The Robust Repeated Median Velocity System
Working Paper October 2005
Copyright © 2004 Dennis Meyers

In a previous article we examined a trading system that used the velocity of prices fit by a Least Squares straight line, through "N" past prices, to determined buy and sell points. The reasoning behind this type of system was to only trade when the price trend velocity was above a certain threshold. Many times prices are meandering around without a notable trend. At these times we do not wish to trade because of the whipsaws losses that occur from this type of price action. When a price trend finally starts, the velocity of that price trend moves above some minimum value. Thus the velocity system would only issue a trade when certain velocity barriers were crossed.

The Least Squares line is determined by minimizing the sum of the squares of the difference between the N prices and the value of the straight line.

$r^2(i) = [Price(i)-(a+b*i)]^2$ = error squared

$$\text{Minimize(a,b)} \sum_{i=1}^{i=N} r^2(i)$$

This mathematical technique has an exact solution and dates back to Gauss in the 1800's. At the time of it's development there were no computers so this method became a cornerstone of statistics.

Recently (1980's) interest has developed in what is called robust regression and outlier detection techniques, Ref [1]. Regression techniques are now defined by a measure called the "breakdown point". The breakdown point is loosely defined as the smallest amount of bad data points that can cause the regression solutions to take on values arbitrarily far from their true values. Unfortunately the Least Squares technique has a breakdown point of 1/N. In other words only one bad data point can significantly change the computation of the velocity or slope of a straight line. The median of a set of numbers has a breakdown point of 50% . This is because when 50% of the numbers are bad then there is no way of telling which are the bad numbers and which are the good numbers. 50% is the highest breakdown point.

The least absolute deviation (LAD) regression estimator which is

$$\text{Minimize(a,b)} \sum_{i=1}^{i=N} \text{absolute value[} r(i)]$$

has a breakdown point of 29.8% (see Ref 1). For the LAD this means around ¼ of the price points can be bad before the computations of a and b become erroneous. Recently Siegel Ref[2], in his paper "Robust regression using repeated medians", introduced a technique for finding the slope that has a 50% breakpoint. The repeated median is also described in Ref [1].

While the repeated median technique may sound complicated it is quite easy to compute. Here's how. For demonstration purposes let's suppose we have 10 data points on an x,y graph such that,

X	1	2	3	4	5	6	7	8	9	10
Y	1	2	3	4	5	15	12	8	9	10

We've added two bad points at X positions 6 and 7. To calculate the repeated median slope we would take the slope of every pair of y values and then find the median of all the pairs of slopes. For this example we would take

Slope 1 = (y(1)-y(2))/(1-2) = 1
Slope 2 = (y(1)-y(3))/(1-3)= 1
Slope 3 = (y(1)-y(4))/(1-4) = 1
Slope 4 = (y(1)-y(5))/(1-5) = 1
Slope 5 = (y(1)-y(6))/(1-6) = 14/5 = 2.8
Slope 6 = (y(1)-y(7))/(1-7) = 11/6 = 1.83
Slope 7 = (y(1)-y(8))/(1-8) = 1
Slope 8 = (y(1)-y(9))/(1-9) = 1
Slope 9 = (y(1)-y(10))/(1-10) = 1

The median slope of the above is 1. The above process is repeated for (y(2)-y(i))/(2-i) i=1 to 10 i≠2, (y(3)-y(i))/(3-i), i=1 to 10 i≠3, … all the way to (y(10)-y(i))/(10-i), i=1 to 9 .

The final slope is then the median of all the medians calculated above. While the repeated median looks redundant because the very first calculation produced the correct slope, price data is not so nicely distributed as our example and the extra calculations are needed to assure that the outliers are eliminated.

The mathematical formula for the above is

$$\text{Slope}(t) = \underset{i=1 \text{ to } N}{\text{median}_i} \left\{ \underset{j=1 \text{ to } N}{\text{median}_{i \neq j}} \left[(\text{price}(j)-\text{price}(i))/(j-i) \right] \right\}$$

