

# Non-Institutional Market Making Behavior: The Dalian Futures

## Exchange

by

*Òscar Jordà\**  
Department of Economics  
U. C. Davis  
Davis, CA 95616  
e-mail:  
[ojorda@ucdavis.edu](mailto:ojorda@ucdavis.edu)

*Holly Liu*  
KPMG  
e-mail: [hliu@kpmg.com](mailto:hliu@kpmg.com)

*Jeffrey Williams*  
Dept. of Agricultural and  
Resource Economics  
U. C. Davis  
Davis, CA 95616  
e-mail:  
[williams@primal.ucdavis.edu](mailto:williams@primal.ucdavis.edu)

*This version: April, 2002*

### Abstract

This paper contains three useful contributions: (1) it collects a new data-set of electronic transaction data on soybean futures from the Dalian Futures Exchange in China that records, not only the usual elements of each transaction (such as price and size) but also identifies broker and customer identities, variables not usually obtainable; (2) it presents new econometric methods for the analysis of dynamic multivariate count data based on the autoregressive conditional intensity model of Jordà and Marcellino (2000); and (3) together, the new data and econometric methods allow us to investigate, in a manner not available before, the determinants and effects of non-institutional market making (or *scalping*).

*Keywords:* market making, autoregressive conditional intensity, high-frequency data.

*JEL Codes:* G13, G14, C32

---

\* Corresponding author. Copyright 2002 by Òscar Jordà, Holly Liu and Jeffrey Williams. All rights reserved. Readers may take verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

<sup>1</sup>**scalp·er** \ ' skalpé (r), -kaúp -\ n -s ['scalp+er]: ...c(1): a speculator who seeks to make small profits on quick transactions...

– *Webster's Third New International Dictionary Unabridged.*

## **1. Introduction**

In futures exchanges, a scalper is a non-institutional agent who makes frequent purchases and sales yet ends each day without any outstanding position. Collectively, scalpers voluntarily approximate the role of institutional market makers by providing essential liquidity services. This paper investigates this type of voluntary market making behavior and examines its effects on market performance in a competitive, continuous-auction environment in which there is no formal institutional market making. Specifically, we analyze two central issues: (1) the determinants of scalper participation in the market, and (2) the effects that scalping has on market liquidity and price volatility.

The earliest work on scalpers is by Working (1954, 1967, 1977), who studied two months of trading records for “Mr. C”, a leading floor trader on the New York Cotton Exchange in 1952. Working presented descriptive statistics and discussions on scalping behavior in his papers, and in particular suggested that scalping contributes to market “fluidity”, and that scalpers “derive income from hedgers through temporarily absorbing hedging orders that are not immediately absorbed otherwise”. Working (1967) also touched on the issue that scalping profit is positively related to trading frequency and negatively related to trade size. Silber (1984) analyzed six weeks of trading records from “Mr. X”, a “representative scalper” on the New York Futures Exchange during 1982-1983, and derived similar conclusions regarding scalping strategy and profit. By studying the Computerized Trade Reconstruction (CTR) records of twelve active futures from the

Chicago Mercantile Exchange (CME) from July to September 1990, Kuserk and Locke (1993) identified scalpers for each contract, and found that scalping revenue varies across commodities and that scalpers tend to specialize in a particular commodity.

Both Working (1967, 1977) and Silber (1984) base their discussion on only one or two scalpers, and the CTR data analyzed by Kuserk and Locke (1993) are compiled from paper records at the end of a trading day and time-stamped to an accuracy of a 15-minute bracket. By contrast, thanks to the implementation of electronic trading systems, we have obtained all the transaction records of soybean futures from the Dalian Futures Exchange in China for the period June to December, 1999. These data are time-stamped to the second and each transaction record is marked with each trader's identity, the price and the volume of the transaction. The code for the trades reveals not just the broker but also the individual customer. While previous studies of scalpers are mostly descriptive, the richness of these data allows us to present more formal econometric evidence than has been possible before.

Modern analysis of financial, high-frequency, tick-by-tick data (such as the data in this study) relies on new econometric techniques based on the random arrival in time of transactions (for a survey see Engle, 2000). As a result, this literature has favored the approach of setting the problem in *event time* – that is, extracting the information contained in the random durations elapsing between transactions. In this paper, because scalping is likely to depend on cross-contract characteristics (such as scalping profit opportunities, price volatility, trading volume and intensity), multivariate extensions of existing econometric methods in event time quickly become impractical. The explanation resides in these models' inability to condition on information other than that available at

each event node. Since trades on different contracts are unlikely to arrive at exactly the same time, this constitutes a substantial shortcoming for the purposes of our research.

Instead, this paper presents an alternative strategy for analyzing multivariate, high-frequency, financial data based on the dual of any duration problem: the count process associated to it. Specifically, we will aggregate the data into five-minute intervals. During these short intervals over which scalping matters, the number of scalpers and the number of their trades are both small integers or counts. Thus, in order to analyze the effect of past information and other determinants of scalper presence in a dynamic context, we experiment with a new class of dynamic count data models: the autoregressive conditional intensity model (ACI) introduced in Jordà and Marcellino (2000).

Bivariate ACI estimates suggest that, all else equal, volatility and high levels of transaction activity, strongly determine a scalper's incentives to enter the market. In particular, Working's (1967) observations that scalping is negatively related to trading size and positively related to trading frequency are sustained by our findings. Moreover, there are important complementarities to scalping across contracts that seem to be assiduously exploited by scalpers. However, with regard to the effects that scalping has on price volatility and "fluidity" (to paraphrase Working's terminology) we are unable to detect anything of significance although this may be partly due to the fact that scalping represents a relatively small portion of overall trading.

The paper is organized as follows. Section 2 describes the main features of the data that we collected for this study and explains how we identify scalpers. The next section discusses candidate determinants of scalper presence in the market. Section 4 presents the econometric aspects of the ACI model and implements it to examine scalper

participation. Section 5 analyzes the effects of scalping on volatility while the last section concludes and presents directions for further research.

## **2. Data Description**

The data that we use in this study was collected at the source from the Dalian Futures Exchange in China. We were able to obtain seven months worth of electronic transaction records of soybean futures for the period spanning June to December, 1999. Each transaction record is marked with broker and customer identities, transaction price and volume, a buy or sell indicator, an indicator of opening or closing positions, as well as a time stamp.

There are six soybean futures contracts with expiration dates in January, March, May, July, September and November. A contract expires in its delivery month midway through which, trading tends to wind down almost completely. For example, trading on the July 2000 futures contract started in mid July 1999 and ended in mid July 2000. Thus, the trading of a contract gains considerable momentum usually two months after its trading date and remains active for another two months before the volume declines dramatically. However, the level of overall trading activity also depends on the season as well as on trading in the spot market. Figure 1 illustrates these patterns by displaying the average trading volumes of the two most actively traded contracts in our sample. In the Dalian Futures Exchange, soybean futures are most actively traded during September to the following January of each year, during which price volatility is high. The futures prices determined during this period tend to set the tone for the rest of the season (Dalian Futures Exchange, 2000).

A preliminary analysis of these transaction data reveals several key characteristics. There are 140 member firms in total of which 125 are brokerage firms with 6213 distinct customers in all. The remaining 15 member firms only trade for their own accounts and each has a seat in the exchange. A member firm can have several seats on the exchange floor, and it can also apply for a distance-trading seat to trade directly off the exchange. However, regulations of the Dalian Futures Exchange prevent brokerage firms from trading on their own accounts and similarly, prevent member firms trading on their own accounts from conducting brokerage business. Table 1 presents a summary of the relevant demographics for all these member firms.

## **2.1 Identifying Scalpers**

The central topic of this paper consists on explaining non-institutional market making behavior by scalpers. Consequently, we first need to identify which traders act as scalpers from a universe of over six thousand market participants. Scalpers are characterized by trading frequently during the day but rarely holding significant overnight positions. Using these two criteria, we define “frequent” trading to mean making at least one trade every 15 minutes on average. Since trading sessions last for two and one half hours (from 9:00 am to 11:30 am) this is equivalent to placing at least 10 orders a day. In addition, there are approximately 20 trading days in any given month. We will further require that for a trader to be considered a candidate scalper, they must follow the trading pattern described above for at least six days in a month.

