# Trading Fast and Slow:
# Security Market Events in Real Time

**Joel Hasbrouck**

February 19, 1999
Preliminary Draft
Not for Quotation
Comments Welcome

Professor of Finance
Stern School of Business
New York University
Suite 9-190 Mail Code 0268
44 West Fourth St.
New York, NY 10012-1126

Tel: (212) 998-0310
Fax: (212) 995–4901
E-mail: jhasbrou@stern.nyu.edu
Web: http://www.stern.nyu.edu/~jhasbrou

The latest version of this paper, data files and SAS programs are available on my web site. All errors are my own responsibility.

# Trading Fast and Slow: Security Market Events in Real Time


## Abstract

Continuous security markets evolve as a sequence of timed events. This study is a descriptive analysis of NYSE market data in which trades, quote revisions and orders are considered to constitute a stationary multivariate point process, which can be analyzed by standard time- and frequency-domain techniques. There are three principal findings. (1) Although occurrence intensities for different types of events are positively correlated, they are not characterized by the uniform proportionality that a strict sense of time deformation would require. (2) The frequencies and durations of informational epochs (periods of uncertainty and informational asymmetry) are highly variable. (3) The correlation in arrivals of market orders and opposing limit orders is zero or negative over periods of thirty minutes or less.

## 1. Introduction

The specification and interpretation of time is fundamental to theoretical and empirical models of security market dynamics. From a purely descriptive viewpoint, it is obvious that real-world continuous security markets give rise to well-defined events (e.g., orders, trades and quote revisions) that occur randomly in continuous real time. If we seek tractable models of market behavior, however, this observation is not particularly helpful. The difficulties of theoretically modeling or empirically characterizing such processes strongly motivate simpler approaches, such as emphasizing one type of event (usually trades), aggregating data over convenient intervals of real ("wall-clock") time or viewing the events as ordered (but not timed).

While greatly facilitating empirical analysis, these simplifications suppress aspects of the situation that may be economically important. Market operations inherently comprise multiple dependent processes. Time aggregation smears the individual events and aggravates simultaneity. An untimed ordering of the data leaves no place for real-time arrival rates. These occurrence rates lie at the heart of what a practitioner might describe as a "fast market", circumstances that might feature an "avalanche" of orders, "waves" of program trades, and/or quotes that are stale almost as soon as they are posted.

In light of these considerations, this paper pursues a line of inquiry that views the collection of market events as a multivariate point process, and applies the descriptive statistical tools that this perspective affords. The study uses the NYSE's TORQ data, a comprehensive record of not only trades and quotes, but also orders, differentiated according to buy/sell, market/limit and program/nonprogram. The inquiry applies standard time- and frequency-domain analysis to the point and count data associated with these series. I discuss in detail the activity for IBM, but supplement this with results for a broader sample.

This approach might initially appear to be preoccupied with real time to the exclusion of all other aspects of the problem. Indeed, in the econometrics, all attributes ("marks") of the events apart from their occurrence are ignored. The paper analyzes trade occurrence times, for example, but not the prices or volumes of the trades. It does not follow from this admission, however, that the characterization of occurrence rates is irrelevant for the economic and statistical modeling of prices and volumes. In fact, real time plays an important role in these models.

The present analysis is significant in several respects. First, it affords important insights into the microstructure foundations of time deformation. The principle of time deformation in security markets differentiates between real time and operational or informational time. Analyses of mixture of distributions and/or price/volume effects often invoke this distinction. This paper suggests that time deformation be characterized as commonality in the occurrence rates of the events that collectively constitute the market process.

By way of physical analogy, each of the cylinders in an automobile engine contains a spark plug that fires at a certain point in the engine's revolution. Suppose that, knowing nothing of this mechanism, we analyze a dated and timed record of all the spark plugs' firings for a week. Random usage of the automobile will lead to extensive time variation in the occurrence rates (or time between firings) for each and all plugs. Nevertheless, local estimates of these rates for different plugs will be very strongly correlated. This is time deformation, in which operational time is the instantaneous rate of the engine revolution. Analogously, for market data, we seek to determine the nature of correlations between different events. The paper finds that that while correlations between the arrival rates of orders, trades and quote revisions on NYSE are generally positive, they are far less pervasively so than the automobile analogy would suggest. This implies that the relative roles of these processes are different in fast and slow markets.

A second important insight concerns the characterization of private information. Most theoretical models of trading are set in a notional time frame (termed here an information epoch) that begins with a valuation uncertainty and an information asymmetry, and ends with resolution of the uncertainty. In constructing robust empirical specifications for these models, however, we have few guidelines concerning how these informational epochs might map into real time. The models do suggest, however, that the order flow over an epoch is one-sided. Analyzing the point process characteristics of buy and sell orders is a natural way to characterize these epochs. Here, the findings are mostly negative. Auto- and cross-correlations in order flow vary considerably, suggesting that sequences of predominately one-sided runs vary in frequency and duration. This largely rules out a concise summary characterization of information epochs.

A third insight involves the demand and supply of liquidity. In a continuous double auction market such as the NYSE, liquidity hinges on the willingness of outsiders to maintain a flow of new limit orders (to replenish the book). From a public policy viewpoint, one might hope that when market orders arrive rapidly, so do limit orders. The results suggest that this is only partially the case. For opposing market and limit orders (e.g., market buys and limit sells), the relationships between short and medium term components (periods less than thirty minutes) are mostly zero or negative. Thus, small-scale liquidity crises appear to arise frequently.

The paper is organized as follows. The next section discusses how time (both real and notional) figures in market models, and motivates the paper's perspective. Section 3 describes the data. Orders, trades and quotes are analyzed in Section 4, establishing the evidence in support of (and limitations of) time deformation. The paper then turns to a more detailed examination of incoming orders in Section 5. A discussion of implications and summary concludes the paper in Section 6.

## 2. The role of time in security market dynamics: an overview

Most of our knowledge of security price behavior, especially as it relates to broader economic activity, is based on real time observations with data aggregated over periods (daily or longer) that are large relative to trading time frames. Theoretical and empirical analyses of trading, on the other hand, are most conveniently set in a notional time scale that sequences agents' moves or observable market events. Although paper ultimately aims at bridging these two views of time, it is initially simpler to treat them separately. The discussion here first considers time deformation (broadly defined), a device that arises chiefly in the empirical analysis of real-time aggregated data. It next turns to the microstructure literature, and the manner in which the notional time scales favored in that literature might map to real time.

### *a. Time Deformation*

Many models of security prices allow fixed intervals of real time to encompass varying amounts of "information". The intuition here is that since information drives prices, price dynamics are most concisely specified when the rate of information arrival or information intensity is constant. Allowing real time observation periods to embody compressed or expanded informational periods is then a convenient way to account for certain statistical properties of real time observations.

The distinction between the operational time scale in which a system evolves and the real observational time scale is the essential principal of time deformation. Formally, a multivariate latent process $\xi(s)$ evolves in continuous operational time $s$ (here equivalent to information time). A transformation, $t = g(s)$, maps operational time to real time $t$ (Stock (1988), or similarly, Ghysels, Gourieroux, and Jasiak (1998a)). Although few analyses of security market dynamics estimate a fully specified time deformation model, the intuition is nonetheless pervasive.

*Mixture of distributions and price/volume studies*

Historically, the most prominent applications of time deformation have been to empirical studies of short-term time-aggregated stock market data. Most importantly, if regular real-time intervals are allowed to encompass varying spans of informational time, returns over these intervals will be distributed as mixtures of the informational/operational-time distributions. The generation of a complicated distribution via random parameterization of a simpler one is, in and of itself, a purely mathematical device. The idea acquires economic content in this context in its implications for other economic variables (besides prices) that might also presumed to be driven by information.

The most important of these is trading activity. Numerous studies examine returns and trading volume (variously measured as the number of trades or units of the security exchanged) over regular real-time intervals. Statements of this "mixture of distributions" hypothesis are usually formulated in this framework as joint hypotheses over returns and volume. A related and overlapping "price-volume" literature has evolved exploring the broader descriptive aspects of this relationship.[1]

The early economic models of price-volume relationships emphasize agent preferences and information structures (Epps and Epps (1976), Tauchen and Pitts (1983)). This establishes, sensibly in the spirit of Occam's razor, a minimal framework for generating prices and trading volumes. In these models, the price change and volume of trade are usually assumed to arise in the course of the market's response to an informational event. Furthermore, the number of informational events in a day (or other regular interval) is assumed random. This device allows for time deformation, in the sense of variation in the average daily rate of information arrival.

The trend in this literature is toward models that are less time-aggregated and that incorporate the intuitions of microstructure analyses. The analyses of Foster and Viswanathan (1995), Andersen (1996) and Llorente et al. (1998), for example, view volume effects as arising in part from asymmetric information. The market clearing processes in these papers, however, generally involve one-time clearings that are rapid enough not to impinge on any subsequent information event.

As the time intervals of interest become finer, however, the demands on the empirical modeling increase dramatically. Over a day, for example, it is reasonable to measure trading activity by the total volume, which can plausibly be considered a continuous random variable.

---

[1] The formalities of time deformation are usually stated in terms of subordindated stochastic processes (Bochner (1960)). Among the numerous studies of security prices relevant to the present paper are: Clark (1973), Gallant, Rossi, and Tauchen (1992), Harris (1986); Harris (1987), Karpoff (1987), Mandelbrot and Taylor (1967), Richardson and Smith (1994), Tauchen and Pitts (1983)).

At a one-second resolution, trading undeniably consists of discrete events. This motivates a discussion of empirical analyses that explicitly model the point-process character of markets.

*Empirical studies of market events as point processes*

The simplest point process is the Poisson process with constant intensity. Garbade and Lieber (1977) analyze a short sample of NYSE transaction times for two stocks (Potlatch and IBM). They find the Poisson assumption to be roughly appropriate, but violated by bursts of frequent arrivals. From a time deformation perspective, this suggests periods of high information intensity, but in this model, the trade arrival process is independent of the stock's value.

Recent work, facilitated by better data sets and methodological advances, has achieved significant refinements and extensions. Engle and Russell (1998) propose an autoregressive conditional duration (ACD) model, a parametric specification that admits stochastic time-variation in the expected occurrence rate. They find that the model captures the clustering of IBM trades in real time.

Engle and Russell point out that although the ACD specification is not, strictly speaking, a time deformation model (no operational/real time correspondence is established), it nevertheless admits a similar interpretation. Most importantly in this context, clustered trading (low intertrade duration) is associated with high volatility (Engle (1996); Russell and Engle (1998)). Furthermore, persistence in conditional intensity may help explain volatility persistence (Ghysels, Gourieroux, and Jasiak (1998b); Ghysels and Jasiak (1998)).

ACD models are powerful and useful devices for characterizing market point process phenomena. For present purposes, however, the ACD framework is not completely suitable. One limitation concerns the applicability to multivariate point processes. Although Engle and Lunde (1998) present a bivariate point process model for trades and quotes, the specification is not symmetric in the two processes ("The transaction times are the forcing process.") Extending this approach to more than two processes would appear to require the specification of a recursive hierarchy of timing influences. Furthermore, although the parametric simplicity of the ACD model is in many respects a strength, it may obscure features of the data that might be more visible in a descriptive analysis.

*Time deformation in market events*

For empirical purposes, this paper suggests that time deformation in multivariate point-process market data be measured as the variation in occurrence rates that is common across all component point processes. This definition draws on two themes in the above-cited literature. First, from the event-modeling literature, come the perspectives that the collection of market events is a multivariate point process and the occurrence intensities are indicative of the rate of information arrival. Second, from the formal definitions, time deformation implies a *single* directing process (mapping from operational to real time) that is applicable to all components of the system.

This working characterization of time deformation implies positive correlations among the intensities of the component processes. These intensities, however, are not observed. In the absence of a parametric model, the intensities must be locally estimated, either over short windows of time or (as will be discussed later) frequencies. Thus, time deformation defined in this fashion depends fundamentally on horizon.

By way of example, suppose that information intensity is proxied by a nonnegative continuous random variable, $I_t$, which evolves as the Ornstein-Uhlenbeck process $d\log(I_t) = \alpha(\log(I_t) - \theta)dt + \sigma\, dz$. Let the trade and quote point processes be Poisson with intensities proportional to this variable: $\lambda_{Trade} = c_1 I_t$ and $\lambda_{Quote} = c_2 I_t$ where $c_1, c_2 > 0$. Furthermore, assume that conditional on $I_t$, the two processes are independent. This system exhibits the characteristics of time deformation in the present sense: variation in intensities common to both components. This does not imply, of course, that actual occurrences over any arbitrary interval are perfectly correlated. In fact, as the width of the interval shrinks, the correlation vanishes. (Statistical inference is discussed Section 4 and the appendix.) This example easily generalizes to more than two processes.

This example is useful in motivating the empirical analysis. Its treatment of market events, however, is statistical, not economic. Accordingly, I now turn to a discussion of the microstructure literature.

### *b. Microstructure models and their timing characteristics*

In most respects and in most microstructure models, the timing structure is an assumed, stylized framework designed to capture certain real-world features in a tractable fashion. Most models are set in a notional time frame, termed here an "informational epoch", that begins with the existence of value uncertainty (and possibly an information asymmetry), and ends with the resolution of the uncertainty.