Figure 1 shows a plot of the x,y numbers above with the repeated median line and the least squares line on the graph. Notice how the bad points draw the least squares line towards them while the repeated median line is completely unaffected by the outliers. The least Squares line is given by the formula y=0.6 + 1.1455*x . The true line is given by the formula y=x . From this simple example we can observe how noise has distorted the least squares estimates of a and b.
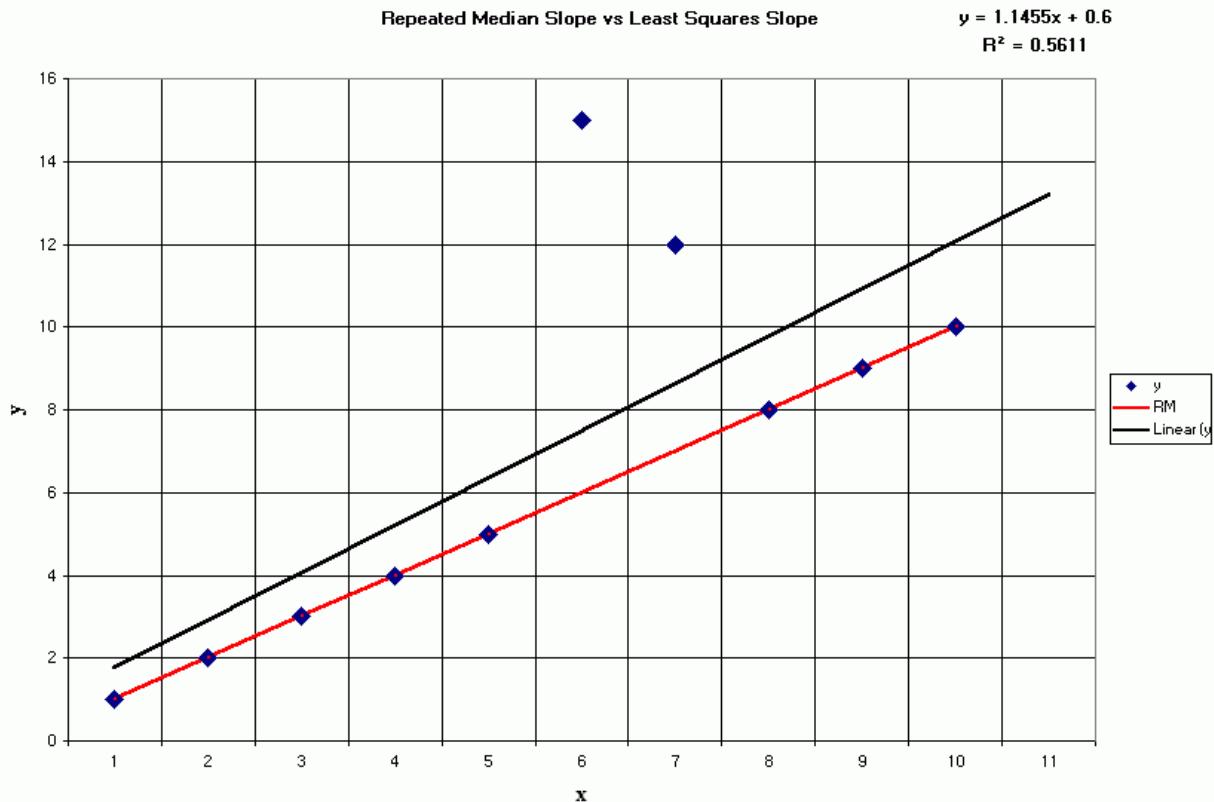
Figure 1 Repeated Median Slope vs Least Squares Slope.

Here we will use the repeated median slope to create a system. For a straight line the velocity is equal to the slope. The repeated median velocity has the advantage that it is a natural random price noise inhibitor. We can create a system such that unless the repeated median velocity using N past price bars is greater than some threshold we will not buy or sell. A large percentage of price noise generates a lot of back and forth movements of small magnitudes over a short period of time. With a lot of systems this back and forth movement creates many false buy and sell signals. However using the repeated median velocity over N past prices, we can filter out many of the small price noise movements by requiring that the velocity to be greater than some threshold before we act.

The Repeated Median Velocity System Defined
At each bar we calculate the repeated median velocity(RMV) from the formula above. When the velocity is greater than the threshold amount *vup* we will go long. When the velocity is less than the threshold amount *-vdn* we will go short.

Buy Rule:
IF RMV is greater than the threshold amount *vup* then buy at the market.

Sell Rule:
IF RMV is less than the threshold amount *-vdn* then sell at the market.

Intraday Bars Exit Rule:
Close the position 5 minutes before the E-Mini close (no trades will be carried overnight).

