DESIGNING AND ESTIMATING MODELS OF HIGH-FREQUENCY DATA

L. C. G. Rogers & O. Zane

University of Bath and First National Bank of Chicago

Research supported in part by EPSRC grant numbers GR/J97281 and GR/L10000.

L. C. G. Rogers
Financial Studies Group
Department of Mathematical Sciences
University of Bath
Bath BA2 7AY
Tel: +44 1225 826224
Fax: +44 1225 826492
email: lcgr@maths.bath.ac.uk

O. Zane
Quantitative Research
First Chicago NBD
1 Triton Square
London NW1 3FN
Tel: +44 171 903 4049
Fax: +44 171 528 7185
email: ozane@uk.fcnbd.com

KEYWORDS: high-frequency data, Markov, intensity, Ornstein-Uhlenbeck process.

DESIGNING AND ESTIMATING MODELS OF HIGH-FREQUENCY DATA

L. C. G. Rogers & O. Zane

University of Bath and First National Bank of Chicago

Summary. The growing availability of complete market data makes it important to have models which realistically describe the behaviour of such tick-by-tick datasets. We shall propose here a general modelling framework for tick data, illustrating it with a number of different specific choices which are reasonably tractable, and presenting estimation procedures which perform effectively on simulated samples.

1. Introduction. The log-Brownian paradigm for share prices is now well established, and has proven highly successful because of its simplicity and tractability. Anyone who has ever worked with it, and considered market data, knows that every one of the assumptions made in this model is open to question, and that it can be considered at best as a first approximation to what is really going on. The assumption of continuous sample paths with independent Gaussian increments appears to get less unrealistic as the time-intervals considered get larger, mainly because of some Central Limit effect; but this means that on shorter time scales, its deficiencies become even more marked, and this is especially true in the extreme where one considers tick data, that is, data which records every quote, or every trade. Such data is routinely available to financial houses and should *in principle* be far more informative than just daily price figures, but it will of course be hard to extract any added information if one insists on modelling it in a way which clearly fails to capture the gross features of the situation.

What then are these gross features? To fix our ideas, let us suppose that we are considering share quote data from a busy market [1] The first thing to remark is that the quote data is essentially *discrete.* Quotes are posted one at a time, and each posted quote carries a time at which it is posted, a price, and a quantity for sale or purchase. The simplest conceivable model would assume that the times at which quotes are posted form a Poisson process of some given intensity, and that the associated prices and quantities are drawn independently from some distribution. Such a model is inevitably unsatisfactory, because of the observed fact that the markets are generally much busier at certain times of day than at others. So we could assume that the Poisson process of times has a non-constant intensity; the simplest thing we could do is to assume that the intensity is a deterministic function of time, rising and falling in line with the historic intra-day level of market activity. However, this also is too simple-minded; such a model would mean that changes in prices of different shares would always be uncorrelated. So we have to allow a non-constant *stochastic* intensity for the process of times at which quotes are posted.

[1] We could as well be considering data on actual trades, or data on foreign exchange; the qualitative features will be broadly similar, and quote data has at least the merit of being more readily available than data on the trades themselves. We would model such trade data in a similar fashion to that discussed for quotes.

We also need to choose a stochastic intensity to model another gross feature of the data, which is that activity in shares tends to be *bursty*; there will be periods when, because of incoming news, the activity in a given share will increase, and this will persist for some time before settling back to a lower 'background' level. The way we propose to model this stochastic intensity is to suppose that there is some underlying Markov process, and that the intensities of different share quote processes are functions of this Markov process. In Section 2, we shall develop some of the theory of such models, concentrating particularly on expressions for means and covariances of readily-updated functionals of the history, under the assumption that the Markov process is ergodic. This simplifying assumption is probably innocent in most applications, and certainly makes the development a lot easier. In Section 3, we shall discuss the estimation of three particularly simple models. Working from simulated data, we tested the proposed simulation method, and the closeness of the estimates to the true values was encouraging.

Thus far, we have said little about the modelling and estimation of the sizes and prices of quotes. This is because we shall decouple the two problems by assuming that the sizes and prices are driven by a random mechanism *independent* of the process of times. This may seem a brave assumption, but some simple-minded analysis of share data from the London Stock Exchange showed rather surprisingly that during periods when an individual share was trading at increased levels of activity, there seemed to be no tendency for the share to be moving clearly up or clearly down. Of course, there will be situations (where some particularly adverse piece of news has been disclosed) when there will be a clear direction to the price moves, but it appears that this is not common. So we shall make the modelling simpler by this assumption; the results of Tauchen & Pitts [18] give some support to this assumption. The features of size and price which we will incorporate are that there should be some underlying 'notional' price process, and the quotes are scattered around that in some way. In particular, quotes for large quantities are likely to be quite keenly priced, close to the notional, as small quotes are often posted as a way of generating trade, or testing the mood of the market, but big quotes mispriced cost money. For a very liquid asset, we could probably dispense with the need to have some underlying 'notional' price process, basing the new quotes on recent history, but to be able to handle a situation where the asset is not very frequently traded, we must allow for the underlying price to be evolving even though no trades are taking place; if it is two days since there was any activity in a minor share, we have to be prepared for the possibility that the share has changed in value even though there were no quotes available to make this visible. We develop one simple example of such a model in Section 4, and present an estimation method for it.

The final (and probably most important) issue in tick-data modelling is to understand the implications of the chosen model for pricing and hedging. We must return to this on another occasion, though we report elsewhere [13] some results on a simplified parody of the models discussed here which serves as a first study on liquidity effects. The models studied here are of course incomplete-market models, which are by their nature hard to work with; there seem to be no models of incomplete markets where one can say much in closed form about pricing and hedging. But there is another feature of these models which sets them apart from other types of market incompleteness (such as stochastic volatility, or transactions costs), namely, that in the current models *it is not possible to choose any* *portfolio you wish at any time*; this feature wipes out arbitrage-pricing theory.

An excellent recent overview of the state of research into high-frequency data has been given by Goodhart & O'Hara [8]. As is clear from that article, there is a diversity of approaches to the topic, and relatively little that can be considered conclusively proven. Some researchers (for example, Kyle [9], Glosten & Milgrom [7], Easley & O'Hara [4], [5], Admati & Pfleiderer [1], [2]) try to explain the formation of prices by modelling the different information of market participants. This approach has considerable intellectual appeal, but seems to be hard to convert into closed-form solutions. The situation is analogous to the pricing of shares by way of a log-Brownian motion; the possible equilibrium justification of such an assumption may be rather difficult, but it certainly allows for fruitful conclusions. Similarly, the approach we take here makes no claims to the high ground of economic theory, but does attempt to model simply the principal features which any high-frequency data model should include.

Another approach (surveyed by Bollerslev, Chou, & Kroner [3]) is to model the price process by some form of GARCH/ARCH process; this should be considered as a fitting exercise, rather than an attempt to explain economic fundamentals. The standard GARCH framework has difficulty in dealing with the different times between trades which characterise high-frequency data, though the recent paper of Engle & Russell [6] proposes a GARCH-type model of the times between trades for a share. Quite how such a modelling framework may deal with many shares at once is not clear to us. Recent work of Rydberg & Shephard [17] studies stochastic intensity models for tick data, and analyses all trades of IBM stock on the NYSE in 1995. The class of processes of times of trades for a single share which they consider is similar to ours, but the model for movement of prices is different, in that they assume that the price jumps are independent with a common distribution. This assumption makes it harder to explain the observation of Roll [16] that there is negative correlation between successive prices.

2. Stochastic-intensity point process models. The aim of this Section is to present a wide class of stochastic-intensity models for the point process of times of quotes, and to develop this theory far enough to derive expressions for means of functionals of the point process which are very simple to update, and can therefore form the basis for a dynamic estimation procedure, using a generalised method of moments approach.

Underlying everything will be a Markov process X [2], which we shall assume to be stationary and ergodic, with invariant distribution π. Next we take standard [3] Poisson processes $\tilde{N}^1, \dots \tilde{N}^K$ independent of X, and consider the counting processes

$$N_t^i \equiv \tilde{N}^i\Big(\int_0^t f_i(X_s)ds\Big), \quad i = 1, \dots, K, \tag{1}$$

[2] Basic facts and definitions about Markov processes are covered in various texts, for example, Rogers & Williams [10], and we refer the reader there for any unexplained terminology.

[3] That is, the number of points of the point process which occur in an interval of length h is Poisson with mean h, and the numbers of points occurring in disjoint intervals are independent.

where the functions f_i are bounded and non-negative [4] . The process N^i is our model for the times at which quotes for the ith share are posted. If we are observing the market and trying to estimate parameters of some model, our estimates necessarily change (albeit slowly) as time passes, and it is clearly impractical to use estimation procedures which require frequent review of the entire history of the share so far. We therefore propose to concentrate on simple functionals of the share history, which are easy to update:

$$Y_t^i(\alpha) \equiv \int_{-\infty}^t \alpha e^{\alpha(s-t)} dN_s^i, \tag{2}$$

$$\eta_t^{ij}(\alpha,\beta) \equiv \int_{-\infty}^t \alpha e^{\alpha(s-t)} Y_{s-}^j(\beta) dN_s^i. \tag{3}$$

$$\xi_t^{ijk}(\alpha,\beta,\gamma) = \int_{-\infty}^t \alpha e^{\alpha(s-t)} Y_{s-}^j(\beta) Y_{s-}^k(\gamma) dN_s^i \tag{4}$$

The parameters α, β and γ are positive. There may be benefit in considering other functionals, but these are good enough to be starting with, and are certainly easy to update; the processes $Y^i(\alpha)$, $\eta^{ij}(\alpha,\beta)$ and $\xi^{ijk}(\alpha,\beta,\gamma)$ decay exponentially between quotes, and when a new quote comes in they jump up again, by α in the case of Y^i, by $\alpha Y_{t-}^j(\beta)$ in the case of η^{ij}, and by $\alpha Y_{t-}^j(\beta) Y_{t-}^k(\gamma)$ in the case of ξ^{ijk}. Thus with suitable choice of the parameters, some finite collection $\{Y^i, \eta^{ij}, \xi^{ijk}, i, j, k = 1, \ldots, K\}$ is sufficient to update itself.

Let us introduce the abbreviated notation $(\pi, g) \equiv \int g(x)\pi(dx)$, the integration taking place over the statespace of the Markov process X. The first result we need is the following.

PROPOSITION 1.

$$EY_0^i(\alpha) = (\pi, f_i) \tag{5}$$

$$E\eta_0^{ij}(\alpha,\beta) = \beta(\pi, f_j R_\beta f_i) \tag{6}$$

$$E\xi_0^{ijk}(\alpha,\beta,\gamma) = \beta\gamma\big[(\pi, f_k R_\gamma f_j R_{\beta+\gamma} f_i) + (\pi, f_j R_\beta f_k R_{\beta+\gamma} f_i) + \delta_{jk}(\pi, f_j R_{\beta+\gamma} f_i)\big] \tag{7}$$

Proof. See Appendix.

Remarks. The approach of the proof of Proposition 1 yields expressions for higher moments. We record here a few higher moments which are useful in computing the covariance of the various quantities:

$$EY_0^i(\alpha) Y_0^j(\beta) = \frac{\alpha\beta}{\alpha+\beta}\,(\pi, \delta_{ij} f_i + f_i R_\alpha f_j + f_j R_\beta f_i) \tag{8}$$

[4] The functions f_i only need to satisfy mild integrability conditions to guarantee that later expressions are finite-valued; it is certainly sufficient that each f_i is bounded, but that is far from necessary. Rather than try to give sufficient conditions of broad applicability, we shall leave the integrability check to the user in each application.

$$
\begin{aligned}
(9)\qquad E\eta_0^{ij}(\alpha,\beta)Y_0^k(\gamma) = \frac{\alpha\beta\gamma}{\alpha+\gamma}\ &(\pi, f_jR_\beta f_iR_\alpha f_k + f_jR_\beta f_kR_{\gamma+\beta}f_i \\
&+ f_kR_\gamma f_jR_{\gamma+\beta}f_i + \delta_{ik}f_jR_\beta f_k + \delta_{jk}f_kR_{\gamma+\beta}f_i)
\end{aligned}
$$

$$
\begin{aligned}
(10)\ E\eta_0^{ij}(\alpha,\beta)\eta_0^{kl}(\lambda,\mu) = \frac{\alpha\beta\lambda\mu}{\alpha+\lambda}\ &(\pi, f_lR_\mu f_kR_\lambda f_jR_{\lambda+\beta}f_i + f_lR_\mu f_jR_{\beta+\mu}f_kR_{\lambda+\beta}f_i \\
&+ f_jR_\beta f_lR_{\beta+\mu}f_kR_{\lambda+\beta}f_i + f_jR_\beta f_iR_\alpha f_lR_{\alpha+\mu}f_k \\
&+ f_jR_\beta f_lR_{\beta+\mu}f_iR_{\alpha+\mu}f_k + f_lR_\mu f_jR_{\beta+\mu}f_iR_{\alpha+\mu}f_k \\
&+ \delta_{ik}\delta_{jl}f_lR_{\beta+\mu}f_i + \delta_{il}f_jR_\beta f_iR_{\alpha+\mu}f_k + \delta_{jk}f_lR_\mu f_jR_{\beta+\lambda}f_i \\
&+ \delta_{ik}\{f_lR_\mu f_jR_{\beta+\mu}f_i + f_jR_\beta f_lR_{\beta+\mu}f_i\} \\
&+ \delta_{jl}\{f_lR_{\beta+\mu}f_kR_{\lambda+\beta}f_i + f_lR_{\beta+\mu}f_iR_{\alpha+\mu}f_k\})
\end{aligned}
$$

where $(R_\lambda)_{\lambda>0}$ is the resolvent of X, and δ_{ij} is the Kronecker delta, 1 if $i=j$ and 0 else.

