The Predictive Power of Zero Intelligence in Financial Markets

J. Doyne Farmer,[1] Paolo Patelli,[1,2] Ilija I. Zovko[1,3,*]

[1]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
[2]Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà 33, Pisa, Italy
[3]CENDEF, University of Amsterdam, Roetersstraat 11, Amsterdam, The Netherlands

[*]Correspondence should be addressed to zovko@santafe.edu

Using data from the London Stock Exchange, we test a model that treats the statistical mechanics of price formation and the accumulation of stored supply and demand under the simple assumption that people place orders to trade at random. The model makes excellent predictions for transaction costs, price diffusion rates, and a quantity closely related to supply and demand. Thus, it appears that the price formation mechanism strongly constrains the market, playing a more important role than the strategic behavior of agents. The remarkable success of this approach suggests a new and unorthodox approach to economics.

Since the nineteenth century one of the classic questions in economics has been, "What determines supply and demand?". Similarly, since Bachelier (*1*) introduced the random walk model for prices in 1900, another important question has been "What determines the price diffusion rate?". Standard models in economics, which are based on rational utility maximizing agents, have had only limited success in addressing these questions. In this paper we demonstrate that a model built on the opposite approach – that agents are of zero intelligence, and simply place orders to trade at random – can successfully address these questions and others, providing one properly models the statistical mechanics of price formation.

Traditionally economics has devoted considerable effort to modeling the strategic behavior and expectations of agents. While no one would dispute that this is important, it has also been pointed out that some aspects of economics are independent of the agent model. For example, Becker (*2*) showed that a budget constraint is sufficient to guarantee the proper slope of supply and demand curves, and Gode and Sunder (*3*) demonstrated that if one replaces the students in a standard classroom economics experiment by zero-intelligence agents, price setting and other properties match better than one might expect. In this paper we show that this principle can be dramatically more powerful, and can make surprisingly accurate quantitative predictions. In particular, we test a zero-intelligence statistical mechanics model due to Daniels et al. (*4, 5*), which builds on earlier work in financial economics (*6–9*) and physics (*10–16*). This added to the prior literature by constructing and approximately solving a simple model for price setting that makes quantitative, testable predictions about fundamental market properties, many of which can be expressed as simple algebraic formulae.

The model of Daniels et al. (*4*) assumes a continuous double auction, which is the most widely used method of price formation in modern financial markets (*5*). There are two fundamental kinds of trading orders: Impatient traders submit *market orders*, which are requests to buy or sell a desired number of shares immediately at the best available price. More patient traders submit *limit orders*, which include the worst allowable price for the transaction. Limit orders may fail to result in an immediate transaction, in which case they are stored in a queue called the *limit order book*, illustrated in Fig. 1. As each buy order arrives it is transacted against accumulated sell limit orders that have a lower selling price, in priority of price and arrival time. Similarly for sell orders. The lowest selling price offered in the book at any point in time is called the *best ask*, $a(t)$, and the highest buying price the *best bid*, $b(t)$.

The model assumes that two types of zero intelligence agents place and cancel orders randomly (see Fig. 1). Impatient agents place market orders of size σ, which arrive at a rate μ *shares per time*. Patient agents place limit orders of the same size σ, which arrive with a constant rate density α *shares per price per time*, and queued limit orders are cancelled at a rate δ, with dimensions of *1/time*. Prices change in discrete increments called *ticks*, of size dp. To keep the model as simple as possible, there are equal rates for buying and selling, and order placement and cancellation are Poisson processes. All of these processes are independent except for coupling through their boundary conditions: Buy limit orders arrive with a constant density over the semi-infinite interval $-\infty < p < a(t)$, where p is the logarithm of the price, and sell limit orders arrive with constant density on the semi-infinite interval $b(t) < p < \infty$. As new orders arrive they may alter the best prices a and b, which in turn changes the boundary conditions for subsequent limit order placement. As a result $a(t)$ and $b(t)$ each make random walks, but because of coupling of the buying and selling processes the bid-ask *spread* $s(t) \equiv a(t) - b(t)$ is a stationary random variable. It is this feedback between order placement and price diffusion that makes this model interesting, and despite its apparent simplicity, quite difficult to understand analytically (*5*).

