4

Univariate Time-Series Analysis

The Roll model described in the last chapter is a simple structural model, with a clear mapping to parameters (the variance and autocovariance of price changes) that are easily estimated. There are many interesting questions, though, that go beyond parameter estimation. We might want to forecast prices beyond the end of our data sample or to identify the series of m_t (the unobserved efficient prices) underlying our data. Furthermore, when we suspect that the structural model is misspecified, we might prefer to make assumptions about the data, rather than about the model.

To address these issues, the present chapter examines the Roll model from a different viewpoint. Whereas the previous chapter took a structural economic perspective, the present one adopts a more data-oriented statistical "reduced-form" approach. In the process of going back and forth between the structural and statistical representations, we illustrate econometric techniques that are very useful in more general situations. The chapter begins by describing some useful general properties of time series, proceeds to moving average and autoregressive models, and then discussing forecasting and estimation.

4.1 Stationarity and Ergodicity

Much statistical inference relies on the law of large numbers (LLN) and central limit theorem (CLT). These results establish the limiting properties of estimators as the sample size increases. The usual forms of these theorems apply to data samples consisting of independent observations.

Time-series data are by nature dependent. To maintain the strength of the LLN and CLT when independence doesn't hold, we rely on alternative versions of these results that assume stationarity and ergodicity. The following is an intuitive presentation of these concepts. White (2001) presents a more rigorous discussion.

A time series $\{x_t\}$ with constant mean, $Ex_t = \mu$, and autocovariances $\mathrm{Cov}(x_t, x_{t-k}) = \gamma_k$ that do not depend on t is said to be *covariance stationary*. A time series for which all joint density functions of the form $f(x_t), f(x_t, x_{t+1}), \ldots, f(x_t, x_{t+1}, x_{t+2}), \ldots$ don't depend on t is (*strictly*) *stationary*.

The price changes implied by the Roll model, Δp_t, are covariance stationary: $E\Delta p_t = 0$ and $\mathrm{Cov}(\Delta p_t, \Delta p_{t-k}) = \gamma_k$. The price levels are not covariance stationary. (Among other things, $\mathrm{Var}(p_t)$ increases with t.) Covariance stationarity of the Δp_t would also be violated if we replaced the homoscedasticity assumption $Eu_t^2 = \sigma_u^2$ with something like $Eu_t^2 = 5 + \mathrm{Cos}(t)$ or a similar time-dependent feature.[1]

We sometimes describe a sequence of independent observations by saying that an observation carries no memory of observations earlier in the sequence. This is too restrictive for time-series analysis. We typically assume instead that the effects of earlier observations decay and die out with the passage of time. A time series is *ergodic* if its local stochastic behavior is (possibly in the limit) independent of the starting point, that is, initial conditions. An ergodic process eventually "forgets" where it started. The price *level* in the Roll model is not ergodic: The randomness in the level is cumulative over time. But the price changes are ergodic: Δp_t is independent of Δp_{t-k} for $k \geq 2$. Nonergodicity could be introduced by positing $m_t = m_{t-1} + u_t + z$, where z is a zero-mean random variable drawn once at time zero.

The economic models discussed in later chapters (particularly the asymmetric information models) are often placed in settings where there is a single random draw of the security's terminal payoff and the price converges toward this value. The price changes in these models are not ergodic because everything is conditional on the value draw. Nor are they covariance stationary (due to the convergence). Empirical analyses of these models use various approaches. We sometimes assume that a sample consists of a string of these models placed end to end (for example, a sequence of trading days). In this view the sample is an *ensemble*, a collection of independent sample path realizations. Alternatively, we might view the models as stylized descriptions of effects that overlap in some unspecified fashion that results in covariance stationarity. For example, in each time period, we might have a new draw of some component of firm value.

Domowitz and El-Gamal (1999) note that ergodicity, in the sense of dependence on initial conditions, may be an important attribute of market mechanisms. In the long run, we would expect security prices to reflect fundamentals. A trade mechanism that induces persistent price components might impair this adjustment.

