Slippage March 1996

Doyne

An analysis of our slippage to date on our futures portfolio. An attempt is made to document the software. The results are summarized in the conclusion.

1. Introduction

1.1 Definitions

For a given fill the *slippage* s is

$$s = sign(o)(f - e) ,$$

where o is the order, f is the fill price, and e is the “expected price”. For futures the expected price is the price of the last tick before the order is generated. “Expected price” is perhaps a bad name, in that on average we don't "expect" this price; we expect a price somewhat higher, i.e., we expect our slippage to be negative.

For many purposes it is useful to distinguish between positive and negative orders. For this purpose I use simply the *fill price shift*, i.e. the difference between the fill price and the expected price. Note that we have defined slippage so that when the price goes against us it is positive.

1.2 Software

The S+ programs for producing the results in this report are contained in the S+ unit splus/srcpc/slippage/src. They are all centered around a "new" slippage structure which is oriented around orders. This structure is described in the document in docs/research/jdf/96/slippage.struct, titled "Slippage Data Structure". This describes both the structure of the old data structure and the new data structure.

One of the recurring problems with analyzing data of this type is that in some cases it is desirable to analyze instruments one at a time, and in other cases one wants to analyze results for multiple instruments all at once. The slippage data structure for a single instrument is a list, with primary components "orders", "ticks", and "fills". We also often use a data structure that is a list with one of these lists for each instrument. Some programs work on the single instrument list, and others work on the multiple instrument list. I have dealt with this by using the naming convention that programs that work on multiple instrument lists start with "All". These are often simple programs that iterate through the elements of the list, calling the corresponding single instrument programs.

1.3 Extracting and testing data

The slippage data extraction function (called "SlippageData") has been modified to retrieve slippage data in the new format. However, if you have data that was retrieved in the old format, there are functions to convert it to the new format:

- ConvertSlipList(x), where x is an old-style slippage list.
- AllConvertSlipList(x) where x is a multi-instrument old style slippage list

Some programs that are useful for checking the integrity of the data are:

- AllSlipNulls(x, verbose=F). Gives an account of nulls in the ticks or fills, as well as the total number of orders. x is a multi-instrument slippage list; verbose=T causes the program to give a listing of dates for each null.
- PlotTicksFills(slipdat, start.date). Plots ticks associated with each order, and gives info about the order and fills. slipdat is a single instrument slippage structure; start.date is a starting date, which need not correspond to one of the actual dates in the data set;
- PlotTickTimeSpan(slipData). Plots the time span in mins between the first and last tick for each order.

- PlotNumTicks(slipData). Plots the number of ticks associated with each order.
- PlotTicksPerMin(slipData). Plots the number of ticks per minute associated with each order.
- PlotAckTImeDelay(slipData). Plots the time span (in minutes) between the order time and ack time for each order.

For the experiments described here three different data sets were extracted. They covered the six futures models we are trading on now, in a time range from July 1, 1995 - Feb. 26, 1996.

1. slip.1: 5 minutes before order to last fill time. There are problems here because the last fill time is entered inconsistently; sometimes the value is timely, and other times it is entered long after the order has been filled.
2. slip.2: 0 minutes before order to 15 minutes (1 hour for oil) after. However, the ticks were not truncated at the close, and so in many cases there were aftermarket ticks.
3. slip.3: 0 minutes before order to 15 minutes (1 hour for oil) later, or the close, whichever happened first. However, there were problems caused by portions of the data set in which the interval the ticks were extracted over was inconsistent, e.g. a month of data for oil in which only ten minutes of ticks were extracted even though we were filling the orders as much as an hour later. (I believe these problems are present in all three data sets).

Unless otherwise noted, most experiments were performed on slip.3.