We test the model using data from the London Stock Exchange (LSE) during the period August 1st 1998 - April 30th 2000, which includes a total of 434 trading days and roughly six million events. This data set shows all orders and cancellations, making it possible to measure the parameters of the model. For a more detailed description of the LSE and the dataset see (*19*). We chose 11 stocks with at least 300,000 events in the sample and at least 80 events on any given day. We measure the average value of the five above-defined parameters μ, α, δ, σ, and dp for each day, making the assumption that the parameters of the model are stationary within each day, but change from day to day.

The bid-ask spread is of central interest in financial markets because it is an important component of transaction costs. The mean value of the spread predicted based on a mean field theory analysis of the model is $\hat{s} = (\mu/\alpha) f(\sigma\delta/\mu, dp/p_c)$, where f is a relatively slowly varying non-dimensional function (*5*). To test this relationship, we measure the actual average spread $\bar{s}$ across the full time period for each stock, and compare to the predicted average spread $\hat{s}$ based on order flows. To test our hypothesis that the two values coincide, we perform a regression of the form $\log \bar{s} = A \log \hat{s} + B$. The regression, shown in Fig. 2(a), has $R^2 = 0.96$, with $A = 0.99 \pm 0.10$ and $B = 0.06 \pm 0.29$, in comparison to the model predictions $A = 1$ and $B = 0$. We thus very strongly reject the null hypothesis that $A = 0$, indicating that the predictions are far better than random. Even more surprising, we are unable to reject the null hypotheses that $A = 1$ and $B = 0$, indicating that we match the data extremely well even without fitting any free parameters. See (*19*) for details of the error analysis.

Another quantity of primary interest is the price diffusion rate, which drives the volatility of prices and is the primary determinant of financial risk. If we assume that prices make a random walk, then the diffusion rate measures the size and frequency of its increments. The variance V of an uncorrellated normal random walk after time t grows as $V(t) = Dt$, where D is the diffusion rate. Numerical experiments indicate that the short term price diffusion rate predicted by the model is $\hat{D} = k\mu^{5/2}\delta^{1/2}\sigma^{-1/2}\alpha^{-2}$ where k is a constant. As for the spread, we compare this to the actual price diffusion rate $\bar{D}_i$ for each stock averaged over the 21 month period, and regress the logarithm of the predicted vs. actual values, as shown in Fig. 2(b). This gives $R^2 = 0.76$, with $A = 1.33 \pm 0.25$ and $B = 2.43 \pm 1.75$. Thus, we again strongly reject the null hypothesis that $A = 0$. The results are not quite as good as they are for the spread, but we are still unable to reject the null hypothesis that $A = 1$ and $B = 0$; in each case the measured value is a little more than one standard deviation too high. We have accomplished something that is rather hard to achieve in economics, i.e. we have made testable predictions that are validated without any adjustment of free parameters. However, an important caveat is that problems in measuring the parameters α and δ introduce some arbitrariness into the intercept parameter B (*19*).

Finally, the model makes a prediction about market impact, which is the dominant source of transaction costs for large traders, and is related to supply and demand. When a market order of size ω arrives it causes transactions which can cause a change in the midpoint price $m(t) \equiv (a(t) + b(t))/2$. The average market impact function ϕ is the average logarithmic midpoint price shift Δp conditioned on order size, $\phi(\omega) = E[\Delta p|\omega]$.