4.2 Moving Average Models

We will often assume that a time series like $\{\Delta p_t\}$ is covariance stationary, and now we turn to various ways in which the series can be represented. We start with a *white noise* process: a time series $\{\varepsilon_t\}$ where $E\varepsilon_t = 0$, $\mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2$, and $\mathrm{Cov}(\varepsilon_t, \varepsilon_s) = 0$ for $s \neq t$. This is obviously covariance stationary. In many economic settings, it is convenient and plausible to assume that $\{\varepsilon_t\}$ are strictly stationary and even normally distributed, but these assumptions will be avoided here. White noise processes are convenient building blocks for constructing dependent time series. One such construction is the *moving average* (MA) model. The moving average model of order one (the MA(1) process) is:

$$x_t = \varepsilon_t + \theta\varepsilon_{t-1}. \tag{4.1}$$

The white noise driving a time-series model is variously termed the disturbance, error, or innovation series. From a statistical viewpoint, they all amount to the same thing. The economic interpretations and connotations, however, vary. When randomness is being added to a nonstochastic dynamic structural model, the term *disturbance* suggests a shock to which the system subsequently adjusts. When forecasting is the main concern, *error* conveys a sense of discrepancy between the observed value and the model prediction. *Innovation* is the word that is most loaded with economic connotations. The innovation is what the econometrician learns about the process at time t (beyond what's known from prior observations). Moving forward in time, it is the update to the econometrician's information set. In multivariate models, when x_t comprises a particularly varied, comprehensive, and economically meaningful collection of variables, the innovation series is often held to proxy the update to the *agents'* common information set as well.

The Δp_t in the Roll model have the property that the autocovariances are zero beyond lag one. The MA(1) model in (4.1) also has this property. For this process, the variance and first-order autocovariance are $\gamma_0 = (1+\theta^2)\sigma_\varepsilon^2$, $\gamma_1 = \theta\sigma_\varepsilon^2$, and $\gamma_k = 0$ for $k > 1$. More generally, the moving average model of order K is

$$x_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_K\varepsilon_{t-K}.$$

The MA(K) process is covariance stationary and has the property that $\gamma_j = 0$ for $j > K$. If we let $K = \infty$, we arrive at the infinite-order moving average process.

Now comes a point of some subtlety. If we believe that the $\{\Delta p_t\}$ are generated by the Roll model (a *structural* economic model), can we assert that a corresponding moving average model (a *statistical* model) exists? By playing around with the θ and σ_ε^2 parameters in the MA(1) model, we can obviously match the variance and first-order autocovariance of the structural Δp_t process. But this is not quite the same thing as claiming that the full joint distribution of the Δp_t realizations generated by the structural model could also be generated by an MA(1) model. Moreover, there's a good reason for suspecting this shouldn't be possible. The structural model has two (uncorrelated) sources of randomness, u_t (the efficient price innovations) and q_t (the trade direction indicators). The MA(1) model has only one source of randomness, ε_t.

Is the existence of an MA(1) representation an important issue? Why can't we simply limit the analysis to the structural model, and avoid questions of alternative representations? There are several answers. In the first place, the full structural model involves unobserved variables. The econometrician observes neither the u_t nor q_t, so he or she doesn't know the efficient price. The moving average representation is a useful tool for constructing an estimate of the efficient price, as well as for forecasting. Moreover, a moving average representation may be valid even if the structural model is misspecified.

Fortunately, an MA(1) representation for price changes in the Roll model does exist. In this assertion, we rely on the Wold (not Wald) theorem. The Wold theorem states that any zero-mean covariance stationary process $\{x_t\}$ can be represented in the form

$$x_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} + \kappa_t$$

where $\{x_t\}$ is a zero-mean white noise process, $\theta_0 = 1$ (a normalization), and $\sum_{j=0}^{\infty} \theta_j < \infty$. κ_t is a linearly deterministic process, which in this context means that it can be predicted arbitrarily well by a linear projection (possibly of infinite order) on past observations of x_t. For proofs, see Hamilton (1994) or Sargent (1979). For a purely stochastic series, $\kappa_t = 0$, and we are left with a moving average representation.