1.4 Slippage summary

A summary of the slippage across the full data set is given in the following table:

- SlipSummary(slip.3)

	mean M$/yr	mean tk/cn	mean sz wt	mean $/$	stdv	Tstat	mn/ std	HR	vol/sp	trnovr
CL	**1.1**	**1.5**	**1.9**	**743.6**	**3.1**	**9.8**	**7.6**	**73.6**	**9.7**	**1.1**
DM	**0.2**	**0.8**	**0.9**	**119.7**	**2**	**7**	**6.8**	**71.2**	**26.9**	**1.8**
FDB	**0.2**	**0.4**	**0.4**	**38.5**	**3.3**	**2.1**	**1.9**	**57.4**	**40.3**	**1.2**
JY	**0.1**	**0.8**	**1.2**	**65.4**	**4.1**	**2.3**	**3**	**62.4**	**45.2**	**2.5**
SP	**0.3**	**4**	**4**	**80.2**	**29.3**	**2.8**	**2.2**	**64.9**	**34.8**	**1.4**
US	**0.7**	**0.5**	**0.5**	**172.6**	**1.9**	**6**	**4.7**	**66.4**	**15.5**	**1.2**
	2.5	**1.3**	**1.5**	**203.4**	**7.3**	**5**	**4.4**	**66**	**28.7**	**1.5**

The data in each column is explained below:

- **mean M$/yr:** This is the mean slippage, measured in millions of dollars/year, and adjusted to match the present ramp level by multiplying orders in other periods by the appropriate scaling factor. However, this does not adjust for possible changes in the slippage per contract due to larger order sizes (for this is necessary to model the increase in slippage/contract; see Section 5.4). For this column the last row is the sum of the values; for the other columns it is the mean.
- **mean tk/cn:** This is the mean of the slippage associated with each order, in ticks/contract. There is no weighting for different order sizes.
- **mean sz wt:** The mean slippage for each order in ticks/contract, weighted by order size. The weighted mean is the more relevant number, but it is also more sensitive to outliers. Since there is a tendency for slippage to increase with order size, the simple mean tends to be biased downward. Both mean and weighted mean are quoted to give a feeling for the robustness of the results.
- **mean $/$:** The mean slippage in nondimensional units, obtained by multiplying the slippage in ticks/contract by the number of dollars/tick divided by the number of dollars/contract. Values are multiplied by a million to make them easier to read. This has the advantage that it makes a sensible comparison between the slippage on different instruments.
- **stdv:** The standard deviation of the slippage associated with each order.
- **t-stat:** The t-statistic for the mean, i.e. the mean/sqrt(var/num), where var is the variance and num is the number of orders.
- **mn/stdv:** The ratio of the mean to the standard deviation. This is useful because it is a dimensionless quantity to measure slippage. It depends on the overall slippage, and also gives insight into the execution strategy being used by the broker. (The ratio will be lower for a broker who works the orders more). The result is multiplied by 16 (which is roughly sqrt(252)) to annualize it.
- **HR:** The "hit rate" for slippage, i.e. the percentage of slippage values that are unfavorable.
- **vol/spd:** The ratio of the close to close market volatility to the spread. The spread is computed by multiplying our mean slippage by two. Note the volatility here is taken from a different period (chapter 2 of the portfolio backtest report). It would be good to incorporate close prices into the volatility structure to simplify this.
- **trnovr:** The ratio of the mean absolute position to the mean absolute order size. This has units of time (days in this case), and corresponds to the typical time it takes for us to turn over our positions. It was previously called holding period.

This table illustrates several things:

- A disproportionate fraction of the slippage measured in dollars, roughly 70%, comes from two models, CL and US.
- The mean slippage in ticks/contract looks very high for SP. This is because we are taking the tick size for the SP contract to be the nominal size of $5. If we instead use the effective tick size of $25, then the mean slippage for SP is 0.8 ticks/contract,

which is in the normal range with everything else. This suggests that it would be worth the trouble to change our software so that it is more convenient to use effective tick size.