Economic considerations suggest considerable variation in the timing of informational epochs. Most microstructure models were originally motivated as applying primarily to corporate equity securities. The informational asymmetries generally relate to "terminal" value, i.e., value at the end of or beyond the horizon of trading. Informed trading is sometimes taken as synonymous with "inside" trading. These considerations suggest a time frame approximating that of major corporate information development, quarterly, perhaps, excepting the odd merger announcement or restatement of past earnings.

Other considerations suggest a much briefer period. If traders who place limit orders in the market do not monitor public information closely, their orders may not accurately reflect public information. "Private information" exists in this market, in the sense that traders' information sets as reflected in their market interactions differ (Hasbrouck (1991)). Of course, the information epochs associated with this structure would tend to be of short duration. Furthermore, private information (and the attendant price impacts) need not be related to terminal value. It may instead arise from transient market imbalances or uncertainty about the trading population, both of which may be short term (Lyons (1997); Lyons (1996); Lyons (1995); Saar (1998)).

The concept of an informational epoch, though somewhat artificial, is a useful one. It is pervasive in the microstructure literature. It is common in the price/volume literature, as well, embracing the random periods of trading that often arise in these models. If we broaden the concept slightly to refer the period over which there is existence and resolution of the uncertainty/asymmetry in one component of the security's value (as opposed to the totality), it becomes realistic as well. That is, we can view the trading process as consisting of informational epochs of varying duration, possibly overlapping. The device is useful empirically because the strongest content of most microstructure models pertains to activity within an epoch.

*Empirical timing implications of the sequential trade models*

In the sequential trade models, an information epoch is generally characterized by a one-sided order flow, which arises from informed trading over the epoch. In some analyses, this is an implication of the model deriving from assumptions about preferences, endowments and information (e.g., Glosten (1989); Glosten (1994); Glosten and Milgrom (1985)). In other models, this behavior is simply assumed.

If the distinguishing feature of an informational epoch is a one-sided order flow, we should observe over the epoch a high occurrence rate on one side of the market. This may be distinguished from a fast market due to time deformation in that the increase in occurrence intensity is not pervasive across all components of the system. In particular, the sequential trade models suggest no necessary surge in orders on the opposite side of the market.

The setup in Easley et al. (1996) is particularly suggestive. If an information event has occurred, then the informed agents trade on the side of the signal with Poisson intensity $\lambda$. Liquidity traders buy and sell with intensity $\varepsilon$. The unconditional probability of an information event in a particular epoch is $\alpha$, and the probability of a low signal is $\delta$, which is here assumed to be one-half. Under these assumptions, the unconditional buy and sell order intensities both equal $\varepsilon + (\alpha/2)\lambda$.

Easley and O'Hara assume that the market evolves as a sequence of independent information epochs of one day's duration placed end-to-end. This is roughly appropriate in the case where news consists of an after-hours corporate announcement that is leaked early in the day. From a broader empirical perspective, however, this restriction seems unrealistically confining. As noted above, there are good reasons to suspect that the durations of information epochs vary considerably, from the relatively brief (a few minutes, in the case of stale limit orders) to the lengthy (many days, in the case of advance knowledge of an earnings announcement). There are also, of course, few reasons to presume a priori that epochs occur back-to-back with no intervening trading, or that they begin at regular times.

In the absence of plausible assumptions concerning the exact timing of epochs, more general descriptive analyses may be useful. In the present context, the statistical point process approach offers a good characterization of the horizons over which occurrence rates of orders on opposite sides of the market are correlated.[2]

The sequential trade models have additional implications for event processes within an information epoch. The mapping from notional to real time is largely irrelevant as long as the sequencing of moves is preserved. In the basic versions of these models, these moves correspond to observable events and the market operation follows a regular cycle. A market maker posts quotes; an order arrives; a trade occurs; and the market maker updates the quotes. In such a market, we would expect the point process of trades and quotes to be identical to that of

---

[2] The principal alternative to the sequential trade line of modeling employs auction models (Kyle (1985); Kyle (1984) and related developments). Although these are in the limit set in continuous time, the point-process character of market events is not modeled. At each instant, for example, equilibrium is characterized by an order flow that is normally distributed. These models have, therefore, no direct implications for trading events per se. The models possess the additional property that the cumulative signed order flow is a Martingale, which implies an absence of autocorrelation.

orders, possibly displaced slightly by processing delays. Consistent with the characterization of time deformation in the last section, variation in the occurrence rate of any single component (all orders, quotes or trades) would be common to all components.[3]

More sophisticated models, however, do not necessarily imply such regularity. If there is uncertainty about whether an information event has occurred, Easley and O'Hara (1992) show that a "no trade" move conveys information (specifically, that an information event has most likely *not* occurred). Observable events thus constitute a censored sample of agents' moves. The former won't be evenly spaced in time (even if the latter are).[4]

The Easley and O'Hara (1992) model also suggests that upon receiving a large order, the dealer increases her conditional probability that an information event has occurred, and widens the spread accordingly. In a limit order market, this same effect may arise as a reluctance of limit order traders to refill the book after a large order. This hypothesis may be examined indirectly by comparing the time patterns in arrival rates of market orders and opposing limit orders.

### *c. Summary of empirical implications*

The preceding discussion suggests specific issues that are usefully framed in a point-process perspective:

---

[3] It is interesting that from a microstructure viewpoint, the empirical price-volume relation is problematic *within an informational epoch*. O'Hara (1995) notes that relation does not arise in the basic continuous auction or sequential trade models. These models do not in any obvious way, however, rule out the devices used in the price/volume literature, such as allowing variable numbers of epochs within a real time interval.

[4] It is worth noting that in Easley and O'Hara (1992), the market maker knows the times at which trade might potentially occur. That is, a no-trade observation is equivalent knowing that a potential trader has arrived in the market, observed the quotes, and has declined to trade. This might be approximated in real markets where the public quotes are merely indicative and trading requires negotiations that might break down. Most equity markets, however, publicize quotes that are firm (at least for small trades). In these situations, the market maker would seem to be in the position of inferring the information event from local order arrival rates. In the empirical implementation, Easley and O'Hara adopt a fixed rule that five minutes without a trade constitutes a "no-trade" move.

- To what extent can variation in market activity be characterized as time deformation? That is, can different sorts of market events be characterized as time-homogeneous point processes subject to the same deformation? This question will be investigated by considering the extent of correlation between the occurrence intensities of different events at various horizons.
- What are the frequencies and durations of information epochs? This can be investigated by determining the horizons over which incoming buy and sell orders are not positively correlated.
- How is liquidity affected within informational epochs? This is a broader hypothesis and will be investigated by examining the timing relation between market orders and opposing limit orders.

The paper now turns to an investigation of these relations.

## 3. Data

The data are drawn from the TORQ dataset (Hasbrouck (1992)). This is a three-month sample of orders, trades and quote revisions for consolidated trading in selected NYSE stocks. The trade and quote data are comprehensive. They contain all transactions and quote revisions in all trading venues covered by the consolidated information systems. The order data, however, cover only orders handled on the NYSE's automated systems (chiefly the SuperDOT system). Harris and Hasbrouck (1996) provide an extensive characterization of these orders.

The time-stamp associated with a trade is imposed by the Consolidated Trade System (CTS). It marks when the trade was reported (not when it actually occurred). For present purposes a "quote" is considered to be a revision in the bid price, ask price, bid size (number of shares "sought" at the bid) or ask size (number of shares "available" at the ask). The time stamp is fixed by the Consolidated Quote System (CQS).

Among the many attributes of orders, the most important are side (buy or sell) and market vs. limit. A market order demands immediate execution at the best available price. Market orders are commonly spoken of as consuming liquidity. A limit order is a priced order (e.g., buy 500 shares at a price of 40 or better) for which execution is not certain. Limit orders are sometimes said to supply liquidity. The time stamp on an order is the time when the order was received by the NYSE's Common Message Switch (CMS). If the price of a buy limit order is above the prevailing ask price, the limit order is said to be marketable. Unless the market is rapidly moving, the order is functionally equivalent to a market order. Marketable limit orders are grouped with market orders for present purposes.

The systems that time-stamp the data are identified above by name simply to emphasize that they are different. Although they are all synchronized to the same clock, processing latencies may induce errors.

Most of the statistical analysis focuses on IBM (the same firm analyzed in Engle and Russell (1998)). There are sixty-three days in the TORQ sample. Of these, four were deleted.[5] Also investigated were the sixteen most actively traded stocks in the TORQ sample.

Figure 1 plots occurrences of trades, quotes and orders of various types for IBM for the first day in the sample. Each occurrence is plotted as a point on one of six time lines (depending on the type of event). Points are vertically "jittered" (i.e., shifted up or down by a small random amount) to visually spread out observations that would otherwise (at plotting resolution) coincide. The jittering is solely in the vertical dimension: a point's horizontal position is an accurate time coordinate.

The graph exhibits several interesting features. First, despite the general rule that most aspects of market activity are "U-shaped", i.e., elevated at the beginning and end of trading, this is not strikingly obvious in the picture, except perhaps for a general slowing down over the lunch hour. This simply reflects the fact that random variations in intensities are much larger than the deterministic intraday variations. The U-shape is much more apparent when intensities are

---

[5] There was an early closing on December 24, 1990. System failures on November 23, 1990 and December 27, 1990 impaired trading for parts of those days. On January 17, 1991, IBM experienced a delayed opening.

averaged across days. Secondly, each type of event appears to exhibit random clustering (times of greater occurrence density). Thirdly, the clustering periods are not apparently uniform across types of events.

## 4. Orders, Trades and Quotes

This section investigates microstructure time deformation using three essential events related to market activity: incoming orders (of all types), trades and quote revisions. I examine the processes from both time and frequency-domain perspectives. The former is more intuitive, while the latter offers a sharper characterization of effects at various horizons. I outline in this section the features of the statistical framework. The appendix contains a more complete summary.

### *a. Multivariate point processes in the time domain.*

The time-stamped record of market events studied here is assumed, after removal of intraday "seasonalities", to be a sample from a stationary vector point process (Brillinger (1974); Brillinger (1975)). The interpretation of the stationarity condition is similar to that of ordinary time series: joint probabilities of occurrences at times $t$ and $t+s$ depend only on $s$ (and not on $t$). Although events occur in continuous time, we can investigate the process using discrete-time averages of the occurrence rates, as long as the averaging intervals are short relative to the phenomena of interest. The appendix discusses the relation between the properties of these local averages and those of the underlying continuous process. This study averages events over 30-second windows.[6]

---

[6] A point process is said to be orderly if there is a vanishing probability that a time interval of infinitesimal duration contains more than one event. The present data are not, at the one-second resolution of the time stamps, "orderly". In the sixty-two day TORQ sample for IBM, there were over 14,000 one-second intervals that contained multiple events (orders, trades and/or quote revisions). The largest number of "simultaneously-occurring" events was twenty-five events (on January 2, 1991 at 13:03:59).

To facilitate the computations, the study focuses primarily on intraday stochastic behavior. The 30-second average intensities were demeaned (by removal of the mean for the day). Intraday deterministic components ("seasonalities") were removed by regressing the average intensities against time-of-day polynomials (and using the residuals in subsequent analysis).

### *b. Results for IBM in the time domain*

Figure 2 depicts the auto- and cross-correlation for order, trade and quote intensities. Average intensities for all series are positively autocorrelated. Intuitively, all events tend to be clustered in time.

If the market followed a regular cycle such that an occurrence of any one type of event were always associated (possibly at a fixed lag or lead) with occurrences of the other types, then the cross correlations would be unity at the given lag or lead and zero otherwise. Not surprisingly, the cross-correlations refute this simplistic sort of time deformation. The cross-correlations are, however, pervasively positive. This suggests the presence of common components that are (at least relative to the 30-second windows) long term. Thus, the analysis must turn to considering the horizon of the effects. In exploring such issues, the frequency domain perspective is useful.

### *c. Multivariate point processes in the frequency domain*

Like an ordinary time series, a vector point process possesses a spectral (frequency domain) representation, in which the randomness is viewed as arising in the amplitudes of the component sinusoid (sine and cosine) functions that constitute the process. Because a sinusoid has a well-defined period of repetition, the frequency domain perspective is useful for characterizing variances and covariances over different time horizons. In the present application, it is economically reasonable to conjecture that market event intensities are driven in part by short-term effects, such as a trader's delay in getting to the phone or (nowadays) the latency time in an Internet connection. Nevertheless, there are also long term components arising from relatively infrequent news announcements. The question of "horizon" is fundamental to the present characterization of time deformation, defined as positive covariation among components of a multivariate point process. The frequency domain perspective permits a concise characterization of these effects.

Although spectral analysis is often used in the physical sciences to uncover deterministic regularities (seasonalities), and could be used to capture the intraday patterns in market data, this is not the primary purpose of the present analysis. The principal intent is to provide an alternative perspective on the *stochastic* properties of the data series. Although a market practitioner might refer nowadays to "waves of program orders", I suspect that the metaphor is meant to convey strength and inevitability (once under way) rather than cyclic predictability. In any event, no presumption of cyclic regularity underlies the present analysis.

