

# How Good Is Your Data?

---

by Sunny J. Harris

## Introduction

It was brought to my attention recently, in a competitive comment made by a stock market charting software vendor, that perhaps not all data is equal.

I had never given it much thought. I have been using one or two vendors nearly exclusively for about 20 years. I have never had reason to question their data. My fills are “good enough.” My closing prices seem to match what I see on TV or hear on the radio, or even find online. As long as the profits continue to roll in, there has been no reason to question the data. And, so I haven’t. Along the way, I have used ancillary data, before my software vendor became a data vendor, and even then there was no reason to question any data discrepancies.

But, this particular vendor told me that my vendor’s data is off by just enough to generate a side income for the vendor, through the slippage from actual price to the price I am presented. I questioned the remark, knowing that all the vendors compete heavily with each other on many fronts. It was probably just one of those chip-on-the-shoulder kind of remarks.

Nevertheless, my curiosity was peaked, and I began an investigation. My first effort was to start a spreadsheet and compare the two vendors in question. I decided to consider only the past 5 years of data, to keep it simple. My data experiment runs from 6/30/2005 to 6/29/2010.

## Running Experiments

I began my experiment by simply exporting the data for a single symbol from each software application to a cvs (comma separated values) text file. The instrument I chose was the Russell 2000 Index, which has different symbols in different software—like ^RUT, \$RUT, and RU2000. I selected the Russell 2000 because of its high liquidity, ease of use, and because it’s something little guys like us can trade.

The Figure below (Figure 1.01) shows the beginning of the spreadsheet, with the two vendors’ data in the columns. At first glance it would appear that everything is in order, with a few small discrepancies here and there. Not much to worry one. The differences in the data, where there is one, seem to be out in the hundredths place, like 600.01 vs 600.02. It doesn’t seem like that would make much difference over the course of time, with some errors to the positive and some to the negative. Seems like it should pretty much be a wash. At least, I don’t think deviations that small would affect my profit bottom-line.

Next, I put columns in the spreadsheet to calculate the differences between the OHLC of each vendor to see if I could make any sense of that. A piece of that spread sheet is shown in Figure 1.02. At the top of each column, in the first row of data, is the result of calculating the sum of all the differences between the two vendors’ OHL and C data. This small piece of information is of interest to me. I wouldn’t have

been surprised it each component had been consistently lower or consistently higher than the other. But, these summation numbers show that the data is all over the map. The closes are 52 points lower, the opens are 40 points higher, the highs are 65 points lower and the lows are 48 points higher. It seems kind of fishy to me that the spread between the numbers is alternating positive and negative. Could it be, as one vendor suggested, that there is enough spread in there for Vendor T to cash in on the spread alone?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2	Vendor T						Vendor M						YAHQ	
3	Date	Time	Open	High	Low	Close	<Date>	<Time>	<Open>	<High>	<Low>	<Close>	Date	
4	9/30/2005	1300	665.04	667.78	663.02	667.78	9/30/2005	11:59:00 PM	665.03	667.8	663.02	667.80	9/30/2005	
5	10/3/2005	1300	667.75	673.24	667.75	670.55	10/3/2005	11:59:00 PM	667.8	673.24	667.75	670.55	10/3/2005	
6	10/4/2005	1300	670.99	674.17	663.83	663.84	10/4/2005	11:59:00 PM	670.55	674.17	663.83	663.84	10/4/2005	
7	10/5/2005	1300	663.84	663.84	644.98	644.98	10/5/2005	11:59:00 PM	663.84	663.84	644.98	644.98	10/5/2005	
8	10/6/2005	1300	645.13	648.54	633.38	639.45	10/6/2005	11:59:00 PM	644.98	648.54	633.38	639.45	10/6/2005	
9	10/7/2005	1300	640.79	645.64	640.4	644.34	10/7/2005	11:59:00 PM	639.45	645.64	639.45	644.33	10/7/2005	
10	10/10/2005	1300	644.98	645.61	637.32	637.97	10/10/2005	11:59:00 PM	644.33	645.61	637.32	637.97	10/10/2005	
11	10/11/2005	1300	638.55	642.09	630.09	630.09	10/11/2005	11:59:00 PM	637.97	642.09	630.08	630.08	10/11/2005	
12	10/12/2005	1300	629.9	631.13	616.78	621.57	10/12/2005	11:59:00 PM	630.08	631.13	618.68	621.57	10/12/2005	
13	10/13/2005	1300	620.94	624.1	614.76	623.28	10/13/2005	11:59:00 PM	621.57	624.1	614.76	623.28	10/13/2005	
14	10/14/2005	1300	624.87	633.14	621.16	633.14	10/14/2005	11:59:00 PM	623.28	633.15	621.16	633.15	10/14/2005	
15	10/17/2005	1300	633.49	635.86	626.74	633.36	10/17/2005	11:59:00 PM	633.15	635.86	626.74	633.37	10/17/2005	
16	10/18/2005	1300	633.23	633.23	625.19	625.37	10/18/2005	11:59:00 PM	633.37	633.37	625.19	625.36	10/18/2005	
17	10/19/2005	1300	624.09	638.24	618.06	638.24	10/19/2005	11:59:00 PM	625.36	638.28	618.06	638.28	10/19/2005	
18	10/20/2005	1300	638.16	638.74	623.14	627.54	10/20/2005	11:59:00 PM	638.28	638.74	623.14	627.54	10/20/2005	
19	10/21/2005	1300	627.24	636.26	627.24	632.73	10/21/2005	11:59:00 PM	627.54	636.26	627.24	632.73	10/21/2005	
20	10/24/2005	1300	632.99	646.6	632.99	646.6	10/24/2005	11:59:00 PM	632.73	646.6	632.73	646.60	10/24/2005	
21	10/25/2005	1300	646.6	646.6	636.82	642.73	10/25/2005	11:59:00 PM	646.6	646.6	636.82	642.73	10/25/2005	
22	10/26/2005	1300	642.24	648.01	637.37	638.41	10/26/2005	11:59:00 PM	642.73	648.01	637.37	638.41	10/26/2005	
23	10/27/2005	1300	636.9	636.9	623.54	624.03	10/27/2005	11:59:00 PM	638.41	638.41	623.54	624.03	10/27/2005	
24	10/28/2005	1300	624.18	635.56	624.18	635.31	10/28/2005	11:59:00 PM	624.03	635.56	624.03	635.33	10/28/2005	
25	10/31/2005	1300	635.39	648.26	635.39	646.54	10/31/2005	11:59:00 PM	635.33	648.26	635.33	646.61	10/31/2005	
26	11/1/2005	1300	645.4	646.09	640.18	643.03	11/1/2005	11:59:00 PM	646.61	646.61	640.18	643.02	11/1/2005	
27	11/2/2005	1300	643.09	657.05	642.88	657.04	11/2/2005	11:59:00 PM	643.02	657.05	642.88	657.05	11/2/2005	
28	11/3/2005	1300	659.1	663.9	656.56	658.78	11/3/2005	11:59:00 PM	657.05	663.9	656.56	658.77	11/3/2005	
29	11/4/2005	1300	659.38	660.13	653.29	658.16	11/4/2005	11:59:00 PM	658.77	660.13	653.29	658.16	11/4/2005	
30	11/7/2005	1300	659.17	663.07	656.31	661.24	11/7/2005	11:59:00 PM	658.16	663.07	656.31	661.24	11/7/2005	
31	11/8/2005	1300	660.17	660.17	654.67	656.23	11/8/2005	11:59:00 PM	661.24	661.24	654.67	656.23	11/8/2005	
32	11/9/2005	1300	656.73	663.75	654.98	659.81	11/9/2005	11:59:00 PM	656.23	663.75	654.98	659.83	11/9/2005	