From a practical point of view, it is very easy to keep the updated values of the Y^i, η^{ij} and ξ_{ijk} if one has a real-time data-feed; in order to make use of these for estimation, we shall need to have explicit forms for the moments (5)-(7) for the particular example under consideration, and in the rest of this Section, we give three examples where (5)-(7) can be evaluated explicitly: in some cases, we can make progress in evaluating (9) and (10), but the expressions quickly become unmanageable, and are best handled with an algebra manipulation package.

Example 1: finite Markov chain. If the Markov process is a finite-state chain with state-space F and jump-rate matrix Q, then we may consider the function $f_i : F \to \mathbb{R}^+$ as the vector $(f_i(x))_{x\in F}$, and in this identification, the resolvent R_α acts as multiplication by the matrix $(\alpha - Q)^{-1}$. The expressions (5)-(7) are immediately evaluated by substitution.

In the next Section, we shall study the simplest possible case of a single share and a Markov chain on the two states 0 and 1 as the underlying Markov process. The behaviour of the share is determined by four parameters: $f(0)$, $f(1)$, $q_0 \equiv q_{01}$ and $q_1 \equiv q_{10}$. We shall demonstrate the feasibility of estimation in this context. For this particularly simple chain, the expressions (4)–(5) are not hard to evaluate in closed form: we obtain

$$
E\ Y_0(\alpha) = \frac{q_1f_0 + q_0f_1}{q_0+q_1}
$$

$$
E\ \eta_0(\alpha,\beta) = \frac{q_1{f_0}^2(\beta+q_1) + 2\,q_1f_0q_0f_1 + q_0{f_1}^2(\beta+q_0)}{(q_0+q_1)\,(\beta+q_1+q_0)}
$$

Introducing the moments $m_k \equiv (q_1f_0^k + q_0f_1^k)/(q_1+q_0)$ of the invariant law, and the shorthand $\tau \equiv q_0+q_1$ for the trace of Q, we have more compactly the expressions

$$
(11)\qquad E\ Y_0(\alpha) = m_1
$$

$$
(12)\qquad E\ \eta_0(\alpha,\beta) = \frac{\tau m_1^2 + \beta m_2}{\beta+\tau}
$$

$$
(13)\qquad (\pi, fR_\beta fR_\alpha f) = \frac{\tau^2m_1^3 + \alpha\beta m_3 + (\alpha+\beta)\tau m_1m_2}{\alpha\beta(\alpha+\tau)(\beta+\tau)}
$$

$$(\pi, fR_\mu fR_\lambda fR_\beta f) =$$

$$= \frac{\tau^4 m_1^4 + m_4\mu\lambda\beta\tau + \mu\beta\tau^2 m_2^2 + (\lambda+\beta+\mu)m_1^2 m_2\tau^3 + \lambda(\beta+\mu)\tau^2 m_1 m_3}{\lambda\beta\mu\tau(\lambda+\tau)(\mu+\tau)(\beta+\tau)} \tag{14}$$

These expressions allow one to build up the values (5)–(10) of the first few moments of the estimation functionals Y and η.

Note that though a Markov chain is a simple process, it may not be ideal for modelling in that if it has more than a very few states, the jump-rate matrix Q will have a large number of entries, and these will typically be very hard to estimate reliably.

Example 2: Ornstein-Uhlenbeck process. As another example, we take the underlying Markov process to be the n-dimensional Ornstein-Uhlenbeck process which is the solution to

$$dX_t = dW_t - BX_t dt,$$

where W is a standard n-dimensional Brownian motion, and B is an $n\times n$ matrix which we assume without much loss of generality to be diagonal, $B = \text{diag}(b_i)$, and positive definite. At this level of generality, we shall take the functions f_i to be non-negative quadratic forms $f_i(x) \equiv (x-c_i)\cdot K_i(x-c_i)$. This family of examples is just about tractable using an algebra manipulator such as Maple or Mathematica, but the formulae which result are not worth writing down; one simply converts them into their Fortran or C equivalents and embeds them in a program designed to do all the hard work. The point is that the operator $\lambda - \mathcal{G}$ acts on the space of polynomials as a linear operator, whose inverse may also be expressed as a linear operator, and this allows the expressions appearing in (5)–(10) above to be computed.

To illustrate this, we reduce to the situation $n = 1$, which is studied in depth in the next section. The expressions (5)–(10) are needed in the estimation procedure, and to evaluate them, we shall need to take expectations of polynomials of degree up to 8 in the variable X_0, which has a $N(0, 1/(2b))$ distribution. If we identify the polynomial $p(x) \equiv \sum_{r=0}^{8} a_r x^r$ with the vector $(a_0, a_1, \ldots, a_8)^T$, then the operator $\lambda - \mathcal{G}$ acts as the matrix

$$\begin{bmatrix} \lambda & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda+b & 0 & -3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda+2\,b & 0 & -6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda+3\,b & 0 & -10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda+4\,b & 0 & -15 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda+5\,b & 0 & -21 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda+6\,b & 0 & -28 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda+7\,b & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda+8\,b \end{bmatrix}$$

whose inverse is also upper triangular, and can be easily computed with Maple.

We can likewise express the action of multiplying a polynomial of degree at most 6 by the quadratic $\theta_0 + \theta_1 x + \theta_2 x^2$ by way of the matrix

$$
\begin{bmatrix}
\theta_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\theta_1 & \theta_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\theta_2 & \theta_1 & \theta_0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \theta_2 & \theta_1 & \theta_0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \theta_2 & \theta_1 & \theta_0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \theta_2 & \theta_1 & \theta_0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \theta_2 & \theta_1 & \theta_0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \theta_2 & \theta_1 & \theta_0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \theta_2 & \theta_1 & \theta_0
\end{bmatrix}
$$

Taking expectations with respect to the invariant measure is equivalent to taking the inner product with the vector

$$(1, 0, \frac{1}{2b}, 0, \frac{3}{4b^2}, 0, \frac{15}{8b^3}, 0, \frac{105}{16b^4})$$

Using these components, it is possible with patience to build up the expressions (5)–(10). We obtain

$$EY_0(\alpha) = Kc^2 + \frac{K}{2b}$$

$$E\eta_0(\alpha,\beta) = \frac{\left((2\,b+3\,\beta)\,(\beta+b) + 4\,b\,(b+3\,\beta)\,(\beta+2\,b)\,c^2 + 4\,b^2\,(\beta+b)\,(\beta+2\,b)\,c^4\right)K^2}{4b^2\,(\beta+b)\,(\beta+2\,b)}$$

We also need for the computation of the covariance of $\eta_0(\alpha,\beta)$ and $Y_0(\gamma)$ expressions of the form

$$
\begin{aligned}
(\pi, fR_\alpha fR_\beta f) = \Big[&(\beta+b)\,(\alpha+b)\,(4\,b^2+6\,\alpha\,b+6\,\beta\,b+15\,\beta\,\alpha) + 2b(45\,\beta^2\,\alpha^2 \;+87\,b\,\beta\,\alpha^2 \\
&+169\,b^2\,\beta\,\alpha + 12\,b^4 + 38\,b^3\,\alpha + 18\,\alpha^2\,b^2 + 87\,b\,\beta^2\,\alpha + 38\,b^3\,\beta \\
&+18\,\beta^2\,b^2)c^2 \;+4\,b^2\,(\beta+2\,b)\,(\alpha+2\,b)\,(3\,b^2+7\,\alpha\,b+7\,\beta\,b+15\,\beta\,\alpha)\,c^4 \\
&+8\,b^3\,(\beta+b)\,(\beta+2\,b)\,(\alpha+b)\,(\alpha+2\,b)\,c^6\Big]K^3\,. \\
&\Big[8\alpha\,(\alpha+2\,b)\,\beta\,b^3\,(\alpha+b)(\beta+b)\,(\beta+2\,b)\Big]^{-1}
\end{aligned}
$$

The expression for $(\pi, fR_\alpha fR_\beta fR_\lambda f)$ which results is too lengthy to be worth writing down here.

For the multidimensional situation, the analysis is much more involved, and the dimension of the problem quickly becomes prohibitive (there are 120 different terms of the form $x_1^{n_1}x_2^{n_2}x_3^{n_3}$ when $n_1+n_2+n_3=8$, for example) but we shall discuss a 'full' version of the estimation procedure below, as well as a 'quick-and-dirty' version, and we shall see that the latter really does quite well.

Example 3: A simple interacting system. Though this example is a special case of Example 1, its additional structure makes it possible to analyse further. We take a continuous-time

Markov chain with statespace $I = \{0,1\}^N$, an element $x = (x_1, x_2, \ldots, x_N) \in I$ being thought of as a listing of the states of N shares, where we interpret $x_i = 1$ to mean that share i is *excited*, and $x_i = 0$ to mean that share i is *quiet*. We suppose that each excited share becomes quiet at rate μ, and that the invariant distribution of the chain is given by

$$\pi(x) \propto \exp\Big[\tfrac{1}{2} b \sum\sum_{i \neq j} x_i x_j + a \sum_i x_i\Big],$$

where a and b are constants. Assuming reversibility, we are able to deduce the rates at which shares become excited; if $x_i = x_i'$ for all $i \neq j$, and $0 = x_j < x_j' = 1$, we conclude that jumps from x to x' take place at rate

$$\frac{\pi(x')}{\pi(x)}\,\mu = \mu \exp\Big[a + b \sum_i x_i\Big].$$

We can often usefully think of the process $\nu(t) \equiv \sum_i x_i(t)$ as a birth-and-death chain on $\{0, 1, \ldots, N\}$ with jump rates

$$\begin{aligned} k \mapsto k+1 \quad & \text{at rate} \quad (N-k)\exp(a+bk) \\ k \mapsto k-1 \quad & \text{at rate} \quad k\mu \end{aligned}$$

and with equilibrium distribution

$$P[\nu(0) = k] \propto \binom{N}{k} \exp\{ak + \tfrac{1}{2} bk(k-1)\}.$$

In this example, we naturally take $f_i(x) = c_i + c_i' I_{\{x_i = 1\}}$, where c_i and c_i' are positive constants, and now need to compute at least the moments (5) and (6). We have the explicit expressions

$$\begin{aligned} P(x_i(0) = 1) &= \frac{\sum_{k=0}^N \frac{k}{N} \binom{N}{k} \exp\{ak + \frac{1}{2}bk(k-1)\}}{\sum_{k=0}^N \binom{N}{k} \exp\{ak + \frac{1}{2}bk(k-1)\}} \\ &= \frac{\sum_{k=0}^{N-1} \binom{N-1}{k} \exp\{ak + a + \frac{1}{2}bk(k+1)\}}{\sum_{k=0}^N \binom{N}{k} \exp\{ak + \frac{1}{2}bk(k-1)\}} \\ &= \pi(N; a, b), \end{aligned}$$

say. For small N, we can work with this expression explicitly, but for larger N, the asymptotic

$$\begin{aligned} P[\nu(0) = k] &\sim c\ \exp[ak + \tfrac{1}{2}bk(k-1) - (k+\tfrac{1}{2})\log\Big(\frac{k}{N}\Big) - (N-k+\tfrac{1}{2})\log\Big(1 - \frac{k}{N}\Big)] \\ &= c\ \exp[N\{a\,\frac{k}{N} + \tfrac{1}{2}bN\,\frac{k}{N}\frac{k-1}{N} - \frac{k+\frac{1}{2}}{N}\log\Big(\frac{k}{N}\Big) - (1 - \frac{k-\frac{1}{2}}{N})\log\Big(1 - \frac{k}{N}\Big)\}] \end{aligned}$$

so that the proportion k/N of excited states must be approximately the value ξ which maximises this expression; ξ must solve

$$\frac{1-\xi}{\xi} = \exp(-a - bN\xi),$$

though there is not in general a unique solution to this equation (there are always either 1 or 3 roots).