A related result due to Ansley, Spivey, and Wrobleski (1977) establishes that if a covariance stationary process has zero autocovariances at all orders higher than K, then it possesses a moving average representation of order K. This allows us to assert that an MA(1) representation exists for the Roll model.

Empirical market microstructure analyses often push the Wold theorem very hard. The structural models are often stylized and underidentified (we can't estimate all the parameters). The data are frequently non-Gaussian (like the trade indicator variable in the Roll model). Covariance stationarity of the observations (possibly after a transformation) is often a tenable working assumption. For many purposes, as we'll see, it is enough.

Section 3.4 derived the autocovariances of the Roll model (γ_0 and γ_1) in terms of the structural parameters (c and σ_u^2). The parameters of the corresponding MA(1) model in Equation (4.1) are θ and σ_ε^2. The MA(1) has autocovariances $\gamma_0 = (1+\theta^2)\sigma_\varepsilon^2$ and $\gamma_1 = \theta\sigma_\varepsilon^2$. From the autocovariances (or estimates thereof) we may compute the moving average parameters:

$$\theta = \frac{\gamma_0 - \sqrt{\gamma_0^2 - 4\gamma_1^2}}{2\gamma_1} \quad \text{and} \quad \sigma_\varepsilon^2 = \frac{\gamma_0 + \sqrt{\gamma_0^2 - 4\gamma_1^2}}{2} \tag{4.2}$$

This is actually one of two solutions, the so-called invertible solution. It has the property that $|\theta| < 1$, the relevance of which shortly becomes clear. The other (noninvertible) solution is $\{\theta^*,\ \sigma_\varepsilon^{2*}\}$. The relation between the two solutions is given by $\theta^* = 1/\theta$ and $\sigma_\varepsilon^{2*} = \theta^2\sigma_\varepsilon^2$. For the noninvertible solution, $|\theta^*| > 1$.

4.3 Autoregressive Models

A moving average model expresses the current realization in terms of current and lagged disturbances. These are not generally observable. For many purposes (particularly forecasting) it is useful to express the current realization in terms of past realizations. This leads to the autoregressive form of the model.

To develop this for the MA(1) case, note that we can rearrange $\Delta p_t = \varepsilon_t + \theta\varepsilon_{t-1}$ as $\varepsilon_t = \Delta p_t - \theta\varepsilon_{t-1}$. This gives us a backward recursion for $\varepsilon_t : \varepsilon_{t-1} = \Delta p_{t-1} - \theta\varepsilon_{t-2}, \varepsilon_{t-2} = \Delta p_{t-2} - \theta\varepsilon_{t-3}$, and so forth. Using this backward recursion gives

$$\begin{aligned}\Delta p_t &= \theta\left(\Delta p_{t-1} - \theta\left(\Delta p_{t-2} - \theta\left(\Delta p_{t-3} - \theta\varepsilon_{t-4}\right)\right)\right) + \varepsilon_t \\ &= \theta\Delta p_{t-1} - \theta^2\Delta p_{t-2} + \theta^3\Delta p_{t-3} - \theta^4\varepsilon_{t-4} + \varepsilon_t;\end{aligned} \tag{4.3}$$

or, with infinite recursion:

$$\Delta p_t = \theta\Delta p_{t-1} - \theta^2\Delta p_{t-2} + \theta^3\Delta p_{t-3} + \cdots + \varepsilon_t. \tag{4.4}$$

This is the autoregressive form: Δp_t is expressed as a linear function of its own lagged values and the current disturbance. Although the moving average representation is of order one, the autoregressive representation is of infinite order.