Figure 1.01—Data Comparison: Vendor T vs Vendor M

J	AV	AW	AX	AY	AZ	BA
	(51.93)	39.58	(64.65)	48.46		(61)
	DIFFERENCES T v M					Diff
	Close	Open	High	Low		Ch
	(0.02)	0.01	-0.02	0		
	-	-0.05	0	0		
	-	0.44	0	0		
	-	0	0	0		
	-	0.15	0	0		
	0.01	1.34	0	0.95		
	-	0.65	0	0		
	0.01	0.58	0	0.01		
	-	-0.18	0	-1.9		
	-	-0.63	0	0		
	(0.01)	1.59	-0.01	0		
	(0.01)	0.34	0	0		
	0.01	-0.14	-0.14	0		
	(0.04)	-1.27	-0.04	0		
	-	-0.12	0	0		
	-	-0.3	0	0		
	-	0.26	0	0.26		
	-	0	0	0		
	-	-0.49	0	0		
	-	-1.51	-1.51	0		
	(0.02)	0.15	0	0.15		
	(0.07)	0.06	0	0.06		
	0.01	-1.21	-0.52	0		
	(0.01)	0.07	0	0		
	0.01	2.05	0	0		
	-	0.61	0	0		
	-	1.01	0	0		
	-	-1.07	-1.07	0		
	(0.02)	0.5	0	0		

Figure 1.02—Total Differences

With these two pieces of information in hand, I now am curious and want to compare the data to another well-known vendor, to see whether their data matches either Vendor T or Vendor M. So, it's back to the Export facility in software G, to create another set of columns in the spreadsheet. I'm hoping (the eternal optimist) that data from Vendor G will match one or the other of the first two vendors and I'll have an answer.

The next figure (Figure 1.03) shows the data from Vendor T and Vendor G, the new set for comparison.