In order to compute the expectations (6), we need to be able to evaluate $P(x_j(T) = 1|x_i(0) = 1)$, where T is an exponential variable of rate β independent of x. Firstly, we deal with the case $i \neq j$. By considering the first time that share i becomes quiet, it is not hard to convince oneself that

$$\begin{aligned} P(x_j(T) = 1|x_i(0) = 1) &= \frac{\beta}{\beta+\mu}\, P(x_j(T) = 1|\, x_i(s) = 1\ \forall s \leq T) \\ &\quad + \frac{\mu}{\beta+\mu}\, P(x_j(T) = 1|\, x_i(0) = 0) \\ &= \frac{\beta}{\beta+\mu}\, \pi(N-1; a+b, b) + \frac{\mu}{\beta+\mu}\, P(x_j(T) = 1|x_i(0) = 0). \qquad (15) \end{aligned}$$

We combine this equation with the simpler equation

$$\pi(N; a, b) = \pi(N; a, b)P(x_j(T) = 1|x_i(0) = 1) + (1 - \pi(N; a, b))P(x_j(T) = 1|x_i(0) = 0)$$

to deduce an expression for $P(x_j(T) = 1|x_i(0) = 1)$, namely

$$P(x_j(T) = 1|x_i(0) = 1) = \frac{\mu\pi(N; a, b) + \beta\pi(N-1; a+b, b)(1 - \pi(N; a, b))}{\mu + \beta(1 - \pi(N; a, b))}.$$

For $i = j$, a similar expression results with the term in $\pi(N-1; a, b)$ replaced by 1. The analysis which leads to these expressions is not entirely correct, however; if τ is the first time that x_i becomes 0, the law of $x(\tau)$ is *not* the same as the invariant law of x given that $x_i = 0$. Indeed, if we know that x_i has been 1 for the whole of the time interval $[0, \tau)$, then the other sites are more likely to be excited. An exact numerical computation of the true value shows that for realistic parameter values the discrepancy is actually very small, so we will ignore it for the sake of computational speed. The resulting estimates reported in the next section are good enough that the extra effort required to do the computations exactly (which would be prohibitive for many shares) is unnecessary.

3. Estimation of the parameters of the timing Markov process. We suppose we have for each $j = 1, \dots, K$ a sequence $(\tau_i^j)_{i=1}^N$ of times at which quotes in the jth share were posted. In applications, these would come from a market data feed; in our analysis they were simulated (of which we will say more presently), but either way we take these as the inputs for the estimation procedure. We now generate a sequence $(Z_i)_{i=1}^N$ of d-vector observables, where for each i and each $j = 1, \dots, d$, the observable Z_i^j is the value at time τ_i either of some $Y^l(\alpha)$ or of some $\eta^{lm}(\beta,\lambda)$, where the parameters and the particular shares are chosen suitably. The main points to bear in mind are that small values of α will give relatively stable estimates of the long-run average behaviour, but will be less successful in probing properties such as mean-reversion in the Ornstein-Uhlenbeck example, say, whereas values of α which are too big will result in estimate instability. We also should choose observables Z^j which give information about all the shares of interest. The observables update quite simply, as we have

$$\begin{aligned} Y^l_{\tau(n)}(\alpha) &= \alpha + e^{-\alpha(\tau_n-\tau_{n-1})}Y^l_{\tau(n-1)}(\alpha), \\ \eta^{lm}_{\tau(n)}(\beta,\lambda) &= \alpha(Y^m_{\tau(n)}(\lambda)-\lambda) + e^{-\alpha(\tau_n-\tau_{n-1})}\eta^{lm}_{\tau(n-1)}(\beta,\lambda). \end{aligned}$$

As is usual in this sort of work, the initial values of Y^l and η^{lm} are not determined, and one way to proceed would be to take them to be zero, then update all the way through the sequence, and then repeat, using the final estimates as the initial. What we actually did is to take the first several thousand prices to 'burn in' the Y's and η's, and then begin to estimate from there. The mean μ_Z and covariance V_Z of the vector Z_i, based on the assumption that the underlying Markov process is in equilibrium, can now be expressed via the equations (5)–(10); they will depend on the parameters Θ of the Markov process. So in order to estimate those parameters, we perform the minimisation

$$\min_{\Theta}(Z_n-\mu_Z)\cdot V_Z^{-1}(Z_n-\mu_Z). \tag{16}$$

This gives a parameter estimate which will be updated with each new quote. The function to be minimised is typically highly nonlinear, and a general minimisation routine will be needed for this step. It is reasonable in practice to fix V_Z^{-1} by evaluation at the last minimising value of Θ. However, this minimisation does not prove to be as laborious as it sounds, since when a new quote comes in we simply take the last estimate of the parameter Θ as the starting point for the minimisation; as the minimising value is unlikely to have changed much from one quote to the next, the minimisation is usually achieved very rapidly.

One possible practical problem with this approach is that the expressions for the covariances in V_Z are frequently quite unwieldy, as the examples of the previous section show. So a 'quick and dirty' approach to the problem is to replace the quadratic form in (16) by a diagonal quadratic form, whose diagonal elements are just $(Z^j_1)^{-2}$. We display in the Figures the results of doing this 'quick and dirty' minimisation on the two-state Markov chain example with parameters $f_0 = 1$, $f_1 = 18$, $q_0 = 0.05$ and $q_1 = 0.5$. The functionals we chose for the estimation were $Y(0.01)$, $Y(0.005)$, $\eta(0.0051, 80)$, $\eta(0.00501, 5)$, $\eta(0.00505, 0.5)$, $\xi(0.005, 20, 20)$, and $\xi(0.01, 10, 10)$. Next we display the results of the 'quick and dirty' minimisation on Example 2, the one-dimensional Ornstein-Uhlenbeck example, with parameters $b = 0.5$, $K = 20$, and $c = 1$, and using estimation functionals $Y(0.005)$, $\eta(0.00499, 40)$, $\eta(0.00498, 0.2)$, $\xi(0.00502, 0.8, 0.8)$, and $\xi(0.00503, 8, 8)$. And the final figures refer to the estimation of Example 3 where the parameters are $\alpha = \beta = 0.05$, $\mu = 1$, and the c_i and c'_i take the same values for all shares, namely 1.5 and 10 respectively. The estimation was based on $Y(0.04)$ and seven η processes involving different shares and parameter values. In all cases, the simulation 'beds in' with a few thousand events before the estimation begins, and then we take 10000 events in each example. The results reported are expressed as the estimated value divided by the true value, so we are looking for paths which stay close to the dashed line at level 1. The x-axis is numbered in blocks of 50, so corresponds in each case to the whole 10000 events. As can be seen, the estimation of the parameters of Example 1 is remarkably good. The estimates of the parameters of Example 2 is acceptable, but this is not an example that one would feel too confident about using in practice. This is because it may be very hard to determine the parameter values from the data *whatever* estimation procedure is used. Consider what would happen if b were very large; this would mean that the diffusion remains close to 0 most of the time, so that the intensity $f(X_t) = K(X_t - c)^2$ will most of the time be approximately Kc^2. Thus we should expect to be able to recover an estimate of Kc^2 with high accuracy, but the separate estimates of K and c are likely to be very poor. Turning to example 3, we find that the estiamtes of α and $c' \equiv f_1$ are quite acceptable, but that the estimates of β and μ wander quite a lot. One difference between this example and the other two is that here we did not use any ξ processes, and including these would undoubtedly improve the estimation. Part of the difficulty of using ξs is that there is only an approximate expression for the mean values, and the closeness of approximation is hard to assess. As part of the numerical study, we computed the covariance matrix of a collection of Ys and ηs using the approximate argument, and found that it was not non-negative-definite. Thus the approximation may not be very good here, but at least one may say that for larger numbers of shares we would expect these approximations to give better answers.

4. Estimation of the distribution of sizes and prices. We now turn to the structure and estimation of the sizes and prices of quotes. For simplicity, we shall assume that, conditional on the times of the quotes for all shares, the sizes and prices for *different* shares are *independent*. This assumption reduces the situation to that of a single share, which allows us to simplify notation by omitting the index for the share under consideration.

So we shall assume that we have a single share, and that quotes for this share are posted at times $(\tau_i)_{i \in \mathbb{Z}}$ generated according to the mechanism introduced in Section 2. The quote at time τ_i will consist of an observed log-price y_i and an amount a_i, whose joint distribution we shall soon specify. Notice that the amount may be negative or positive, according as the quote is an offer to buy or to sell. Firstly, we shall suppose that *the amounts* (a_i) *are independent identically-distributed (IID) random variables.* To estimate the distribution of the a_i, we simply take market data and inspect it. We might take N quotes, note the values $a_1, \dots, a_N$, and form the empirical distribution (or some smoothed version of it); alternatively, if the empirical distribution looks close to some simple parametric family, we may assume that the distribution is from that parametric family, and proceed to estimate the parameters. In short, the problem we are dealing with here is very well studied; we are attempting to estimate the distribution of an IID sequence.

Next we shall specify the behaviour of the y_i, in terms of an 'underlying' price process z, which we shall assume for the present discussion is a drifting Brownian motion $z_t = \sigma W_t + \mu t$ (though see later comments on other alternatives). We shall suppose that

$$y_i = z(\tau_i) + f(a_i)\varepsilon_i, \tag{17}$$

where the ε_i are IID zero-mean normals with unit variance, and the function f reflects the effect that if the size a_i is smaller, the variance is likely to be larger. How do we go about estimating the unknowns μ, σ and f of this specification, as well as the underlying price process z?

To begin with, we propose to estimate f from data on frequently-traded shares. If f is supposed to explain the tendency for larger deals to be more accurately priced, it is reasonable to suppose that this effect is common across shares; and if this is so, we may estimate it from the more liquid shares. The unobservability of z complicates this estimation, but if we were to form some exponentially-weighted average of past log-price values,

$$\bar{y}_i \equiv \frac{\sum_{j<i} e^{-\lambda(\tau_i - \tau_j)} y_j}{\sum_{j<i} e^{-\lambda(\tau_i - \tau_j)}}, \tag{18}$$

where $\lambda > 0$ is not too small, $\bar{y}_i$ should be a reasonable approximation to $z(\tau_i)$. Thus if we considered the sample of pairs $(a_i, y_i - \bar{y}_i)$ we should be seeing something which looks like $(a_i, f(a_i)\varepsilon_i)$, and from this we may estimate the function f. This part of the estimation needs to be done from historical data, but would not need to be updated very frequently.