If $|\theta| < 1$, then the autoregressive representation is convergent: the coefficients of the lagged Δp_t converge to zero. Intuitively, the effects of lagged realizations eventually die out. When a convergent autoregressive representation exists, the moving average representation is said to be invertible. Convergence is determined by the magnitude of θ. The condition $|\theta| < 1$ thus defines the invertible solution for the MA(1) parameters (compare equation (4.2) and the related discussion). Hamilton (1994, p. 64) discusses general criteria for invertibility.

To move between moving average and autoregressive representations, it's often convenient to use the lag operator, L (sometimes written as the backshift operator, B). It is defined by the relation $Lx_t = x_{t-1}$. Multiple applications work in a straightforward fashion ($L^2 x_t = x_{t-2}$, etc.). The operator can also generate "leads" (e.g., $L^{-3} x_t = x_{t+3}$). Using the lag operator, the moving average representation for Δp_t is $\Delta p_t = \varepsilon_t + \theta L \varepsilon_t = (1 + \theta L)\varepsilon_t$. The autoregressive representation is:

$$\Delta p_t = \left(\theta L - \theta^2 L^2 + \theta^3 L^3 + \cdots\right) \Delta p_t + \varepsilon_t.$$

We derived this by recursive substitution. But there is an alternative construction that is particularly useful when the model is complicated. Starting from the moving average representation, we may write $(1 + \theta L)^{-1} \Delta p_t = \varepsilon_t$, where we've essentially treated the lag operator term as an algebraic quantity. If L were a variable and $|\theta| < 1$, we could construct a series expansion of the left-hand side around $L = 0$. This expansion, through the third order, is $[1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + O(L^4)]\Delta p_t = \varepsilon_t$, where $O(L^4)$ represents the higher order terms. This can be rearranged to obtain the autoregressive representation given in equations (4.4) or (4.5).

In summary, we have modeled a time series by assuming covariance stationarity, proceeding to a moving average representation (via the Wold theorem), and finally to the autoregressive representation. The last two representations are equivalent, but in any particular problem, one might be considerably simpler than the other. For example, the Roll model is a moving average of order one, but the autoregressive representation is of infinite order.

Sometimes, though, the autoregressive representation is the simpler one. An autoregressive representation of order one has the form $x_t = \phi x_{t-1} + \varepsilon_t$, or in terms of the lag operator, $(1 - \phi L)x_t = \varepsilon_t$. The moving average form is:

$$\begin{aligned} x_t &= (1 - \phi L)^{-1} \varepsilon_t = \left(1 + \phi L + \phi^2 L^2 + \cdots\right) \varepsilon_t \\ &= \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \cdots \end{aligned}$$

Here, we have used a power series expansion of $(1 + \phi L)^{-1}$. Recursive substitution would give the same result. The moving average representation is of infinite order.

The following exercise develops an autoregressive representation for a persistent discretely valued series. It demonstrates the generality of the Wold theorem by showing that such a process can be modeled using zero-mean uncorrelated disturbances. It also illustrates the limitations of the theorem by showing that the disturbances are dependent.

Exercise 4.1. For the trade direction indicator q_t in the Roll model, Madhavan, Richardson, and Roomans (1997) allow for serial dependence. Suppose that $q_t \in \{-1, +1\}$, and that $\Pr(q_{t+1} = +1 | q_t = +1) = \Pr(q_{t+1} = -1 | q_t = -1) = \alpha$ (and, of course, $\Pr(q_{t+1} = +1 | q_t = -1) = \Pr(q_{t+1} = -1 | q_t = +1) = (1 - \alpha)$. α is called the continuation probability. If $\alpha = 1/2$, trade directions are uncorrelated. If $1/2 < \alpha < 1$, trade directions are persistent (buys tend to follow buys, etc.). With this structure, q_t may be expressed as the AR(1) process $q_t = \phi q_{t-1} + v_t$ where $Ev_t = 0$, $Ev_t^2 = \sigma_v^2$, and $Ev_t v_{t-k} = 0$ for $k \neq 0$. The model may be analyzed by constructing a table of the eight possible realizations (paths) of $(q_t,\ q_{t+1},\ q_{t+2})$.