These two criteria discriminate approximately 10 scalper candidates each month for the period June to December, 1999. In addition, we complement our selection criteria

to rule out “day traders.” Specifically, day traders are characterized by placing numerous one-sided orders and then liquidating these positions with reverse orders a half hour to an hour later. Traders that exhibit this trading pattern are excluded from the candidate scalper pool. To complete our selection, we expanded the scalper pool to include traders that have been consistently selected as scalpers in previous months but in a given month only trade actively for four or five days. The resulting number of scalpers each month in the sample is reported in Table 2 whereas the intra-day pattern of scalper presence in the two most actively traded contracts is displayed in the top panel of Figure 2.

### **3. The Determinants of Scalping Activity**

Given the scalpers identified by the procedures described in the previous section, this section investigates what determines a scalper’s decision to enter the market. As we shall see shortly, we will explore several candidate factors that include scalping profit opportunities, price volatility, transaction intensity, transaction volume and scalping competition. However, because we are using high-frequency, electronic transaction data, it is worth discussing more fundamental methodological issues first.

The availability of ultra-high frequency, tick-by-tick data (such as the one we study in this paper) has revolutionized the area of financial econometrics (for a survey see Engle, 2000). These data inherently arrive at random times and thus require of new econometric modeling techniques that deal with the random durations that elapse between transactions in a dynamic manner. One way to approach this problem is given by the autoregressive conditional duration (ACD) model proposed by Engle and Russell

(1998). The ACD model is a dynamic duration model in which the conditioning on past and exogenous or predetermined variables takes place in *event time*.

This last consideration constitutes a significant disadvantage for our study: often times we will want to condition on variables whose realizations do not coincide precisely with the event-arrival times of the dependent variables. Consequently, we have decided to work instead with the dual of any duration problem – the count process associated to it. We choose to aggregate the data into 5-minute intervals so that each trading day will contain 30 such intervals. Because scalpers close their positions usually within a few minutes, a five-minute interval is long enough to measure relevant statistics, but short enough to preserve the microstructure of trading. For instance, as the top left panel of Figure 2 reveals, the trading-day average number of scalpers present organized into five-minute intervals ranges from a high of 3 to a low of 1 with an average value of about 2.

It is well known that when the count variable associated to a duration process attains the value of one over the interval of observation, then there is no loss of information in estimating the count model instead of the duration model. In Section 4 we will estimate a dynamic count data model for the number of scalpers present which, as we have just seen, is a count variable whose mean is close to 2. This suggests that by aggregating the data into five-minute intervals, the loss of information due to this aggregation can be regarded as trivial, particularly when compared to the gains resulting from conducting the analysis in *calendar time* instead of in *event time*. Section 4 will therefore present in detail the econometric methods we have designed to attack the problem in this fashion. Before then however, we discuss the variables we will use to describe scalper's entry decisions.

### 3.1 Scalper's Spread

Market-makers on NASDAQ, specialists on NYSE, and scalpers in futures markets can all be considered as providers of a matching service between buyers and sellers who need immediate execution of their orders. The “price of immediacy” (Demsetz, 1968; Grossman and Miller, 1988) thus becomes the mechanism by which scalpers derive their profit. A natural way to determine this implicit cost of trading traditionally relies on measures of the bid-ask spread: the price difference paid by urgent buyers and urgent sellers. Unfortunately, our data do not provide bid-ask quotes (however, O’Hara, 1995 suggests that the bid-ask spread may be sensitive trade size, thus making across market comparisons difficult).

Instead, we construct a measure of scalping profit opportunities, which we denote *scalper's spread*, as follows. Let  $v_{\tau}^k(j)$  and  $p_{\tau}^k(j)$  denote the volume and price of a given transaction respectively, where  $\tau \in (t-1, t]$  indicates that the transaction belongs to the five-minute interval  $t$ ;  $k \in K$  denotes that  $k$  belongs to the set of scalpers  $K$ , and  $j \in \{B, S\}$  denotes that the transaction was a buy if  $j = B$  (from the point of buy of the scalper), and a sale if  $j = S$ . Therefore, we define scalper's spread as the difference between the volume weighted average of *buy* and *sale* prices, that is,

$$SS_t = \sum_{\substack{\tau \in (t-1, t] \\ k \in K}} \left[ \left( \frac{v_{\tau}^k(S)}{\sum_{\substack{\tau \in (t-1, t] \\ k \in K}} v_{\tau}^k(S)} \right) p_{\tau}^k(S) - \left( \frac{v_{\tau}^k(B)}{\sum_{\substack{\tau \in (t-1, t] \\ k \in K}} v_{\tau}^k(B)} \right) p_{\tau}^k(B) \right] \quad (1)$$

The bigger the spread  $SS_t$ , the bigger the profit opportunities are for scalping, which may be taken as an indicator of market liquidity conditions (see Grossman and Miller, 1988).

### **3.2 Volatility**

Intimately related to a scalper's profit opportunities measured by  $SS_t$ , is the notion of price volatility. Price changes typically represent shocks to order imbalance and therefore, the larger these imbalances, the more plentiful scalping profit opportunities are. Working (1967) noted that there is "price jiggling" in futures markets by which he meant small price changes in one direction immediately followed by a price change in the opposite direction. Working (1967) attributes this price pattern to the imbalance of market orders at a point in time and suggests that scalping tends to restrict the size of price jiggling. Garbade and Silber (1979) also suggest that scalper participation in a continuous-auction market reduces the volatility of transaction prices around the equilibrium price. Daigler and Wiley (1999) find that clearing members in futures markets, who observe order flow, reduce the volatility of their own trades.

These views suggest that price volatility will be an important predictor of scalper presence while also suggesting that scalping has positive externalities on price volatility. We will investigate the latter issue in more detail below when we investigate the response of volatility to scalping and trading volumes across the two most actively traded contracts in our sample (the May 2000 and July 2000 contracts). Here, however, our first concern is that of measuring price volatility itself.

The traditional approach to measuring volatility in financial markets typically relies on some GARCH based measure since it is infeasible to compute price volatility

from single price observations for each period. However, because we are analyzing the data in five-minute intervals, it is straightforward to compute a volume-weighted sample estimate of price volatility. In particular and following the notation in subsection 3.1, let  $v_\tau$  denote the volume associated with the  $\tau^{\text{th}}$  transaction in the interval  $(t-1, t]$  (note that here we do not distinguish whether or not the transaction involves a scalper), and let  $p_\tau$  be the associated price. Similarly, let  $V_t$  denote the total transaction volume in interval  $t$  and define the volume-weighted price average during the  $t^{\text{th}}$  interval as  $P_t = \sum_{\tau \in (t-1, t]} \frac{v_\tau}{V_t} p_\tau$ .

Therefore, we can measure volatility as

$$\sigma_t = \sqrt{\sum_{\tau \in (t-1, t]} \frac{v_\tau}{V_t} (p_\tau - P_t)^2} \quad (2)$$

The top right panel of Figure 2 displays the intra-day seasonal patterns for this measure of volatility. These exhibit the usual “volatility smile” patterns common to many financial price time series: volatility tends to be highest at the opening and close of the market.

### 3.3 Other determinants

In addition to a scalper’s spread and price volatility, we will investigate two groups of variables that relate to the level of transaction activity in the market: transaction volume and transaction intensity. For each group, we will classify the variable into scalper and non-scalper transactions. Specifically,  $V_t^j$  will denote total volume transacted in the  $t^{\text{th}}$  five-minute interval, where  $j = k$  will denote scalper transactions and  $j = n$  will denote

non-scalper transactions. Similarly,  $I_t^j$  will denote the transaction intensity (number of transactions) during the  $t^{\text{th}}$  interval for  $j = k, n$ . Although the joint inclusion of volume and intensity variables may appear redundant (they both measure the level of market activity), a small number of high volume transactions may present different scalping incentives than a high number of small volume of transactions (this was Working's, 1967 observation). The distinction between scalper and non-scalper transaction activity owes to the fact that while the first is associated with scalping opportunities, the latter reflects scalping competition. With regard to the intra-day patterns of transaction activity, the bottom panels in Figure 2 display the average variation in the total number of transactions throughout the trading day along with the total number of scalper transactions.