The next stage of the estimation is to form estimates of the parameters μ and σ of the underlying process z. Since the effect of z on a small time scale is likely to be small compared to the fluctuations of $f(a_i)\varepsilon_i$, we have to look on longer time-scales to estimate μ and σ. On these longer time-scales, the effect of the error terms $f(a_i)\varepsilon_i$ may safely be ignored, and we are effectively in the situation of estimating the mean and variance of an IID sequence of Gaussian random variables. This we propose to do using a Bayesian procedure, since it is important to keep track of the errors in our estimates of μ and σ, especially for μ, where the errors in estimation are notoriously large. If we have a sequence $(Z_i)_{i\geq 1}$ of IID Gaussian variables with mean α and variance $v \equiv 1/w$, we shall suppose we have a prior density for (α, w) of the form

$$p_0(\alpha, w) = c_0 \exp(-\tfrac{1}{2}k_0 w(\alpha - a_0)^2 - \rho_0 w - \theta w^{-1} - \gamma_0 \log(w))$$

for some constants $\rho_0 > 0$, $\theta > 0$, $k_0 > 0$, γ_0, a_0 and the normalising constant

$$c_0 = \sqrt{\frac{k_0}{2\pi}}\, 2 \left(\frac{\theta}{\rho_0}\right)^{(2\gamma_0+3)/4} K_{\gamma_0+3/2}(2\sqrt{\rho_0\theta}).$$

If we now assume inductively that the posterior density of (α, w) given $Z_1, \ldots, Z_n$ has the form

$$p_n(\alpha, w) = c_n \exp(-\tfrac{1}{2}k_n w(\alpha - a_n)^2 - \rho_n w - \theta w^{-1} - \gamma_n \log(w)),$$

then the joint density of (Z_{n+1}, α, w) will be

$$c_n\sqrt{2\pi}\ \exp(-\tfrac{1}{2}w(z-\alpha)^2 - \tfrac{1}{2}k_n w(\alpha - a_n)^2 - \rho_n w - \theta w^{-1} - (\gamma_n - \tfrac{1}{2})\log(w)),$$

which is easily worked into the form

$$c_n\sqrt{2\pi}\ \exp(-\tfrac{1}{2}w(1+k_n)(\alpha - \frac{z + k_n a_n}{1+k_n})^2 - \tfrac{1}{2}w\ \frac{k_n(z-a_n)^2}{1+k_n} - \rho_n w - \theta w^{-1} - (\gamma_n - \tfrac{1}{2})\log(w)).$$

Hence we deduce the updating rules:

$$k_{n+1} = 1 + k_n \tag{19}$$

$$a_{n+1} = \frac{z + k_n a_n}{1 + k_n} \tag{20}$$

$$\rho_{n+1} = \rho_n + \frac{k_n (z - a_n)^2}{2(1 + k_n)} \tag{21}$$

$$c_{n+1} = \frac{\sqrt{k_{n+1}}}{2\pi}\, 2 \left(\frac{\theta}{\rho_{n+1}} \right)^{(2\gamma_{n+1}+3)/4} K_{\gamma_{n+1}+3/2}(2\sqrt{\rho_{n+1}\theta}) \tag{22}$$

$$\gamma_{n+1} = \gamma_n - \tfrac{1}{2} \tag{23}$$

Applying this filtering procedure to (say) historical weekly data will generate a posterior distribution for the unknown μ and σ^2.

The final step of the estimation procedure is different in that we need to estimate the 'underlying' price $z(\tau_i)$, rather than any parameter. For this, we propose a Kalman filter, working on the assumption that the value of σ^2 is the posterior mean value. This is a bold assumption, but in view of the fact that one typically knows the volatility of a share with much higher precision than the mean return, it may suffice. In any case, the theory of the Kalman filter is sufficiently well known that we need only report the final form of the updating; if conditional on observations up to time τ_{n-1} we have

$$\begin{pmatrix} z(\tau_{n-1}) \\ \mu \end{pmatrix} \sim N\left(\begin{pmatrix} \hat{z}_{n-1} \\ \hat{\mu}_{n-1} \end{pmatrix}, \begin{pmatrix} v_{zz} & v_{z\mu} \\ v_{\mu z} & v_{\mu\mu} \end{pmatrix} \right)$$

and the nth quote is for amount a_n at time s_n later, then conditional on all observations up to time τ_n we have

$$\begin{pmatrix} z(\tau_n) \\ \mu \end{pmatrix} \sim N\left(\begin{pmatrix} \hat{z}_n \\ \hat{\mu}_n \end{pmatrix}, \begin{pmatrix} v'_{zz} & v'_{z\mu} \\ v'_{\mu z} & v'_{\mu\mu} \end{pmatrix} \right)$$

where

$$\hat{z}_n = \hat{z}_{n-1} + \hat{\mu}_{n-1} s_n + \nu_n \, \frac{\theta}{f(a_n)^2 + \theta}$$

$$\hat{\mu}_n = \hat{\mu}_{n-1} + \nu_n \, \frac{\psi}{f(a_n)^2 + \theta},$$

where $s_n \equiv \tau_n - \tau_{n-1}$, a_n is the amount of the nth quote, $\nu_n \equiv y_n - \hat{z}_{n-1} - \hat{\mu}_{n-1} s_n$ is the innovation, and

$$\theta = \sigma^2 s_n + v_{zz} + 2 s_n v_{z\mu} + s_n^2 v_{\mu\mu}$$

$$\psi = v_{z\mu} + s_n v_{\mu\mu}.$$

The covariance matrix is

$$\begin{pmatrix} v'_{zz} & v'_{z\mu} \\ v'_{\mu z} & v'_{\mu\mu} \end{pmatrix} \equiv \begin{pmatrix} \frac{\theta f(a_n)^2}{f(a_n)^2+\theta} & \frac{\psi f(a_n)^2}{f(a_n)^2+\theta} \\ \frac{\psi f(a_n)^2}{f(a_n)^2+\theta} & v_{\mu\mu} - \frac{\psi^2}{f(a_n)^2+\theta} \end{pmatrix}.$$

If we want to step away from the assumption that the increments of z are Gaussian, we could allow z to be a Lévy process with jumps, and this would give a likelihood for the changes in z which would not be of simple quadratic form. To handle this, we propose an approximate Kalman filter, as explained in Rogers & Zane [12] (a sketch of the method is summarised in Rogers [14]); the minimiser of the conditional likelihood provides the new estimate of the mean, and the second derivative of the conditional likelihood provides the new estimate of the inverse of the covariance.

6. Conclusions. We have presented in this paper a very general framework for modelling high-frequency data, we have derived general estimation methods, and have in three concrete examples of interest developed the estimation of simulated data to show that the methods proposed really can recover the parameter values satisfactorily. Much remains to be done; the most obvious directions for further research are in testing such models against actual data, and in studying the consequences of such models for pricing and hedging. Nevertheless, a start has been made; the techniques required are often quite demanding compared to some other approaches, but the modelling framework seems robust enough to be able to answer further questions put to it, and also rich enough to be able to model many of the qualitative features of high-frequency data. These are promising properties, and we look forward to further developments.

Appendix. We collect here several proofs and other discussions of a technical nature which may have interrupted the flow at an earlier point. To follow the proofs, it is necessary to have an understanding of martingales at about the level of Williams [19], and of finite-variation stochastic integration at about the level of Part 3 of Chapter IV in Rogers & Williams [11].

Proof of Proposition 1. The starting point is the fact that

$$(A1) \qquad N^i_t - \int_0^t f_i(X_s)ds \equiv N^i_t - \Lambda^i_t \text{ is a finite-variation martingale}$$

for each i. This means then that the processes

$$\begin{aligned} M^i_t(\alpha) &\equiv \int_{-\infty}^t \alpha e^{\alpha s}(dN^i_s - d\Lambda^i_s) \\ &= e^{\alpha t} Y^i_t(\alpha) - \int_{-\infty}^t \alpha e^{\alpha s} f_i(X_s)ds \end{aligned} \qquad (A2)$$

are also finite-variation martingales, so

$$\begin{aligned} E\Big[M^i_t(\alpha) M^j_t(\beta)\Big] &= E\Big[\sum_{s\le t} \Delta M^i_t(\alpha) \Delta M^j_t(\beta)\Big] \\ &= 0 \end{aligned} \qquad (A3i)$$

if $i \neq j$, because the Poisson processes $\tilde{N}^i$ are independent, and therefore have no jumps in common. If $i = j$, then we have instead

$$\begin{aligned} E\Big[M^i_t(\alpha) M^i_t(\beta)\Big] &= E\Big[\sum_{s\le t} \alpha\beta e^{\alpha s} e^{\beta s} \Delta N^i_s\Big] \\ &= E\int_{-\infty}^t \alpha\beta e^{(\alpha+\beta)s} dN^i_s \\ &= E\int_{-\infty}^t \alpha\beta e^{(\alpha+\beta)s} f_i(X_s)ds, \end{aligned} \qquad (A3ii)$$

again using (A1). Now from (A2) we use the fact that $EM^i_t(\alpha) = 0$ to learn that

$$\begin{aligned} E[Y^i_0(\alpha)] &= E\int_{-\infty}^0 \alpha e^{\alpha s} f_i(X_s)ds \\ &= \int_{-\infty}^0 \alpha e^{\alpha s}(\pi, f_i)ds \\ &= (\pi, f_i), \end{aligned}$$

which is equality (4).

Turning to equality (6), we need the fact that

$$E[M_t^i(\alpha)|\mathcal{F}^X] = 0 \tag{A4}$$

for all i, t, and α, where $\mathcal{F}^X \equiv \sigma(X_t, t \in \mathbb{R})$ is the σ-field generated by the process X. This is because the Poisson processes $\tilde{N}^i$ are independent of X. Using (A4), we therefore have from (A2) and (A3) that

$$\begin{aligned} E\Big[Y_0^i(\alpha)Y_0^j(\beta)\Big] &= E\Big[M_t^i(\alpha)M_t^i(\beta)\Big] + E\Big[\int_{-\infty}^0 \alpha e^{\alpha s} f_i(X_s)ds \int_{-\infty}^0 \beta e^{\beta u} f_j(X_u)du\Big] \\ &= \delta_{ij}\, \frac{\alpha\beta}{\alpha+\beta}\, (\pi, f_i) + E\Big[\int_{-\infty}^0 \alpha e^{\alpha s} f_i(X_s)ds \int_{-\infty}^0 \beta e^{\beta u} f_j(X_u)du\Big], \end{aligned} \tag{A5}$$

leaving just the final term to understand. The product of the integrals breaks into two similar terms, depending on which of s or u is the larger. One of the terms is

$$\begin{aligned} E\Big[\int_{-\infty}^0 &\alpha e^{\alpha s} f_i(X_s)\Big(\int_s^0 \beta e^{\beta u} f_j(X_u)du\Big)ds\Big] \\ &= \int_{-\infty}^0 \alpha e^{\alpha s}\Big(\int_s^0 \beta e^{\beta u}(\pi, f_i P_{u-s} f_j)du\Big)ds \\ &= \int_0^\infty \alpha e^{-\alpha t} \int_0^t (\pi, f_i P_v f_j)\beta e^{\beta(v-t)}dvdt \\ &= \int_0^\infty e^{\beta v}(\pi, f_i P_v f_j)\, \frac{\alpha\beta}{\alpha+\beta}\, e^{-(\alpha+\beta)v}dv \\ &= \frac{\alpha\beta}{\alpha+\beta}\, (\pi, f_i R_\alpha f_j), \end{aligned} \tag{A6}$$

where $(P_t)_{t\geq 0}$ is the transition semigroup of the Markov process X, and

$$R_\alpha \equiv \int_0^\infty \alpha e^{-\alpha t} P_t dt$$

is the resolvent. Together with the similar expression from the integral where $s > u$ we combine (A5) and (A6) to obtain (6).

Next, for (5), we have

$$\begin{aligned} E\eta_0^{ij}(\alpha,\beta) &= E\Big[\int_{-\infty}^0 \alpha e^{\alpha s} Y_{s-}^j(\beta)dN_s^i\Big] \\ &= E\Big[\int_{-\infty}^0 \alpha e^{\alpha s} Y_{s-}^j(\beta) f_i(X_s)ds\Big] \\ &= E\Big[\int_{-\infty}^0 \alpha e^{\alpha s}\Big(\int_{-\infty}^s \beta e^{\beta(u-s)} f_j(X_u)du\Big) f_i(X_s)ds\Big] \\ &= \int_{-\infty}^0 \alpha e^{\alpha s}\Big(\int_{-\infty}^s \beta e^{\beta(u-s)}(\pi, f_j P_{s-u} f_i)du\Big)ds \\ &= \beta(\pi, f_j R_\beta f_i) \end{aligned}$$

after a few calculations, which is (5).

