Response to “The cost of latency,” by Ciamac Moallemi and Mehmet Sağlam 
Robert Almgren, NYU and Quantitative Brokers

The effect observed in this paper is due to time discretization, not to latency. The authors have ignored the price change within the interval $T_i$ to $T_{i+1}$, which is of the same asymptotic size $O(\sqrt{\Delta t})$ as the price change from $T_i$ to $T_{i+1}$. Correctly calculated, the cost for the discretization-only model in Section 4.3 is of the same order as the overall cost for the latency model.

Consider the model of Section 4.3, in which time is discretized into intervals of length $\Delta t$, and the limit price $\ell_i$ may be determined instantaneously at $T_i$ but is required to be held constant until $T_{i+1}$. The authors say (second bullet point) that the limit order will execute at $\ell_i$ if an impatient buyer appears between $T_i$ and $T_{i+1}$ and if “$\ell_i \leq S_{T_i} + \delta$, i.e., the limit price $\ell_i$ is within a margin $\delta$ of the bid price at the start of the interval.” In this model the optimal limit price level is $\ell_i^* = S_{T_i} + \delta$, and the value obtained is close to the value of the continuous process. (With their approximate Poisson arrival probability $\mu \Delta t$ in place of the true value $1 - e^{-\mu \Delta t}$, the value of the discrete problem is actually slightly higher than the value of the continuous problem, a clue that something is wrong; with the exact probability the values are exactly equal.) But using the bid price at the start of the interval is an incorrect discretization of the continuous problem.

A correct discretization would use the bid price $S_{\tau}$ at the time $\tau$ that the impatient buyer appears, since the market price does not stop moving just because of this trader’s latency. Then, if the limit price $\ell_i$ were to be set at $\ell_i$ if an impatient buyer appears between $T_i$ and $T_{i+1}$ and if “$\ell_i \leq S_{T_i} + \delta$, then when the buyer appears, with probability $1/2$ we would have $S_{\tau} < S_{T_i} = \ell_i - \delta$ and the limit order would not fill. In effect, with that strategy the arrival rate would be $\mu/2$ rather than $\mu$, and the value under this strategy would be worse by $O(1)$ than the continuous result.

To reduce this cost, exactly as clearly described in 4.1, the limit order price could be lowered by $C \sigma \sqrt{\Delta t}$, increasing the fill probability by a finite amount but reducing the payoff by a cost $O(\sqrt{\Delta t})$. As in 4.1, the optimal strategy will reduce the limit order price by an asymptotically slightly larger amount, achieving the same near-certain fill probability as in the continuous problem, with cost $O(\sqrt{\Delta t} \log \Delta t)$. Of course the problem including latency will have a higher cost than this, since as observed in Appendix B any strategy for the problem with latency is admissible for the discretization-only problem. But the costs are of the same asymptotic order in $\Delta t$, demonstrating that the essential phenomenon is time discretization rather than latency.

To summarize, the effect observed in this paper is primarily due to the fact that the limit order cannot be modified to adjust to price motion during a finite time interval. If the discrete-time model is constructed so that the price does not move during a time interval $\Delta t$ but only at the end of that interval, then of course it is found that latency is necessary to the result. But if price motion during the interval $\Delta t$ is included, then the effect will be seen even without latency.