#### **4. Predicting Scalper Presence**

This section investigates which are the best predictors of scalper's presence. Although our dataset contains six soybean active futures contracts with differing maturing dates, we find that scalpers tend to limit their participation to the two most actively traded contracts during any given day. Because of the maturity structure of these contracts, we refer to the most actively traded contract as the *dominant contract* and similarly, denote the second most actively traded contract as the *subsidiary contract*. In particular, the May 2000 contract will be the dominant contract in our sample, with July 2000 as the subsidiary. The sample that we investigate includes 100 days of trading that took place between August to December 1999, in other words, 3000, 5-minute interval observations.

#### 4.1 The Autoregressive Conditional Intensity Model

The number of scalpers present in any given 5-minute interval is a count variable that shares many characteristics with variables such as the number of customers that arrive at a service facility, the arrival of phone calls at a switchboard, and other analogous variables that describe rather infrequent events that occur at random times within the interval of observation. Therefore, it is natural to treat the number of scalpers as a count variable that we denote as  $y_t$ , where the subscript  $t$  refers to the five-minute interval associated to it. The benchmark for count data is the Poisson distribution (see Cameron and Trivedi, 1998 for an excellent survey on count data models), with density

$$f(y_t | \mathbf{x}_t) = \frac{e^{-\lambda_t} \lambda_t^{y_t}}{y_t!} \quad y_t = 0,1,2,\dots \quad (3)$$

and conditional expectation

$$E(y_t | \mathbf{x}_t) = \lambda_t = \exp(\mathbf{x}_t' \boldsymbol{\gamma}) \quad (4)$$

so that  $\log \lambda_t$  depends linearly on  $\mathbf{x}_t$ , a vector of explanatory variables which will primarily include the variables discussed in section 3, along with lags of the dependent variable  $y$ . Expression (4) is called the exponential mean function and together with expression (3) they form the Poisson regression model, the workhorse of count data models. The model can be easily estimated by maximum likelihood and conventional numerical techniques since the likelihood is globally concave.

However, unlike most applications of the Poisson regression model, the variable  $y_t$  is a time series that exhibits remarkable persistence. For example, the Ljung-Box statistic for the May 2000 contract attains a value of  $Q_5 = 1228$  at five lags and  $Q_{10} = 7409$  at ten lags, well above the conventional 5% critical value (in fact, the first five values of the autocorrelation function are above 0.5 for both the May 2000 and the July 2000 contracts). One solution to this problem is offered in Jordà and Marcellino (2000) in which the intensity of the Poisson process described in expression (4) is formulated to have a time series representation similar to that of an ACD or GARCH models. In particular, Jordà and Marcellino (2000) replace expression (4) with

$$\log(\lambda_t) = \alpha \log(\lambda_{t-1}) + \beta y_{t-1} + \mathbf{x}_t' \boldsymbol{\gamma} \quad (5)$$

Thus, the model given by expressions (3) and (5) is referred to as the *autoregressive conditional intensity* model of order (1,1) or ACI(1,1). Extensions of expression (5) to general lags ACI( $p, q$ ) are trivial but to focus the discussion consider the simple (1,1) case. Expression (5) ensures that the intensity parameter  $\lambda_t$  remains strictly positive for all values of the parameters  $\alpha, \beta, \boldsymbol{\gamma}$  while allowing the dependence of  $\log \lambda_t$  to be linear in its past values. The process described in (5) will be stationary as long as  $\alpha + \beta < 1$ . Note that the ACI(1,1) endows the original expression in (4) with rather rich dynamics in a parsimonious manner: the process  $\log \lambda_t$  depends on lags of  $y_{t-1}$  and  $\mathbf{x}_t$  at a geometrically decaying rate  $\alpha$ . Estimation of the ACI(1,1) can be done by conditional maximum likelihood techniques by setting  $\lambda_0$  to the unconditional mean of  $y$  (alternatively,  $\lambda_0$  can be estimated as an additional parameter if the model is nonstationary, for example) and is

disarmingly simple. For example, one can specify the following three lines of code in the LogL object in EViews, version 4.0 (see EViews manual, chapter 18):

```
@log ll
log(lambda) = c(1) + c(2)*log(lambda(-1)) + c(3)*y(-1) + c(4)*x
ll = log(@dpoisson(y, lambda))
```

The empirical analysis that we report below involves scalper presence in the May 2000 and July 2000 contracts, which we will denote by  $y_t^M$  and  $y_t^J$  respectively. It is common practice for scalpers to concentrate their trading on the dominant contract (May 2000) for one or two hours, and then switch to the subsidiary contract (July 2000). Consequently, in order to determine scalper participation in the market, it is important to model the joint behavior of  $y_t^M$  and  $y_t^J$  conditional on past information so that we take into account in complementarity/substitutability between contracts.

Specifically,  $y_t^M$  and  $y_t^J$  are both count variables and let  $\mathbf{x}_t$  now be a vector of explanatory variables that includes the scalping opportunities in both contracts. The object of interest is now the joint density for  $y_t^M$  and  $y_t^J$ , which can be decomposed, without loss of generality, as

$$f(y_t^M, y_t^J \mid \check{y}_{t-1}^M, \check{y}_{t-1}^J, \mathbf{x}_t) = g(y_t^M \mid \check{y}_{t-1}^M, \check{y}_{t-1}^J, \mathbf{x}_t)h(y_t^M \mid \check{y}_t^M, \check{y}_{t-1}^J, \mathbf{x}_t) \quad (6)$$

where  $\check{y}_t^i = \{y_t^i, y_{t-1}^i, \dots\}$  for  $i = M, J$ ;  $g(\cdot)$  is the marginal density of  $y_t^M$  and  $h(\cdot)$  is the conditional density of  $y_t^J$  given  $y_t^M$ . Fully parametric approaches based on the joint

distribution of non-Gaussian random variables given a set of covariates are difficult to apply because analytically and computationally tractable expressions for such joint distributions are available for special cases only, and in fact, applications of multivariate count models are relatively uncommon (see chapter 8, Cameron and Trivedi, 1988). However, because of the nature of the contracts that we are investigating, it is straightforward to assign a hierarchical structure to the dominant and subsidiary contracts to the decomposition in expression (6). In particular, we will assume that  $g(\cdot)$  and  $h(\cdot)$  are univariate Poisson densities whose conditional means follow the bivariate version of the ACI(1,1) process in (5),

$$\begin{bmatrix} 1 & 0 \\ \alpha_{JM}^0 & 1 \end{bmatrix} \begin{bmatrix} \log(\lambda_t^M) \\ \log(\lambda_t^J) \end{bmatrix} = \begin{bmatrix} \alpha_{MM}^1 & \alpha_{MJ}^1 \\ \alpha_{JM}^1 & \alpha_{JJ}^1 \end{bmatrix} \begin{bmatrix} \log(\lambda_{t-1}^M) \\ \log(\lambda_{t-1}^J) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \beta_{JM}^0 & 0 \end{bmatrix} \begin{bmatrix} y_t^M \\ y_t^J \end{bmatrix} + \begin{bmatrix} \beta_{MM}^1 & \beta_{MJ}^1 \\ \beta_{JM}^1 & \beta_{JJ}^1 \end{bmatrix} \begin{bmatrix} y_{t-1}^M \\ y_{t-1}^J \end{bmatrix} + \begin{bmatrix} \gamma_M' \\ \gamma_J' \end{bmatrix} \mathbf{x}_t \quad (7)$$

Note that expression (7) imposes a similar semi-structural restriction to that in the analysis of the structural-VAR literature: namely, that of imposing a Wold-causal order which in this case consists on having the contemporaneous values of the dominant contract's dependent variables appear in the expression of the conditional mean for the subsidiary contract. The bivariate ACI(1,1) model expanded by expression (7) can be easily estimated by maximizing the joint conditional likelihood expression in (6) by numerical techniques.

## 4.2 ACI Estimates

This section presents estimates obtained with the bivariate ACI model in expressions (3), (6) and (7) but it complements the analysis with univariate ACI and simple Poisson regression estimates to provide a benchmark with which to compare the benefits of the ACI methodology. To reiterate the main elements of the analysis, the data is organized in five-minute intervals, each trading day containing 30 such intervals (9:00am to 11:30am) and there are 100 such days in our sample (which correspond to the period August to December 1999). However, since there is no trading during the morning break (10:15am to 11:00am) we drop observations 16 through 18 in each trading day (all the variables attain zero values during these periods). The dependent variables  $y_t^i$  for  $i = M, J$  denote the number of scalpers present in trading for the May 2000 (denoted by  $i = M$ ) and the July 2000 (denoted by  $i = J$ ) contracts.