- The weighted mean is typically close to the mean, and when it is different it is larger. The biggest difference occur on the shortest series.
- The slippage measured in nondimensional units varies greatly from instrument to instrument. The largest value of 743 for oil is almost 20 times larger than the smallest value of 39 for the Bund. There are large variations even among instruments that should be comparable, e.g. 173 for the T-bond and 80 for the SP. It is difficult to understand why the slippage should be four and half times as large on the T-bond as on the Bund.
- All of our slippage values are significantly positive.
- The ratio 16 mean/stdv ranges from 1.9 for the Bund to 7.6 for oil. This reflects some combination of the difference in the way the brokers work the orders, and a difference in the slippage. It is particularly instructive to compare this for the SP and US, which one would think would be similar; the ratio is twice as large for the SP as for the US. This corresponds to the fact that the SP trader feels that working the orders gives us an advantage, whereas the US traders have told us they do not believe it is helpful (and we have reduced the time before the close correspondingly).
- The volatility/spread ratio ranges from 45 for the yen to 10 for oil, with a mean value of about 30. Of course, it is somewhat misleading to call this a "spread", since it is really an inferred spread from our slippage, which depends on the skill of our broker as well as characteristics of the market.
- Our turnover periods range from 1.1 for oil to 2.5 for the yen. Oil and the T-bond, where we have the most several slippage problems, are also the two models with the smallest turnover period. These results are an average across the whole period. From plotting the positions as a function of time it is apparent that the holding period has been lengthened in later releases; the holding period of the current models is not as short as is shown here. The lengthening of the holding period in the most recent release is particularly apparent for the S&P model.

2. Variation of slippage with time

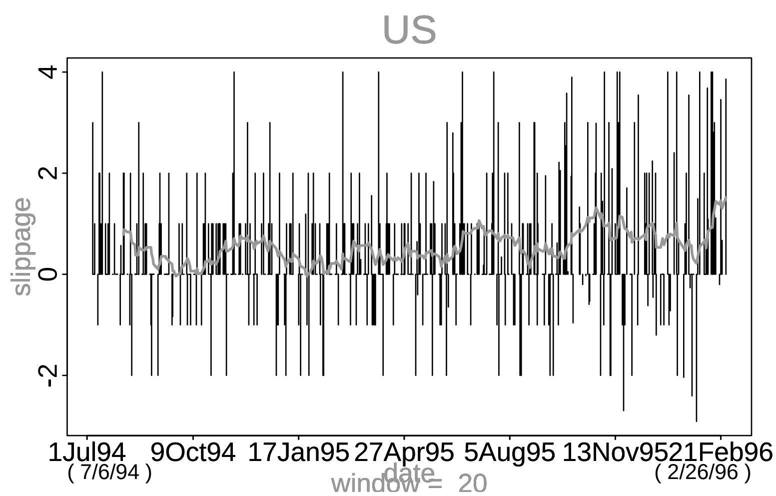
It is often useful to plot slippage vs. time. We are already using this in our weekly slippage reporting, based on the old slippage structure. These reports plot the 20 day running slippage, weighted by fill size. This is done by slippage/src/RunningFillPlot.

2.1 T-bond

A counterpart of this which works on the order based slippage data is

- RunningSlipPlot(slip.3[[“US”], window=20, main.tit=”US”, style=”running”, trim=.05)

The output of this, applied to the US T-bond, is shown in the following figure



Notice that this running plot differs from the weekly reports in several ways. Each data point in the plot shows the mean daily slippage (weighted by fill size for that day); however, the 20 day running mean is the mean of the daily values. This is more convenient than computing the weighted mean of the fills, and it is more robust, but it is not as representative of the true slippage. Another difference is the argument “trim”, which makes it possible to trim outliers. In this case 5% of the data is trimmed. This is done by eliminating the largest 2.5% of positive outliers and 2.5% of negative outliers.