To compute the mean of $\xi_0^{ijk}(\alpha,\beta,\gamma)$, we proceed similarly to obtain

$$\begin{aligned}
E\xi_0^{ijk}(\alpha,\beta,\gamma) &= E\Big[\int_{-\infty}^{0} \alpha e^{\alpha s} Y_{s-}^{j}(\beta) Y_{s-}^{k}(\gamma) dN_s^i\Big] \\
&= E\Big[\int_{-\infty}^{0} \alpha e^{\alpha s} Y_{s-}^{j}(\beta) Y_{s-}^{k}(\gamma) f_i(X_s) ds\Big] \\
&= E[Y_{0-}^{j}(\beta) Y_{0-}^{k}(\gamma) f_i(X_0)] \\
&= E[\{M_{0-}^{j}(\beta) + \int_{-\infty}^{0} \beta e^{\beta s} f_j(X_s) ds\}. \\
&\qquad \{M_{0-}^{k}(\gamma) + \int_{-\infty}^{0} \gamma e^{\gamma s} f_k(X_s) ds\} f_i(X_0)] \\
&= \delta_{jk} E[\int_{-\infty}^{0} \beta\gamma e^{(\beta+\gamma)s} f_j(X_s) ds f_i(X_0)] \\
&\qquad + E\Bigg[\Big(\int_{-\infty}^{0} \beta e^{\beta s} f_j(X_s) ds\Big)\Big(\int_{-\infty}^{0} \gamma e^{\gamma s} f_k(X_s) ds\Big) f_i(X_0)\Bigg] \\
&= \delta_{jk} \beta\gamma (\pi, f_j R_{\beta+\gamma} f_i) \\
&\qquad + E\Bigg[\Big(\int_{-\infty}^{0} \beta e^{\beta s} \big(\int_{-\infty}^{s} \gamma e^{\gamma u} f_k(X_u) du\big) f_j(X_s) ds\Big) f_i(X_0)\Bigg] \\
&\qquad\qquad + E\Bigg[\Big(\int_{-\infty}^{0} \gamma e^{\gamma s} \big(\int_{-\infty}^{s} \beta e^{\beta u} f_j(X_u) du\big) f_k(X_s) ds\Big) f_i(X_0)\Bigg] \\
&= \beta\gamma\big(\delta_{jk}(\pi, f_j R_{\beta+\gamma} f_i) + (\pi, f_j R_\beta f_k R_{\beta+\gamma} f_i) + (\pi, f_k R_\gamma f_j R_{\beta+\gamma} f_i)\big)
\end{aligned}$$

as stated at (7).

References

[1] ADMATI, A. R., and PFLEIDERER, P., A theory of intraday patterns: volume and price variability, *Rev. Fin. Studies* **1**, 3–40, 1988.

[2] ADMATI, A. R., and PFLEIDERER, P., Divide and conquer: a new theory of intraday and day-of-the-week effects, *Rev. Fin. Studies* **2**, 189–223 1989.

[3] BOLLERSLEV, T., CHOU, R. Y., and KRONER, K. F., ARCH modelling in finance: a review of the theory and empirical evidence, *J. International Econ.* **36**, 355–372, 1992.

[4] EASLEY, D., and O'HARA, M., Price, trade size and information in securities markets. *J. Fin. Econ.* **19**, 113–138, 1987.

[5] EASLEY, D., and O'HARA, M., Time and the process of security price adjustment, *J. Finance* **47**, 577–605, 1992.

[6] ENGLE, R. F., and RUSSELL, J. R., Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model, *J. Empirical Finance* **4**, 187–212, 1997.

[7] GLOSTEN, L. R., and MILGROM, P. R., Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders, *J. Fin. Econ.* **14**,71–100, 1985.

[8] GOODHART, C. A. E., and O'HARA, M. High frequencey data in financial markets: issues and applications. *J. Empirical Finance* **4**, 73–114, 1997.

[9] KYLE, A. S., Continuous auctions and insider trading, *Econometrica* **53**, 1315–1335, 1985.

[10] ROGERS, L. C. G., and WILLIAMS, D., *Diffusions, Markov Processes, and Martingales, Vol. 1*, Wiley, Chichester, 1994.

[11] ROGERS, L. C. G., and WILLIAMS, D., *Diffusions, Markov Processes, and Martingales, Vol. 2*, Wiley, Chichester, 1987.

[12] ROGERS, L. C. G., and ZANE, O., Fitting potential models to interest rate and foreign exchange data, *Vasicek and beyond*, ed. L. Hughston, RISK Publications, London 1997, 327–342.

[13] ROGERS, L. C. G., and ZANE, O., A study of liquidity effects: II, Preprint, 1998.

[14] ROGERS, L. C. G., One for all, *RISK* **10**, 57–59, March 1997.

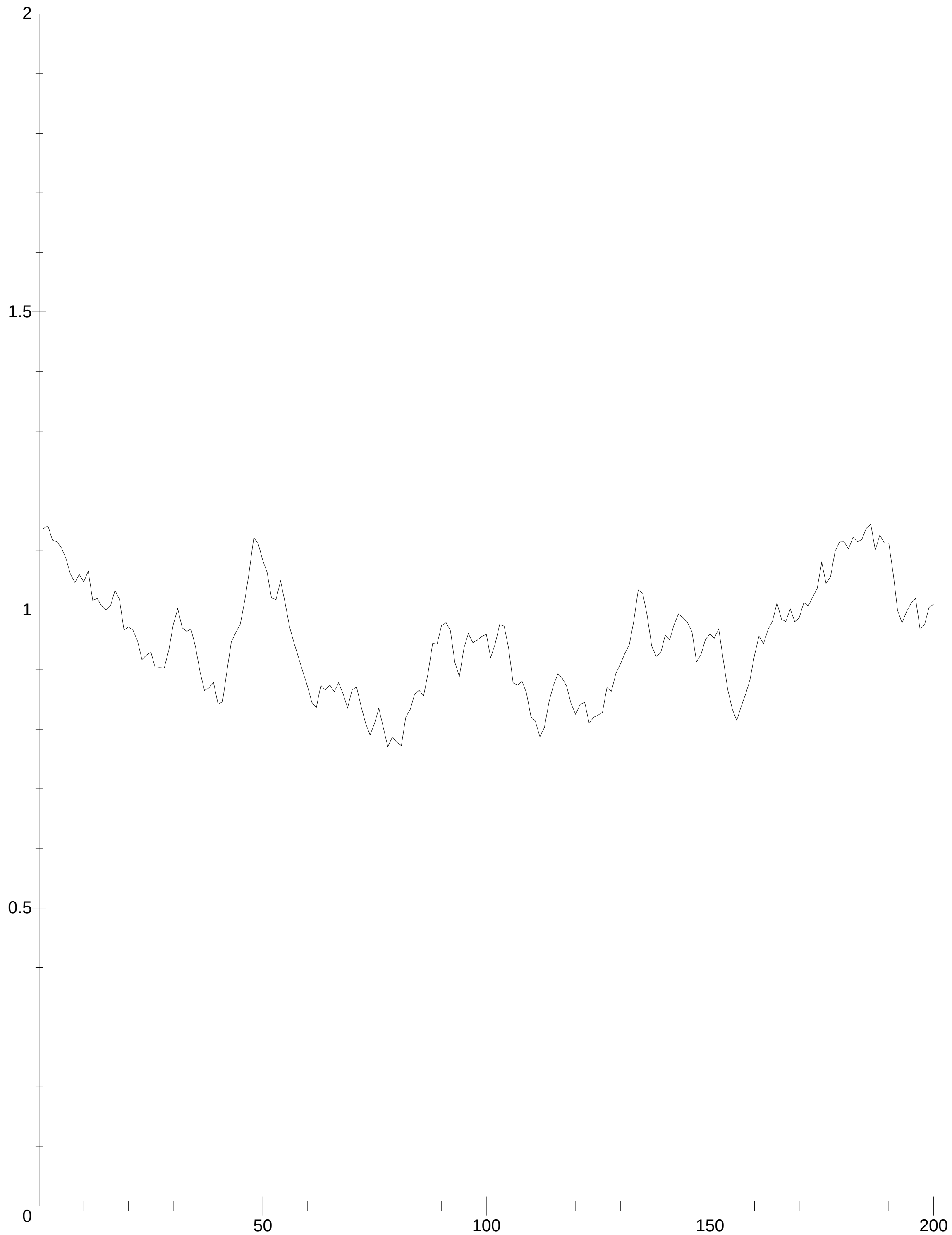
[15] ROGERS, L. C. G., Volatility forecasting in a tick-data model, Preprint 1998.

[16] ROLL, R., A simple implicit measure of the effective bid-ask spread in an efficient market, *J. Finance,* **39**, 1127–1139, 1984.

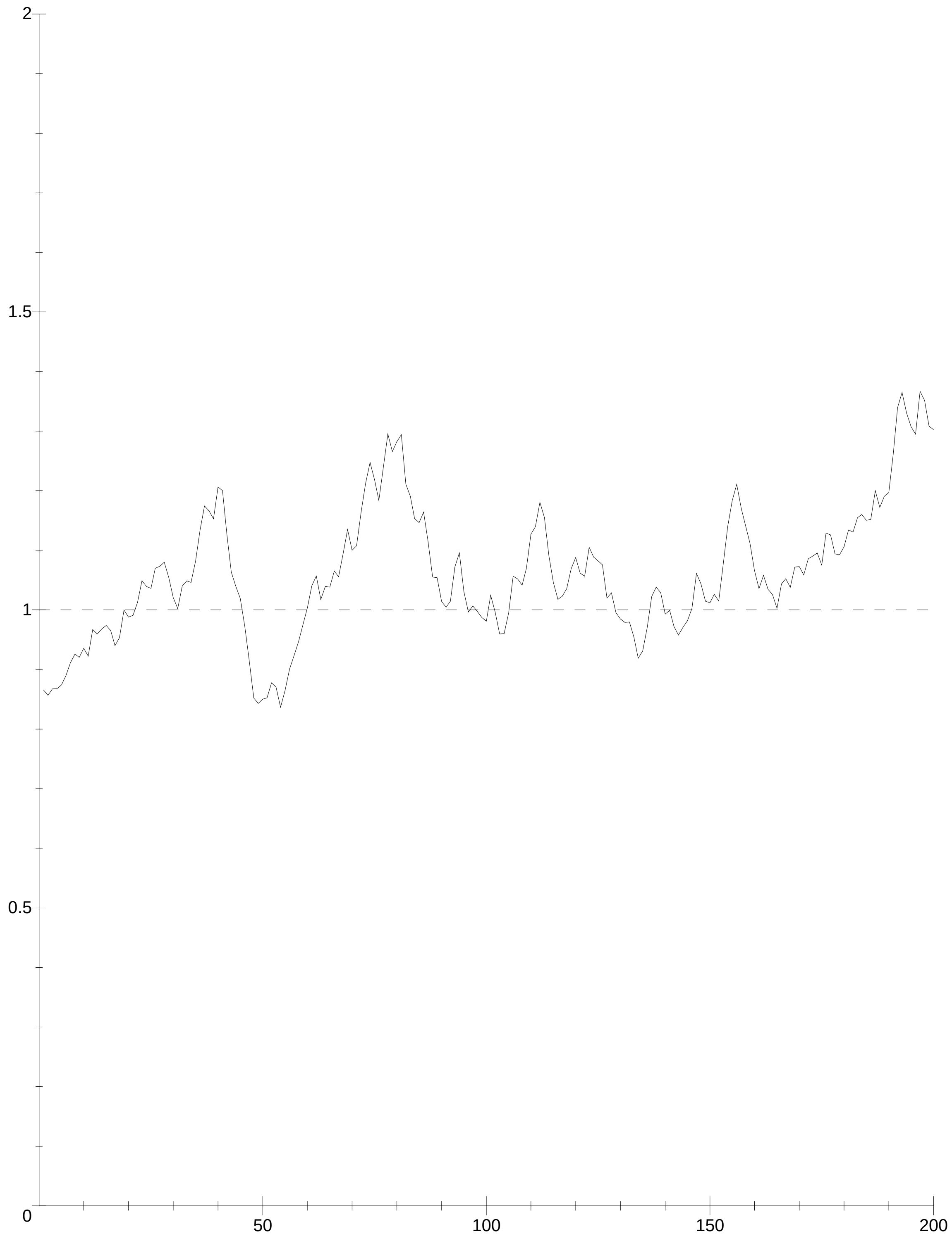
[17] RYDBERG, T., and SHEPHARD, N., A modelling framework for the prices and times of trades made on the NYSE, Preprint, 1998.