The vector of explanatory variables  $\mathbf{x}_t$  will contain the following variables. The intra-day seasonal pattern will be captured with a set of 27 dummy variables (one for each five-minute interval in a trading day) denoted  $D_t' = (d_{1t}, \dots, d_{15t}, d_{19t}, \dots, d_{30t})$  with  $d_{it} = 1$  if  $t \in i^{th}$  interval, 0 otherwise. In addition, the vector  $\mathbf{z}_{it}' = (SS_t, \sigma_t, V_t^k, V_t^n, I_t^k, I_t^n)$  for  $i = M, J$  will contain the variables described in section 3 for each contract, (we abuse notation in favor of simplifying the presentation by forgoing the indicator subscript for the type of contract in each individual element of  $\mathbf{z}_{it}$ ).

We begin by discussing univariate estimates of the Poisson and ACI regression models (expressions (3)-(5)). These are reported in Table 3. The Poisson regression model contains up to four lags of the explanatory variables so that  $\mathbf{x}_t' = (D_t', \mathbf{z}_{t-1}', \dots, \mathbf{z}_{t-4}', y_{t-1}, \dots, y_{t-4})$ . On the other hand, the ACI only contains one lag so that  $\mathbf{x}_t' = (D_t', \mathbf{z}_{t-1}')$ ,

(note that  $y_{t-1}$  appears explicitly in expression (5) and is therefore omitted from  $\mathbf{x}$ ).

Because we are estimating univariate models, note that we have dropped the subscript that indicates to what contract do the covariates belong to since the model for each contract will only contain regressors specific to that contract. The parameter estimates for the Poisson regression model are omitted and only the basic regression statistics are reported at the bottom of Table 3.

The first result worth remarking is the substantial efficiency with which the ACI model deals with the dynamics. Note that, although the Poisson regression model contains 20 additional regressors, it attains a lower likelihood value than the ACI estimates (note that the likelihood values are directly comparable). The reason for this result is made evident by the parameter estimates for the variables  $y_{t-1}$  and  $\log(\lambda_{t-1})$  ( $\alpha$  and  $\beta$  in expression (5)): their sum is approximately 0.85 for both contracts suggesting a substantial amount of persistence (although this value is well below the canonical value of one, suggesting that the model is stationary). With regard to estimates for particular variables, we prefer to discuss the results in the context of the bivariate estimates. Here we simply remark that, contrary to our expectations, the scalper spread variable,  $SS_{t-1}$ , did not appear to explain scalper presence. Generally speaking, volatility and transaction activity levels of non-scalpers appear to increase the likelihood of additional scalper presence while scalper activity has the opposite effect.

Table 4 contains the estimates for the bivariate ACI specification in expression (7). Note that now  $\mathbf{x}_t' = (D_t', \mathbf{z}_{M,t-1}', \mathbf{z}_{J,t-1}')$ , that is, the explanatory variables for each of the contracts appear in the regression of each contract (this was not the case with the univariate counterparts). As before, note that the amount of persistence detected is quite

substantial: for the May 2000 contract,  $\alpha + \beta = 0.95$ ; and for the July 2000 contract,  $\alpha + \beta = 0.89$ . The scalper-spread variable is not significant overall except in the May 2000 equation, where a significant and negative coefficient on the July 2000 scalper spread suggests that when profit opportunities are good in this contract, they decrease the likelihood of additional scalpers entering the May 2000 contract, as one would expect. As before, volatility is perhaps the most consistent predictor of scalper presence, always entering positively and significantly.

With respect to the interaction between the May 2000 and the July 2000 contracts, the former appears to be more sensitive to characteristics of the latter than vice versa (in terms of the overall significance level of the cross-contract covariates). In what follows, self-contract variables denote regressors that belong to the contract whose dependent variable is being explained, while cross-contract variables will denote regressors that belong to the alternative contract being explained. Hence, with respect to the volume variables, self-contract scalper and non-scalper volume enter negatively but only significantly in the July 2000 contract equation. The self-contract intensity variables also enter with the correct signs, so that together with the volume variable estimates, we are able to replicate Working's (1967) findings on the negative effects of trading size and the positive effects of transaction frequency on scalping. The cross-contract effects of the volume variables and the scalper intensity variables suggest that there is significant substitution of scalping across contracts. However, the signs on the cross-contract, non-scalper intensity variables suggest complementarity rather than substitution.

The overall fit of the bivariate ACI is substantially better than the fit of the individual univariate ACI models (the log-likelihood increases from  $-7932.062$  to  $-$

7664.709 but with 18 additional regressors). The p-value of the likelihood ratio statistic for the additional regressors in the bivariate model is essentially zero, suggesting the bivariate ACI is a substantial improvement.

We conclude this section by summarizing the main findings. First, between the two *liquidity* proxy variables, scalper spread and volatility, the latter is an important and more significant determinant of scalper presence. Second, the interaction between volume and trading intensity suggests that they contain specific information regarding trading activity levels and that they should be jointly included when determining scalper presence. Third, methodologically speaking we demonstrate conclusively the virtues of the ACI model relative to classical Poisson regression when the dependent variable is an autocorrelated time series.

### **4.3. Robustness: The Tale of Mr. D**

The empirical investigation in the previous subsection is based on the aggregate behavior of those traders that, in a given month, belong to the class of scalpers we have defined in Section 2.1. However, one could ask how well does this aggregative approach represent individual scalper behavior. To answer this question, we recorded the transaction activity of an anonymous trader – whom we called Mr. D – who belonged to the class of scalpers during most of the months in our sample.

Mr. D is a rather active trader, having engaged in transactions about 20% of the time in the May 2000 contract and one-third of the time in the July 2000 contract. In the 5-minute intervals in which he is active, he engages in two transactions on average although sometimes it can be as many as 21. His pattern of trades corresponds rather

closely to any conventional notion of scalping: any position he holds is typically reversed within the same or one 5-minute interval after it is acquired and he rarely finishes the day with any significant positions.

We investigate Mr. D's trading patterns using the same approach (the ACI model) and the same explanatory variables as in the previous subsection with the exception of the dependent variable – the variable  $y_t^i$ ,  $i = M, J$ , although a still a count variable, now records Mr. D's transactions in a given 5-minute interval instead. Table 5 directly reports the bivariate ACI(1,1) estimates of this model, which is based on the specification in expression (7) and is estimated over the same sample as in Table 4 to make the results between these tables easily comparable.

Several results deserve comment. Overall, there is a remarkable qualitative similitude between the aggregate results in Table 4 and the results for Mr. D in Table 5: the coefficient estimates are different in magnitude because the dependent variables are different but the sign and the significance levels are very similar. Once again, the ACI(1,1) specification captures a significant amount of persistence with  $\alpha + \beta = 0.77$  for the May 2000 contract and  $\alpha + \beta = 0.73$  for the July 2000 contract.

However, there are some differences between models in that Mr. D appears to be more preoccupied with the May 2000 contract variables than with the July 2000 contract. Thus, for example, when volatility is high in the May 2000 contract, Mr. D reduces his engagement in July 2000 trading and conversely, when the volume of transactions and the transaction intensity of scalpers in the May 2000 contract increases, Mr. D becomes more engaged in July 2000 contract trading. But other than the different hierarchy between the contracts, as these results show, they confirm the essential findings

of the aggregative model and provide a robustness check to our conclusions. The next section investigates some of the interactions between volatility, scalper transacted volume and non-scalper volume that will complement the results in this section.

## **5. Volatility Responses to Scalper and Non-Scalper Transaction Activity**

The previous section allowed us to investigate the determinants of scalper presence in the market for soybean futures in the Dalian Futures Exchange. However, while we were able to ascertain that, among other factors, volatility is an important determinant of scalper presence, our discussion in Subsection 3.2 suggests that scalping may have important positive externalities on volatility. Scalping activity and competition are likely to reduce the magnitude of order imbalances (the main source of scalping profit), therefore mitigating price volatility. On the other hand, the overall level of volatility is mostly correlated with the general level of transaction activity: thin markets tend to display higher levels of volatility. Finally, the interaction between the dominant and subsidiary contracts is likely to contribute to fluctuations in the relative levels of volatility experienced in each of these markets.