This plot suggests that our slippage has tended to be higher since about June 95. We can test whether this is statistically significant by running the function Pop2MeanTest, e.g.

- SlipSplitTest(slip.3, “US”, “1Jun95”)

This functions is just a special purpose repackaging of the function Pop2MeanTest. The first argument is a multi-instrument slippage structure, the second is the date to split the sample, and the third is the instrument name. It produces the output

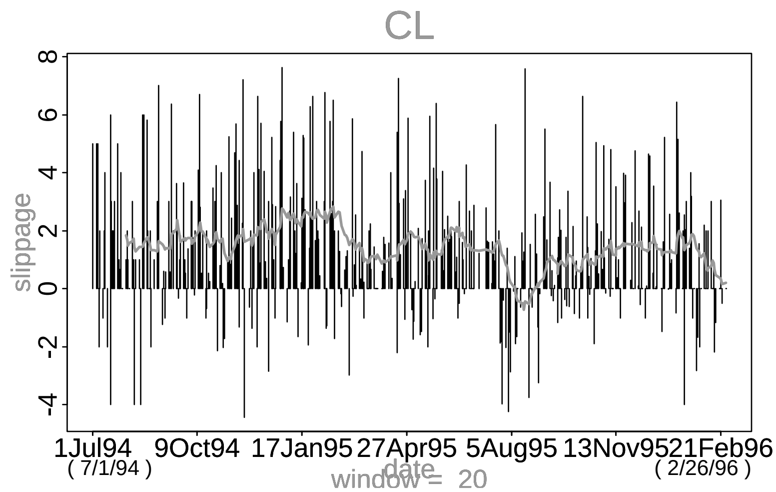
sample	**n**	**mean**	**stdv**	**skew**	**kurt**
1	**230**	**0.325**	**1.299**	**-0.198**	**1.492**
2	**182**	**0.829**	**2.245**	**0.118**	**0.336**

Note that unfavorable slippages are taken as positive. The first sample (before June 95) has a mean that is about 40% that of the mean since then. This is a 2.6 standard deviation result, with a p-value of 0.01, i.e. the odds of this occurring at random are about 1 in 100. Thus the data so far strongly suggest that our slippage on the T-bond has risen significantly starting in June 95.

2.2 Oil

The slippage for oil as a function of time is shown in the next figure:

Figure 1. Slippage vs. time for oil



Some relevant dates for oil are:

- July 1, 1994 - May 6, 1995: Fired model 13:05
- Feb. 27, 1995: Ramp
- May 9 - July 24, 1995: Fired model 13:00
- July 25, 1995 - August 29, 1995 (roughly): Monitored order closely.
- July 25 - Nov. 20, 1995: Fired model 11:00
- Nov. 21 - present: Random firetime, 10:00 - 12:00
- Jan. 11, 1996: Ramped oil model down by factor of 10

One of the features that is quite noticeable on the graph is the dip in slippage, which occurs roughly in the period where we were monitoring our order. To test if this is indeed a statistically significant effect, the period from July 25 - August 29 is compared with the rest of the sample using

- SlipSplitTest(slip.3, “CL”, “25Jul95”, “29Aug95”)

which gives:

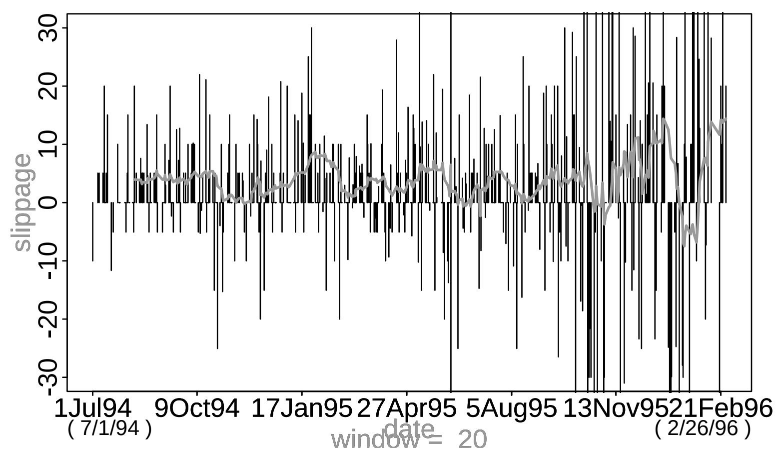
sample	**n**	**mean**	**stdv**	**skew**	**kurt**
1	**25**	**-0.512**	**2.696**	**-0.639**	**1.289**
2	**403**	**1.611**	**3.093**	**-1.008**	**5.983**