Intraday Bars First Trade of Day Entry Rule:
Ignore all trade signals before 10:00 EST (30 minutes after the open). Buy and Sell rules above we have included a first trade of the day entry rule. We've included this rule because often there are gaps in the open creating immediate system buys and sells. Many times these gaps are closed creating a losing whipsaw trade. In order to avoid the opening gap whipsaw trade problem we've delayed the first trade of the day for 30 minutes until after10:00 EST

Testing The Repeated Median Velocity (RMV) System Using Walk Forward Optimization
There are three system parameters to determine:
1. *N*, is the lookback period to calculate the **RMV.**
2. *vup*, the threshold amount that RMV has to be greater than to issue a buy signal
3. *vdn*, the threshold amount that RMV has to be less than to issue a sell signal

To test this system we will use 5 minute bar prices of the E-Mini futures contract traded on the CME and known by the symbol ES from **02/4/2004 to 6/25/2004**.

In order to use the walk forward testing method, we will break the test data up into sixteen 30 calendar day test sections and 7 colander day out-of-sample sections as shown in Figure 2, Page 8. We will use the Power Walk forward Optimizer (PWFO) add-on to TradeStation to run the system optimization and generate those sixteen test and out-of-sample sections. We will then use the Walk Forward Performance Explorer (WFPE) software, described below, on the PWFO optimization output on each of those sixteen test sections to find the best performance variables to determine the input parameters *N, vup,* and *vdn* that will produce good results in the out of sample section immediately following each test section*.*

The WFPE eliminates all cases in the PWFO test files that do not meet certain Profit Factor (PF), and Losers in a row (LR) criteria . The WFPE explores a range of PF and LR criteria generating a number of searches in one run and automatically displays the best PF-LR filter in Excel..

For each test section we will run a TradeStation optimization on:
1. vup from 0.02 to 0.40 steps of 0.02
2. vdn from 0.02 to 0.40 in steps of 0.02
3. N from 10 to 75 in steps of 5.

Why use walk forward analysis? Why not just optimize over the total amount of data once and choose the inputs from the best performance parameters?

Out-of-sample testing is the norm in all mathematical economic disciplines other than technical analysis. The basis of walk forward out-of-sample testing is to take a test sample of data, perform a combinatorial parameter search (dubbed optimization by current trading platforms) on the test data, choose the input parameters from the combinatorial search over the input parameters based upon some performance statistics selection methodology (net profit, profit

factor, etc) in the test sample and then apply those chosen input parameters to data they haven't been tested on to see if they work.

Whenever you run an combinatorial search over many different combinations of input parameters on noisy data on a fixed time interval, the best performance statistics found are guaranteed be due to "curve fitting" the noise and signal. Unfortunately it is human nature to extrapolate past performance to project future trading results and so results from curve fitting give the illusion of future trading profits. A fixed number of prices on a fixed time interval has many spurious movements which are also called noise. When we run, say, 5000 different input parameter combinations , the best performance parameters will be from those system input variables that are able to produce profits from the price or signal dynamics *and* the spurious movements . While the signal dynamics, *if there*, will repeat, the same spurious price movements will not. If the spurious movements that were captured by a certain set of input parameters were a large part of the net profit, then choosing these input parameters would produce losses when traded on future data. These losses occur because the spurious movements will not be repeated in the same way. This is why optimization or combinatorial searches with no out-of-sample testing cause loses in the out-of-sample section from something that looked great in the test section.

Why perform the walk forward out-of-sample analysis so many times? Why not just do the analysis once? Well just as in poker, there is considerable vagaries in hand to hand luck. In our case there is considerable vagaries in week to week out-of-sample profit luck. That is, by pure chance we may have chosen some input parameters that did well in the test section data *and* the out-of-sample section data. In order to minimize this type of "luck" we repeat the walk forward out-of-sample (oos) analysis over many test/oos sections and take an average of our weekly results over all test/oos sections. This average gives us an expected weekly return and a standard deviation of weekly returns which allows us to estimate the expected equity and it's range for N weeks in the future. More on this later.

One thing we do know is that data mining or the curve fitting of the signal and noise will always produce the best performance results. Thus if we eliminate the optimization cases with the very best performance results we are sure to eliminate many of the data mining system input parameters that fitted the past spurious movements. We know from experience that very few good intraday systems can sustain profit factors above 2.5 over time. If we eliminate, from our optimization, all cases that have profit factors greater than 2.5 we will probably eliminate most of the cases that are due to curve fitting the noise.   In addition, since in real time it is tough to sustain more that five losses in a row and still keep trading, we will take the extra precaution and eliminate all those cases that have more than three losses in a row.

