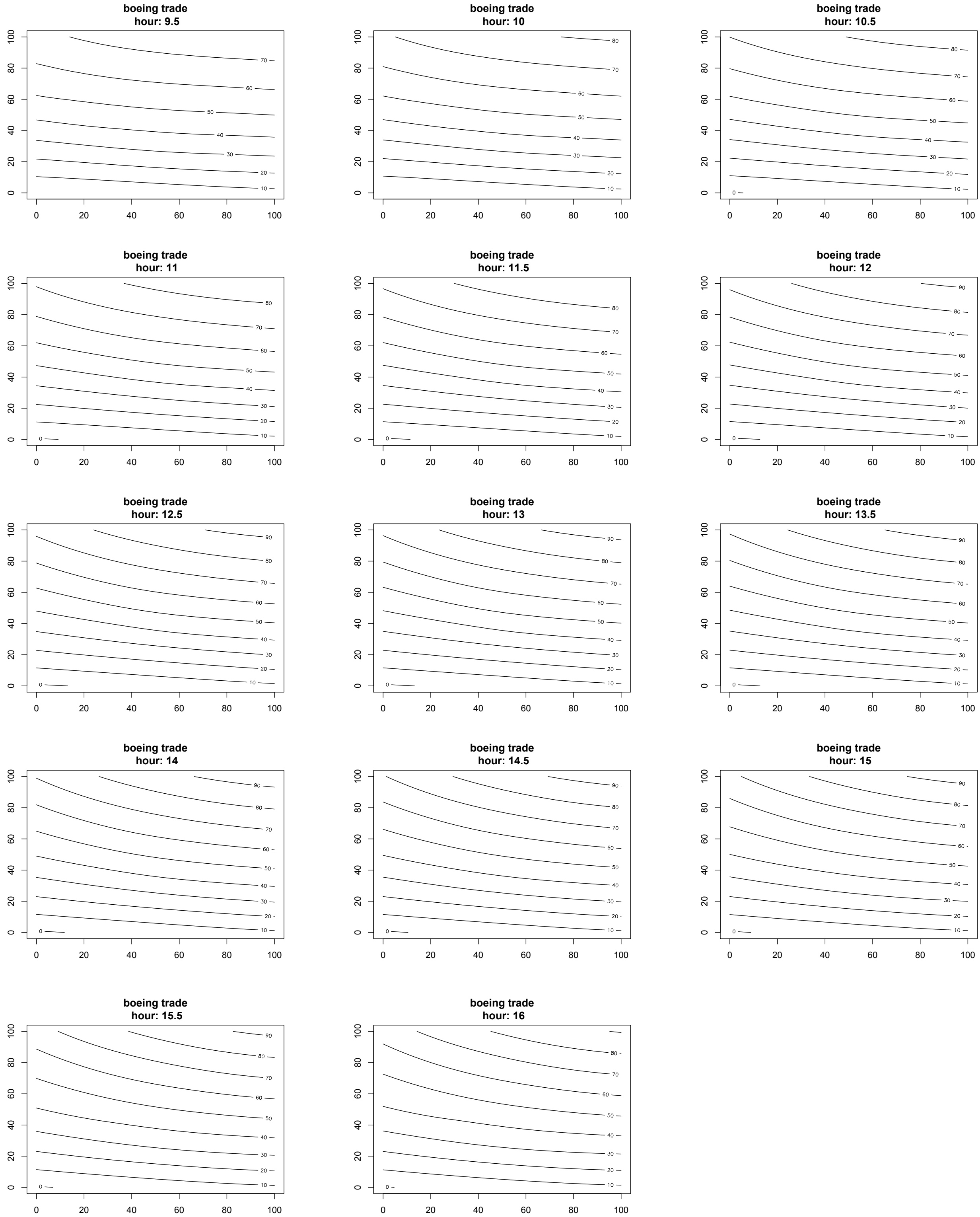


Figure 10: Contour plots of the estimated surface for Boeing trade durations computed at 30-minute intervals.