	A	B	C	D	E	F	G	T	U	V	W	X	Y	Z
1														
2	Vendor T							Vendor G						
3	Date	Time	Open	High	Low	Close		Open	High	Low	Close			
4	9/30/2005	1300	665.04	667.78	663.02	667.78		9/30/2005	665.03	667.8	663.02	667.8		
5	10/3/2005	1300	667.75	673.24	667.75	670.55		10/3/2005	667.8	673.24	667.75	670.55		
6	10/4/2005	1300	670.99	674.17	663.83	663.84		10/4/2005	670.55	674.17	663.83	663.84		
7	10/5/2005	1300	663.84	663.84	644.98	644.98		10/5/2005	663.84	663.84	644.98	644.98		
8	10/6/2005	1300	645.13	648.54	633.38	639.45		10/6/2005	644.98	648.54	633.38	639.45		
9	10/7/2005	1300	640.79	645.64	640.4	644.34		10/7/2005	639.45	645.64	639.45	644.33		
10	10/10/2005	1300	644.98	645.61	637.32	637.97		10/10/2005	644.33	645.61	637.32	637.97		
11	10/11/2005	1300	638.55	642.09	630.09	630.09		10/11/2005	637.97	642.09	630.08	630.08		
12	10/12/2005	1300	629.9	631.13	616.78	621.57		10/12/2005	630.08	631.13	616.78	621.57		
13	10/13/2005	1300	620.94	624.1	614.76	623.28		10/13/2005	621.57	624.1	614.76	623.28		
14	10/14/2005	1300	624.87	633.14	621.16	633.14		10/14/2005	623.28	633.15	621.16	633.15		
15	10/17/2005	1300	633.49	635.86	626.74	633.36		10/17/2005	633.15	635.86	626.74	633.37		
16	10/18/2005	1300	633.23	633.23	625.19	625.37		10/18/2005	633.37	633.37	625.19	625.36		
17	10/19/2005	1300	624.09	638.24	618.06	638.24		10/19/2005	625.36	638.28	618.06	638.28		
18	10/20/2005	1300	638.16	638.74	623.14	627.54		10/20/2005	638.28	638.74	623.14	627.54		
19	10/21/2005	1300	627.24	636.26	627.24	632.73		10/21/2005	627.54	636.26	627.24	632.73		
20	10/24/2005	1300	632.99	646.6	632.99	646.6		10/24/2005	632.73	646.6	632.73	646.6		
21	10/25/2005	1300	646.6	646.6	636.82	642.73		10/25/2005	646.6	646.6	636.82	642.73		
22	10/26/2005	1300	642.24	648.01	637.37	638.41		10/26/2005	642.73	648.01	637.37	638.41		
23	10/27/2005	1300	636.9	636.9	623.54	624.03		10/27/2005	638.41	638.41	623.54	624.03		
24	10/28/2005	1300	624.18	635.56	624.18	635.31		10/28/2005	624.03	635.56	624.03	635.33		
25	10/31/2005	1300	635.39	648.26	635.39	646.54		10/31/2005	635.33	648.26	635.33	646.61		
26	11/1/2005	1300	645.4	646.09	640.18	643.03		11/1/2005	646.61	646.61	640.18	643.02		
27	11/2/2005	1300	643.09	657.05	642.88	657.04		11/2/2005	643.02	657.05	642.88	657.05		
28	11/3/2005	1300	659.1	663.9	656.56	658.78		11/3/2005	657.05	663.9	656.56	658.77		
29	11/4/2005	1300	659.38	660.13	653.29	658.16		11/4/2005	658.77	660.13	653.29	658.16		
30	11/7/2005	1300	659.17	663.07	656.31	661.24		11/7/2005	658.16	663.07	656.31	661.24		
31	11/8/2005	1300	660.17	660.17	654.67	656.23		11/8/2005	661.24	661.24	654.67	656.23		
32	11/9/2005	1300	656.73	663.75	654.98	659.81		11/9/2005	656.23	663.75	654.98	659.83		

Figure 1.03—Data Comparison: Vendor T vs Vendor G

Again, giving the data a quick, cursory glance, nothing really seems amiss. The variations are again out in the hundredth place for the most part.

AY	AZ	BA	BB	BC
(61.10)	26.61	(64.37)	47.27	
DIFFERENCES T v G				
Close	Open	High	Low	
-0.02	0.01	-0.02	0	
0	-0.05	0	0	
0	0.44	0	0	
0	0	0	0	
0	0.15	0	0	
0.01	1.34	0	0.95	
0	0.65	0	0	
0.01	0.58	0	0.01	
0	-0.18	0	0	
0	-0.63	0	0	
-0.01	1.59	-0.01	0	
-0.01	0.34	0	0	
0.01	-0.14	-0.14	0	
-0.04	-1.27	-0.04	0	
0	-0.12	0	0	
0	-0.3	0	0	
0	0.26	0	0.26	
0	0	0	0	
0	-0.49	0	0	
0	-1.51	-1.51	0	
-0.02	0.15	0	0.15	
-0.07	0.06	0	0.06	
0.01	-1.21	-0.52	0	
-0.01	0.07	0	0	
0.01	2.05	0	0	
0	0.61	0	0	
0	1.01	0	0	
0	-1.07	-1.07	0	
-0.02	0.5	0	0	

Figure 1.04—Total Differences T vs G

Again, I have found small discrepancies that lead to large numbers when summed over time. On its own an error of 0.01 doesn't seem like much. But, when you add that up over 5 years of data, that's 1257 trading days, and an accumulated error of \$12.57. Remember, now, that each point is worth \$100 on the RUT. This is where it starts looking scary. Multiplying \$12.57 \* 100 we get \$1,257.00. That's over a thousand dollars out of the trader's pocket. It still isn't huge, but if you are the vendor and you have 20,000 clients at \$1,000 each, that comes to \$20,000,000—twenty million dollars over five years. Now I'm beginning to understand what the first vendor meant. Still, I can't go anywhere with this inkling of an idea.

This situation is now akin to having a clock-shop where each clock tells time a little bit off from every other clock. There's no way to tell what time it really is. Which of these three clocks is telling me the correct time?

This situation now demands that I compare the data from Vendor G to Vendor T and also to Vendor M. I'm not sure what I will know if none of it matches, but if one matches one other, then will know something about the veracity of the vendor's data who doesn't match.

Here's the spreadsheet I have for 3 vendors' data, so far. See Figure 1.04.