To find the system input parameters we will use the Walk Forward Performance Explorer on each of the sixteen test section optimization runs as described above.

Below are the best ten results of the WFPE run sorted by the Student t statistic that is calculated from the weekly out-of-sample net profits generated by the given PF-LR filter. The Student t statistic gives us the probability of the average out-of-sample profit for each PF-LR filter being due to pure chance. The higher the t statistic the lower the probability that the results are due to chance.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	es5rmed30x	1040312	1040625	16														
2	Filter	Var	tOnp	mOnp	aOnp	aanp	aOnT	B0	%P	t	std	llw	eqDD	lr	#	b00	Blw	BE
3	pf20-3:16	r22	7037	482	440	59.6	7.4	-9.9	0.69	3.68	478.6	-188	-239	3	16	439	3	4.7
4	pf20-4	wr	6214	176	388	41.4	9.4	1.5	0.75	2.77	561.4	-475	-475	1	16	352	1	8.4
5	pf20-a:16	r22	6551	562	409	26	15.8	-31.1	0.75	2.67	614.2	-850	-1038	2	16	441	4	9
6	pf20-a:16	mlos	6426	588	402	19.4	20.8	7.1	0.81	2.49	645.3	-1025	-1025	1	16	362	2	10.3
7	pf20-a	wr	4553	194	285	17.2	16.5	30.7	0.75	2.39	477.1	-475	-625	2	16	203	2	11.2
8	pf20-a:16	wr	4553	194	285	17.2	16.5	30.7	0.75	2.39	477.1	-475	-625	2	16	203	2	11.2
9	pf40-5:16	std	6049	363	378	17	22.2	-49.8	0.62	2.28	664.5	-688	-1214	2	16	477	6	12.4
10	pf30-5:16	std	6049	363	378	17	22.2	-49.8	0.62	2.28	664.5	-688	-1214	2	16	477	6	12.4

From this WFPE output we will select the filter on row 3.

Filter: PF>1 and PF<=2 and LR<=3, #Trds>=16 and highest r22.

Where:
PF = Profit Factor in Test optimization section
LR=Maximum consecutive loses in a row in test optimization section
#Trds = The number of trades in the test run. We want our system to trade almost every day. There are 20 to 22 trading days a month, so we only want to look at input parameters that are able to generate at least 16 trades in the one month test sample.
R22 = is the correlation coefficient between the trade Equity line and the 2nd Order Polynomial Line that is fitted to the trade equity line.

Results
Figure 2 presents a table of the sixteen test and out-of-sample windows and the selected optimum parameters and out-of-sample results using the WFPE filter described above.

Figure 3 presents a specialized percentage trade by trade summary of the two one week out-of-sample data segments ending on 6/7/04 through 6/25/04. where the out-of-sample input values for those out-of-sample weeks were put at the last week values of N=20, vup=0.06 and vdn=0.18.

Figure 4 presents the out-of-sample 5 minute bar chart of ES from 6/21/04 to 6/25/04. with the Repeated Median Velocity Indicator and all the buy and sell signals for those dates from the trade by trade summary of Figure 3 indicated on the charts.

Discussion of System Performance
At the bottom of Figure 2 are some statistics that are of interest. The first statistic ave=439.8 is the average weekly net profit for the sixteen oos weeks. Std=478.6 is the standard deviation of the weekly return. Given a process that generates the above average and standard deviation, it would be interesting to know, statistically, the probabilistic outcome from trading this system for 13 weeks (one quarter).   There is a statistical technique that we can use to determine this called bootstrap Ref [3]. Using the bootstrap technique we would randomly choose 13 of the sixteen weekly profits in figure 2, with replacement and take the sum of those 13 randomly chosen oos net profits. "With replacement" means we don't eliminate the randomly chosen weekly profit

from being chosen again. Let us suppose that we repeat this random choosing of 13 weeks and summing the results 200 times. We would then have 200 different 13 week net profit summations. If we take the average and standard deviation of those 200 different 13 week net profit summations we would have an estimate of what to expect from this system by trading it for 13 weeks.

For this system the 200 bootstrap average of 13 week net profits is $5574 and the standard deviation is $1715. This means that at two standard deviations we can expect our 13 week return from this system to be between $2144 and $9004, 95% of the time. Please note that slippage and commissions were not taken into account so these numbers would be higher that could be attained.