To illustrate the sort of inferences facilitated by spectral representations, I consider two examples.

*Example 1: Poisson orders leading to trades with a random delay.*

In this stylized simulation, I assume that in each discrete time period, the number of arriving "orders" is Poisson with mean 0.2. The orders that arrive in time period $t$ result in an equal number of "trades", but due to various random latencies in the order handling process, the trades are reported with a delay. This delay is assumed to be exponentially distributed with a mean of three periods. I simulate 10,000 time periods and compute time- and frequency-domain statistics.

Auto- and cross-correlations are graphed in Figure 3 (Panel A). The univariate autocorrelations in the counts are near zero (as one would expect for a Poisson process). The cross-correlations reflect the mechanism linking orders and trades. Due to the random delay, the positive association is smeared over several lags. Estimated spectra and cross-spectra are graphed in Figure 3 (Panel B). Both univariate spectra (on the diagonal) are virtually flat due to the constant Poisson assumption. The cospectrum decomposes the covariance in the arrival probabilities into components associated with various frequencies (periods or horizons). At high frequencies (short horizons), the cospectrum is near zero; at low frequencies (long horizons), the cospectrum is strongly positive. The coherence is similar to an $R^2$. It is a coefficient of determination between the arrival probabilities, also decomposed by frequency. It is near zero at high frequencies and near unity at low frequencies. Both the cospectrum and coherence suggest strong long-run dependency. An economic interpretation is that time deformation is a reasonable conjecture at periods of roughly fifty or longer, but at shorter horizons the relationship is corrupted by random delays.[7]

*Example 2: Trades and quotes that jointly depend on persistent stochastic arrival intensity.*

An example that arose in the discussion of time deformation in Section 2.a involved an information intensity variable, $I_t$, which followed a persistent Ornstein-Uhlenbeck process. Here, I assume a discrete-time version with persistence parameter $\alpha = 0.995$.[8] The Poisson intensity parameters for trades and quotes are assumed to be $\lambda_{Trades} = I_t$ and $\lambda_{Quotes} = I_t/2$. Conditional on $I_t$, the numbers of trades and quotes in a period are independent. As in the previous example, I simulate this for 10,000 periods and compute time- and frequency-domain statistics.

Figure 4 graphs the results. In Panel A, both trades and quotes evince declining autocorrelations. The cross-correlations are also declining and symmetrical. In Panel B, the univariate spectra of both series show concentrated power at low frequencies (long periods). The cospectrum is near zero up to approximately fifty periods, whereupon it rises sharply. Over this same interval, the coherence (upper right corner) rises to unity.

As in the last example, we see strong low-frequency dependence in the counts associated with time deformation. The mechanisms in the two examples, differ markedly however. In the

---

[7] The order processing delay could be investigated in the frequency domain by examining the phase spectrum. But this feature seems more visible in the time domain analysis.

[8] The full specification is: $\log(I_t) = \log(I_{t-1}) - 0.995(\log(I_{t-1}) - 0.5) + \varepsilon_t$ where $\varepsilon_t$ is normal with zero mean and a standard deviation of 0.1.

order/trade example, one order always leads to exactly one trade. The counts would be perfectly coherent at all frequencies were it not for the random short-term delay. In the present trade/quote example, there is no deterministic relation between the two count series. The long-run coherence arises from a common persistent component.

### *d. Results for IBM in the frequency domain*

Figure 5 depicts the spectral properties of IBM's orders, trade and quote point processes (corresponding to the time domain analysis depicted in Figure 2). The spectral statistics (like the autocovariances) are based on the "deseasonalized" counts (i.e., after removal of daily mean and intraday patterns). The frequency- and time-domain analyses are therefore directly comparable.

On the diagonal of Figure 5, the univariate spectra of all three series exhibit strong long horizon components. These presumably mostly reflect persistent stochastic components, but they might also arise from deterministic components left over from a misspecified deseasonalization process.

The off-diagonal graphs in Figure 5 describe the joint properties of the event occurrence rates. All pairs of series show the same general features: weak dependence at low periods and stronger dependence at longer periods. The cospectra are near zero at periods of ten minutes or below, and gradually rise thereafter. The coherencies are low (roughly 10%-20%) up to periods of ten minutes and then rise to approximately 60%. Although the positive dependence in the occurrences is consistent with time deformation, these coherencies are considerably lower than those induced by time deformation in the examples. The "speeding up" and "slowing down" of the market is not uniform across events. This suggests that the roles of the component processes are changing as the overall pace of activity varies.

## 5. Orders in Detail

The discussion of the asymmetric information models in Section 2 suggests that the intensities of orders differentiated by sign (buy or sell) are useful in characterizing information epochs. This section begins with an example that illustrates the time- and frequency- domain properties of such models, and then turns to the results for IBM.

***a. Example***

Consider a simulation that is similar in some respects to Easley et al. (1996). The arrival intensity of uninformed buys and sells is $\varepsilon = 0.2$ per period. Informed trading intensity is $\lambda = 10$. An information epoch lasts twenty periods. Information epochs arise randomly. If the market is not currently in an epoch, there is a 0.01 probability that one will start this period. If an epoch starts, there is an even chance of positive or negative private information.

Figure 6 graphs the auto- and cross-correlations of the simulated buy and sell series. The duration of the information epoch shows up most clearly in the auto-correlations, which are positive and decaying up to lag 20. Although the cross-correlations are negative, confirming the intuition that the hallmark of informed trading is a net order imbalance, they are small in magnitude and diffuse.

The difficulty with relying solely on the autocorrelations of the buy and sell orders to characterize the informational epochs is that the pattern (positive and decaying) is also consistent with a world in which there is no asymmetric information, but there is time deformation, as in Example 2 of Section 4.

Certain features of the model emerge somewhat more clearly in the frequency domain. Panel B of Figure 6 graphs the spectra and cross-spectra. As in the case of time deformation, the univariate spectra are concentrated at long periods. The cospectra (lower left) are quite different, however. Whereas time deformation predicts a positive cospectrum, that associated with asymmetric information is at long periods negative.[9]

[9] The range of periods over which the cospectrum is negative in Panel B reflects both the duration and occurrence frequency of the information epochs.

***b. Market Buy and Sell Orders***

It will be recalled that the term "order" in the asymmetric information models generally refers to market orders. The present analysis accordingly differentiates orders not only as to side, but also as to market or limit.

Figure 7 depicts time-domain (Panel A) and frequency-domain (Panel B) analyses for orders classed by market or limit and buy or sell. The own-effects (autocovariances and spectra on the diagonals of the graph panels) are consistent with persistence in occurrence rates. The autocorrelations for limit buy orders, however, are weaker (smaller and quicker to fade) than the other order types.

The cross-correlations (Panel A) are mostly positive, but not as strong as those in the orders/trades/quotes analyses in the last section. For some pairs, cross-correlations at low leads and lags are zero or negative. This is mirrored in the cross-spectra (Panel B), as the cospectra are negative at certain frequencies. The maximum coherencies are generally about 10%-20%, substantially lower than the corresponding values for the orders/trades/quotes set.

The cross-analysis that is most appropriate for assessing asymmetric information effects is market buys vs. market sells. The cross-correlations (Figure 7, A, "Market Buys(t), Market Sells(t-k)") are generally positive, but near zero or negative at lead/lags under a minute or so. In addition, cross-correlations are asymmetric in that the tendency for market sells to follow market buys after a few minutes is stronger than the tendency for market buys to follow market sells. The predominance of positive correlations is broadly consistent with time deformation. That the cross-correlations are zero or negative only at leads or lags up to a minute or so suggests that the duration of the information epochs is extremely brief.

The cross-spectra (Figure 7, B, "Market Buys, Market Sells") suggest a slightly different story. In moving from low to high periods, the cospectrum is initially zero, then becomes negative up to about thirty minutes, then becoming positive. Thus, the long period (low frequency) components look like time deformation, while the shorter components suggest asymmetric information. Recall that a "period" contains a full cycle of a sinusoid, one "high" interval and one "low" level. The twenty-minute period that marks the most negative point of the cospectrum, therefore, reflects order surges lasting ten minutes.

Both the time- and frequency- domain analyses are consistent with a model in which orders are jointly driven by time deformation and asymmetric information effects. If we identify the latter by the interval over which the cross-correlations are zero/negative, the implied duration is on the order of a minute or two. The frequency-domain analysis, however, suggests that information epochs are longer, up to thirty minutes.

To further illuminate these findings, the analysis turns to the sixteen most actively traded issues in the TORQ data. Figure 8 presents for these stocks the cospectrum between market buy and market sell orders, with two-standard-error confidence bands. As in the case of IBM, most of the cospectra are positive at the longest period considered; many are negative in some intermediate range. Particularly in view of the confidence bands, however, these features can hardly be considered definitive.

This lack of uniformity suggests that the timing features of informational epochs defy simple characterization. The difficulty is not that buy and sell orders aren't clustered: the univariate spectra of buys and sells for all firms (not reproduced here) exhibit dominant low-frequency components. Rather, the problem is that buy and sell surges "sometimes" occur in tandem (consistent with time deformation effects) and at other times surge on one side or the other. It might be supposed that we could resolve informational epochs by examining simple transformations, e.g. "net" order flow. In fact, any operational definition of net order flow is contingent on fixing a horizon over which to cumulate or average. The present results suggest that there is no obvious choice of horizon.

The negative finding here relates to a basic construct of virtually all asymmetric information models. This does not, most emphatically, imply that these models are fundamentally flawed. It does, however, suggest caution in taking the timing assumptions of these models literally.

***c. Market orders and limit orders***

While the strongest content of the sequential trade models pertains to market buy and sell orders, other predictions relate to the provision of liquidity. In these models, liquidity is measured by size of the spread and/or the slope of the price/quantity function posted by the market maker. In the present analysis, it is useful to take as a proxy for liquidity the arrival rates of limit orders.

We first consider for IBM the relationships between market orders and limit orders on the same side of the market. Figure 7 (panel A) graphs the cross-correlations between market buys and limit buys, and between market sells and limit sells. These are positive and declining. Furthermore, the corresponding cospectra (panel B) are positive at all but the briefest periods. From a purely statistical perspective, this suggests that these relations are primarily driven by time-deformation effects. The findings may also arise from the economic substitutability of market and limit orders.

Next, consider the relationships between market and limit orders on opposing sides of the market, i.e., market buys vs. limit sells and market sells vs. limit buys. The cross-correlations (panel A) are mostly positive, but zero or negative at low lags and leads. As in the market order analysis, the buy and sell sides of the market are somewhat different. High intensity in market buy orders is associated with high intensity in limit sell orders over the next fifteen minutes or so; the corresponding relation between market sell orders and limit buy orders is also positive, but of lower magnitude. In the frequency domain (panel B), the corresponding cospectra are zero or negative up to approximately thirty minutes.[10]

---

[10] In Figure 7 (panel B), the quadrature spectrum graphed in "Market Buys, Limits Sells" is generally positive; that of "Limit Buys, Market Sells" is generally negative. This does not indicate a corresponding difference in economic relationships, but merely arises from the ordering of the variables. (The quadrature spectrum of *x* on *y* is the negative of the quadrature spectrum of *y* on *x*.)

The cospectra for the broader sample of sixteen actively traded stocks are plotted in Figure 9 (market buys and limit sells) and Figure 10 (market sells and limit buys). Interestingly, the cospectra are more similar across stocks than those for market buys and market sells (cf. Figure 8). The cospectra at the longest periods are all positive (as are most of the two-standard-error confidence bands). They are near zero at the shortest periods, and turn positive only at moderate periods (twenty to forty minutes).

### *d. Program and nonprogram orders*

Program orders pertain to a list of stocks. The list is submitted as a single order, but once received by the NYSE, the components are sent to the posts for individual execution. The most common use of program orders is in index arbitrage trading or dynamic trading strategies that attempt to replicate a stock index derivative. Harris, Sofianos, and Shapiro (1994) discuss the relation between these orders and volatility; Hasbrouck (1996) analyzes the effect of these orders on price evolution in a VAR framework. Program orders are of particular interest because they are often perceived as highly clustered and as demanding liquidity in a destabilizing fashion.

Figure 11 depicts the time- and frequency-domain results for program and nonprogram orders also differentiated by buy or sell. (Only market and marketable limit orders are considered here.) The autocorrelations (on the diagonal in panel A) are positive and declining; those of program orders are stronger than those of nonprogram orders. Most cross-correlations are positive. A striking exception arises in the cross-correlations between program buys and sells, which are negative at low lags. This is economically reasonable because program orders are almost by nature one-sided. The market conditions that give rise to an index-arbitrage related buy order would cause the reverse arbitrage to lose money. Similarly, if most users of dynamic trading strategies are trying to replicate protective puts, buy programs should be driven by positive recent market returns and sell programs by negative recent returns.

These results are amplified by the spectral analysis (Panel B). The cospectrum between program buys and sells is predominately negative, even at the longest period considered. In summary, program orders are distinctively positively autocorrelated and markedly one-sided..