The nondimensional coordinates dictated by the model are very useful for understanding the average market impact function. There are five parameters of the model and three fundamental dimensional quantities (*shares*, *price*, and *time*), leading to only two independent degrees of freedom. Since the order flow rates μ, α, and δ are more important than the discreteness parameters σ and dp, it is natural to construct non-dimensional units based on the order flow parameters alone. There are unique combinations of the three order flow rates with units of *shares*, *price*, and *time*. These define a characteristic number of shares $N_c = \mu/\delta$, a characteristic price interval $p_c = \mu/\alpha$, and a characteristic timescale $t_c = 1/\delta$. These characteristic values can be used to define nondimensional coordinates $\hat{p} = p/p_c$ for price, $\hat{N} = N/N_c$ for shares, and $\hat{t} = t/t_c$ for time.

Each market order ω_i causes a possible price change Δp_i, defining an impact event $(\omega_i, \Delta p_i)$. If we bin together events with similar ω and plot the mean order size as a function of the mean price impact Δp, we typically see highly variable behavior for different stocks, as shown in Fig. 3(b). However, if we plot the data in nondimensional units, we see a collapse of the data onto roughly a single curve, as shown in Fig. 3(a). The variations from stock to stock are quite small; on average the corresponding bins for each stock deviate from each other by about 8%, roughly the size of the statistical sampling error. We have made an extensive analysis, but due to problems caused by the long-memory property of these time series, it remains unclear whether these differences are statistically significant (*19*). In contrast, using standard coordinates the differences are highly statistically significant. This collapse illustrates that the non-dimensional coordinates dictated by the model provide substantial explanatory power, and that we understand how the market impact varies from stock to stock by a simple transformation of coordinates. This sheds light on empirical results for the average market impact for the New York Stock Exchange (*17*).

If we fit a function of the form $\phi(\omega) = K\omega^{\beta}$ to the market impact curve, we get $\beta = 0.26 \pm 0.02$ for buy orders and $\beta = 0.23 \pm 0.02$ for sell orders, as shown in Fig. 4. The functional form of the market impact we observe here is not in agreement with a recent theory by Gabaix et al. (*18*), which predicts $\beta = 0.5$. There is an interesting underlying debate: Their theory follows traditional thinking in economics, and postulates that agents optimize their behavior to maximize profits, while the theory we test here assumes that they behave randomly, and that the form of the average market impact function is dictated by the statistical mechanics of price formation.

The market impact function is closely related to the more familiar notions of supply and demand. At any instant in time the stored queue of sell limit orders reveals the quantity available for sale at each price, thus showing the supply, and the stored buy orders similarly show the revealed demand. The price shift caused by a market order of a given size depends on the stored supply or demand through a moment expansion (*5*). Thus, the collapse of the market impact function reflects a corresponding property of supply and demand. Normally one would assume that supply and demand are functions of human production and desire; in this case, their form is dictated by the dynamical interaction of order accumulation, removal by market orders and cancellation, and price diffusion.

These results have several practical implications. For market practitioners, understanding the spread and the market impact function is very useful for estimating transaction costs and for developing algorithms that minimize their effect. And for regulators they suggest that it may be possible to make prices less volatile and lower transaction costs by creating incentives for limit orders and disincentives for market orders.

The model we test here was constructed before looking at the data (*4,5*), and was designed to be as simple as possible for analytic analysis. A more realistic (but necessarily more complicated) model would more closely mimic the properties of real order flows, which are strongly correlated, and would hopefully be able to capture even more features of the data, such as the power law tails of prices. Nonetheless, as we have shown above, this simple model does a remarkable job of explaining important fundamental properties of markets, such as transaction costs, price diffusion and supply and demand. The model captures the statistical mechanics of the market quite well, and in particular, the way order placement and price formation interact to alter the accumulation of stored supply and demand. For the phenomena studied here this appears to be the dominant effect. We do not mean to claim that market participants are unintelligent: Indeed, one of the virtues of this model is that it provides a benchmark to separate properties that are driven by the statistical mechanics of the market institution from those that are driven by conditional strategic behaviour. It is surprising that such a simple model can explain so much about a system as complex as a market, and shed light on century-old questions about the rate of price diffusion and the form of supply and demand

References and acknowledgments

1. L. Bachelier, “Théorie de la spéculation”, 1900. Reprinted in P.H. Cootner, *The Random Character of Stock Prices*, 1964, MIT Press Cambridge.