a. Assuming that q_t is equally likely to be ± 1, compute the probabilities of each path. Show that $\phi = 2\alpha - 1$.
b. Compute v_{t+1} and v_{t+2}. Verify that $Ev_{t+1} = Ev_{t+2} = 0$ and $\mathrm{Cov}(v_{t+1},\ v_{t+2}) = Ev_{t+1} v_{t+2} = 0$.
c. Demonstrate that the v_t values are not serially independent by verifying that $\mathrm{Cov}(v_{t+1},\ v_{t+2}^2) \neq 0$.

4.4 Forecasting

A crucial calculation in agents' trading decisions is their forecast of the security's future value. It is convenient to construct these forecasts by taking expectations of MA and AR representations, but there is an important qualification. The assumption of covariance stationarity suffices only to characterize a restricted form of the expectation. An expectation (e.g., $E[x_t | x_{t-1}, x_{t-2}, \ldots]$) generally involves the full joint distribution $f(x_t, x_{t-1}, x_{t-2}, \ldots)$, not just the means and covariances. Considerable simplification results, however, if we approximate the true expectation by a linear function of the conditioning arguments, that is, $E[x_t | x_{t-1}, x_{t-2}, \ldots] \approx \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots$. This approximate expectation is technically a linear projection. When the difference is important, it will be denoted E^* to distinguish it from the true expectation. The following material summarizes results on linear forecasting discussed at greater length in Hamilton (1994, pp. 72–116).

The technique of linear projection is especially compatible with AR and MA representations because the AR and MA representations have no more and no less information than is needed to compute the projection. It is quite conceivable that a more complicated forecasting scheme, for example, one involving nonlinear transformations of $\{x_{t-1}, x_{t-2}, \ldots\}$, might be better (have smaller forecasting errors) than the linear projection, but such a forecast could not be computed directly from the AR or MA representation. More structure would be needed.

We'll first consider the price forecast in the Roll model. Suppose that we know θ and have a full (infinite) price history up the time t,

$\{p_t, p_{t-1}, p_{t-2}, \ldots\}$. Using the autoregressive representation, we can recover the innovation series $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$. Then:

$$E^*\left[\Delta p_{t+1} | p_t, p_{t-1}, \ldots\right] = E^*\left[\varepsilon_{t+1} + \theta\varepsilon_t | p_t, p_{t-1}, \ldots\right] = \theta\varepsilon_t. \tag{4.6}$$

Therefore, the forecast of next period's price is: $f_t \equiv E^*[p_{t+1}|p_t, p_{t-1}, \ldots] = p_t + \theta\varepsilon_t$. How does f_t evolve?

$$\begin{aligned}\Delta f_t &= f_t - f_{t-1} = p_t + \theta\varepsilon_t - \left(p_{t-1} + \theta\varepsilon_{t-1}\right)\\ &= (\varepsilon_t + \theta\varepsilon_{t-1}) + \theta\varepsilon_t - \theta\varepsilon_{t-1} = (1+\theta)\,\varepsilon_t.\end{aligned} \tag{4.7}$$

That is, the forecast revision is a constant multiple of the innovation. The innovations process is uncorrelated, so the forecast revision is as well.

Now we raise a more difficult question. A martingale has uncorrelated increments, so f_t might be a martingale. Can we assert that $f_t = m_t$, that is, have we identified the true implicit efficient price? It turns out that there is a bit of problem. If $f_t = m_t$, then $p_t = f_t + cq_t$ and $\Delta p_t = \Delta f_t + c\Delta q_t$. But this implies

$$\Delta p_t = \varepsilon_t + \theta\varepsilon_{t-1} = (1+\theta)\,\varepsilon_t + c\Delta q_t \Leftrightarrow -\theta\,(\varepsilon_t - \varepsilon_{t-1}) = c\Delta q_t. \tag{4.8}$$

In other words, all of the randomness in the model is attributable to the q_t. But this is structurally incorrect: We know that changes in the efficient price, u_t, also contribute to the ε_t. Thus, we have not identified m_t. It will later be shown that $f_t = E^*[m_t|p_t, p_{t-1}, \ldots]$, that is, that f_t is the projection of m_t on the conditioning variables.