Evaluating how volatility responds to scalping therefore requires that we consider all of these issues simultaneously in the context of a general dynamic model. In essence, the type of question we want to answer is: What is the dynamic response volatility to scalping, conditional on past information and, on transaction activity levels and cross-contract interactions? To answer this question we follow a two-pronged approach based on a structural vector autoregression (VAR). The first prong consists on analyzing the dynamic response of volatility to structural shocks of the variables in the system (which

will be defined shortly) while the second prong consists on counter-factual experimentation. Specifically, we will analyze the response of volatility to transaction activity shocks under two scenarios: the first allows for the usual response of scalping activity to such a shock while the second shuts down scalping responses instead. The difference in the dynamic response of volatility to these two scenarios will reflect scalper externalities.

We turn now to the specific implementation of these experiments. Consider the vector of variables,  $\mathbf{w}_t' = (\mathbf{x}_{Mt}', \mathbf{x}_{Jt}')$  where  $\mathbf{x}_{it}' = (\sigma_t, I_t^n, I_t^k)$  for  $i = M, J$  (once more, we abuse notation slightly by omitting subscripts on the component variables of  $\mathbf{x}_{it}$  for the sake of clarity). All of the component variables in  $\mathbf{x}$  have been defined previously in Section 3 and they are respectively: volatility, non-scalper transaction intensity, and scalper transaction intensity.

Consequently, the vector  $\mathbf{w}$  will be modeled as a finite order VAR. To endow the model with a structural interpretation, we follow the conventional assumptions in the extensive VAR literature (see chapter 11, Hamilton 1994) based on the Cholesky decomposition and the specification of a Wold-causal ordering of the vector of contemporaneous variables,  $\mathbf{w}_t$ . The configuration of our problem confers a natural hierarchy that is based on the maturity of the contracts. Recall that the May 2000 contract is the dominant contract with the July 2000 contract as the subsidiary. Subsequently, the block of variables  $\mathbf{x}_{Mt}$  will be ordered Wold-causally prior to the block  $\mathbf{x}_{Jt}$ . Furthermore, within each block, we will assume that scalper transaction intensity depends on the contemporaneous values of the level of volatility and non-scalper transaction intensity.

Therefore, within each of the blocks  $\mathbf{x}_{it}$  the Wold-causal order will place volatility ahead of non-scalper transaction intensity and ahead of scalper transaction intensity.

### **5.1 The Dynamic Responses of Volatility**

Based on the VAR described in the previous section, Figure 3 reports a selection of the impulse response functions to a one standard deviation shock to each of the variables in the system, along with the associated asymptotic two standard error bands. These responses are traced for up to one hour after the shock (12 five-minute intervals) and constitute the first of the two prongs described in our research strategy. The next section will discuss the results for the counterfactual experiments. Figure 3 is organized as follows. Each row displays impulse responses of different variables to the shocks specified at the beginning of each row (in order displayed: non-scalper transaction intensity, scalper transaction intensity and volatility). The two headers at the top of the different panels subdivide the display into two groups of columns according to whether the shock was to a May 2000 contract variable or to a July 2000 variable instead. All the shocks considered are positive by design.

We begin by examining the responses of scalper transaction activity to shocks in volatility, which are displayed in the last row of Figure 3. These responses offer confirmatory evidence of the results found in Section 4: an increase in volatility generates increased scalping activity, both in the self-contract responses but also in the cross-contract responses. The congruence of these results with those in Section 4 increase our confidence both on the findings themselves and on the merits of the bivariate ACI specification. Next, consider the responses displayed in the first row, which correspond to

shocks in non-scalper transaction intensity. As one would expect, an increase in this type of activity generates higher self-contract volatility and scalper transaction intensity.

Finally, the second row of figures displays the more interesting albeit less intuitive results. In response to a positive shock in scalper transaction intensity, we would expect two effects: (1) cross-contract increases in scalper transaction intensity due to the reduction of scalper competition, and (2) reductions in price volatility due to increases in scalping services. However, the reaction in cross-contract scalping is clearly nil and while volatility does not appear to be very sensitive to scalping transaction intensity levels, there is some evidence that the initial response is significantly positive rather than negative. Because of this last result, we examined how robust this finding is to alternative measurements of scalping activity on volatility through the counterfactual experiment reported in the next subsection.

## **5.2 Scalping Effects on Volatility: A Counterfactual Experiment**

This section asks the question: How would have volatility responded to increased levels of non-scalper transaction intensity if scalper activity were constrained to remain fixed rather than respond by also increasing? If the hypothesis that scalping makes the market more fluid is true, we should expect that volatility increase by a larger extent when scalping is constrained than otherwise. Due to the apparent complementarity between the May and July 2000 contracts found in both the ACI and VAR estimates reported in subsections 4.2 and 5.1, we decided to implement our counterfactual experiment by restricting the responses of the scalping intensity variables in *both* contracts to be simultaneously constant in response to a shock in non-scalping trading intensity.

Figure 4 therefore displays the results of this experiment and is organized as follows. Each row corresponds to the responses of volatility due to a shock in non-scalping transaction in the May 2000 contract (top row) and the July 2000 contract (bottom row). Each column in turn displays the May 2000 volatility response (left column) and the July 2000 response (right column). The panels on the diagonal of Figure 4 contain the same responses displayed in Figure 3 and essentially coincide with our expectation that increased self-contract transaction activity generates more volatility. However, note that when the scalping response is shut down, there is no statistical and discernible difference in the response of volatility. This suggests that, at best, scalping is not contributing positively to reducing volatility and, at worst, it is inducing higher volatility a half hour after the shock in both contracts. The off-diagonal panels of Figure 4 contain the cross-contract responses of volatility. These are positive, further confirming the complementarity that exists between contracts discussed in Subsection 5.1, but here too, scalping has no effect on volatility.

## **6. Conclusion**

This paper contains three main contributions. The main focus of the paper has been that of studying non-institutional market making but to do an effective job we have obtained an exclusive electronic transaction data set with customer and broker identities and we have developed the appropriate econometric methods to deal with the challenges that these data presented. Our results support earlier conclusions made by Working (1967) regarding the effects of trade volume and trade size on scalper participation, but perhaps more importantly, we have provided new light on the interaction of competing scalping

opportunities in determining scalper presence in a particular contract. However, because scalping represents a small fraction of total trading, we have been unable to establish that scalping reduces price volatility. We suspect this result reflects the idiosyncrasies of our data rather than more fundamental scalping externalities (or lack thereof).

In future work, the detailed nature of the data will allow us to investigate whom are scalpers trading with and how scalper's profit opportunities are affected by the heterogeneous nature of other traders. In preliminary results, we have reason to believe that there exist substantial information asymmetries among different groups of traders, which affect their overall market performance. Nevertheless, although scalpers appear to belong to the class of uninformed traders, they tend to perform substantially better than other uninformed but non-scalping traders.

Finally, it is worth highlighting the advantages of modeling high frequency data by aggregating the data into short intervals rather than by using the random arrival times in which it is reported. By examining the count process associated with these dynamic duration processes, we have been able to simplify the analysis in several dimensions. The fixed-interval aggregation process allows the problem to be recast back into calendar time rather than in event time. This allows for seasonal effects to be filtered in an effective, general and straightforward manner by means of appropriate sets of seasonal dummies. By contrast, dynamic duration models require of more sophisticated, non-parametric smoothers. More fundamentally however, studying electronic transaction data in their native state makes the conditioning on variables whose inter-arrival times do not coincide with that of the endogenous variables a prohibitively complicated problem. We argue that, as long as the interval of aggregation is sufficiently small, there is little loss of the

micro-structure effects financial economists may wish to study while the payoff in simplicity is quite substantial. The ACI model demonstrates one effective way of taking advantage of knowledge on the varying rates of transaction activity levels that have characterized the newest family of dynamic duration models.