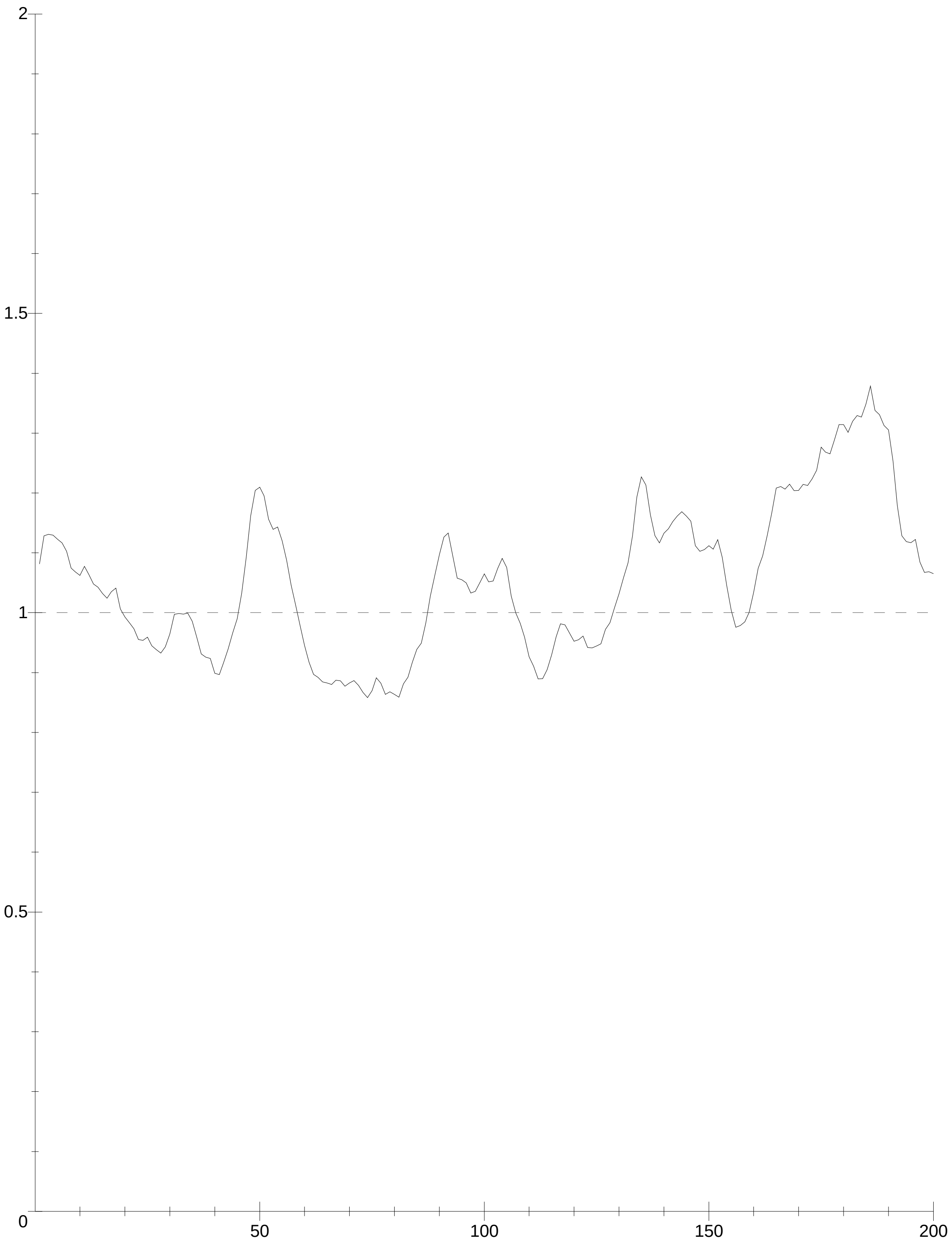
[18] TAUCHEN, G. E., and PITTS, M. The price variability-volume relationship on speculative markets. *Econometrica* **51**, 485–505, 1983.

[19] WILLIAMS, D., *Probability with Martingales*, Cambridge Univ. Press, Cambridge 1991.