	A	B	C	D	E	F	G	H	I	J	K	L	R	S	T	U	V	W	X	Y
1																				
2	Vendor T						Vendor M						Vendor G							RUSSELL
3	Date	Open	High	Low	Close		<Date>	<Open>	<High>	<Low>	<Close>		Open	High	Low	Close				Date
4	9/30/2005	665.04	667.78	663.02	667.78		9/30/2005	665.03	667.8	663.02	667.80		9/30/2005	665.03	667.8	663.02	667.8			9/30/2005
5	10/3/2005	667.75	673.24	667.75	670.55		10/3/2005	667.8	673.24	667.75	670.55		10/3/2005	667.8	673.24	667.75	670.55			10/3/2005
6	10/4/2005	670.99	674.17	663.83	663.84		10/4/2005	670.55	674.17	663.83	663.84		10/4/2005	670.55	674.17	663.83	663.84			10/4/2005
7	10/5/2005	663.84	663.84	644.98	644.98		10/5/2005	663.84	663.84	644.98	644.98		10/5/2005	663.84	663.84	644.98	644.98			10/5/2005
8	10/6/2005	645.13	648.54	633.38	639.45		10/6/2005	644.98	648.54	633.38	639.45		10/6/2005	644.98	648.54	633.38	639.45			10/6/2005
9	10/7/2005	640.79	645.64	640.4	644.34		10/7/2005	639.45	645.64	639.45	644.33		10/7/2005	639.45	645.64	639.45	644.33			10/7/2005
10	10/10/2005	644.98	645.61	637.32	637.97		10/10/2005	644.33	645.61	637.32	637.97		10/10/2005	644.33	645.61	637.32	637.97			10/10/2005
11	10/11/2005	638.55	642.09	630.09	630.09		10/11/2005	637.97	642.09	630.08	630.08		10/11/2005	637.97	642.09	630.08	630.08			10/11/2005
12	10/12/2005	629.9	631.13	616.78	621.57		10/12/2005	630.08	631.13	618.68	621.57		10/12/2005	630.08	631.13	616.78	621.57			10/12/2005
13	10/13/2005	620.94	624.1	614.76	623.28		10/13/2005	621.57	624.1	614.76	623.28		10/13/2005	621.57	624.1	614.76	623.28			10/13/2005
14	10/14/2005	624.87	633.14	621.16	633.14		10/14/2005	623.28	633.15	621.16	633.15		10/14/2005	623.28	633.15	621.16	633.15			10/14/2005
15	10/17/2005	633.49	635.86	626.74	633.36		10/17/2005	633.15	635.86	626.74	633.37		10/17/2005	633.15	635.86	626.74	633.37			10/17/2005
16	10/18/2005	633.23	633.23	625.19	625.37		10/18/2005	633.37	633.37	625.19	625.36		10/18/2005	633.37	633.37	625.19	625.36			10/18/2005
17	10/19/2005	624.09	638.24	618.06	638.24		10/19/2005	625.36	638.28	618.06	638.28		10/19/2005	625.36	638.28	618.06	638.28			10/19/2005
18	10/20/2005	638.16	638.74	623.14	627.54		10/20/2005	638.28	638.74	623.14	627.54		10/20/2005	638.28	638.74	623.14	627.54			10/20/2005
19	10/21/2005	627.24	636.26	627.24	632.73		10/21/2005	627.54	636.26	627.24	632.73		10/21/2005	627.54	636.26	627.24	632.73			10/21/2005
20	10/24/2005	632.99	646.6	632.99	646.6		10/24/2005	632.73	646.6	632.73	646.60		10/24/2005	632.73	646.6	632.73	646.6			10/24/2005
21	10/25/2005	646.6	646.6	636.82	642.73		10/25/2005	646.6	646.6	636.82	642.73		10/25/2005	646.6	646.6	636.82	642.73			10/25/2005
22	10/26/2005	642.24	648.01	637.37	638.41		10/26/2005	642.73	648.01	637.37	638.41		10/26/2005	642.73	648.01	637.37	638.41			10/26/2005
23	10/27/2005	636.9	636.9	623.54	624.03		10/27/2005	638.41	638.41	623.54	624.03		10/27/2005	638.41	638.41	623.54	624.03			10/27/2005
24	10/28/2005	624.18	635.56	624.18	635.31		10/28/2005	624.03	635.56	624.03	635.33		10/28/2005	624.03	635.56	624.03	635.33			10/28/2005
25	10/31/2005	635.39	648.26	635.39	646.54		10/31/2005	635.33	648.26	635.33	646.61		10/31/2005	635.33	648.26	635.33	646.61			10/31/2005
26	11/1/2005	645.4	646.09	640.18	643.03		11/1/2005	646.61	646.61	640.18	643.02		11/1/2005	646.61	646.61	640.18	643.02			11/1/2005
27	11/2/2005	643.09	657.05	642.88	657.04		11/2/2005	643.02	657.05	642.88	657.05		11/2/2005	643.02	657.05	642.88	657.05			11/2/2005
28	11/3/2005	659.1	663.9	656.56	658.78		11/3/2005	657.05	663.9	656.56	658.77		11/3/2005	657.05	663.9	656.56	658.77			11/3/2005
29	11/4/2005	659.38	660.13	653.29	658.16		11/4/2005	658.77	660.13	653.29	658.16		11/4/2005	658.77	660.13	653.29	658.16			11/4/2005
30	11/7/2005	659.17	663.07	656.31	661.24		11/7/2005	658.16	663.07	656.31	661.24		11/7/2005	658.16	663.07	656.31	661.24			11/7/2005
31	11/8/2005	660.17	660.17	654.67	656.23		11/8/2005	661.24	661.24	654.67	656.23		11/8/2005	661.24	661.24	654.67	656.23			11/8/2005
32	11/9/2005	656.73	663.75	654.98	659.81		11/9/2005	656.23	663.75	654.98	659.83		11/9/2005	656.23	663.75	654.98	659.83			11/9/2005

Figure 1.05—Three Vendors' Data

Back to the differences spreadsheet, I insert columns for calculating the new spreads: Vendor T vs Vendor G; Vendor T vs Vendor M; and Vendor G vs Vendor M. That setup will have each compared against the other, and maybe some clarity will arise. The differences section of the spreadsheet is shown in Figure 1.06.



AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH
(51.93)	39.58	(64.65)	48.46		(61.10)	26.61	(64.37)	47.27		(9.17)	(12.97)	0.28	(1.19)	
DIFFERENCES T v M				DIFFERENCES T v G				DIFFERENCES M v G						
Close	Open	High	Low	Close	Open	High	Low	Close	Open	High	Low			
(0.02)	0.01	-0.02	0	-0.02	0.01	-0.02	0	-	0	0	0	0	0	
-	-0.05	0	0	0	-0.05	0	0	-	0	0	0	0	0	
-	0.44	0	0	0	0.44	0	0	-	0	0	0	0	0	
-	0	0	0	0	0	0	0	-	0	0	0	0	0	
-	0.15	0	0	0	0.15	0	0	-	0	0	0	0	0	
0.01	1.34	0	0.95	0.01	1.34	0	0.95	-	0	0	0	0	0	
-	0.65	0	0	0	0.65	0	0	-	0	0	0	0	0	
0.01	0.58	0	0.01	0.01	0.58	0	0.01	-	0	0	0	0	0	
-	-0.18	0	-1.9	0	-0.18	0	0	-	0	0	0	0	1.9	
-	-0.63	0	0	0	-0.63	0	0	-	0	0	0	0	0	
(0.01)	1.59	-0.01	0	-0.01	1.59	-0.01	0	-	0	0	0	0	0	
(0.01)	0.34	0	0	-0.01	0.34	0	0	-	0	0	0	0	0	
0.01	-0.14	-0.14	0	0.01	-0.14	-0.14	0	-	0	0	0	0	0	
(0.04)	-1.27	-0.04	0	-0.04	-1.27	-0.04	0	-	0	0	0	0	0	
-	-0.12	0	0	0	-0.12	0	0	-	0	0	0	0	0	
-	-0.3	0	0	0	-0.3	0	0	-	0	0	0	0	0	
-	0.26	0	0.26	0	0.26	0	0.26	-	0	0	0	0	0	
-	0	0	0	0	0	0	0	-	0	0	0	0	0	
-	-0.49	0	0	0	-0.49	0	0	-	0	0	0	0	0	
-	-1.51	-1.51	0	0	-1.51	-1.51	0	-	0	0	0	0	0	
(0.02)	0.15	0	0.15	-0.02	0.15	0	0.15	-	0	0	0	0	0	
(0.07)	0.06	0	0.06	-0.07	0.06	0	0.06	-	0	0	0	0	0	
0.01	-1.21	-0.52	0	0.01	-1.21	-0.52	0	-	0	0	0	0	0	
(0.01)	0.07	0	0	-0.01	0.07	0	0	-	0	0	0	0	0	
0.01	2.05	0	0	0.01	2.05	0	0	-	0	0	0	0	0	
-	0.61	0	0	0	0.61	0	0	-	0	0	0	0	0	
-	1.01	0	0	0	1.01	0	0	-	0	0	0	0	0	
-	-1.07	-1.07	0	0	-1.07	-1.07	0	-	0	0	0	0	0	
(0.02)	0.5	0	0	-0.02	0.5	0	0	-	0	0	0	0	0	

Figure 1.06—Differences T vs M vs G

Ahah! Look at the zeros in columns BD through BG. Reading the description in row B over those columns (shaded green on my screen), I see that the zeros show up when comparing Vendor M to Vendor G. We're getting closer to a revelation. Still, looking at the numbers over the header "Differences M v G" we see that despite all the zeros there are discrepancies along the way, giving us (9.17) among the closing values. As I scan down the columns of this spreadsheet comparison I find that on 9/17/2008 there was a difference of (8.83) between the close of Vendor M and the close of Vendor G. That's where most of the error comes in.

How could these vendors have such differences among their data? Isn't the close the close, no matter who vends it? Seems to me it should be. But, this profession is fraught with more "that's the way it is" instead of "should be's."

Next I decided I should get it from the horse's mouth. So, I called the Russell 2000 exchange and got the data from them. Now that data should be correct, don't you think? It's their own index, so they should know. Interestingly, however, the only had closing prices for the first part of my experiment, dates

09/30/2005 through 6/22/2007. Closes will just have to do for comparing data vendors versus the Russell 2000 itself. Of note, is that their closing prices are accurate down to 6 decimal places, while all the others only have 2 digits after the decimal.

Adding yet another set of columns to the spreadsheet, I placed the RU2000 from the exchange in place. While I was at it, I introduced another column with the calculation for the range of each day. As I look at the data, I'm beginning to wonder whether the errors in the data work themselves out by having the same range for the day, even if the open and close are different. That's why you'll now see a column labeled Range. Figure 1.07 shows three data vendors and the Russell 2000 exchange data side-by-side.