## 6. Summary and implications.

This study views trades, quote revisions and orders (variously classified) as timed events that can be summarily described as a stationary multivariate point process. Although such processes may be exceedingly difficult to model in a parsimonious fashion, their essential descriptive features readily fall out of conventional time- and frequency-domain analyses. The point process perspective offers numerous insights into how the notional time of the theoretical microstructure models might correspond to real time. The analysis is used here to characterize the intraday properties of actively-traded NYSE stocks (most extensively, IBM), based on the TORQ data.

The principle of time deformation, which differentiates between real and operational/informational time scales, is frequently invoked in security market studies. The present analysis suggests that it be characterized as positive commonality in the occurrence rates of the diverse events that collectively comprise the market process, i.e., a proportional speeding up or slowing down of order frequency, quote revision frequency, etc. Because events occur only at a few points in the time continuum, time deformation (in this sense) can only be assessed conditional on a time horizon, a characterization that is most visible from frequency-domain analyses.

The analysis of IBM's undifferentiated orders, trades and quote revisions suggests that over brief time horizons (say, up to ten minutes), there is little evidence of time deformation. (Event occurrence rates are not highly correlated.) Over long horizons, however, the correlation/coherence increases. At a one-trading-session period (the longest considered here), the coherence between event pairs is typically about 50%. This is certainly suggestive of time deformation. It nevertheless falls considerably short of the 100% coherence that would arise from a simple mechanical market process, such as the regular order/trade/quote revision cycle found in the basic sequential trade microstructure models.

This finding points to the limitations of empirical microstructure specifications that assume stationarity in real time (e.g., Hasbrouck (1996)) or in event time (e.g., Hasbrouck (1991)). Real-time stationarity is refuted by the clustering of events. Event-time stationarity is rendered suspect by the absence of uniformity in the intensities of various events. A fast market, in other words, is not merely a normal market that is speeded up, but one in which the relationships between component events differ. Engle and Lunde (1998) find, for example, that the price impacts of trades depends (negatively) on intertrade durations.

The joint properties of (market) buy and sell orders are also useful in characterizing a feature that is used nearly universally in theoretical microstructure models. Termed here an informational epoch, this is simply the notional time span that begins with a valuation uncertainty and an informational asymmetry, and ends with a resolution of the uncertainty. From an empirical viewpoint, it is demarcated by a one-sided order flow. This would normally be expected to give rise to clustering (positive autocorrelation) in buys and sells over lags roughly commensurate with the average durations of the information epochs. Information-related clustering is distinguished from that due to time deformation in that the former gives rise to clustering in buys that is uncorrelated or negatively correlated with that in sells (and vice versa).

The empirical results suggest that orders reflect a complicated mix of informational and time deformation effects. The picture for IBM is relatively straightforward and suggests that information effects dominate up to periods of thirty minutes or so, and time deformation holds sway for longer horizons. But the results for the broader sample exhibit considerable variation across stocks.

The failure of the present study to establish a strong, stable and plausible empirical characterization of informational epochs is, although a negative finding, an important one. This conclusion does not vitiate the usefulness of informational epochs as modeling devices. But it does suggest caution in interpreting the results of empirical specifications in which assumptions about the frequencies and durations of these epochs figure prominently. Logically, the results also suggest the importance of investigating microstructure models that admit uncertainty about the characteristics of informational epochs. As inconclusive as the present statistical results are, it is worth remembering that they are based on ex post analysis of an entire data sample. The difficulties faced by market participants making inferences solely from realized data are even more formidable.

Interestingly, and contrary to street lore, program buy and sell orders are not more highly clustered (autocorrelated) than nonprogram orders. They are, however, more one-sided. A surge in program buy orders, for example, is accompanied by a marked slowdown or withdrawal of program sells at periods of several minutes.

Point process analysis is also useful in characterizing the relationship between market orders and opposing limit orders. Most importantly, at periods up to approximately thirty minutes, market buy orders and limit sell orders are either uncorrelated or slightly negatively correlated, and similarly for market sells and limit buys. Strong positive correlation comes in only for components longer than thirty minutes. These results are fairly uniform across the sample. The findings suggest that while a surge of market (liquidity demanding) orders is matched by an increase in limit (liquidity supplying) orders, the response of the latter is slow. It is common to characterize monumental liquidity crises as arising from the simultaneous arrival of sellers and withdrawal of buyers. The results from the present study suggest that far from being a rare phenomenon, this occurs (albeit on a smaller scale) as a matter of routine.

# References

Andersen, Torben G., 1996, Return volatility and trading volume: An information flow interpretation of stochastic volatility, *Journal of Finance* 51, 169-204.

Bochner, S., 1960, *Harmonic analysis and the theory of probability,* (Berkeley: University of California Press).

Brillinger, David R., 1974, Fourier analysis of stationary processes, *Proceedings of the IEEE* 66. Reprinted in Brillinger, D.R., Time Series: Data Analysis and Theory, (Holden Day, 1981).

Brillinger, David R., 1975, Statistical inference for stationary point processes, in M. L. Puri, eds.: *Stochastic Processes and Related Topics, Volume 1*, (New York: Academic Press).

Brillinger, David R. 1981. *Time Series: Data Analysis and Theory*. San Francisco: Holden Day.

Clark, Peter K., 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* 41.

Easley, David, Nicholas M. Kiefer, Maureen O'Hara, and Joseph Paperman, 1996, Liquidity, information and infrequently traded stocks, *Journal of Finance* 51, 1405-36.

Easley, D., and M. O'Hara. 1992. Time and the Process of Security Price Adjustment. *Journal of Finance* 47, no. 2: 576-605.

Engle, Robert F., 1996, The econometrics of ultra-high-frequency data, Working Paper, Department of Economics, UCSD.

Engle, Robert F. and Lunde, Asger, 1998, Trades and quotes: a bivariate point process, Working Paper, Department of Economics, University of California, San Diego.

Engle, Robert F., and Jeffrey E. Russell, 1998, Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica* 66, 1127-62.

Epps, T. W., and M. L. Epps, 1976, The stochastic dependence of security price changes and transactions volumes: implications for the mixture of distributions hypothesis, *Econometrica* 44, 305-21.

Foster, F. D., and S. Viswanathan, 1995, Can speculative trading explain the volume-volatility relation?, *Journal of Business and Economic Statistics* 13, 379-96.

Fuller, Wayne A. 1996. *Introduction to Statistical Time Series, Second Edition*. John Wiley.

Gallant, A. R., Peter E. Rossi, and George Tauchen, 1992, Stock prices and volume, *Review of Financial Studies* 5, 199-242.

Garbade, Kenneth, and Zvi Lieber, 1977, On the independence of transactions on the New York Stock Exchange, *Journal of Banking and Finance* 1, 151-72.

Ghysels, Eric, Christian Gourieroux, and Joanna Jasiak, 1998a, Market time and asset price movements: Theory and estimation, in D. Hand and S. Jacka, eds.: *Statistics in Finance*, (London: Edward Arnold).

Ghysels, Eric, Gourieroux, Christian, and Jasiak, Joanna, 1998b, Stochastic volatility duration models, Working Paper, Department of Economics, Pennsylvania State University.

Ghysels, Eric, and Joanna Jasiak, 1998, GARCH for irregularly-spaced financial data: the ACD-GARCH model, *Studies in Nonlinear Dynamics and Econometrics* 2, 133-50.

Glosten, L. R. 1989. Insider Trading, Liquidity, and the Role of the Monopolist Specialist. *Journal of Business* 62, no. 2: 211-35.

Glosten, Lawrence R., 1994, Is the electronic open limit order book inevitable?, *Journal of Finance* 49, 1127-61.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71-100.

Granger, Clive W. J. and Robert F. Engle, 1983, Applications of spectral analysis in econometrics, in D. R. Brillinger and P. R. Krishnaiah, eds.: *Handbook of Statistics, Volume 3: Time Series in the Frequency Domain*, (Amsterdam: Elsevier - North Holland).

Hamilton, James D., 1994, *Time Series Analysis*, (Princeton: Princeton University Press).

Harris, Lawrence E., 1986, Cross-security tests of the mixture of distributions hypothesis, *Journal of Financial and Quantitative Analysis* 21, 39-46.

Harris, Lawrence E., 1987, Transaction data tests of the mixture of distributions hypothesis, *Journal of Financial and Quantitative Analysis* 22, 127-41.

Harris, Lawrence E., and Joel Hasbrouck, 1996, Market vs. limit orders: the SuperDOT evidence on order submission strategy, *Journal of Financial and Quantitative Analysis* 31, 213-31.

Harris, L. E., G. Sofianos, and J. E. Shapiro. 1994. Program trading and intraday volatility. *Review of Financial Studies* 7, no. 4: 653-85.

Hasbrouck, Joel, 1991, Measuring the information content of stock trades, *Journal of Finance* 46, 179-207.

Hasbrouck, Joel, 1992, Using the TORQ Database, Working Paper, New York Stock Exchange.

Hasbrouck, Joel, 1996, Order characteristics and stock price evolution: An application to program trading, *Journal-of-Financial-Economics* 41, 129-49.

Karpoff, Jonathan M., 1987, The relation between price changes and trading volume: a survey, *Journal of Financial and Quantitative Analysis* 22, 109-26.

Kyle, Albert S., 1984, Market structure, information, futures markets, and price formation, in Gary G. Storey, Andrew Schmitz, and Alexander H. Sarris, eds.: *International Agricultural Trade: Advanced Readings in Price Formation, Market Structure, and Price Instability*, (Boulder, Colorado and London: Westview Press).

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315-36.

Llorente, Guillermo, Michaely, Roni, Saar, Gideon, and Wang, Jiang, 1998, Dynamic volume-return relation of individual stocks, Working Paper, Johnson School, Cornell University.

Lyons, Richard K., 1995, Tests of microstructural hypotheses in the foreign exchange market, *Journal of Financial Economics* 39, 321-51.

Lyons, Richard K., 1996, Optimal transparency in a dealer market with an application to foreign exchange, *Journal of Financial Intermediation* 5, 225-54.

Lyons, Richard K., 1997, A simultaneous trade model of the foreign exchange hot potato, *Journal of International Economics* 42, 275-98.

Mandelbrot, Benoit, and H. Taylor, 1967, On the distribution of stock price differences, *Operations Research* 15, 1057-62.

O'Hara, Maureen. 1995. *Market Microstructure Theory*. Cambridge, USA: Blackwell Publishers.

Richardson, Matthew, and Tom Smith, 1994, A direct test of the mixture of distributions hypothesis: Measuring the daily flow of information, *Journal of Financial and Quantitative Analysis* 29, 101-16.

Russell, Jeffrey R. and Engle, Robert F., 1998, Econometric analysis of discrete-valued irregularly-spaced financial transactions data using a new Autoregressive Conditional Multinomial model, Working Paper, University of Chicago, Graduate School of Business.

Saar, Gideon, 1998, Information and the pricing of assets when orders arrive one at a time, Working Paper, Johnson School, Cornell University.

Stock, James H., 1988, Estimating continuous-time processes subject to time deformation, *Journal of the American Statistical Association* 83, 77-85.

Tauchen, George E., and Mark Pitts, 1983, The price variability volume relationship in speculative markets, *Econometrica* 51, 485-505.

## Appendix: Summary of Statistical Results for Multivariate Point Processes

I summarize here some results from Brillinger (1974); Brillinger (1975) relating to vector point processes. Let $X(t)$ be the vector of cumulative event counts through time $t$: $X_j(t)$ denotes the number of events of the $j$th type that occur in the interval $(0,t]$. Let $dX_j(t)$ be the number of events of the $j$th type that occur in $(t,t+dt]$. The process is stationary if joint distributions of $dX(t_1), dX(t_2), \ldots,$ etc. are invariant to translation. The mean intensity of the process is $c_X$, where $c_X\, dt = E\, dX(t)$. The autocovariance function is $C_{XX}(u)$: $dC_{XX}(u)\, dt = Cov\left[dX(t+u), dX(t)\right]$.

The process is orderly if there is a vanishing probability that an interval of width $dt$ contains more than one event. In this case, the intensity may be interpreted as an occurrence probability. For the $j$th component of $X$:

$$c_j\, dt \approx \Pr\left[\text{There is an event of type } j \text{ in } (t,t+dt]\right]. \qquad (1)$$

If the proces is orderly, the autocovariance function is related to the joint occurrence probability:

$$dC_{jk}(u)\, dt + c_j c_k\, dt\, du = \Pr\begin{bmatrix}\text{event of type } j \text{ in } (t+u, t+u+du] \\ \textit{and} \text{ event of type } k \text{ in } (t,t+dt]\end{bmatrix}. \qquad (2)$$

The most intuitive statistical inference for a stationary point process is based on sample frequency counts (i.e., histograms) of events over short time windows. We estimate the intensity and covariance functions by replacing the right-hand probabilities in equations (1) and (2) with their corresponding sample frequencies computed using a small window $\beta$ in lieu of the infinitesimal $du$ and $dt$. Brillinger (1975) establishes asymptotic distribution properties for these estimators when $\beta \to 0$ as $T \to \infty$. Note that since time in this situation is continuous, this approach requires, in addition to the usual presumption of increasing sample size, an asymptotic diminution of $\beta$. This is often (and in the present case) constrained by the time resolution of the data.