To see the effect of walk forward analysis take a look at Figure 2. Notice how the input parameter N takes sudden jumps from high to low and back . This is the walk forward process adapting to changing volatility conditions.

In observing the chart we can see that the system did very well in catching most major intraday trend of the E-Mini. On those days when the morning trend changed in midday (6/22, 6/23, 6/25, and 6/25) the RMV system changed direction quickly allowing for the capture of the profits from the trend change to the close of trading.   Overall the RMV system did a good job in minimizing the losses due to the inevitable whipsaws that will occur in any trading system and maximizing the profits from the major intraday trend moves of the E-Mini.

References
1. Rousseeuw, P.J., Leroy, A.M., (1987) "Robust Regression and Outlier Detection", New York, John Wiley & Sons.
2. Siegel, A.F. (1982), "Robust Regression using Repeated Medians." *Biometrika*. 69, pp242-244.
3. Efron, B., Tibshirani, R.J., (1993), "An Introduction to the Bootstrap", New York, Chapman & Hall/CRC.

Info on Dennis Meyers
Dennis Meyers (info@MeyersAnalytics.com.) has a doctorate in applied mathematics in engineering. He is a private trader, and president of Meyers Analytics (www.MeyersAnalytics.com). His firm specializes in Financial Engineering consulting for financial institutions and developing publicly available algorithmic trading software.

Figure 2   Walk Forward Out-Of-Sample Performance Summary
E-Mini 5 min bars Repeated Median Velocity System

ES-5 min bars 02/4/2004 - 6/25/2004 using the below filter on each test segment. The input values N, vup, and vdn are the values found from applying the filter to the test Sample optimization run.

Filter= PF>=1; PF<=2; LR<=3; #Trds>=16; Largest Equity 2nd Order Polynomial Correlation Coefficient.

Week #	Test Dates		oos Dates		osnp	Equity	ont	N	vup	vdn
1	02/04/04	03/05/04	03/08/04	03/12/04	**1600**	1600	8	55	0.12	0.06
2	02/11/04	03/12/04	03/15/04	03/19/04	**350**	1950	18	10	0.32	0.08
3	02/18/04	03/19/04	03/22/04	03/26/04	**713**	2663	17	10	0.32	0.18
4	02/25/04	03/26/04	03/29/04	04/02/04	**-138**	2525	14	10	0.32	0.14
5	03/03/04	04/02/04	04/05/04	04/09/04	**-63**	2462	3	40	0.24	0.12
6	03/10/04	04/09/04	04/12/04	04/16/04	**-38**	2424	5	35	0.26	0.16
7	03/17/04	04/16/04	04/19/04	04/23/04	**600**	3024	3	25	0.38	0.28
8	03/24/04	04/23/04	04/26/04	04/30/04	**213**	3237	8	20	0.26	0.36
9	03/31/04	04/30/04	05/03/04	05/07/04	**425**	3662	6	20	0.18	0.38
10	04/07/04	05/07/04	05/10/04	05/14/04	**675**	4337	7	20	0.2	0.38
11	04/14/04	05/14/04	05/17/04	05/21/04	**675**	5012	5	35	0.4	0.06
12	04/21/04	05/21/04	05/24/04	05/28/04	**538**	5550	3	20	0.18	0.38
13	04/28/04	05/28/04	05/31/04	06/04/04	**-188**	5362	5	20	0.04	0.38
14	05/05/04	06/04/04	06/07/04	06/11/04	**825**	6187	3	20	0.1	0.4
15	05/12/04	06/11/04	06/14/04	06/18/04	**-50**	6137	3	65	0.16	0.1
16	05/19/04	06/18/04	06/21/04	06/25/04	**900**	7037	10	20	0.06	0.18
			OSNP AVE =		**439.8**					
			OSNP STD =		**478.6**					
		200 Boot	13wk Sum AVE =		**5574**					
		200 Boot	13wk Sum STD =		**1715**					

Where
PF = Profit Factor in Test optimization section
LR= Consecutive loses in a row in test optimization section
WR= Consecutive winners in a row in test optimization section
osnp = Weekly Out-of-sample net profit
ont = The number of trades in the out-of-sample week.
Equity = running sum of the weekly out-of-sample net profits

FIGURE 3 Specialized Trade By Trade Summary
E-Mini 5min Bars Repeated Median Velocity system   6/7/2004 - 6/25/2004