2. Becker, G., Irrational behavior and economic theory, *J. of Political Economy* **70**, 1 (1962)

1-13.

3. D. Gode and S. Sunder, Allocative efficiency of markets with zero intelligence traders: Markets as a partial substitute for individual rationality. *J. of Political Economy*, **101**, 119 (1993).

4. Daniels, M, Farmer, J.D., Gillemot, L., Iori, G., and Smith, D.E. Quantitative model of price diffusion and market friction based on trading as a mechanistic random process. *Physical Review Letters* **9019**, 10 2003.

5. Smith, E., Farmer, J.D., Gillemot, L., and Krishnamurthy, S., Statistical theory of the continuous double auction. To appear in *Quantitative Finance*, 2003.

6. H. Mendelson, Market behavior in a clearing house, *Econometrica* **50**, 1505-1524 (1982).

7. K.J. Cohen, R.M. Conroy and S.F. Maier, Order flow and the quality of the market, in: Y. Amihud, T. Ho and R. Schwartz, eds., *Market Making and the Changing Structure of the Securities Industry*, 1985, Lexington Books, Lexington MA.

8. I. Domowitz and Jianxin Wang, Auctions as algorithms, *J. of Econ. Dynamics and Control* **18**, 29 (1994).

9. T. Bollerslev, I. Domowitz, and J. Wang, Order flow and the bid-ask spread: An empirical probability model of screen-based trading, *J. of Econ. Dynamics and Control* **21**, 1471 (1997).

10. Potters, M. & Bouchaud, J.-P., "Statistical properties of stock order books: empirical results and models", *Physica A* **324**, 133-140 (2003)

11. J.-P. Bouchaud, M. Mezard, M. Potters, Statistical properties of the stock order books: empirical results and models, *Quantitative Finance* **2** (2002) 251-256.

12. P. Bak, M. Paczuski, and M. Shubik, Price variations in a stock market with many agents, *Physica A* **246**, 430 (1997).

13. D. Eliezer and I.I. Kogan, Scaling laws for the market microstructure of the interdealer broker markets, http://xxx.lanl.gov/cond-mat/9808240.

14. S. Maslov, Simple model of a limit order-driven market, *Physica A* **278**, 571(2000).

15. F. Slanina, Mean-field approximation for a limit order driven market model, *Phys. Rev. E*, **64**, 056136 (2001).

16. D. Challet and R. Stinchcombe, Analyzing and Modeling 1+1d markets, *Physica A* **300**, 285, (2001).

17. Lillo,F. Farmer. J.D. & Mantegna, R.N., Master Curve for Price-Impact Function, *Nature* **421**, 129-130 (2003).

18. Gabaix, X., Gopikrishnan, P., Plerou, V. and Stanley, H.E. A theory of power-law distributions in financial market fluctuations, *Nature* **423**, 267-270 (2003).

19. **Supporting Online Material** www.sciencemag.org Statistical methods and details comments.

20. We would like to thank Credit Suisse First Boston, the McDonnell Foundation, the McKinsey Corporation, Bob Maxfield, and Bill Miller for their support of this project. We would also like to thank Sam Bowles, Supriya Krishnamurthy, Fabrizio Lillo, and Eric Smith for useful discussions, and Mark Bieda, David Krakauer, Harold Morowitz and Elizabeth Wood for comments on the manuscript.