For summary descriptive purposes, we apply this strategy using a window $\beta$ and estimate the autocovariance function $C_{XX}(u)$ for $u = 0, \pm\beta, \pm 2\beta, \ldots$. This is computationally equivalent to estimating the autocovariance function of the local occurrence rates sampled at discrete times $D_\beta(t)$ for $t = 1, \beta, 2\beta, \ldots, (T/\beta)$, where $D_\beta(t) = \beta^{-1}\left[X(t+\beta/2) - X(t-\beta/2)\right]$.

The spectral analysis of point processes closely parallels that of ordinary time series. Excellent presentations of the latter include Hamilton (1994), the survey of Granger and Engle (1983) and the more detailed treatments of Brillinger (1981) and Fuller (1996). The material below specific to point processes is from Brillinger (1974); Brillinger (1975).

The spectral (Cramer) representation of $X(t)$ is:

$$\begin{aligned} X(t) &= \int_{-\infty}^{+\infty}\left[\frac{e^{i\lambda t}-1}{i\lambda}\right] dZ_X(\lambda) \\ dX(t) &= \int_{-\infty}^{+\infty} e^{i\lambda t}\, dZ_X(\lambda)\, dt \end{aligned} \tag{3}$$

where $Z_X$ is a random process defined over frequencies. As one point of contrast, most applications of spectral analysis in econometrics involve discrete regularly-spaced data, wherein the integration range in the integrals is $[-\pi, +\pi]$. The frequencies of a continuous-time process range over the real line. The first and second moments of the $Z_X$ process are given by:

$$\begin{aligned} &E\, dZ_X(\lambda) = \delta(\lambda) c_X\, d\lambda \\ &Cov\left[dZ_X(\lambda), dZ_X(u)\right] = \delta(\lambda - u)\, dF_{XX}(\lambda)\, du \end{aligned} \tag{4}$$

where $\delta(\lambda)$ is the Dirac delta function (a one-unit weighting function concentrated at zero); $c_X$ is the mean intensity and $F_{XX}$ is the spectral measure (loosely speaking, the cumulative distribution) of the process.

The power spectrum (spectral density) is the derivative of the spectral measure (where it exists), and is also the Fourier transform of the covariance density:

$$f_{XX}(\lambda) = \frac{dF_{XX}(\lambda)}{d\lambda} = (2\pi)^{-1}\int_{-\infty}^{+\infty} e^{-i\lambda u} c_{XX}(u)\, du \tag{5}$$

The diagonal elements of $f_{XX}(\lambda)$ summarize $\lambda$-frequency contribution to the variances of the component processes; the off-diagonal elements, to the covariances.

More precisely, we may define the cospectrum and quadrature spectrum of the process as the real and imaginary parts of the spectrum: $\mathrm{Co}_{XX}(\lambda) = \mathrm{Re}\left[f_{XX}(\lambda)\right]$ and $Q_{XX}(\lambda) = \mathrm{Im}\left[f_{XX}(\lambda)\right]$. In the present context, the most useful property of the cospectrum is that it intregrates to the covariance $\left(c_{XX} = \int \mathrm{Co}_{XX}(\lambda)d\lambda\right)$, and therefore suggests a decomposition of the covariance into components associated with different frequencies ("horizons").

It is also useful, particularly in connection with the discussion of time deformation to use a single summary statistic that measures the overall dependence between two point processes at a given frequency. This is provided by the coherence, $K_{XX}(\lambda)$. The coherence between the *i*th and *j*th components of *X* at freqency $\lambda$ is $\left[K_{XX}(\lambda)\right]_{ij} = \left\|\left[f_{XX}(\lambda)\right]_{ij}\right\|^2 \Big/ \left(\left\|\left[f_{XX}(\lambda)\right]_{ii}\right\| \left\|\left[f_{XX}(\lambda)\right]_{jj}\right\|\right)$ where $\|x\| \equiv \sqrt{\mathrm{Re}(x)^2 + \mathrm{Im}(x)^2}$. The coherence is much like a squared correlation coefficient: it lies between zero and one and measures the strength of the overall association. (In some graphs here, for completeness, I also present the phase spectrum $\Phi(\lambda) \equiv \mathrm{Arctan}\left(Q(\lambda)/\mathrm{Co}(\lambda)\right)$, where the Arctan and division operators are element-by-element.)

The sample path of a point process, much like that of an ordinary time series, may be broken down via Fourier analysis into sinusoidal components of various frequencies. Suitably computed and averaged, these components yield consistent estimates of the spectral density and derived quantities.

The consistency properties of the spectral density estimates are cleaner in some respects than those of the time-domain characterizations. When the data are time-stamped to resolution *h* (one second, in the present case), our sample of observations is essentially the ordinary time series $D_h(t)$ collected at *h*-spaced points in time. The condition for consistency of the autocovariance estimates is suspect because even as the sample size increases, *h* can't decline to zero unless the precison of the time stamp increases. Denoting the observed series as $\hat{X}(t) = D_h(t)$, the relation between the spectra of the underlying and observed series is relatively direct:

$$f_{\hat{X}\hat{X}}(\lambda) = \left[ \frac{\sin(h\lambda/2)}{h\lambda/2} \right]^2 f_{XX}(\lambda) \qquad (6)$$

This shows that as $\lambda \to 0$ ("long horizons"), the bias approaches zero. Intuitively, our estimates of hourly components are not much affected by a one-second time resolution. The time resolution also precludes resolving components with periods shorter than $2h$: their effects are impounded ("folded into") longer-term components. Thus, while a frequency domain analysis will not extract more information from a given data sample, the limitations imposed by the finite time resolution are somewhat more clearly defined.

**Figure 1. IBM Events on November 1, 1990.**

NYSE market activity in IBM. The figure plots occurrence times for each of the six event types. The plotted points are vertically "jittered" to spread out overlapping observations.

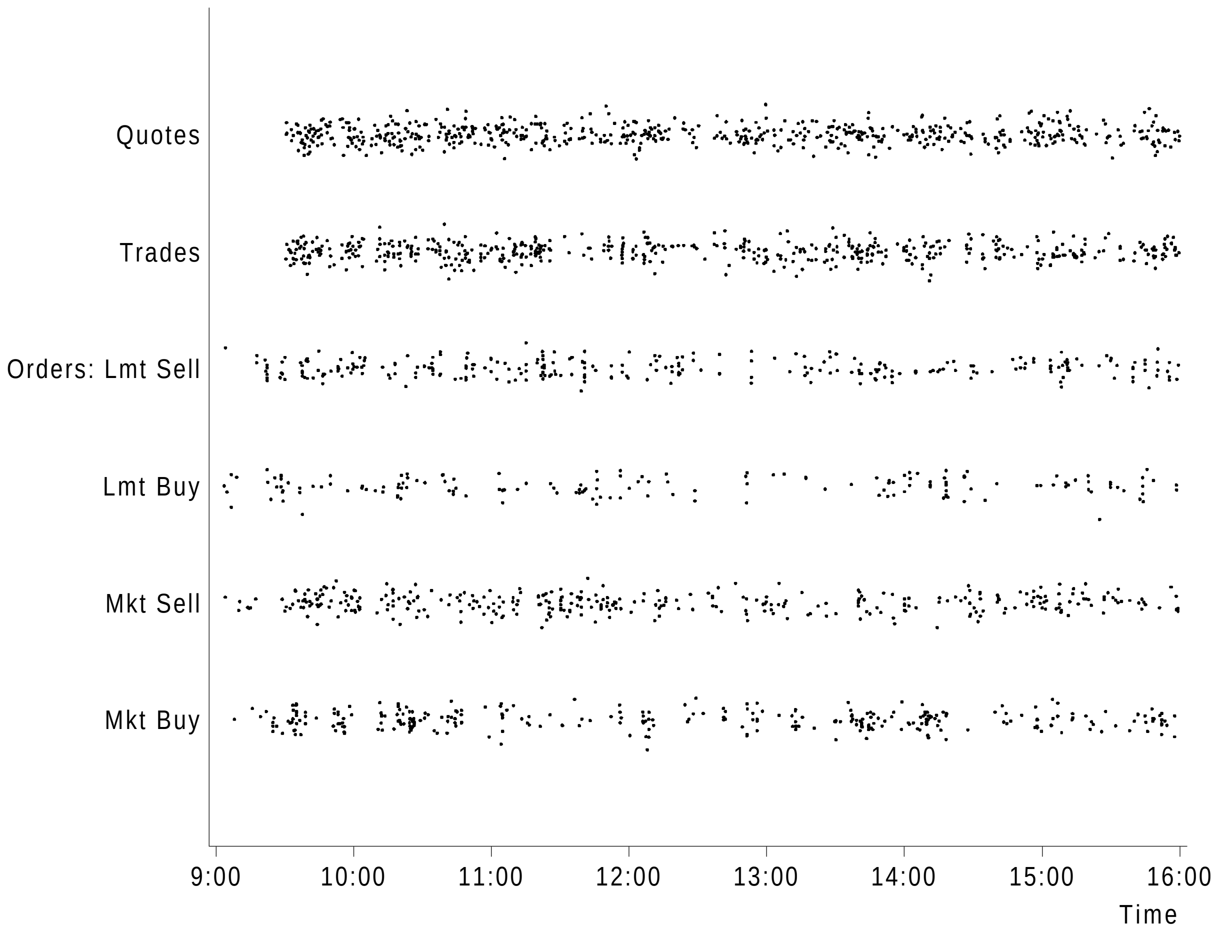


**Figure 2. IBM orders, trades and quotes (auto- and cross-correlations)**

The data comprise NYSE market activity in IBM for fifty-nine trading days from November 1990 through January 1991 (TORQ data). Event occurrence rates (per second) are averages over thirty-second intervals; daily averages are removed; time-of-day effects are removed by polynomial regression. The graphs are auto- and cross-correlations of the average occurrence rates.

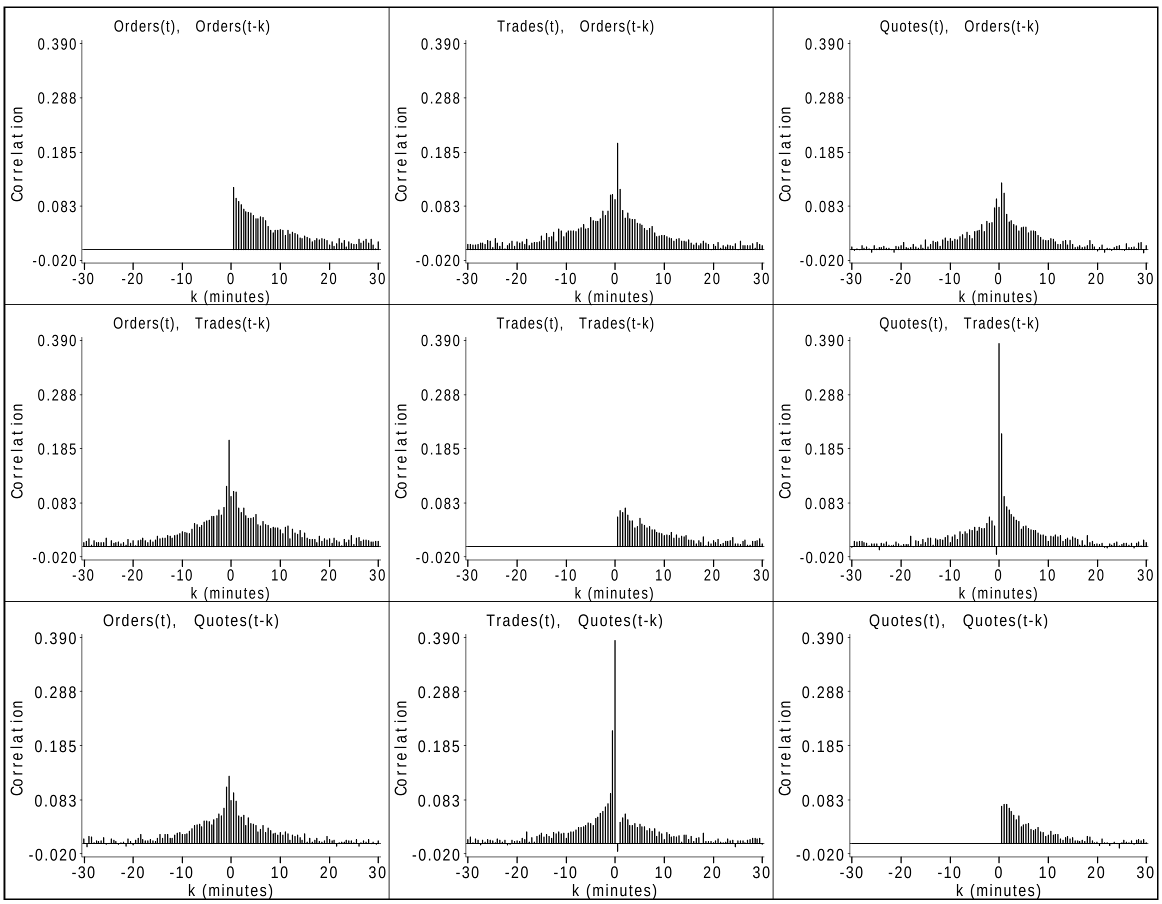
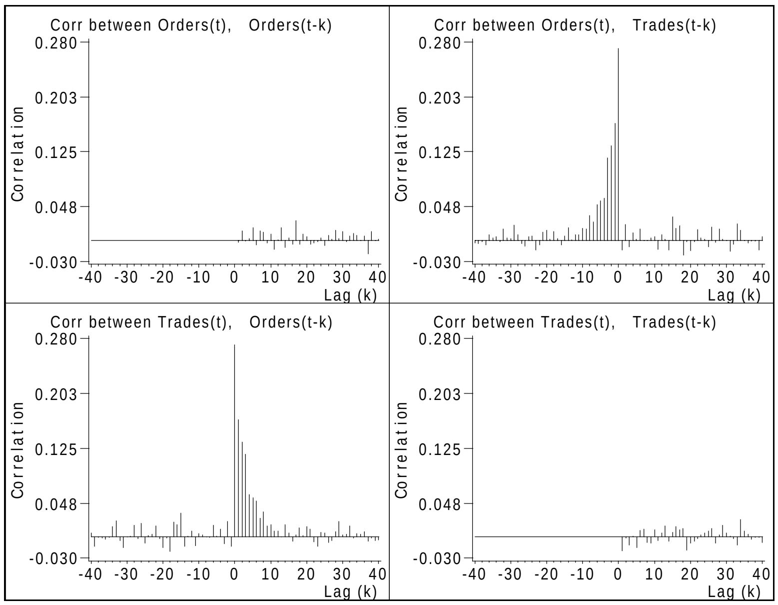


**Figure 3. Simulated Orders and Trades in the Time and Frequency Domains.**

The number of 'orders' in each time interval is assumed Poisson with parameter 0.2. Each 'order' gives rise to a 'trade' *d* intervals later where *d* is an exponential random variable with mean of three. A simulation of 10,000 times was used in the analysis.