## References

- Cameron A. Colin and Pravin Trivedi. Regression Analysis of Count Data. Econometric Society Monographs no. 30. Cambridge University Press. (1998).
- Dalian Futures Exchange. "Introduction to Soybean Futures." May. (2000).
- Daigler, Robert, and Marilyn Wiley. "The Impact of Trader Type on the Futures Volatility-Volume Relation," *Journal of Finance*. Vol. 54, No. 6, 2297-2316. (1999).
- Demsetz, Harold. "The Cost of Transacting," *Quarterly Journal of Economics*, Vol. 82, No. 1, 33-53. (1968).
- Engle, Robert F. "The Econometrics of Ultr-High Frequency Data," *Econometrica*, Vol. 68, no. 1, 1-22. (2000).
- Engle, Robert, and Jeffrey Russell. "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Time Series Data," *Econometrica*, Vol. 66, no. 5, 1127-1162. (1998).
- Eviews 4. User's Guide. Quantitative Micro Software.
- Garbade, Kenneth and William Silber. "Structural Organization of Securities Markets: Clearing Frequency, Dealer Activity and Liquidity Risk," *Journal of Finance*, Vol. 34, 577-93. (1979).
- Grossman, Sanford J. and Merton H. Miller. "Liquidity and Market Structure," *Journal of Finance*, Vol. 43, No. 3, 617-633. (1988).
- Hamilton, James D. Time Series Analysis. Princeton University Press.
- Jordà, Òscar and Massimiliano Marcellino. "Stochastic Process Subject to Time-Scale Transformation: An Application to High-Frequency, FX Data," *U.C. Davis Working Paper 00-02*. (2000).
- Locke, Peter, and Gregory J. Kuserk. "Scalper Behavior in Futures Markets: An Empirical Examination," *The Journal of Futures Markets*. Vol. 13, No. 4, 409-431 (1993).
- O'Hara, Maureen. Market Microstructure Theory. Blackwell Publishers. (1995).
- Silber, William L. "Marketmaker Behavior in an Auction Market: An Analysis of Scalpers in Futures Markets," *Journal of Finance*, Vol. 39, No. 4, 937-953. (1984).
- Working, Holbrook. "Whose Markets? Evidence on Some Aspects of Futures Trading." *Journal of Marketing*, Vol. 19, No. 1, 1-11. (1954).

Working, Holbrook. "Tests of a Theory Concerning Floor Trading on Commodity Exchanges." *Food Research Institute Studies*, Vol. VII. Supplement: Proceedings of a Symposium on Price Effects of Speculation in Organized Commodity Markets. 197-239. (1967).

Working, Holbrook. "Price Effects of Scalping and Day Trading." Selected Writings of Holbrook Working, Chicago Board of Trade. 181-193. (1977).

**Table 1. Demographics of Member Firms in the Dalian Futures Exchange.**

	<b>Non-Brokerage:</b> 15 Firms	<b>Brokerage:</b> 125 Firms
<b>Customers</b>	15	6213
<b>Year-end zero net position</b>	3	5357
<b>Year-end non-zero net position</b>	12	856
<b>Scalpers</b>	0	20

**Table 2. Trader Statistics in the Dalian Futures Exchange.**

	Number of Traders	Number of Scalpers
Jun-99	2545	8
Jul-99	2420	7
Aug-99	3053	8
Sep-99	4119	8
Oct-99	4122	8
Nov-99	5390	9
Dec-99	4371	5

**Table 3. Predicting Scalper Presence: Estimates from Univariate ACI Models**

Parameter	Contract			
	May 2000		July 2000	
	Estimate	p-value	Estimate	p-value
$y_{t-1}$	0.14 (0.01)	0.000	0.12 (0.01)	0.000
$\log(\lambda_{t-1})$	0.71 (0.01)	0.000	0.72 (0.01)	0.000
$SS_{t-1}/100$	3.41 (6.57)	0.605	-8.47 (6.63)	0.201
$\sigma_{t-1}/100$	5.00 (1.03)	0.000	4.91 (0.95)	0.000
$v_{t-1}^k/100$	-5.83 (16.74)	0.728	-113.90 (15.76)	0.000
$v_{t-1}^n/100$	-3.62 (0.95)	0.000	6.12 (1.73)	0.000
$I_{t-1}^k/1000$	-8.97 (2.51)	0.000	2.46 (1.80)	0.171
$I_{t-1}^n/1000$	0.85 (0.16)	0.000	1.44 (0.26)	0.000
Log-likelihood	-4164.212		-3767.850	
No. Coeffs.	35		35	
Avg. LL	-1.545		-1.398	
<b>Poisson Regression Estimates</b>				
Log-Likelihood	-4185.788		-3795.074	
No. Coeffs.	55		55	
Avg. LL	-1.553		-1.408	
Observations	2700			

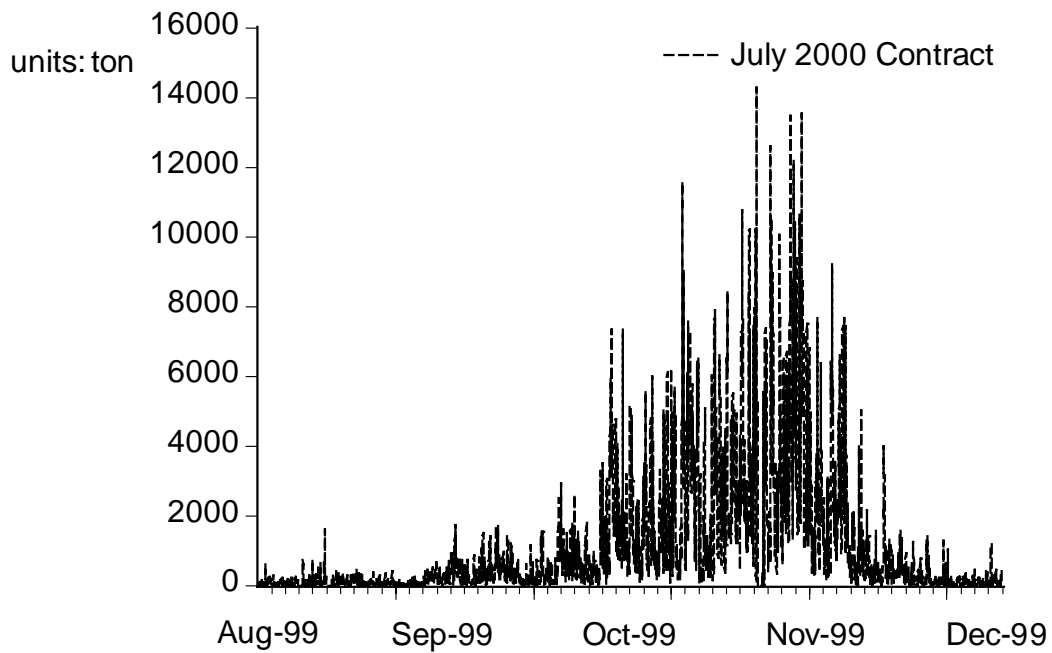
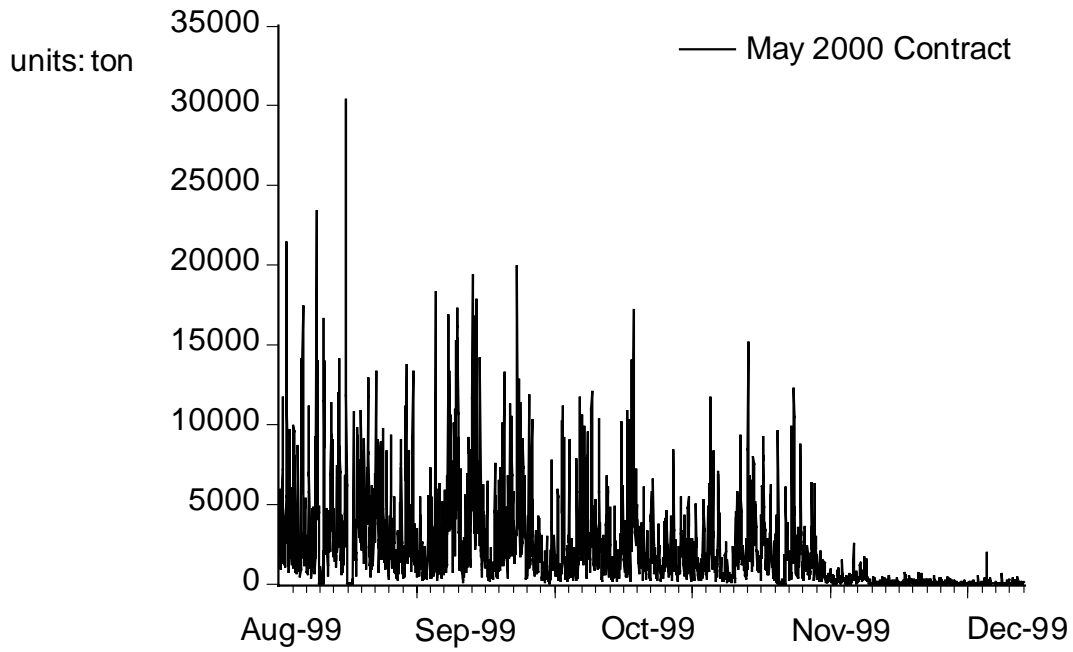
**Table 4. Predicting Scalper Presence: Estimates from the Bivariate ACI Model**