with a t-statistic of 3.8 and a pvalue of 0.0001. Thus, even though this sample is very short, the effect is highly statistically significant. Of course, some of this must be a fluke, since during this period we actually had favorable slippage, which is an effect that is unlikely to persist. Nonetheless, however, the large statistical significance suggests that closely monitoring our fills can give an improvement in our slippage.

Comparisons were also done for the periods following Nov. 21, 1995 (when we started randomizing our order time) and following Jan. 11, 1996 (when we decreased our order size); neither of these were significantly different from the preceding periods. This is not surprising, particularly in the latter case, as the samples are rather small.

2.3 S & P

On September 18 we moved the execution time to 15 minutes rather than 5 minutes before the close, and shortly after changed to a market order “not held”. A plot of the slippage as a function of time is shown below:



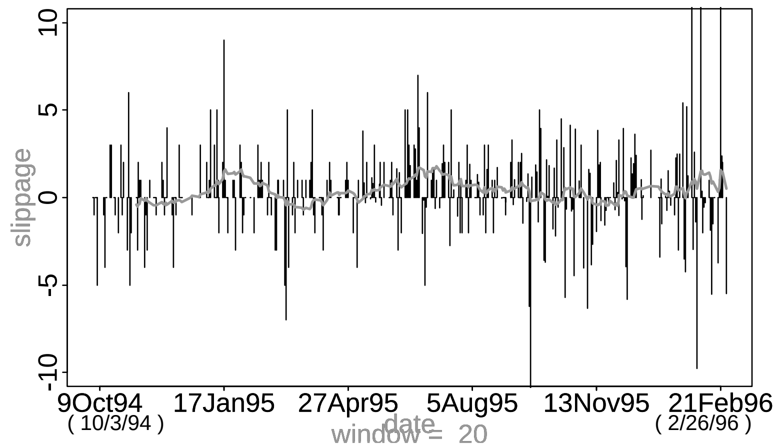
Running SlipSplitTest(slip.3, “SP”, “18Sep95”) produces

	n	mean	stdv	skew	kurt
1	**300**	**3.52**	**12.922**	**0.549**	**28.469**
2	**106**	**5.402**	**36.55**	**0.352**	**0.887**

The standard deviation of the slippage is almost three times as high as it was in the previous period. In addition, the mean slippage is worse in the period after we went the “not held” order. This difference is not statistically significant -- the t statistic is less than 0.5, but this result does call into question whether we are obtaining any benefit from the “not held” order -- it suggests that we should consider returning to the market order that we were employing earlier.

2.4 Bund

A running plot of the Bund slippage is shown below



On Sept. 18. 1995 we moved the order time from 5 to 15 minutes before the close, and changed our order to "not held". Running SlipSplitTest gives the following results:

sample	n	mean	stdv	skew	kurt
1	216	0.437	2.26	0.116	1.254
2	108	0.282	4.11	1.357	6.208

The mean since we went to "not held" is smaller. However, the t statistic is 0.4, with a p value of 0.37, so this result is not statistically significant. What is clearly significant is the increase in the standard deviation.