Exercise 4.2 The Roll model assumes that trade directions are serially uncorrelated: $\mathrm{Corr}(q_t, q_s) = 0$ for $t \neq s$. In practice, one often finds positive autocorrelation (see Hasbrouck and Ho (1987) Choi, Salandro, and Shastri (1988)). Suppose that $\mathrm{Corr}(q_t, q_{t-1}) = \rho > 0$ and $\mathrm{Corr}(q_t, q_{t-k}) = 0$ for $k > 1$. Suppose that ρ is known.

a. Show that $\mathrm{Var}(\Delta p_t) = 2c^2(1-\rho) + \sigma_u^2$, $\mathrm{Cov}(\Delta p_t, \Delta p_{t-1}) = -c^2(1-2\rho)$, $\mathrm{Cov}(\Delta p_t, \Delta p_{t-2}) = -c^2\rho$, and $\mathrm{Cov}(\Delta p_t, \Delta p_{t-k}) = 0$ for $k > 2$.
b. Suppose that $0 < \rho < 1$ describes the true structural model. We compute an estimate of c, denoted $\hat{c}$, assuming that the orginal Roll model is correct. Show that $\hat{c} < c$, that is, that $\hat{c}$ is biased downward.

Exercise 4.3 The basic Roll model assumes that trade directions are uncorrelated with changes in the efficient price: $\mathrm{Corr}(q_t, u_t) = 0$. Suppose that $\mathrm{Corr}(q_t, u_t) = \rho$, where ρ is known, $0 < \rho < 1$. The idea here is that a buy order is associated with an increase in the security value, a connection that will be developed in the models of asymmetric information. Suppose that ρ is known.

a. Show that $\mathrm{Var}(\Delta p_t) = 2c^2 + \sigma_u^2 + 2c\rho\sigma_u$, $\mathrm{Cov}(\Delta p_t, \Delta p_{t-1}) = -c(c + \rho\sigma_u)$, and $\mathrm{Cov}(\Delta p_t, \Delta p_{t-k}) = 0$ for $k > 1$.
b. Suppose that $0 < \rho < 1$ describes the true structural model. We compute an estimate of c, denoted $\hat{c}$, assuming that the orginal Roll model is correct. Show that $\hat{c} > c$, that is, that $\hat{c}$ is biased upward.

4.5 Estimation

In practice, the Roll model parameters are usually estimated as transformations of the estimated variance and first-order autocovariance of the price changes (see section 3.4). It is not uncommon, however, for the estimated first-order autocovariance to be positive. Harris (1990) shows that this can easily happen due to estimation error, even though the model is correctly specified. In these cases, Hasbrouck (2005) suggests a Bayesian approach.

More generally, MA and AR representations can be estimated using a wide variety of approaches. The MA parameters can be obtained from the autocovariances (by solving the set of equations and requiring that the solution be invertible). MA models can be estimated via maximum likelihood (assuming a particular distribution for the disturbances). The MA representation can also be obtained by numerically inverting the AR representation.

The autoregressive representation can often be conveniently estimated using ordinary least squares (OLS). The basic requirement for consistency of OLS estimation is that the residuals are uncorrelated with the regressors. This is true in equation (4.4) because the $\{\varepsilon_t\}$ are serially uncorrelated, and the regressors (lagged price changes) are linear functions of prior realizations of ε_t. For example, $\Delta p_{t-1} = \varepsilon_{t-1} + \theta\varepsilon_{t-2}$ is uncorrelated with ε_t.