	Contract				
	Parameter	May 2000		July 2000	
		Estimate	p-value	Estimate	p-value
<b>MAY 2000 Regressors</b>	$y_t$	-	-	0.10 (0.01)	0.000
	$\log(\lambda_t)$	-	-	-0.11 (0.04)	0.001
	$y_{t-1}$	0.09 (0.01)	0.000	-0.01 (0.01)	0.431
	$\log(\lambda_{t-1})$	0.86 (0.01)	0.000	-0.02 (0.04)	0.589
	$SS_{t-1}/100$	-0.94 (0.61)	0.120	0.33 (0.55)	0.552
	$\sigma_{t-1}/100$	2.44 (1.27)	0.054	3.86 (1.13)	0.001
	$v_{t-1}^k/100$	-27.46 (17.81)	0.123	-23.97 (16.84)	0.155
	$v_{t-1}^n/100$	-1.76 (0.98)	0.073	-1.84 (0.91)	0.042
	$I_{t-1}^k/1000$	-5.02 (2.64)	0.058	-6.28 (2.47)	0.011
	$I_{t-1}^n/1000$	0.36 (0.17)	0.035	0.30 (0.16)	0.058
<b>JULY 2000 Regressors</b>	$y_{t-1}$	0.10 (0.01)	0.000	0.09 (0.05)	0.000
	$\log(\lambda_{t-1})$	-0.22 (0.03)	0.000	0.80 (0.00)	0.000
	$SS_{t-1}/100$	-1.58 (0.58)	0.006	-0.32 (0.52)	0.536
	$\sigma_{t-1}/100$	4.97 (0.93)	0.000	4.88 (0.85)	0.000
	$v_{t-1}^k/100$	-91.74 (16.67)	0.000	-82.43 (15.20)	0.000
	$v_{t-1}^n/100$	-73.24 (17.13)	0.000	-57.78 (15.25)	0.000
	$I_{t-1}^k/1000$	-1.12 (2.01)	0.576	-3.39 (1.79)	0.059
	$I_{t-1}^n/1000$	1.73 (0.25)	0.000	1.45 (0.22)	0.000
Log-Likelihood				-7664.709	
Number of Coefficients				88	
Average Log-Likelihood				-2.843	

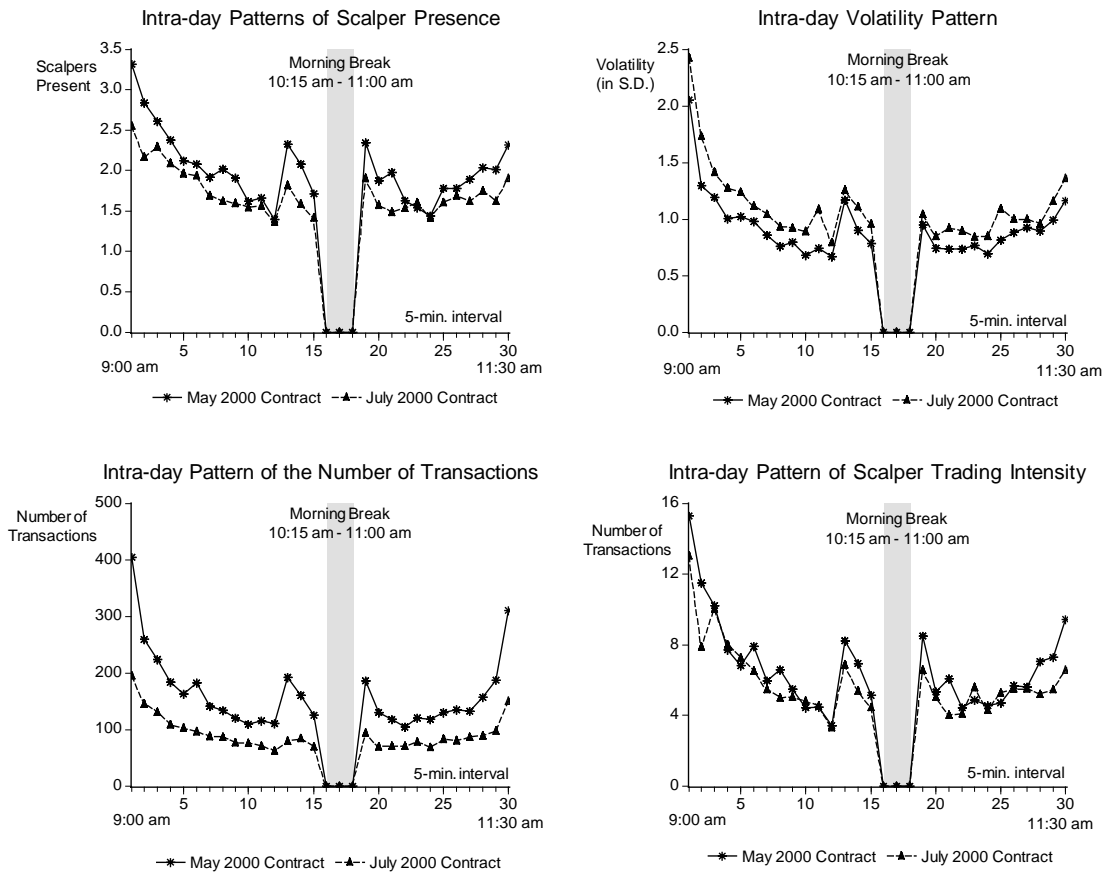
**Table 5. Predicting Mr. D's Transactions: Estimates from the Bivariate ACI Model**

		Contract			
		May 2000		July 2000	
	Parameter	Estimate	p-value	Estimate	p-value
<b>MAY 2000 Regressors</b>	$y_t$	-	-	0.04 (0.02)	0.116
	$\log(\lambda_t)$	-	-	0.16 (0.07)	0.018
	$y_{t-1}$	0.17 (0.02)	0.000	-0.06 (0.03)	0.062
	$\log(\lambda_{t-1})$	0.60 (0.04)	0.000	-	-
	$SS_{t-1}/1000$	-0.007 (0.028)	0.800	0.03 (0.03)	0.202
	$\sigma_{t-1}/100$	0.13 (0.03)	0.000	-0.06 (0.03)	0.024
	$v_{t-1}^k/100$	-0.91 (0.43)	0.036	1.10 (0.36)	0.002
	$v_{t-1}^n/100$	-0.05 (0.03)	0.048	-0.04 (0.02)	0.113
	$I_{t-1}^k/1000$	0.006 (0.006)	0.317	-0.02 (0.00)	0.000
	$I_{t-1}^n/1000$	0.001 (0.000)	0.018	0.001 (0.000)	0.008
<b>JULY 2000 Regressors</b>	$y_{t-1}$	0.02 (0.02)	0.408	0.08 (0.01)	0.000
	$\log(\lambda_{t-1})$	0.06 (0.05)	0.167	0.65 (0.04)	0.000
	$SS_{t-1}/1000$	0.04 (0.02)	0.0963	-0.07 (0.02)	0.000
	$\sigma_{t-1}/100$	-0.00 (0.03)	0.975	0.10 (0.02)	0.000
	$v_{t-1}^k/100$	-0.42 (0.42)	0.314	-0.34 (0.28)	0.230
	$v_{t-1}^n/100$	-0.08 (0.04)	0.031	-0.18 (0.03)	0.000
	$I_{t-1}^k/1000$	-0.007 (0.005)	0.182	0.01 (0.00)	0.000
	$I_{t-1}^n/1000$	0.0003 (0.001)	0.000	0.002 (0.000)	0.000
Log-Likelihood		-4610.234			
Number of Coefficients		87			
Average Log-Likelihood		-1.710			