Figure 1: A random process model of the double continuous auction. Stored limit orders are shown stacked along the price axis, with sell orders (supply) stacked above the axis at higher prices and buy orders (demand) stacked below the axis at lower prices. New sell limit orders are visualized as randomly falling down, and new buy orders as randomly "falling up". New sell orders can be placed anywhere above the best buying price, and new buy orders anywhere below the best selling price. Limit orders can be removed spontaneously (e.g. because the agent changes her mind or the order expires) or they can be removed by market orders of the opposite type. This can result in changes in the best prices, which in turn alters the boundaries of the order placement process. It is this feedback between order placement and price formation that makes this model interesting, and its predictions non-trivial.

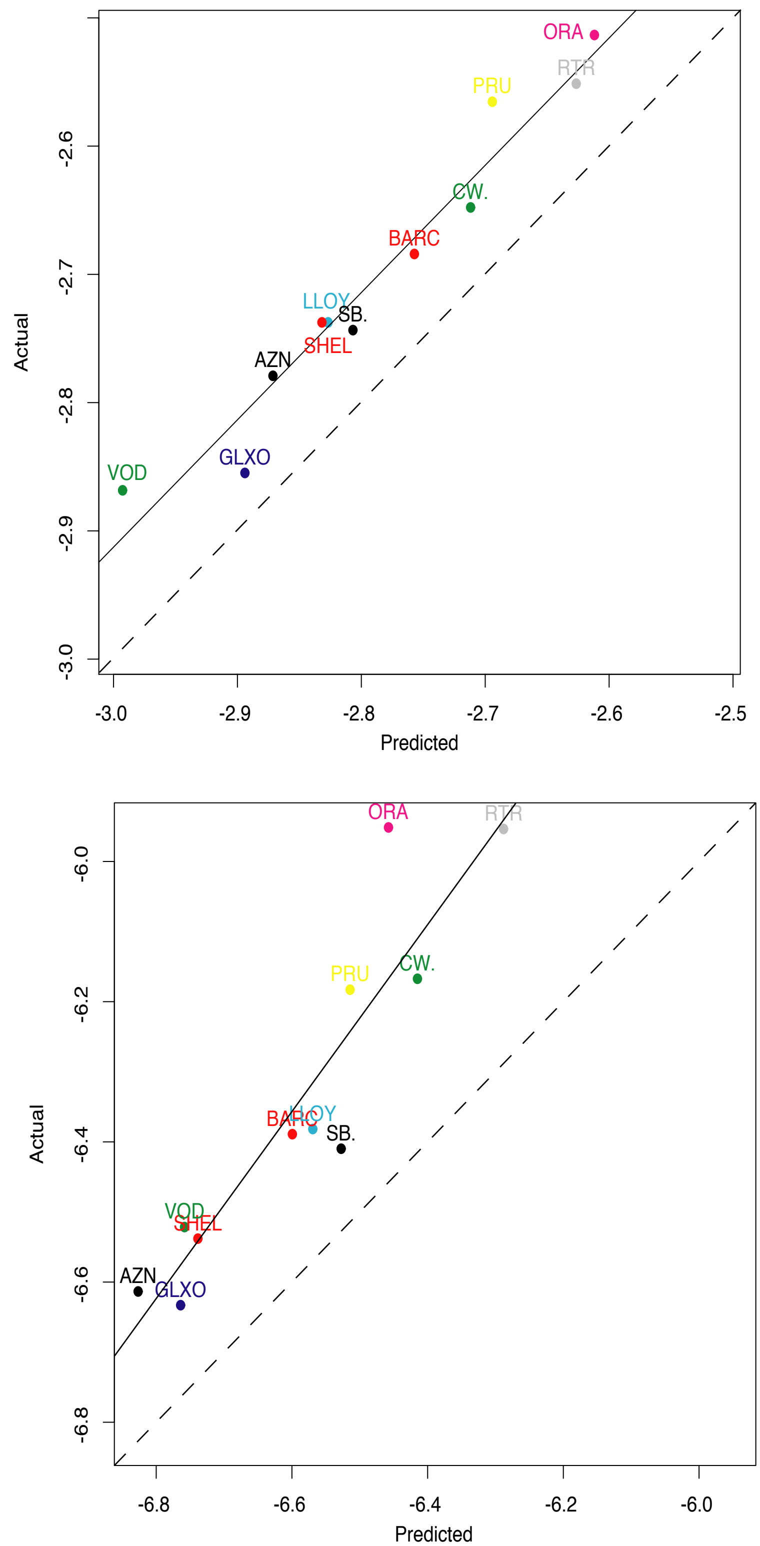
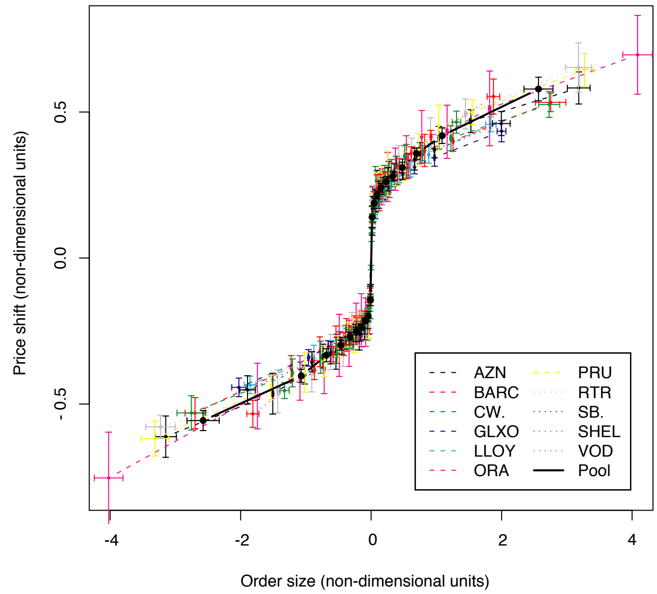