Microstructure data often present particular challenges to statistical software. Samples often contain embedded breaks. In a sample of intraday trade prices that spans multiple days, for example, the closing price on one day and the opening price on the following day will appear successively. The overnight price change between these observations, though, will almost certainly have different properties than the intraday price changes. If the goal is modeling the latter, the overnight price changes should be dropped. This is often accomplished by inserting missing values into the series at the day breaks.

A related issue concerns lagged values realized before the start of the sample. In an autoregression like equation (4.4), if t is the first observation of the sample, none of the lagged values on the right-hand side are known. Most non-microstructure applications take the perspective that the start of sample simply represents the beginning of the record for a process that was already unfolding. For example, when a sample of GDP data begins in 1900, one would assume that the economy had been up and running prior to that date, and that 1900 merely represented the start of the record. In other words, prior to 1900 the process was evolving unobserved. The correct estimation approach is then unconditional, that is, the lagged missing values are viewed as unknown but distributed in accordance with the model. In many microstructure situations, though, the data begin right at the start of trading process. There is no prior unobserved evolution of the trading process. In these cases, conditional estimation, wherein the missing lagged disturbances are set to zero, is more defensible.

4.6 Strengths and Weaknesses of Linear Time-Series Models

This chapter reviews the elements of linear time-series analysis. The development begins with covariance stationarity, which is a plausible and minimal working assumption in many modeling situations. Using the Wold theorem, this leads to a moving average model, then to a vector autoregression, and finally to a forecasting procedure. These are powerful results, but to maintain a balanced perspective, it is now necessary to dwell on some of the framework's limitations.

The characterization of a time series offered by the linear models is not complete. The models do not fully describe the data-generating process. They do not specify how we should computationally simulate the process. Exercise 4.2, for example, posits first-order autocorrelation in the trade directions. We might simulate this as follows. Let $a_t = u_t + \beta u_{t-1}$ where $u_t \sim N(0, \sigma_u^2)$ and $0 < \beta < 1$. Then let $q_t = Sign(a_t)$, where $Sign(x) = +1$ if $x < 0$; -1 if $x < 0$; and 0 if $x = 0$. The resulting $\{q_t\}$ process is covariance stationary; we can specify AR and MA models, identify their parameters, compute forecasts $E^*[q_t | q_{t-1}, q_{t-2}, \ldots]$, and so on. We can't, however, completely reverse the inference and recover the generating mechanism.

As demonstrated in exercise 4.1, the disturbances in MA and AR models are not serially correlated, but may be serially dependent (as manifested by nonzero serial moments of order greater than two). This bears directly on the structural interpretations of these models. The MA and AR representations of a discretely valued process such as q_t are essentially linear models of limited dependent variables. The usual econometric guidelines discourage such specifications on the grounds that the disturbances must possess complicated and nonintuitive properties (see, for example, Greene (2002) p. 665).

These concerns are not misplaced in the present situation if our goal is a fully specified structural model. For example, if we believe that the a_t in the generating process summarize attributes of arriving individual traders, the higher order serial dependencies will generally imply complicated higher order dependencies in the attributes of successively arriving individuals. Rather than model this behavior, it is usually easier to specify a statistical model (such as a logit or probit) for which these features are not needed.

Logit and probit models are certainly part of the microstructure toolbox. We do not, however, use them reflexively simply because some variables of interest happen to be discrete. They are more complicated than linear specifications, and so are more demanding in programming and computation. A larger consideration is that there may be other structural features that are more important than discreteness. Probably paramount among these other concerns is time-varying and persistent volatility.

Even in light of these remarks, linear time-series analysis nevertheless retains strength and utility. It provides logically coherent and computationally simple tools for describing first-order dynamics, forecasting, and forming expectations. The underlying assumptions are minimal (chiefly covariance stationarity), so the analyses may be more robust to misspecification than more refined models. The representations are compatible with a wide range of structural models and so are relatively easy to illustrate and interpret. In short, they are useful aids in developing intuitions of how financial markets work.