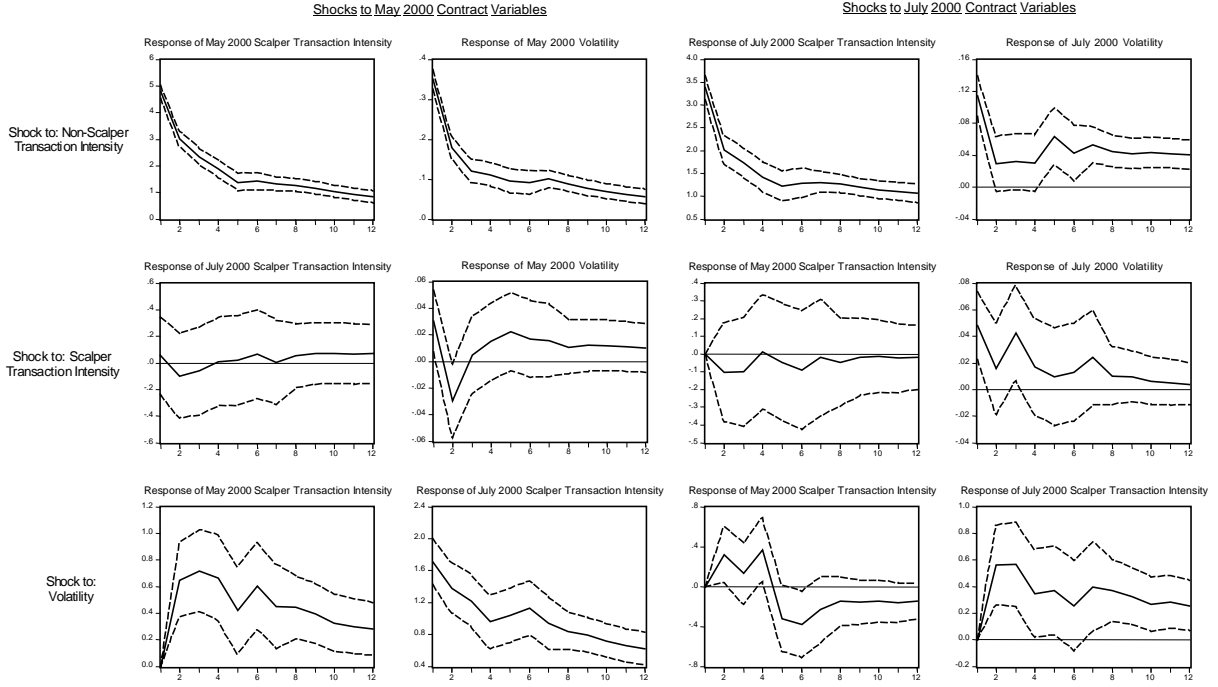
**Figure 1. Trading Volumes for Soybean Futures Contracts**



**Figure 2. Seasonal Intra-Day Patterns**

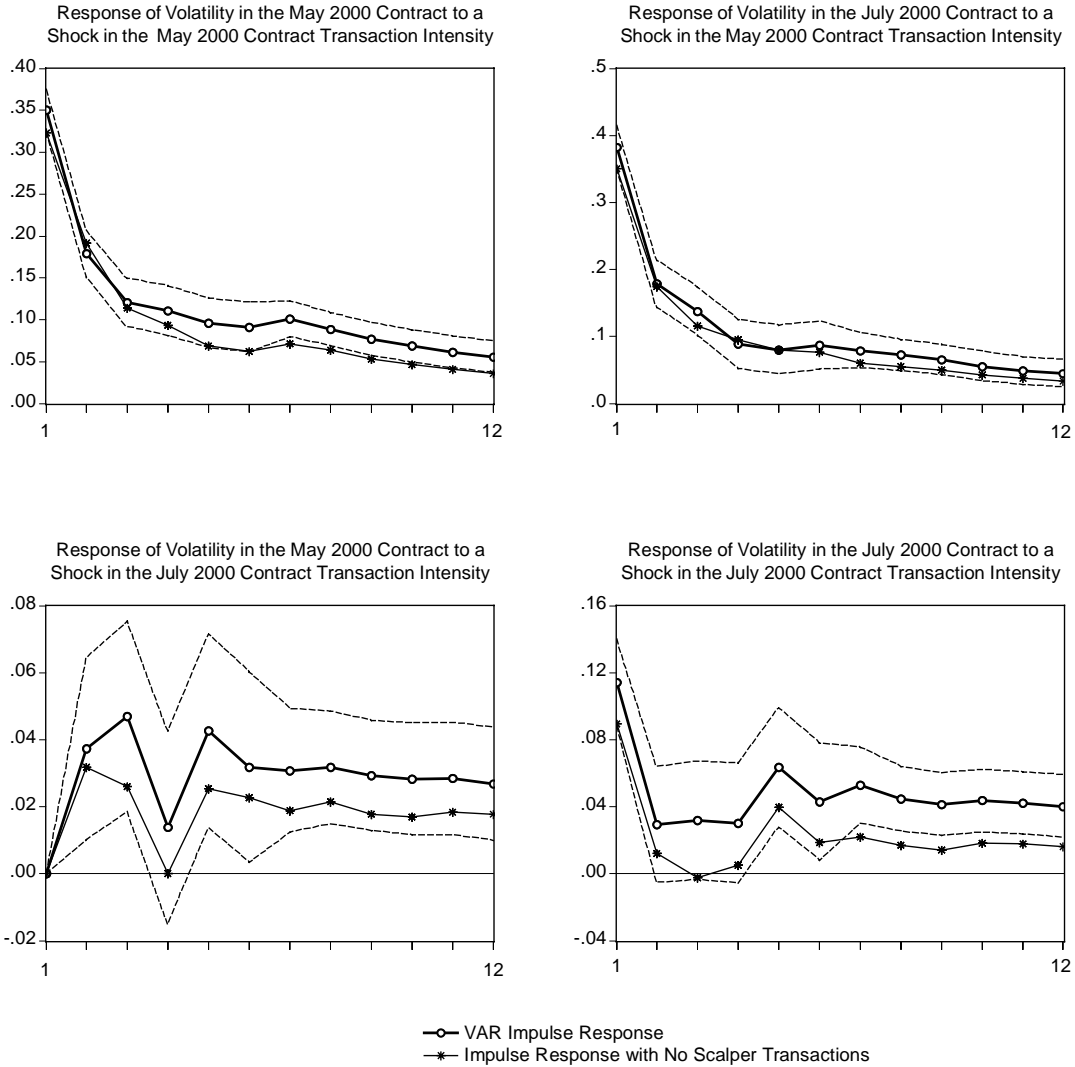


**Figure 3. Impulse Responses from the Structural VAR**



Selection of impulse response functions to a one standard deviation shock to each of the variables in the system:  $\mathbf{w}_t' = (\mathbf{x}_{Mt}', \mathbf{x}_{Jt}')$  where  $\mathbf{x}_{it}' = (\sigma_i, I_t^n, I_t^k)$  for  $i = M, J$ . The component variables in  $\mathbf{x}$  are respectively: volatility, non-scalper transaction intensity and scalper transaction intensity.  $M$  denotes the May 2000 contract variables,  $J$  denotes the July 2000 contract variables. Dotted lines represent the associated asymptotic two standard error bands. These responses are traced for up to one hour after the shock (12 five-minute intervals).

**Figure 4. Counterfactual Experiment: Responses of Volatility to a Shock in Non-Scalper Transaction Intensity.**



Impulse response functions of volatility to a one standard deviation shock to non-scalper transaction intensity from the VAR for  $\mathbf{w}_t' = (\mathbf{x}_{Mt}', \mathbf{x}_{Jt}')$  where  $\mathbf{x}_{it}' = (\sigma_t, I_t^n, I_t^k)$  for  $i = M, J$ . The component variables in  $\mathbf{x}$  are respectively: volatility, non-scalper transaction intensity and scalper transaction intensity.  $M$  denotes variables for the May 2000 contract,  $J$  denotes variables for the July 2000 contract. Bold solid line with circular symbol denotes the usual impulse response. Solid line with star symbols denotes the same response but when the scalper response is shut-down. Dotted lines represent the associated asymptotic two standard error bands. These responses are traced for up to one hour after the shock (12 five-minute intervals).