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	
2	Vendor T						Vendor M							Vendor G							RUSSELL EXCHANGE					DI			
3	Date	Open	High	Low	Close	Range	<Date>	<Open>	<High>	<Low>	<Close>	Range		Date	Open	High	Low	Close	Range		Date	Close	Open	High	Low	Range		CLO	
4	9/30/2005	665.04	667.78	663.02	667.78	4.76	9/30/2005	665.03	667.8	663.02	667.80	4.78		9/30/2005	665.03	667.8	663.02	667.8	4.78		9/30/2005	667.98000							
5	10/3/2005	667.75	673.24	667.75	670.55	5.49	10/3/2005	667.8	673.24	667.75	670.55	5.49		10/3/2005	667.8	673.24	667.75	670.55	5.49		10/3/2005	670.549600							
6	10/4/2005	670.99	674.17	663.83	663.84	10.34	10/4/2005	670.55	674.17	663.83	663.84	10.34		10/4/2005	670.55	674.17	663.83	663.84	10.34		10/4/2005	663.839800							
7	10/5/2005	663.84	663.84	644.98	644.98	18.86	10/5/2005	663.84	663.84	644.98	644.98	18.86		10/5/2005	663.84	663.84	644.98	644.98	18.86		10/5/2005	644.976700							
8	10/6/2005	645.13	648.54	633.38	639.45	15.16	10/6/2005	644.98	648.54	633.38	639.45	15.16		10/6/2005	644.98	648.54	633.38	639.45	15.16		10/6/2005	639.449900							
9	10/7/2005	640.79	645.64	640.4	644.34	5.24	10/7/2005	639.45	645.64	639.45	644.33	6.19		10/7/2005	639.45	645.64	639.45	644.33	6.19		10/7/2005	644.334200							
10	10/10/2005	644.98	645.61	637.32	637.97	8.29	10/10/2005	644.33	645.61	637.32	637.97	8.29		10/10/2005	644.33	645.61	637.32	637.97	8.29		10/10/2005	637.968800							
11	10/11/2005	638.55	642.09	630.09	630.09	12	10/11/2005	637.97	642.09	630.08	630.08	12.01		10/11/2005	637.97	642.09	630.08	630.08	12.01		10/11/2005	630.083100							
12	10/12/2005	629.9	631.13	616.78	621.57	14.35	10/12/2005	630.08	631.13	616.78	621.57	14.35		10/12/2005	630.08	631.13	616.78	621.57	14.35		10/12/2005	621.569300							
13	10/13/2005	620.94	624.1	614.76	623.28	9.34	10/13/2005	621.57	624.1	614.76	623.28	9.34		10/13/2005	621.57	624.1	614.76	623.28	9.34		10/13/2005	623.280900							
14	10/14/2005	624.87	633.14	621.16	633.14	11.98	10/14/2005	623.28	633.15	621.16	633.15	11.99		10/14/2005	623.28	633.15	621.16	633.15	11.99		10/14/2005	633.150700							
15	10/17/2005	633.49	635.86	626.74	633.36	9.12	10/17/2005	633.15	635.86	626.74	633.37	9.12		10/17/2005	633.15	635.86	626.74	633.37	9.12		10/17/2005	633.369900							
16	10/18/2005	633.23	633.23	625.19	625.37	8.04	10/18/2005	633.37	633.37	625.19	625.36	8.18		10/18/2005	633.37	633.37	625.19	625.36	8.18		10/18/2005	625.364500							
17	10/19/2005	624.09	638.24	618.06	638.24	20.18	10/19/2005	625.36	638.28	618.06	638.28	20.22		10/19/2005	625.36	638.28	618.06	638.28	20.22		10/19/2005	638.281800							
18	10/20/2005	638.16	638.74	623.14	627.54	15.6	10/20/2005	638.28	638.74	623.14	627.54	15.60		10/20/2005	638.28	638.74	623.14	627.54	15.6		10/20/2005	627.540600							
19	10/21/2005	627.24	626.26	627.24	632.73	9.02	10/21/2005	627.54	636.26	627.24	632.73	9.02		10/21/2005	627.54	636.26	627.24	632.73	9.02		10/21/2005	632.732900							
20	10/24/2005	632.99	646.6	632.99	646.6	13.61	10/24/2005	632.73	646.6	632.73	646.60	13.87		10/24/2005	632.73	646.6	632.73	646.6	13.87		10/24/2005	646.603500							
21	10/25/2005	646.6	646.6	636.82	642.73	9.78	10/25/2005	646.6	646.6	636.82	642.73	9.78		10/25/2005	646.6	646.6	636.82	642.73	9.78		10/25/2005	642.732400							
22	10/26/2005	642.24	648.01	637.37	638.41	10.64	10/26/2005	642.73	648.01	637.37	638.41	10.64		10/26/2005	642.73	648.01	637.37	638.41	10.64		10/26/2005	638.410100							
23	10/27/2005	636.9	636.9	623.54	624.03	13.36	10/27/2005	638.41	638.41	623.54	624.03	14.87		10/27/2005	638.41	638.41	623.54	624.03	14.87		10/27/2005	624.028200							
24	10/28/2005	624.18	635.56	624.18	635.31	11.38	10/28/2005	624.03	635.56	624.03	635.33	11.53		10/28/2005	624.03	635.56	624.03	635.33	11.53		10/28/2005	635.329100							
25	10/31/2005	635.39	648.26	635.39	646.54	12.87	10/31/2005	635.33	648.26	635.33	646.61	12.93		10/31/2005	635.33	648.26	635.33	646.61	12.93		10/31/2005	646.605400							
26	11/1/2005	645.4	646.09	640.18	643.03	5.91	11/1/2005	646.61	646.61	640.18	643.02	6.43		11/1/2005	646.61	646.61	640.18	643.02	6.43		11/1/2005	643.024100							
27	11/2/2005	643.09	657.05	642.88	657.04	14.17	11/2/2005	643.02	657.05	642.88	657.05	14.17		11/2/2005	643.02	657.05	642.88	657.05	14.17		11/2/2005	657.051900							
28	11/3/2005	659.1	663.9	656.56	658.78	7.34	11/3/2005	657.05	663.9	656.56	658.77	7.34		11/3/2005	657.05	663.9	656.56	658.77	7.34		11/3/2005	658.771800							
29	11/4/2005	659.38	660.13	653.29	658.16	6.84	11/4/2005	658.77	660.13	653.29	658.16	6.84		11/4/2005	658.77	660.13	653.29	658.16	6.84		11/4/2005	658.162800							
30	11/7/2005	659.17	663.07	656.31	661.24	6.76	11/7/2005	658.16	663.07	656.31	661.24	6.76		11/7/2005	658.16	663.07	656.31	661.24	6.76		11/7/2005	661.238500							
31	11/8/2005	660.17	660.17	654.67	656.23	5.5	11/8/2005	661.24	661.24	654.67	656.23	6.57		11/8/2005	661.24	661.24	654.67	656.23	6.57		11/8/2005	656.234700							
32	11/9/2005	656.73	663.75	654.98	659.83	8.77	11/9/2005	656.23	663.75	654.98	659.83	8.77		11/9/2005	656.23	663.75	654.98	659.83	8.77		11/9/2005	659.813300							
33	11/10/2005	659.92	665.14	651.5	664.94	13.64	11/10/2005	659.81	665.14	651.5	664.93	13.64		11/10/2005	659.81	665.14	651.5	664.93	13.64		11/10/2005	664.934500							
34	11/11/2005	664.22	667.82	663.43	666.66	4.39	11/11/2005	664.93	667.82	663.43	666.66	4.39		11/11/2005	664.93	667.82	663.43	666.66	4.39		11/11/2005	666.663900							