3. Measuring execution skill

Slippage depends both on market behavior and on execution skill. For example, if we are buying and the market is trending up, then even if the broker gets one of the best prices in the range, the slippage may be bad. Conversely, if we are buying and the market is trending down, then even if the fill is one of the worst in the range it may result in positive slippage. The purpose of the measure described in this section is to try and measure the component of slippage that depends on execution skill.

3.1 Definition of skill score

To separate market conditions from "skill" it is necessary to examine our fills relative to the market ticks occurring during the relevant period. A "skill score" for a given fill can be defined as follows:

- Append the fill price and tick prices and sort them in order. Rank the values according to their order, e.g. if the prices are (1003, 1001, 1000, 1002, 1004) the ranks are (4,2,1,3,5). For convenience, rescale the ranks so they in the range (-100, 100).

- If there are repeated prices, assign the rank to be the average of the ranks. For example, if the fill prices are (1000, 1001, 1000, 1002, 1000), the ranks are (2,4,2,5,2). This is because 1000 gets the ranks 1, 2, and 3, whose average is 2. This has the advantage of weighting each tick according to how often it occurs.
- If we are buying reverse the sign of the score.
- For multiple fills, compute the score for each fill and average the scores together, weighting them by the size of each fill.

Thus, if we buy at the lowest possible price or sell at the highest possible price we get a score of 100; if we buy at the highest possible price or sell at the lowest possible price we get a score of -100.

3.2 What tick interval?

To compute the score defined above, it is necessary to choose a time period to compare the fills and the ticks. This depends on the way in which we execute our orders and the instructions to the traders. For example, suppose we give the trader a 5 minute period to work the order; the natural point of comparison is the five minutes of ticks following the order time. The three data sets listed in Section 1.3 differ in the time intervals for the ticks. Since these data sets were extracted, I have written functions for selecting ticks within Splus. This allows us to extract the ticks from the database for a comprehensive time range and select them as needed within Splus.

Some relevant functions for computing skill scores are:

- SkillScore(fill.prices, fill.sizes, ticks, order.size) This is the basic function that computes the scores. The first three arguments are vectors giving the information relevant to a given order; order.size is the size of that order.
- SkillScores(x). Calls SkillScore repeatedly on a list of slippage data. x is the list of data for a given instrument.
- AllSkillScores(x). Calls SkillScoreList repeatedly on multiple instruments, and returns a list of scores for each day (which contain null values if either ticks or fills are missing).
- AllCorSlipScore(scores, x). Computes linear and rank correlations between skill score and actual slippage for multiple instruments. Writes results to file “cor.slip.score” that can be sucked into a table.

The following table gives the correlation between the skill score and the actual slippage for the three data sets mentioned in Section 1.3:

data set	**CL**	**DM**	**FDB**	**JY**	**SP**	**US**
1	**39**	**41**	**65**	**23**	**51**	**63**
2	**0**	**11**	**7**	**1**	**18**	**7**
3	**7**	**12**	**25**	**-1**	**20**	**18**

Data set 1, which included ticks before the order but is cut off at the last fill, has the highest correlations. Data set 2, which did not include ticks before the order but also contained some ticks much after the order, had the lowest correlations. Data set 3 was somewhere in between. Given that the main difference between these data sets is the time interval in which the ticks are extracted, this raises the question of why the correlations are so radically different: For data set 1, 50% is a typical value, while for data set 2, 10% is more typical.

One of the difficulties with this analysis is that our instructions to the traders have changed with time. Until roughly mid-95 we were using market orders and typically executing soon after the order hit the floor. After about Sept. 95 we changed our instructions to traders to be “not held” rather than market orders, and gave them 15 minutes rather than 5 minutes to work the order. Thus, one of the problems is that the interval of ticks that is most appropriate for computing the skill score during one era may not be appropriate in others.