A. Auto- and cross-correlations in occurrence rates.




B. Spectra and cross-spectra of occurrence rates.

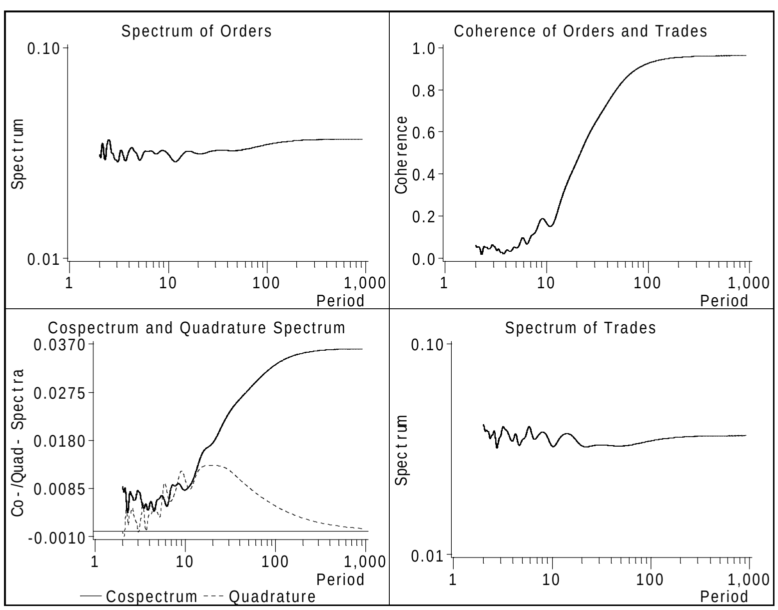
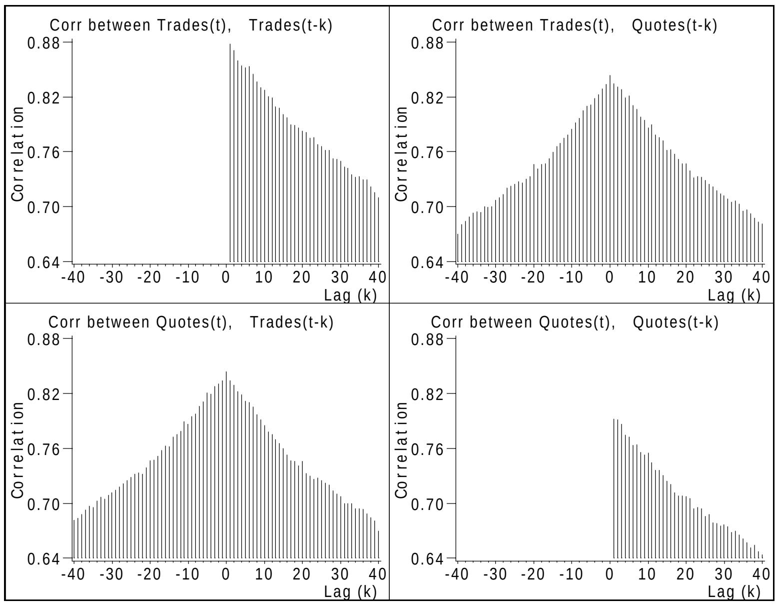


**Figure 4. Simulated Trades and Quotes with Long Run Stochastic Variation**

The simulation postulates a persistent information process:

$\log(I_t) = \log(I_{t-1}) - 0.995(\log(I_{t-1}) - 0.5) + \varepsilon_t$ where $\varepsilon_t$ is normal with zero mean and a standard deviation of 0.1. The Poisson intensity parameters for trades and quotes are assumed to be $\lambda_{Trades} = I_t$ and $\lambda_{Quotes} = I_t/2$. Correlations and spectra are estimated from a single simulation of length 10,000.

A. Auto- and cross-correlations in occurrence rates.




B. Spectra and cross-spectra of occurrence rates.

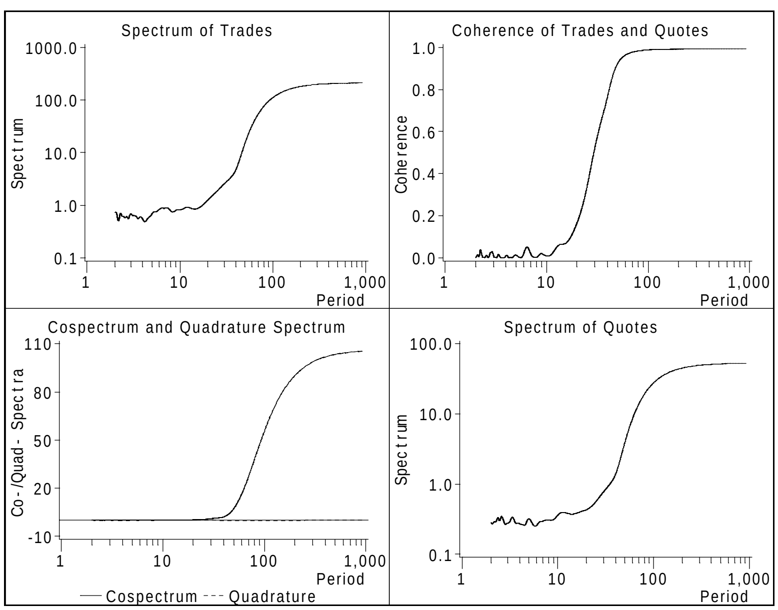


**Figure 5. IBM orders, trades and quotes (spectra and cross-spectra)**

The data comprise NYSE market activity in IBM for fifty-nine trading days from November 1990 through January 1991 (TORQ data). Event occurrence rates (per second) are averages over thirty-second intervals; daily averages are removed; time-of-day effects are removed by polynomial regression. Graphs depict spectra and cross-spectra of the occurrence rates.

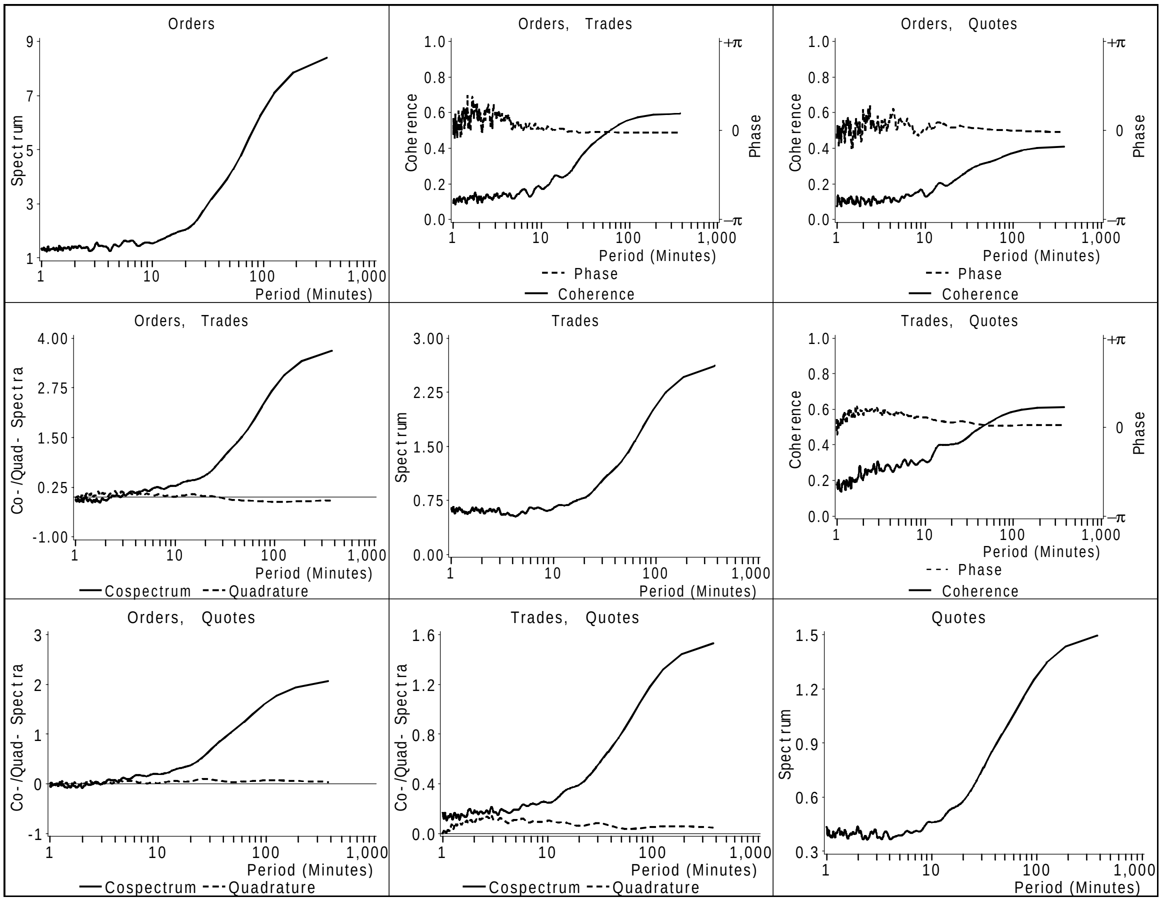
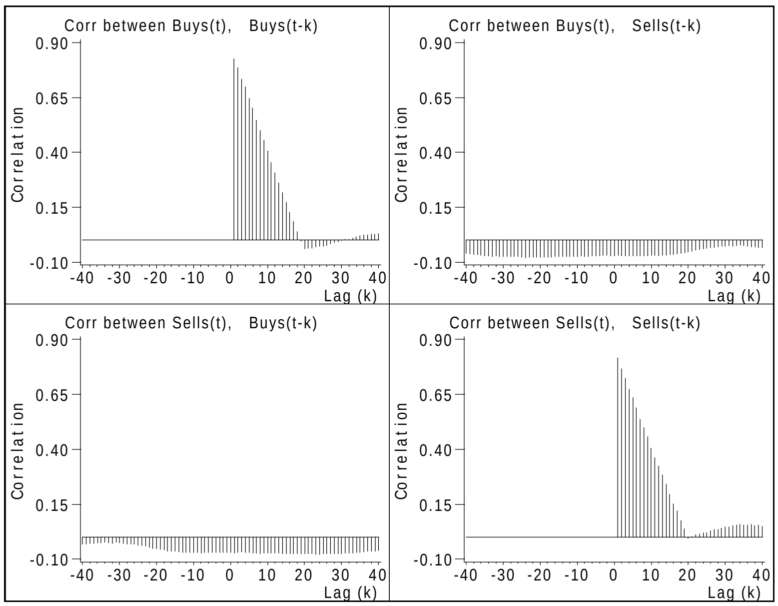


**Figure 6. Simulated Buys and Sells with Periods of 'Informed' Trading**

The simulation assumes that uinformed buys and sells are Poisson with intensity 0.2 per period. Information epochs arise randomly and last for twenty periods. If the market is not currently in an epoch, there is a 0.01 probability that one will start this period. If an epoch starts, there is an even chance of positive or negative private information. Informed agents trade only on the side of the private information (if private information exists) with intensity 10. Estimates are based on a single run of 10,000 periods.