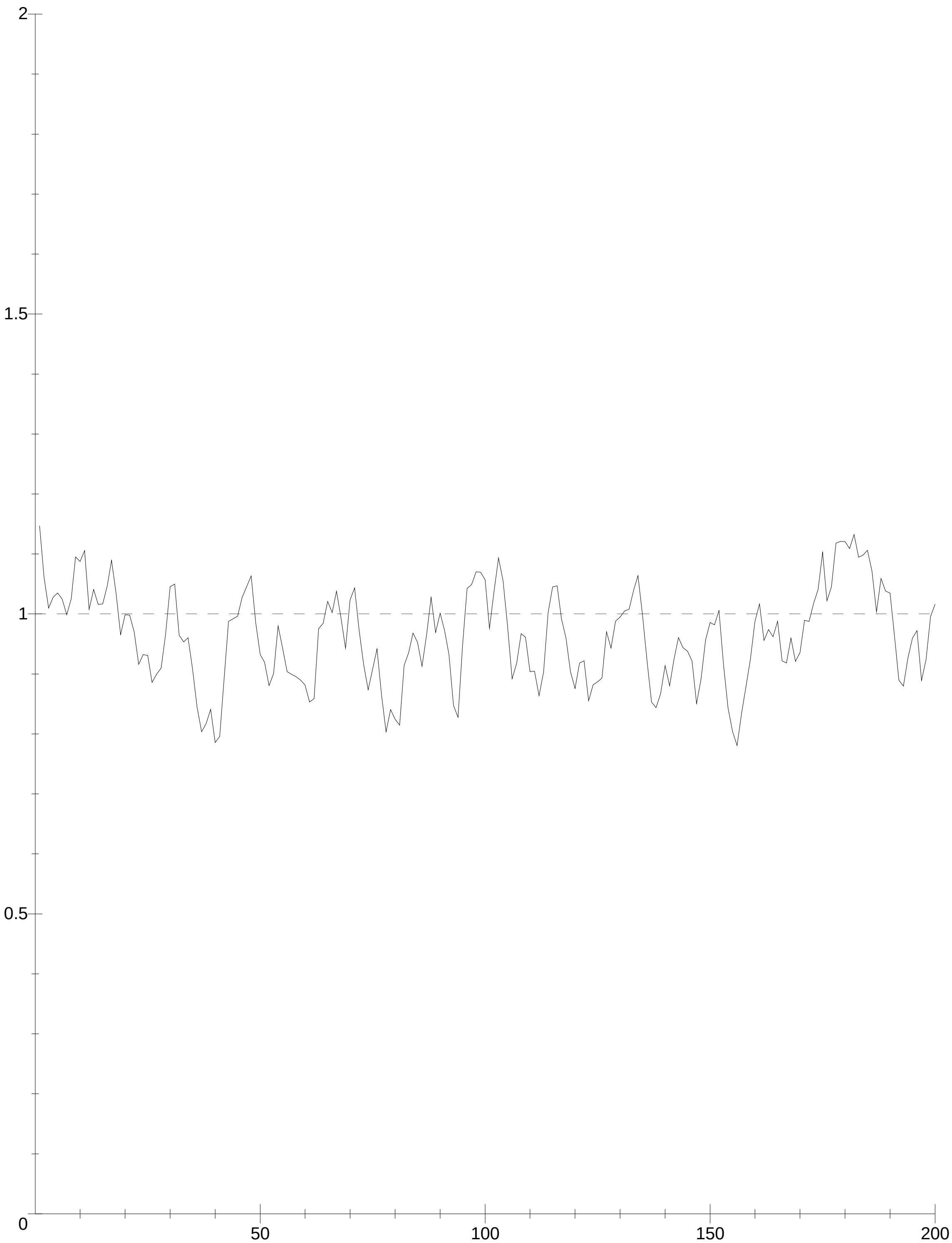


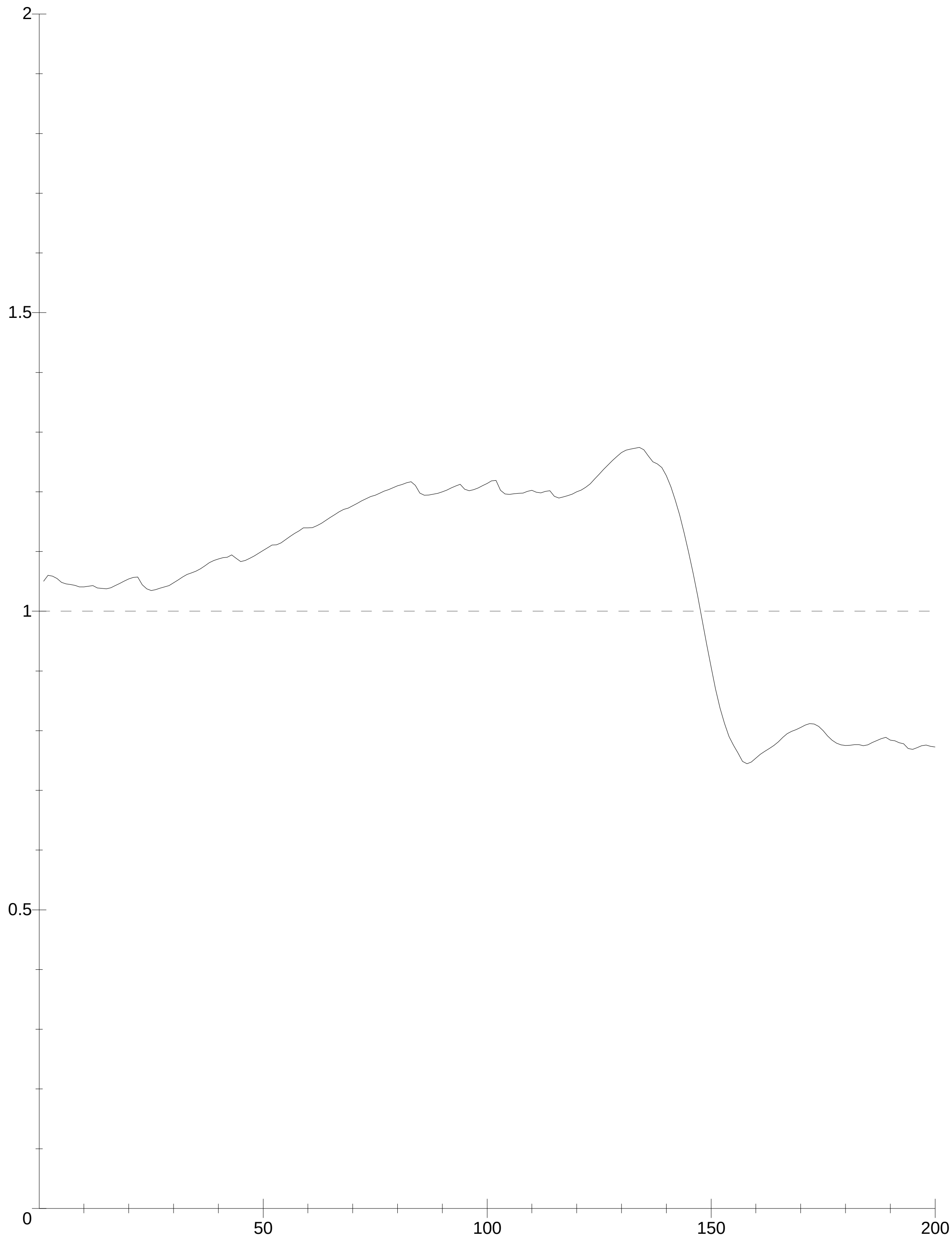
2
1.5
1
0.5
0
50
100
150
200

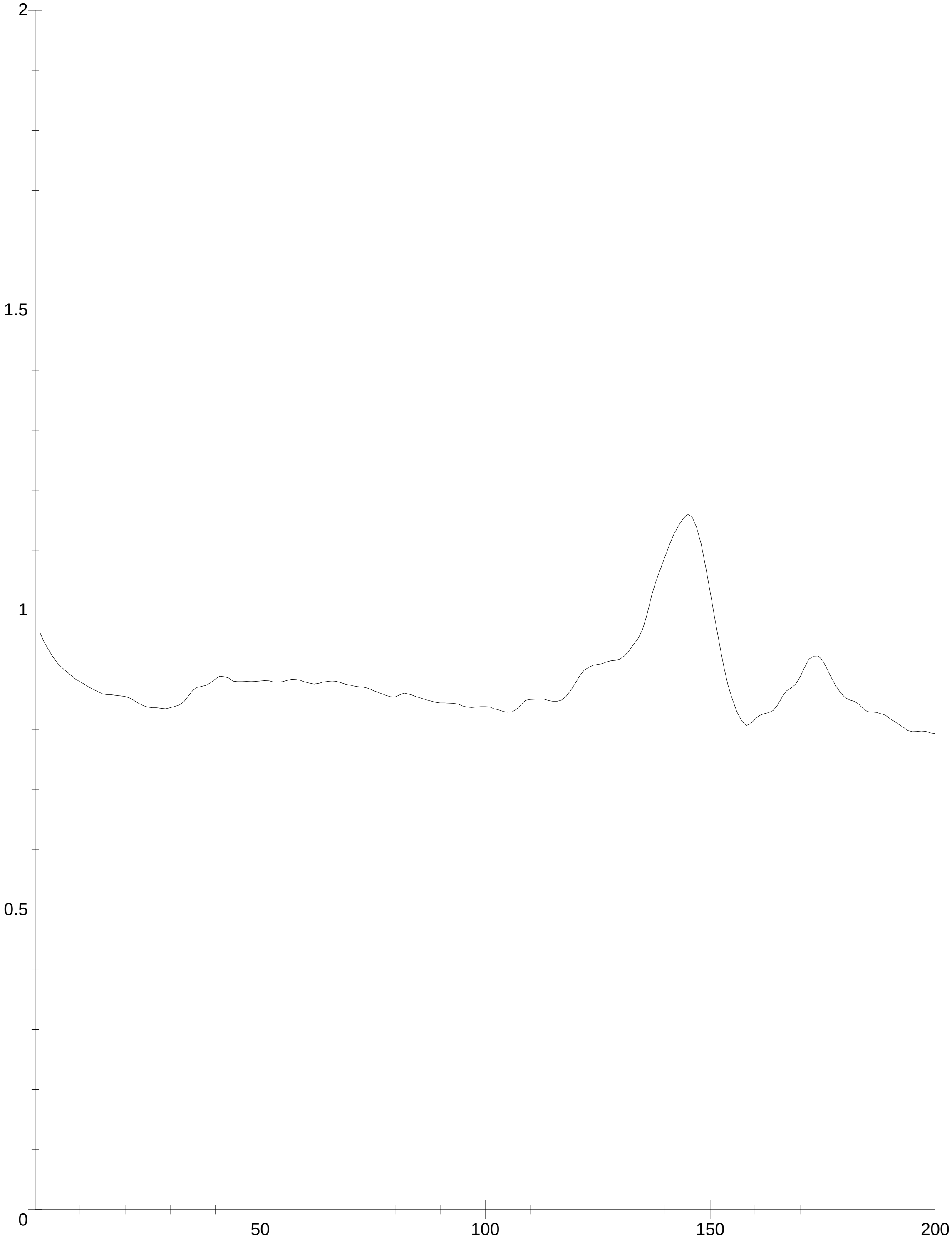

2
1.5
1
0.5
0
50
100
150
200



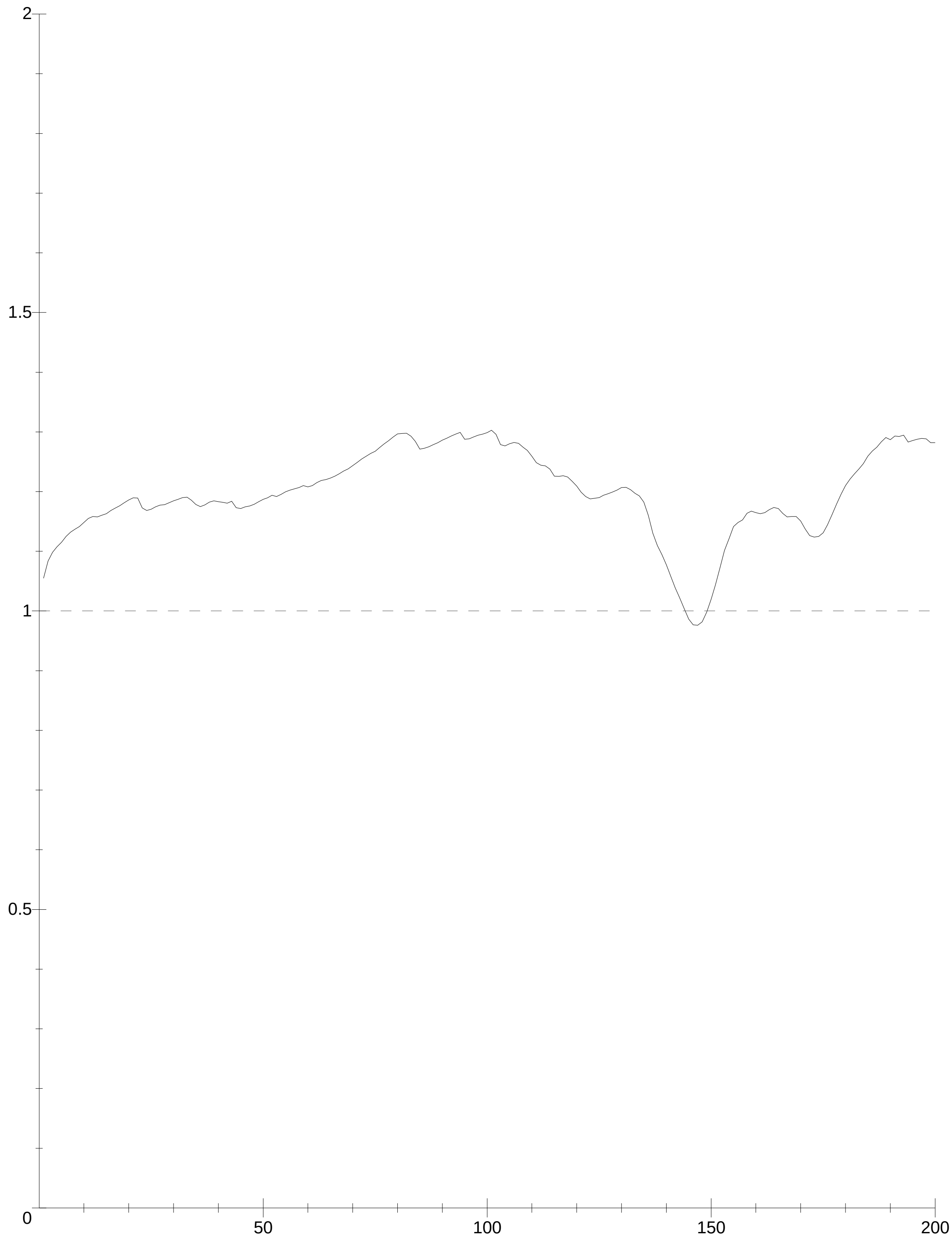
2
1.5
1
0.5
0
50
100
150
200


2
1.5
1
0.5
0
50
100
150
200

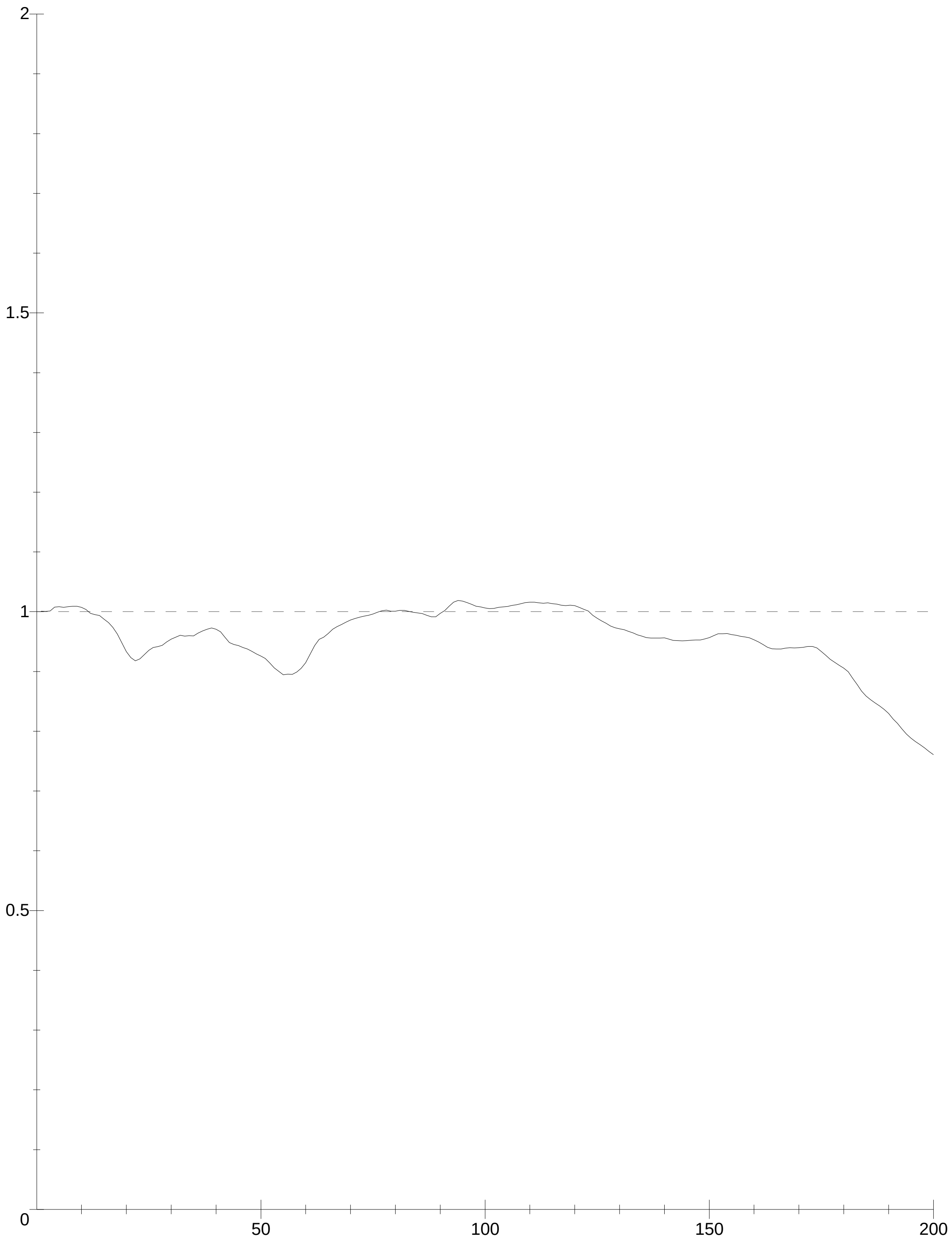

2
1.5
1
0.5
0
50
100
150
200