Entry Date	Entry Time		Entry Price	Exit Date	Exit Time	Exit Price	Bars InTrd	Trade $P&L	Trade Max$Pft	Time	Trade Max$DD	Time
6/7/04	1010	Buy	1130.85	6/7/04	1610	1141.35	72	$525.00	$550.00	1605	($37.50)	1020
6/8/04	1005	Buy	1136.85	6/8/04	1015	1137.35	2	$25.00	$25.00	1015	$0.00	1005
6/8/04	1015	Sell	1137.35	6/8/04	1110	1138.85	11	($75.00)	$25.00	1015	($87.50)	1055
6/8/04	1110	Buy	1138.85	6/8/04	1610	1142.35	60	$175.00	$200.00	1435	($50.00)	1140
6/9/04	1005	Buy	1139.35	6/9/04	1020	1138.85	3	($25.00)	$12.50	1005	($37.50)	1010
6/9/04	1020	Sell	1138.85	6/9/04	1510	1135.10	58	$187.50	$287.50	1410	($37.50)	1025
6/9/04	1510	Buy	1135.10	6/9/04	1550	1133.35	8	($87.50)	$0.00	1510	($125.00)	1540
6/9/04	1550	Sell	1133.35	6/9/04	1610	1131.85	4	$75.00	$75.00	1555	$0.00	1550
6/10/04	1015	Buy	1135.60	6/10/04	1610	1136.10	71	$25.00	$62.50	1600	($112.50)	1105
6/14/04	1005	Buy	1129.85	6/14/04	1020	1129.10	3	($37.50)	$12.50	1005	($50.00)	1010
6/14/04	1020	Sell	1129.10	6/14/04	1610	1124.35	70	$237.50	$287.50	1445	($37.50)	1030
6/15/04	1005	Buy	1132.85	6/15/04	1540	1131.60	67	($62.50)	$262.50	1245	($137.50)	1525
6/15/04	1540	Sell	1131.60	6/15/04	1610	1133.85	6	($112.50)	$25.00	1540	($112.50)	1610
6/16/04	1005	Buy	1131.60	6/16/04	1610	1133.60	73	$100.00	$200.00	1520	($12.50)	1005
6/17/04	1005	Sell	1129.10	6/17/04	1120	1129.60	15	($25.00)	$87.50	1035	($25.00)	1115
6/17/04	1120	Buy	1129.60	6/17/04	1610	1131.85	58	$112.50	$212.50	1415	$0.00	1120
6/18/04	1030	Buy	1137.35	6/18/04	1540	1134.35	62	($150.00)	$87.50	1030	($162.50)	1505
6/18/04	1540	Sell	1134.35	6/18/04	1610	1134.85	6	($25.00)	$0.00	1540	($87.50)	1550
6/21/04	1025	Buy	1137.60	6/21/04	1530	1133.60	61	($200.00)	$0.00	1025	($200.00)	1525
6/21/04	1530	Sell	1133.60	6/21/04	1610	1129.85	8	$187.50	$187.50	1610	$0.00	1530
6/22/04	1005	Sell	1130.85	6/22/04	1225	1127.60	28	$162.50	$312.50	1110	$0.00	1005
6/22/04	1225	Buy	1127.60	6/22/04	1610	1135.85	45	$412.50	$412.50	1605	($25.00)	1230
6/23/04	1210	Buy	1134.35	6/23/04	1610	1145.35	48	$550.00	$550.00	1605	$0.00	1210
6/24/04	1005	Buy	1145.10	6/24/04	1435	1141.85	54	($162.50)	$75.00	1140	($162.50)	1425
6/24/04	1435	Sell	1141.85	6/24/04	1610	1141.35	19	$25.00	$75.00	1510	($87.50)	1440
6/25/04	1005	Buy	1143.85	6/25/04	1225	1141.85	28	($100.00)	$100.00	1035	($100.00)	1220
6/25/04	1225	Sell	1141.85	6/25/04	1600	1138.35	43	$175.00	$225.00	1525	($50.00)	1335
6/25/04	1600	Buy	1138.35	6/25/04	1610	1135.35	2	($150.00)	$25.00	1600	($150.00)	1610

Figure 4 ES 5min Repeated Median Velocity System 6/21/04 to 6/25/04

E-Mini S&P 500 Index - CME-5 min 06/25/2004