Figure 1.07—Russell 2000 Exchange vs T vs M vs G

The more columns I add to this spreadsheet, the more difficult it is for you to read. So, for those interested in the down and dirty, nitty gritty information in these spreadsheets, come visit [www.MoneyMentor.com/Articles.html](http://www.MoneyMentor.com/Articles.html) where you can see enlargements of these figures.

Again, when you compare the closes of the Russell to the closes of Vendor G, or Vendor T, or Vendor M, there are very slight discrepancies. Nothing to shake a stick at, though. I'm going to add one more data vendor, this time one who does not connect a brokerage to the data. Or, not so far as I know, yet. Yahoo makes their data available for free, and lots of people use it. Because it is such a popular data source, they should have pretty clean data. Let's take a look. Once again, as I add another column, you will be able to see even less on the printed page, so hop on over to the website if you want the details.



Vendor T		Vendor M										YAHOO										Vendor G										RUSSELL EXCHANGE					RANGE
Date	Open	High	Low	Close	Range	<Date>	<Open>	<High>	<Low>	<Close>	Range	Date	Open	High	Low	Close	Range	Date	Open	High	Low	Close	Range	Date	Close	Open	High	Low	Range	R V T							
10/5/2007	833	847.3	833	844.6	14.32	10/5/2007	833	847.32	833	844.88	14.32	10/5/2007	833	847.32	833.00	844.88	14.32	10/5/2007	833	847.3	833	844.9	14.32	10/5/2007	844.874822	837.08	847.32	833.00	14.32	-							

Figure 1.08—Yahoo vs all the rest

I’m getting the impression, as I look closer and closer at the data, that some vendors are using the first opened trade for the open value, and some are using the opening range of the first few minutes. And, same with the low, some using the value at the bell and some using the range as all the last few orders trickle in. Take a look at the chart in Figure 1.09 and you will see what I mean. I’ve overlaid two sets of data on top of each other. You can see where the orange tick will be at a different location than the green tick. Orange is for the Russell 2000 data from the exchange, while the underlying blue and green are from Vendor T.



Figure 1.09—Two Data Vendors on One Chart

In Figure 1.09 the picture is clearer than with a bunch of numbers in a spreadsheet. Remember that adage, “a picture is worth a thousand words?” It’s really true. The orange bars on top are the data from the Russell exchange. The blue and green that show from underneath when they don’t match, is from Vendor T. You can see right from the first bar on this chart that the open tick (to the left) has an orange one and a green one, only slightly different, but different.

On the third bar from the left there is quite a bit of difference between the two opens, though the closes are equal. Similarly on the 5<sup>th</sup> bar you can visualize the discrepancy, and on and on across the chart. The differences are subtle, but they are there.

As for my own trading, I enter and exit on market orders and let the front-runners have their pennies and don’t worry about it. Or, I’ll let a stop take me out. In neither case is it crucial that I place orders on the open of the bar, or even on the close of the bar.

However, and this is a great big however, when writing and testing systems ideas, many if not most of the time, coders specify things like:

```
IF condition1 THEN BUY next bar on the OPEN;  
IF condition2 THEN SELL on CLOSE;
```

I’m beginning to see clearly that if I tested such code against these five data sets I would get different results, as in different profits and losses, depending on the data vendor or software vendor. That then begs the question, “Which results are correct?” The answer is, the correct data is the set that would give the same results as actual trades entered into the market would yield. And, that surely brings us to the crux of things, and the edge of the precipice. Do you want to put your real money into the markets in a reversal system like, say, MACD, just to see whether the trades it generates replicate the trades the hypothetical system generates? No, of course not. And, neither do I. So, perhaps we are at an impasse.

I’ll come back to the impasse in a minute, but for now let’s get back to the data comparisons. From Figure 1.06 we can see that Vendor G and Vendor M are very close in the data they provide. Most of the cells in the spreadsheet contain zeros; that means there is no difference in the data point from Vendor G and the data point from Vendor M. Let’s look at the rest of them. Again, this section of the spreadsheet is broad, so in print the data will be in very small print, so to view the whole picture, go to my website. Otherwise, take a look at Figure 1.10.





tests from there. Same with the data from Vendor Y: the data will be imported from a cvs text file into Vendor T's software and tested from there. We will then view the results of the tests by looking at the Performance Reports correlated to each data set.

In order to set up the experiment, we must hold as many variables as possible constant, so that we are testing apples and apples and actually getting meaningful results. Here are the constraints we will employ:

- Trade only one contract
- Constrain the data to the timeframe 10/13/2005 – 10/13/2010
- Do not allow pyramiding
- Limit the input values to 12, 26, 9
- Enter at the Market, not on the open or close of the signal bar

With these values in mind, we will simply run one test with the above constraints, and compare the results. When I first began running the tests, I got wildly different answers. Because of the different philosophies of the software vendors, it was a little challenging finding the locations of the settings of things like trading 1 contract only versus trading 100 contracts at a time. But, with diligence, I got them all set up identically. First, let me display the results from each vendor in the figures to follow.