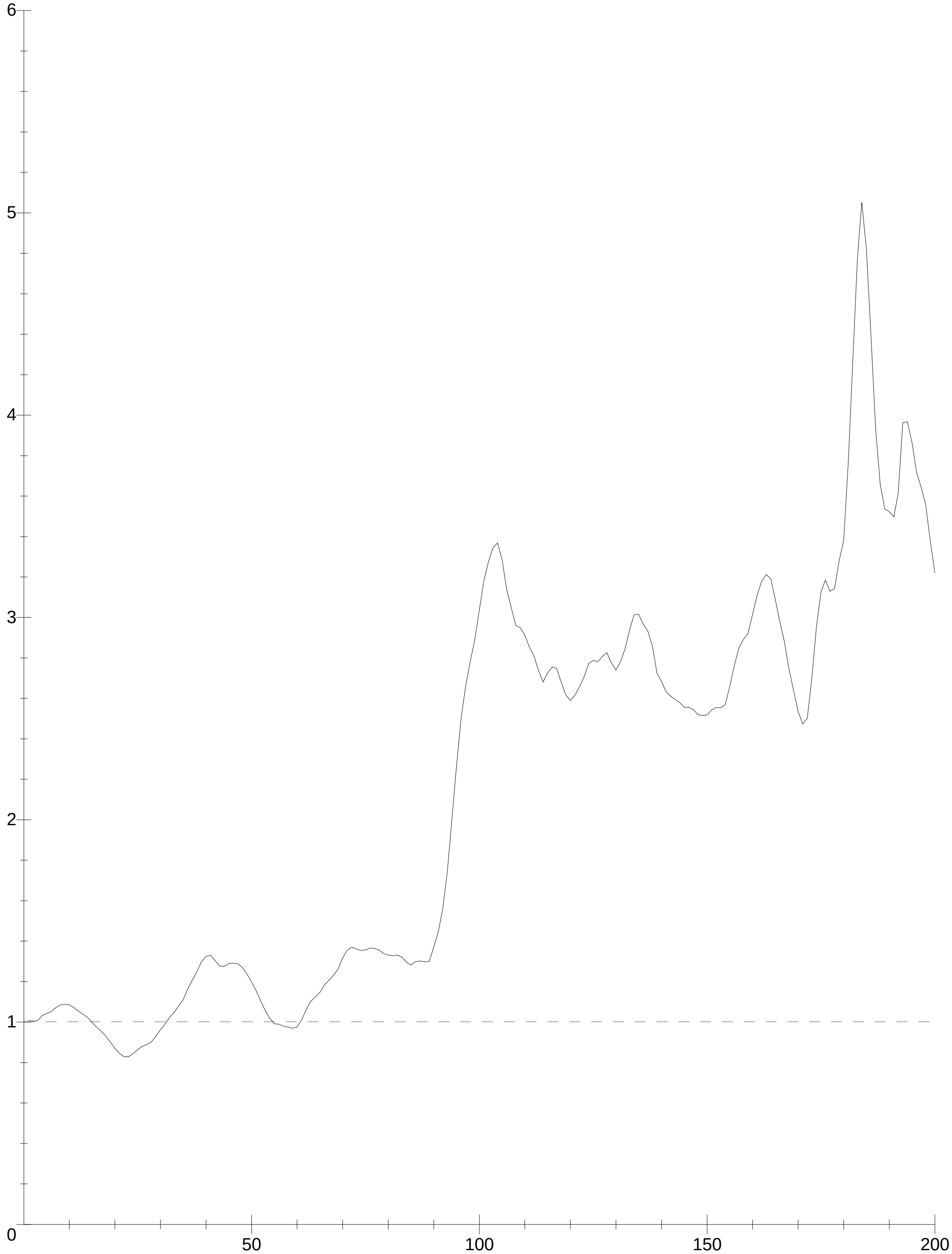
2
1.5
1
0.5
0
50
100
150
200

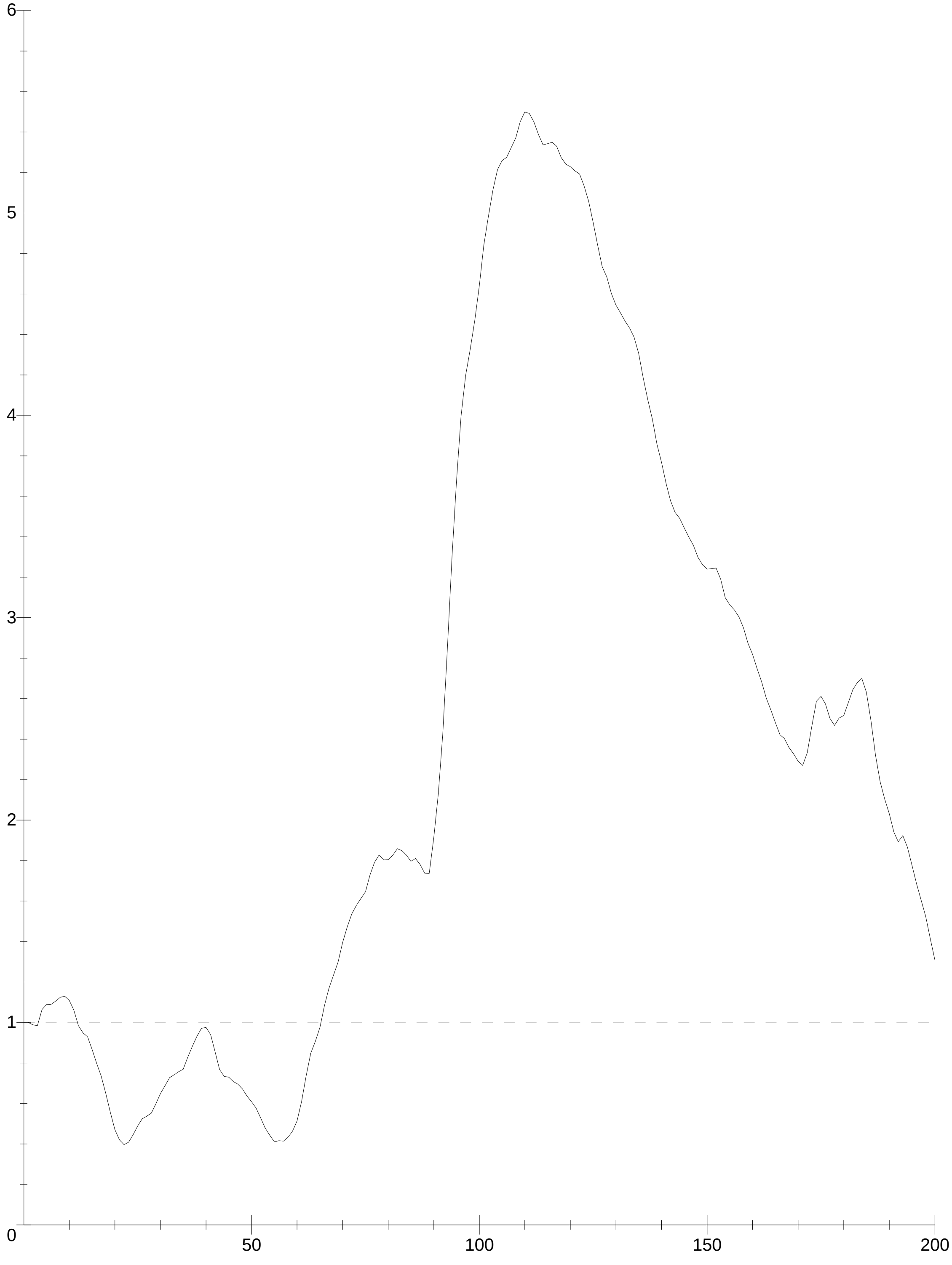


2
1.5
1
0.5
0
50
100
150
200

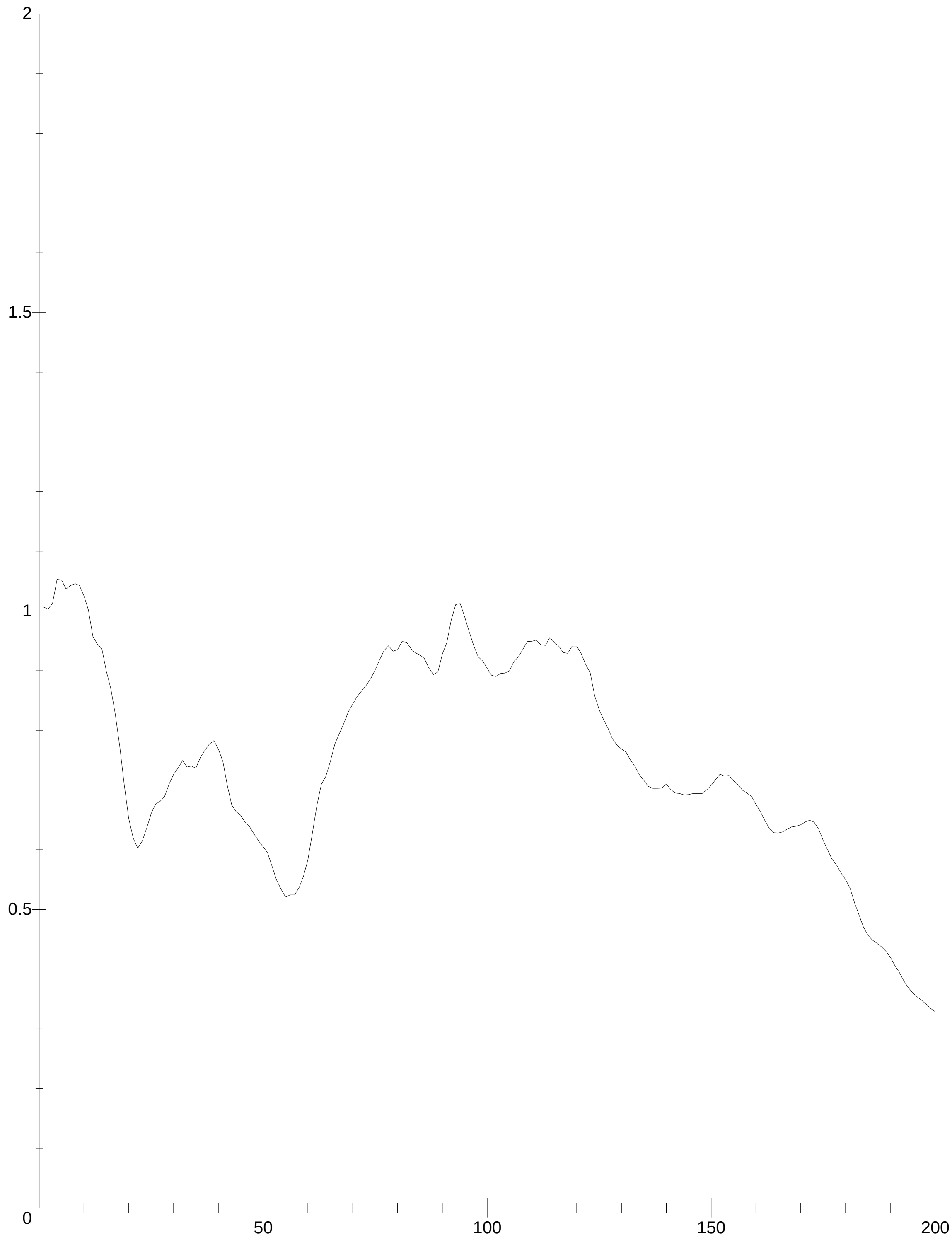


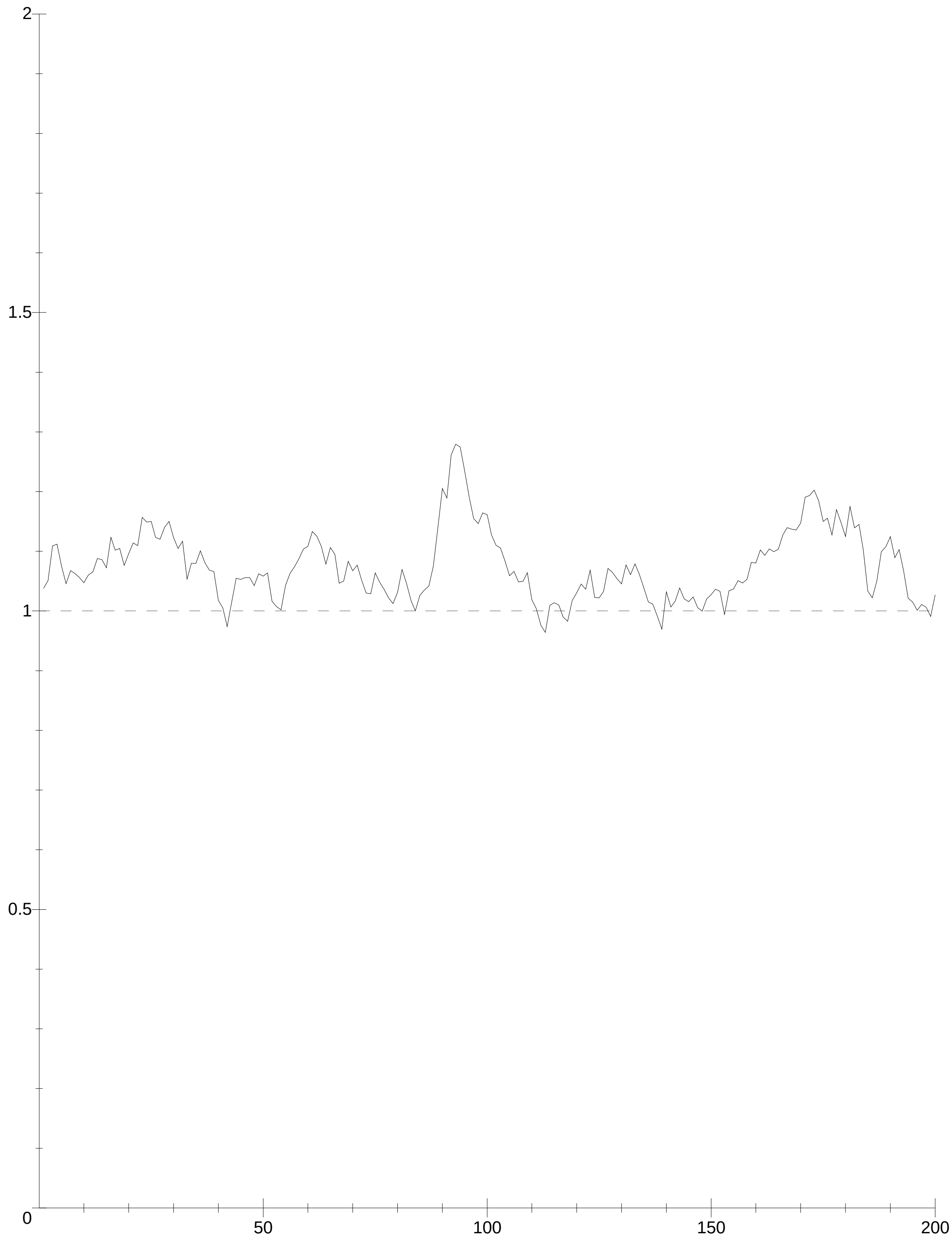
2
1.5
1
0.5
0
50
100
150
200


6
5
4
3
2
1
0
50
100
150
200



6
5
4
3
2
1
0
50
100
150
200


2
1.5
1
0.5
0
50
100
150
200



2
1.5
1
0.5
0
50
100
150
200