A. Auto- and cross-correlations in occurrence rates.




B. Spectra and cross-spectra of occurrence rates.

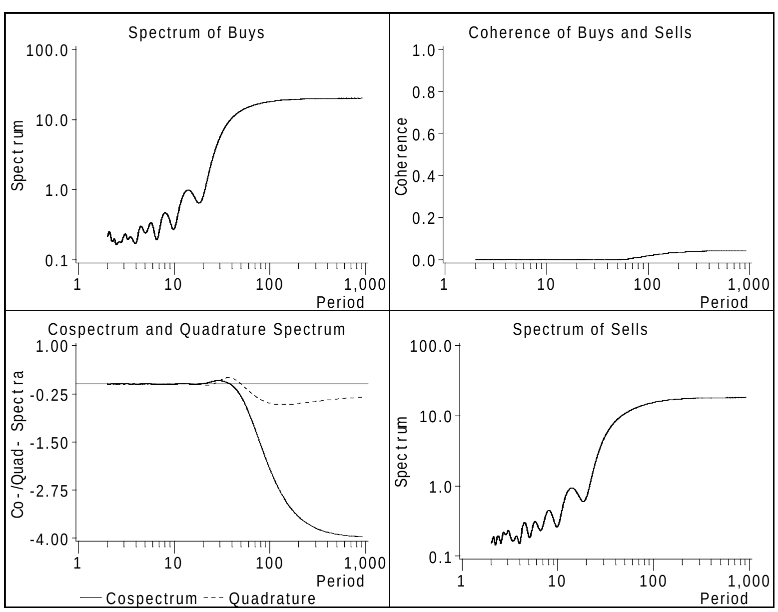
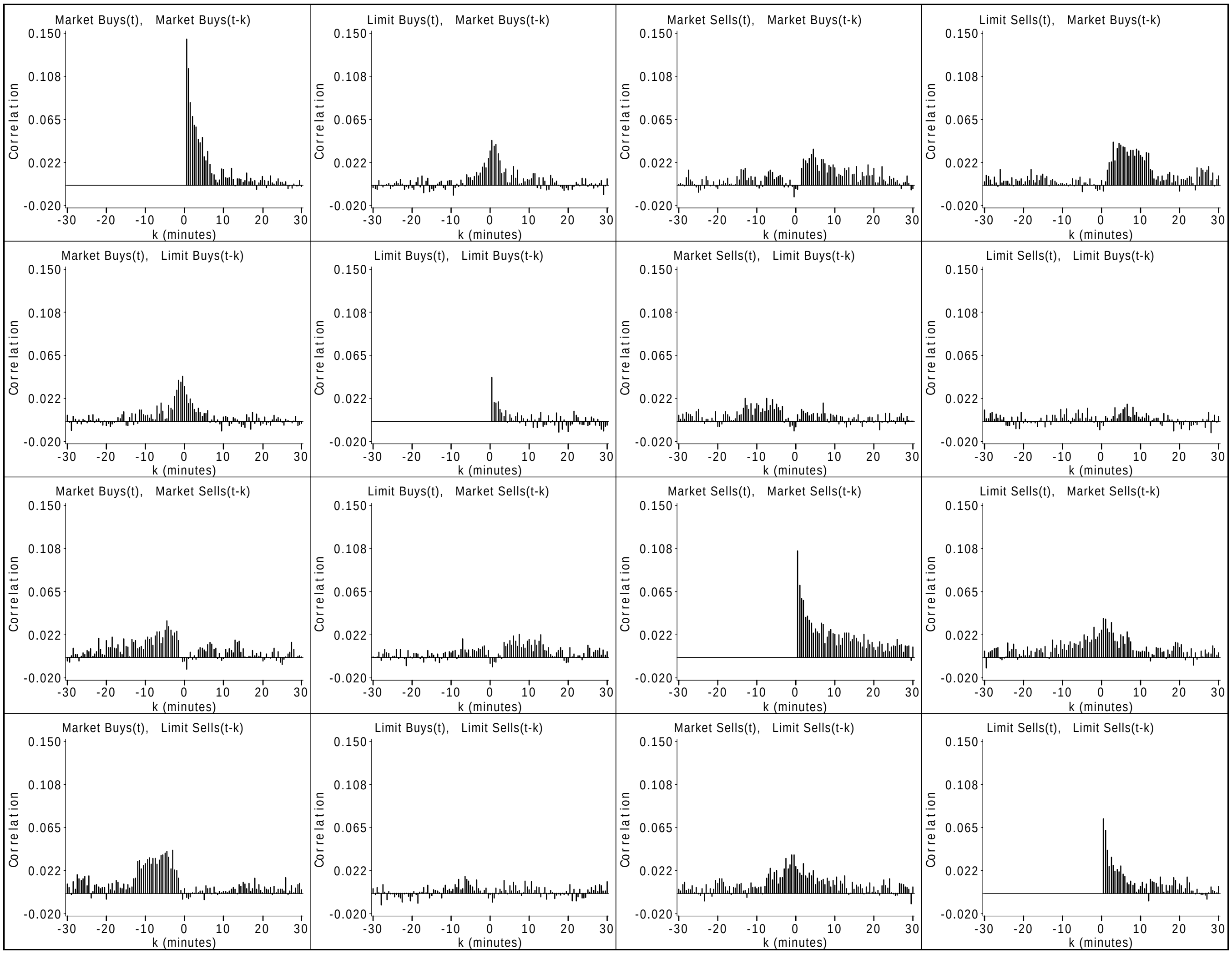


**Figure 7. IBM orders by market/limit and buy/sell.**

The data comprise NYSE market activity in IBM for fifty-nine trading days from November 1990 through January 1991 (TORQ data). Event occurrence rates (per second) are averages over thirty-second intervals; daily averages are removed; time-of-day effects are removed by polynomial regression.

A. Auto- and cross-correlations in occurrence rates.



# Figure 7. IBM orders by market/limit and buy/sell (Continued).

B. Spectra and cross-spectra of occurrence rates.

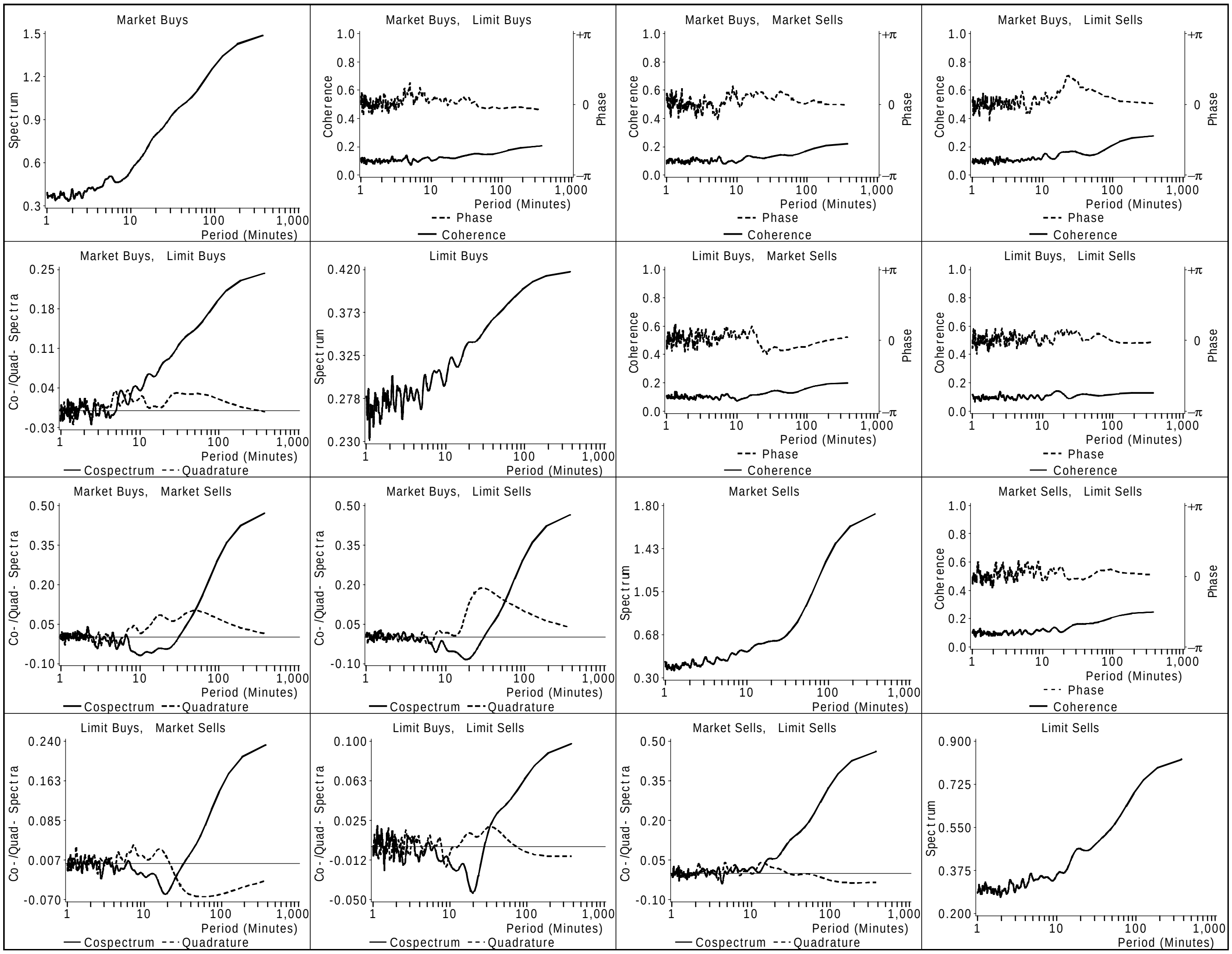


**Figure 8. Market Buys and Sells for Active Stocks**

Spectra and cross-spectra for orders in the sixteen most actively traded stocks in the TORQ database, November 1990 through January 1991. The spectra are estimated for event occurrence rates averaged over thirty-second intervals; daily averages are removed; time-of-day effects are removed by polynomial regression.