Figure 2: Regressions of predicted values based on order flow parameters vs. actual values for the spread (a) and price diffusion rate (b). The dots show the average predicted and actual value for each stock averaged over the full 21 month time period. The solid line is a regression; the dashed line is the diagonal, representing the model's prediction without any adjustment of slope or intercept.



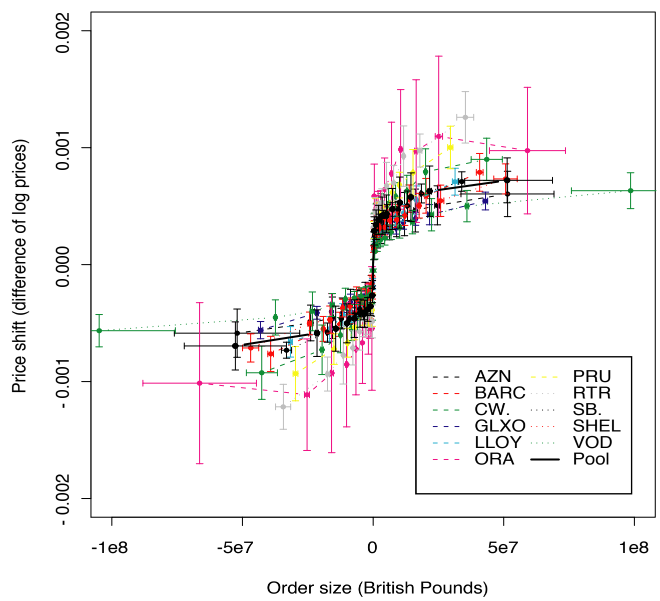


Figure 3: The average market impact as a function of the mean order size. In (a) the price differences and order sizes for each transaction are normalized by the nondimensional coordinates dictated by the model, computed on a daily basis. Most of the stock collapse extremely well onto a single curve; there are a few that deviate, but the deviations are sufficiently small that given the long-memory nature of the data, it is difficult to determine whether these deviations are statistically significant (*19*). This means that we understand the behavior of the market impact, which is closely related to supply and demand, by a simple transformation of coordinates. In (b), for comparison we plot the order size in units of British pounds against the average logarithmic price shift. Other alternative normalizations show similarly large deviations (*19*).