Figure 2.01—Vendor T



Figure 2.02—Vendor M

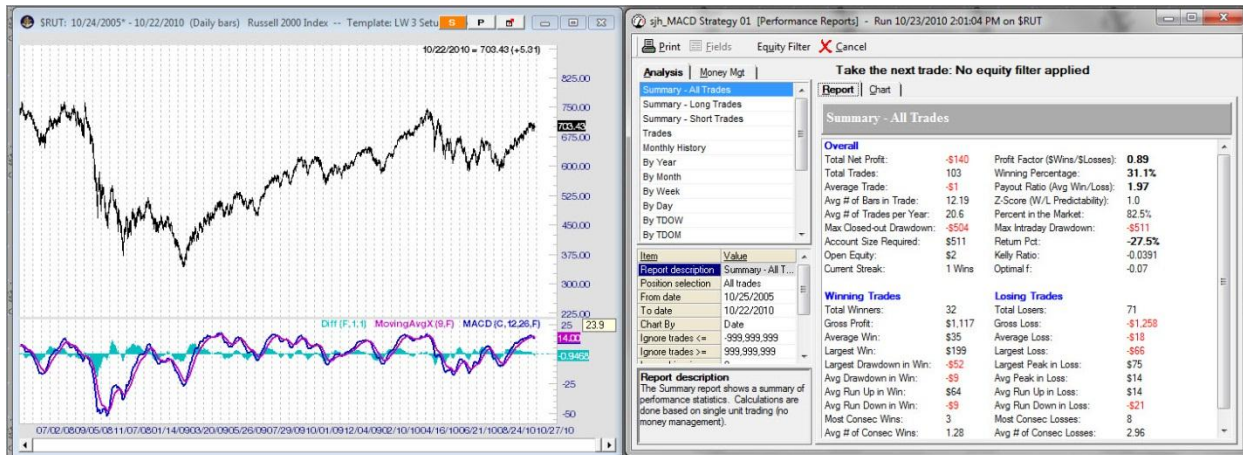


Figure 2.03—Vendor G



Figure 2.04—Vendor Y



Figure 2.05—Vendor R

You can see visually that the results are very much the same. But, we need to inspect the data more closely, and so we will now break it down into tabular format. I am not going to display all of the statistics, only the most important to the analysis.

Stats	Vendor G	Vendor M	Vendor R	Vendor T	Vendor Y
Total Net Profit	<\$221.00>	<\$302.60>	<\$6.17>	<\$272.94>	<\$275.03>
Profit Factor	0.82	0.76	0.99	0.78	0.78
# Trades	102	99	55	100	100
% Profitable	30.4%	29.29%	34.55%	30.00%	30.00%
Avg Trade NP	<\$2.00>	<\$3.06>	<\$0.11>	<\$2.73>	<\$2.75>
Ratio Av Win: Av Loss	1.89	1.83	1.88	1.83	1.82
Avg Bars in Trades	11.98	12.6	14.67	12.72	12.72
Account Size Req'd	\$511.00	\$508.80	\$258.81	\$507.80	\$506.64
CPC Index	0.46	0.41	0.62	0.43	0.43

Table 1.01—Performance Reports from All Five Vendors

Statistically, there is very little difference among the vendors, as far as performance data goes. The one thing that stands out, however, is the difference between the collection of all 4 vendors against the data from the Exchange itself. Notice that Vendor R (the Exchange) has the best performance overall, and is the one dataset that is different from the others.

I am not going to run the full blast of experiments, as I am doing in my next book “Grading the Gurus.” In that book I am testing not only the defaults, but running various optimizations to determine whether there are parameters that work better than the defaults. For the purposes of this article, we are not trying to find whether the MACD system works, but rather are attempting to uncover discrepancies among the data available for analyzing and trading.

Admittedly, this is one set of data on the Russell 2000 Index, and one set of parameters for only the MACD reversal system. It is not a comprehensive test, not by any means a full analysis, but it is useful in answering the question posed by one vendor when touting the accuracy of their data. The outcome is not very dramatic. All of the data vendors present a losing outcome for the standard MACD strategy. The only one that is close to positive is the data from the Exchange; the rest are all about the same amount negative.

It would be interesting to run comprehensive tests with a variety of symbols over a variety of timeframes, optimizing the parameters and using other types of orders besides just buy at the market. But, it would be an extensive test with hundreds of pages of outcomes.

## The Bottom-Line

The bottom-line for these tests is that the trades entered into the real-time markets will vary from each other far more than the data for the trades in this simple experiment. Over approximately 100 trades, there is only a few dollars variation among the data in Table 1.01. It is not significant. In real life trading, your fills will have wider ranges as markets move faster or slower, and the range between bids and asks widen and narrow. In real-life trading, even trading just one share, you can easily lose double the amounts shown above in just a few minutes depending on how fast your order is entered into a fast moving market.



Back-testing is never meant to provide precise replication of what would happen in the real markets. It is meant to give you an overall impression of whether your concepts are viable or not. The markets never again do the very same thing they did yesterday. They may echo similar patterns, but they don't replicate precisely. You can do all the testing in the world, and then when it comes down to entering real trades, the markets will hand you something different and unexpected.

My friend Larry Williams said something that has always stuck with me. He said, "It takes time to make profits." The current propensity to make a few dollars at a time to get \$100 a day is blatantly against my philosophy of trading. I am not pushing buttons real fast to snip off dollars and cents, I believe in mathematical analysis of patterns and detection of setups that are likely to predict important pivot points and turns in the markets.

In the end, all five of these data sources are perfectly adequate for testing and for trading. It's all relative.