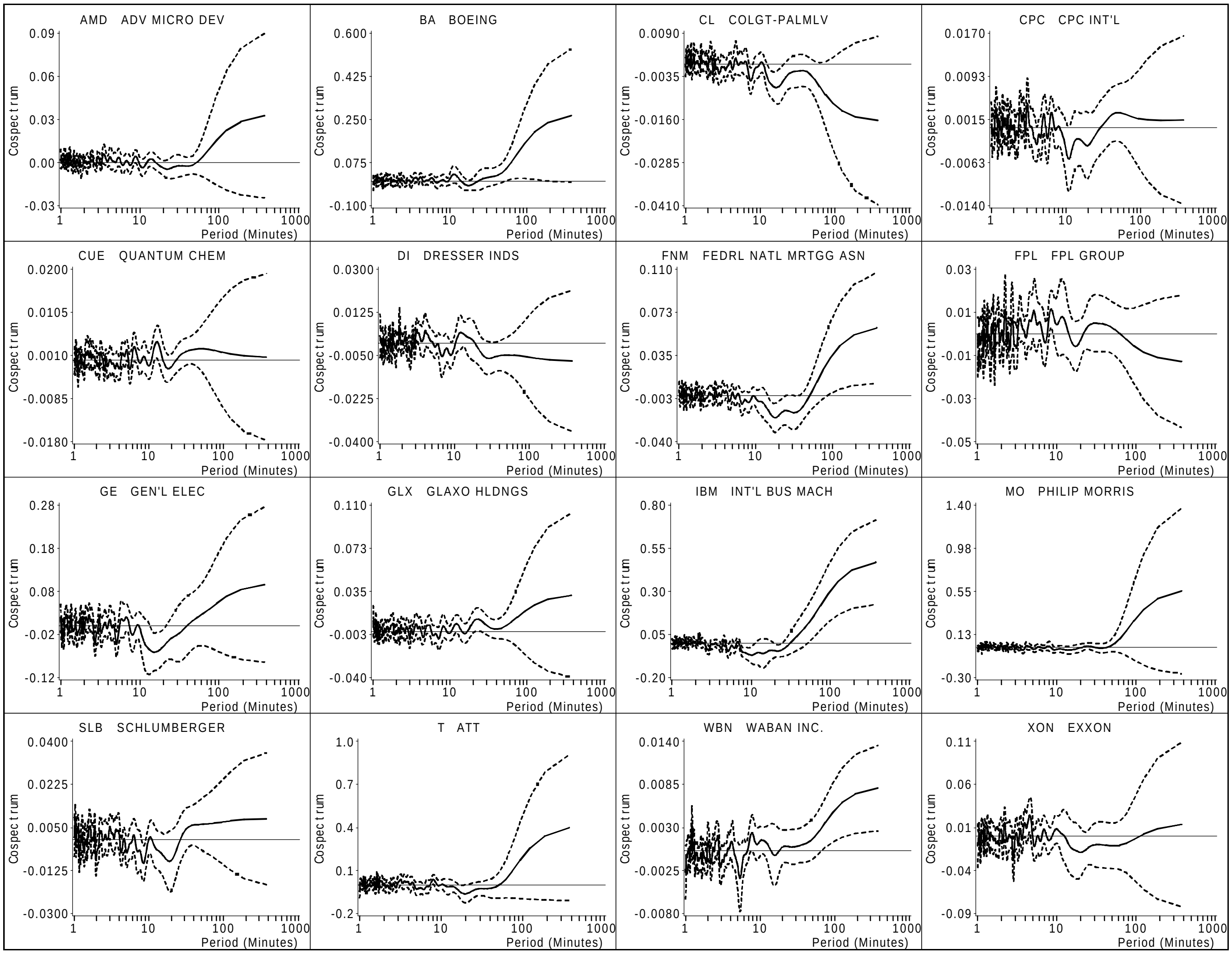
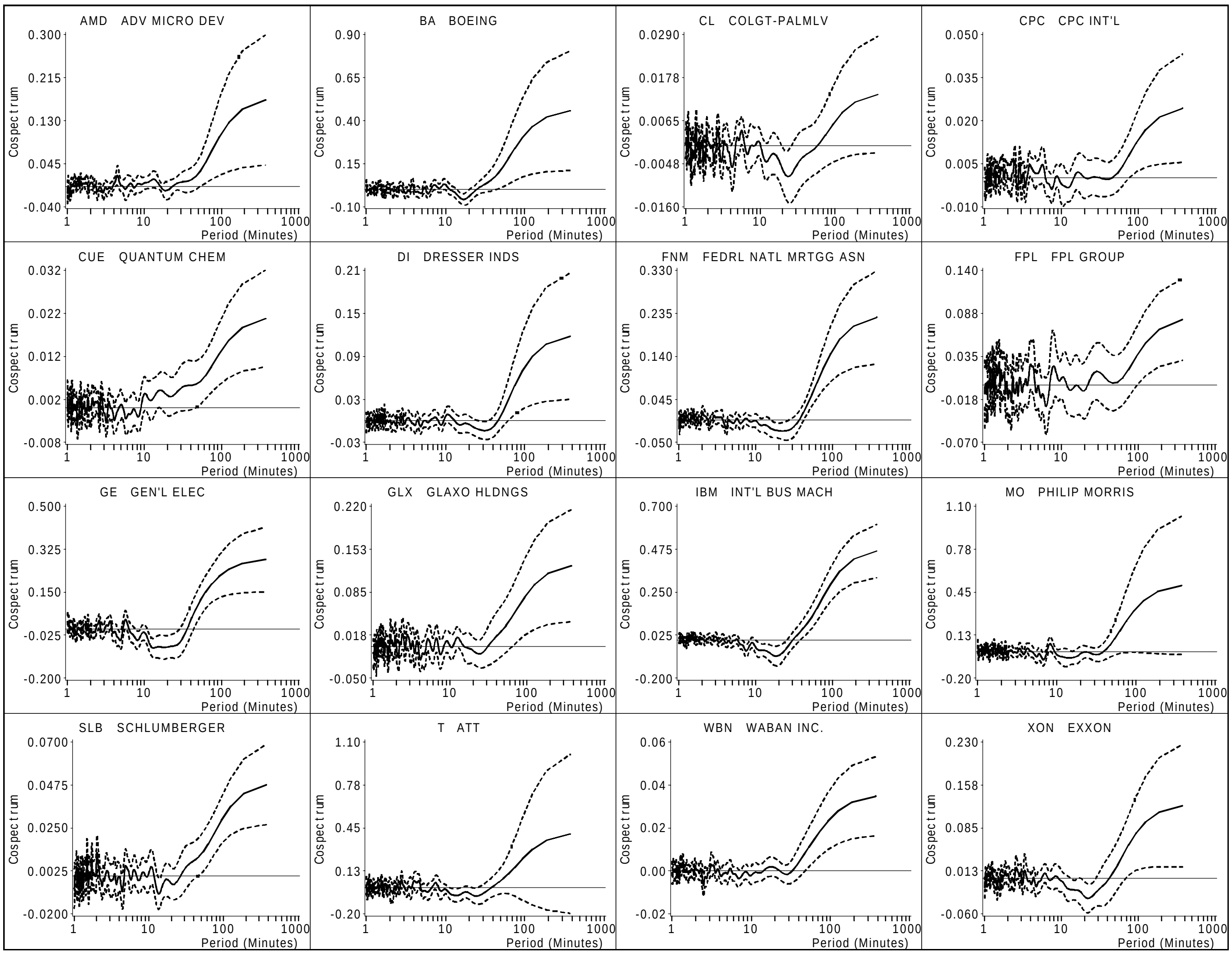


**Figure 9. Market Buys and Limit Sells for Active Stocks**

Spectra and cross-spectra for orders in the sixteen most actively traded stocks in the TORQ database, November 1990 through January 1991. The spectra are estimated for event occurrence rates averaged over thirty-second intervals; daily averages are removed; time-of-day effects are removed by polynomial regression.



## Figure 10. Market Sells and Limit Buys for Active Stocks

Spectra and cross-spectra for orders in the sixteen most actively traded stocks in the TORQ database, November 1990 through January 1991. The spectra are estimated for event occurrence rates averaged over thirty-second intervals; daily averages are removed; time-of-day effects are removed by polynomial regression.

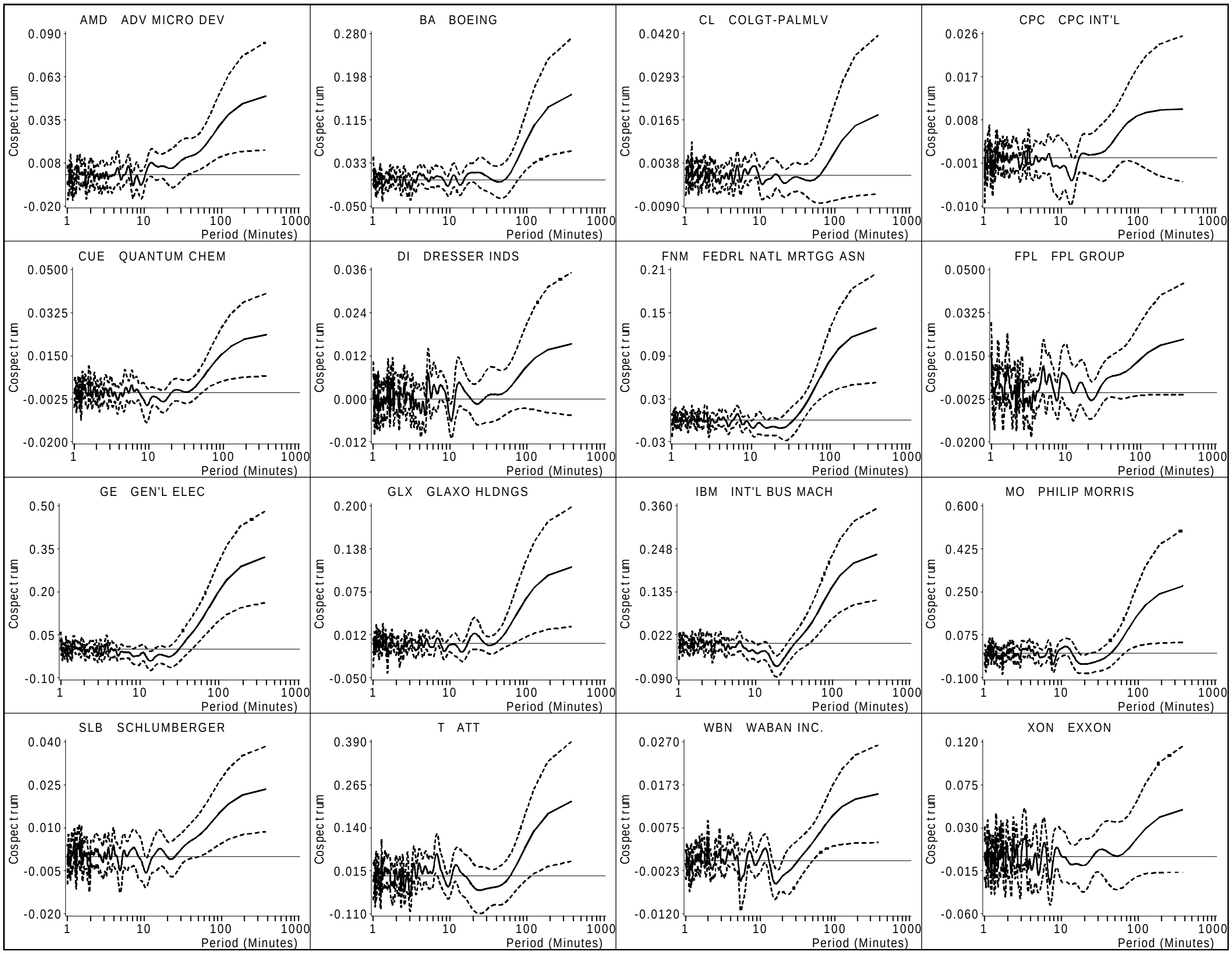
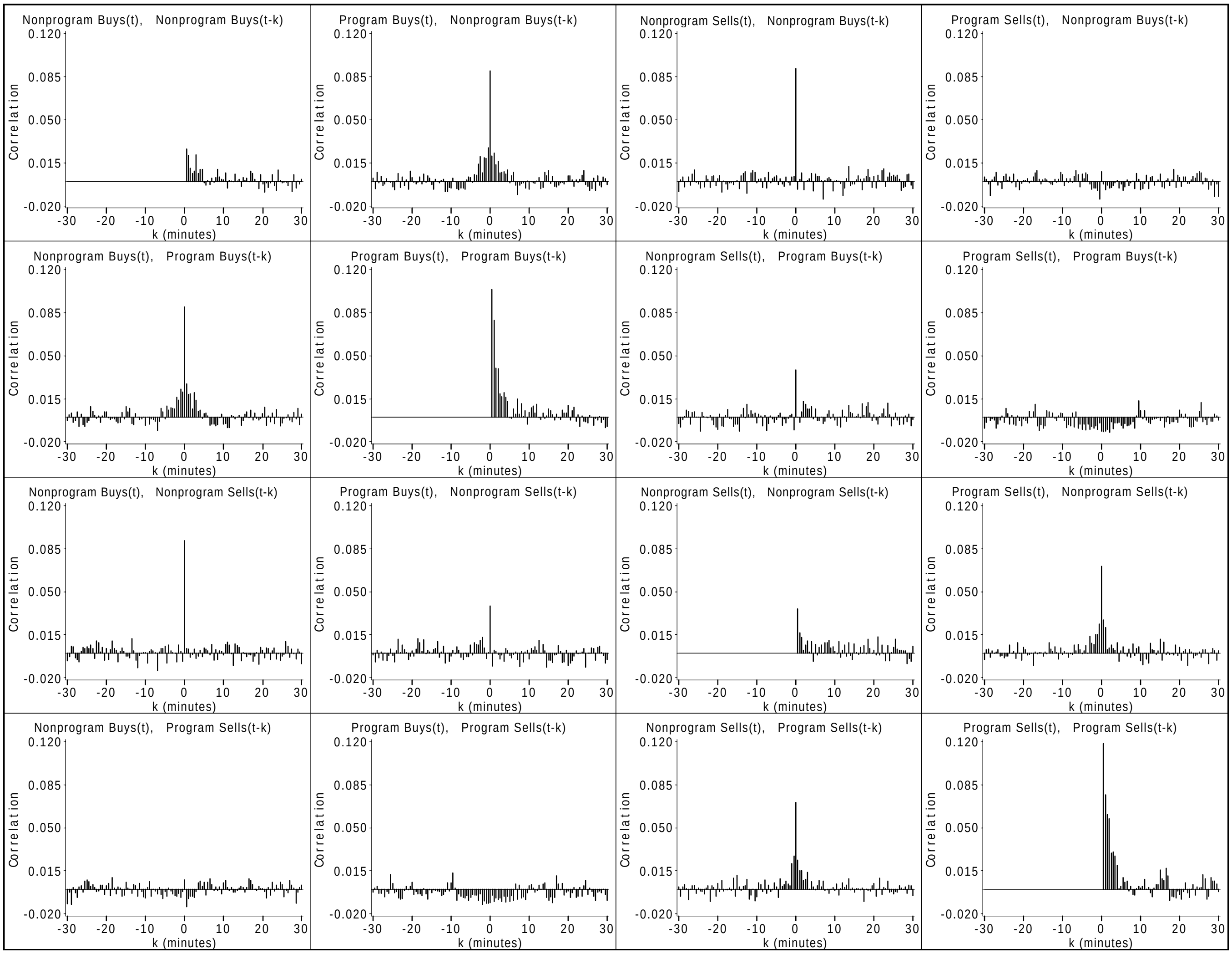


**Figure 11. IBM Market orders (program/nonprogram and buy/sell)**

The data comprise NYSE market activity in IBM for fifty-nine trading days from November 1990 through January 1991 (TORQ data). Event occurrence rates (per second) are averages over thirty-second intervals; daily averages are removed; time-of-day effects are removed by polynomial regression. Program orders are those that comprise a list of stocks. The orders considered here are market and marketable limit orders, only.

A. Auto- and cross-correlations in occurrence rates.



# Figure 11. IBM Market orders (program/nonprogram and buy/sell) (Continued).

B. Spectra and cross-spectra of occurrence rates.