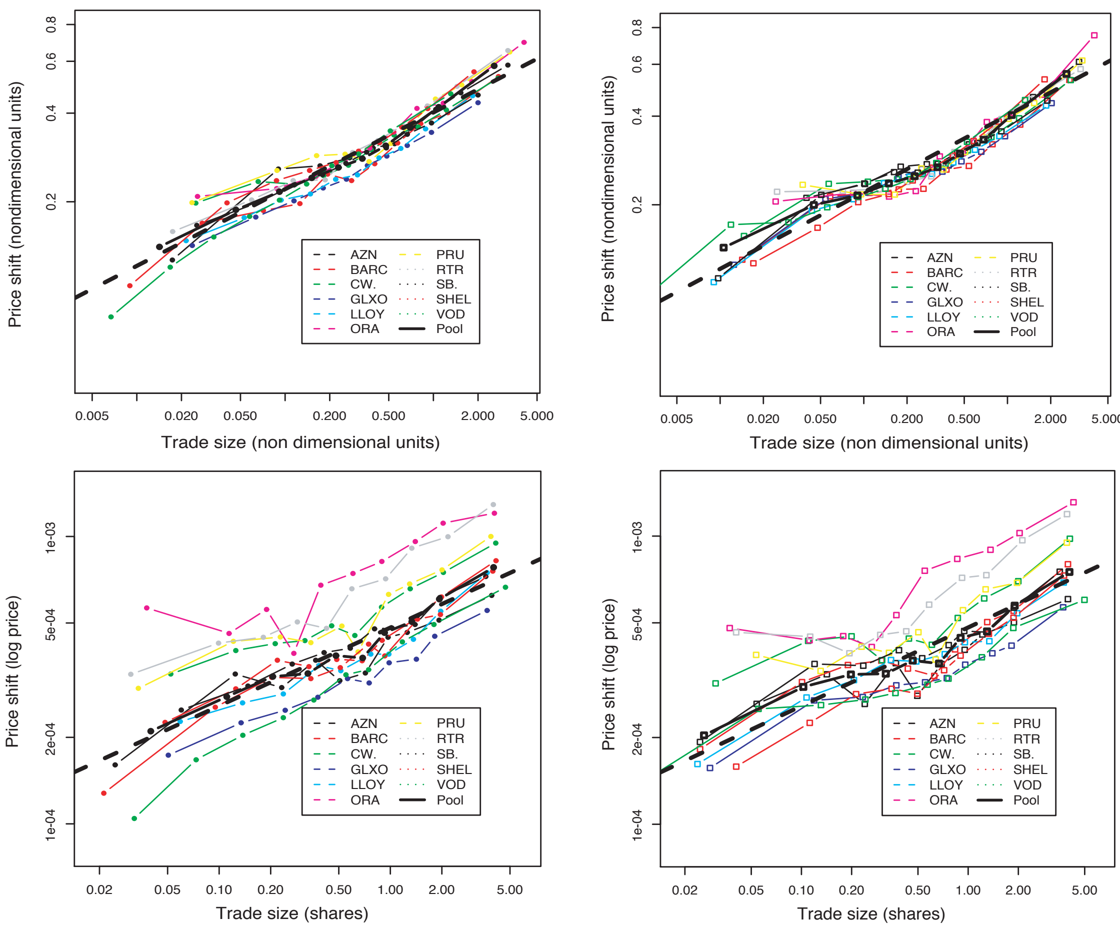


Figure 4: The average market impact vs. order size plotted on log-log scale. The upper left and right panels show buy and sell orders in nondimensional coordinates; the fitted line has slope $\beta = 0.26 \pm 0.02$ for buy orders and $\beta = 0.23 \pm 0.02$ for sell orders (*19*). In contrast, the lower panels shows the same thing in dimensional units, using British pounds to measure order size.