Each interval for the ticks can be characterized by a start and stop time relative to the order time. For example, (start, stop) = (-2, 10) would correspond to an interval from 2 minutes before the order to 10 minutes after the order. The ticks can be selected using the following functions:

- SelectTicks(x, start, stop, last.fill=F, ack.time=F). x is a single instrument slippage structure. (start, stop) are in minutes, as explained above. last.fill=T means that ticks will be discarded after the last fill or the stop time, which ever comes first. ack.time=T means that ticks will be discarded before the acknowledge time or the start time, whichever comes later.

- AllSelectTicks _ function(x, start, stop, last.fill=F, ack.time=F). Runs SelectTicks on a multiple instrument list. start and stop can either be simple numbers, or arrays of length = number of instruments, i.e. length(x). Result is a slippage structure in same form, but with ticks removed.

- CorScoreVsTime(slipdat, start, stop, last.fill=F, ack.time=F). slipdat is a single instrument slippage structure. If we want to sweep start it can be an array of values while stop is a single value. Similarly, if stop is an array of values and start is a single value, stop will be swept. E.g. for the table below for the S&P, stop = c(1,2,3,5,8) and start=0. last.fill and ack.time are as for SelectTicks.

- AllCorScoreVsTime(slipdat, start, stop, last.fill=F, ack.time=F). slipdat is here a multiple instrument slippage structure. Otherwise this is like CorScoreVsTime.

I began by varying the stop time for the S&P and the T-bond. The results are shown in Table 2. The pattern is rather striking: as fewer ticks are included (as the stop time decreases), the correlation goes up. The values for a stop time = 1 minute are roughly comparable to the values for data set 1. The fact that one minute gives the strongest correlation is particularly surprising because, as we will see, based on studies of market impact it looks like our order typically takes more than one minute to arrive on the floor.

	1	2	3	5	8	15
SP	54	54	53	47	-	35
US	62	57	46	27	21	-

I then did an experiment varying the start time in data set 1, and holding the stop time constant at 15 minutes. The rank correlations are almost always higher than the linear correlations, so I used them instead. The results are shown in the next table. The fraction of ticks discarded with start=-1 was typically about 50%; with -6, only about 10% of the ticks were discarded. I was surprised that this is the case, but I suspect that it is due to the cases where the fill time was later than 15 minutes and it suggests that in these cases some aftermarket ticks were picked up.

	-1	-2	-3	-4	-5	-6
CL	34	42	45	48	50	50
DM	36	43	45	46	45	45
FDB	60	66	69	70	70	70
JY	21	26	32	33	35	35
SP	53	61	63	62	61	60
US	61	67	68	67	67	67

These results are surprising. In every case the correlations peak when ticks before the order are included, with the maximum typically occurring about 4 minutes before the order. It is unclear why one would expect this to be the case. A theory is presented at the end of the next section.

3.3 Average skill score

The average skill score for a given instrument can be computed as follows:

- slip.1.stop3 _ AllSelectTicks(slip.1, start=-10, stop=3)
- test _ AllSkillScores(slip.1.stop3)
- AllSkillMeans(test)

The result is:

	CL	DM	FDB	JY	SP	US
mean	-29	-24	-14	-20	-21	-22
std. err	2	3	3	4	3	3

The most striking aspect of this is how good the score is for the FDB. This value is 2 standard deviations better than any of the other scores.

If this analysis is redone using ticks in the interval (-2, 4) the results are:

	CL	DM	FDB	JY	SP	US
mean	**-24**	**-23**	**-18**	**-21**	**-24**	**-26**
std. err	**2**	**3**	**3**	**4**	**3**	**2**

These results contain less future data (data before the order is placed and seem more realistic. Here it is striking that the skill factor is the lowest for the US, at least three standard deviations lower than that for the FDB.

Changing the time interval so there is no future data at all (0, 4) the result become:

	CL	DM	FDB	JY	SP	US
mean	**-20**	**-19**	**-23**	**-24**	**-28**	**-29**
std. err	**2**	**3**	**3**	**4**	**2**	**2**

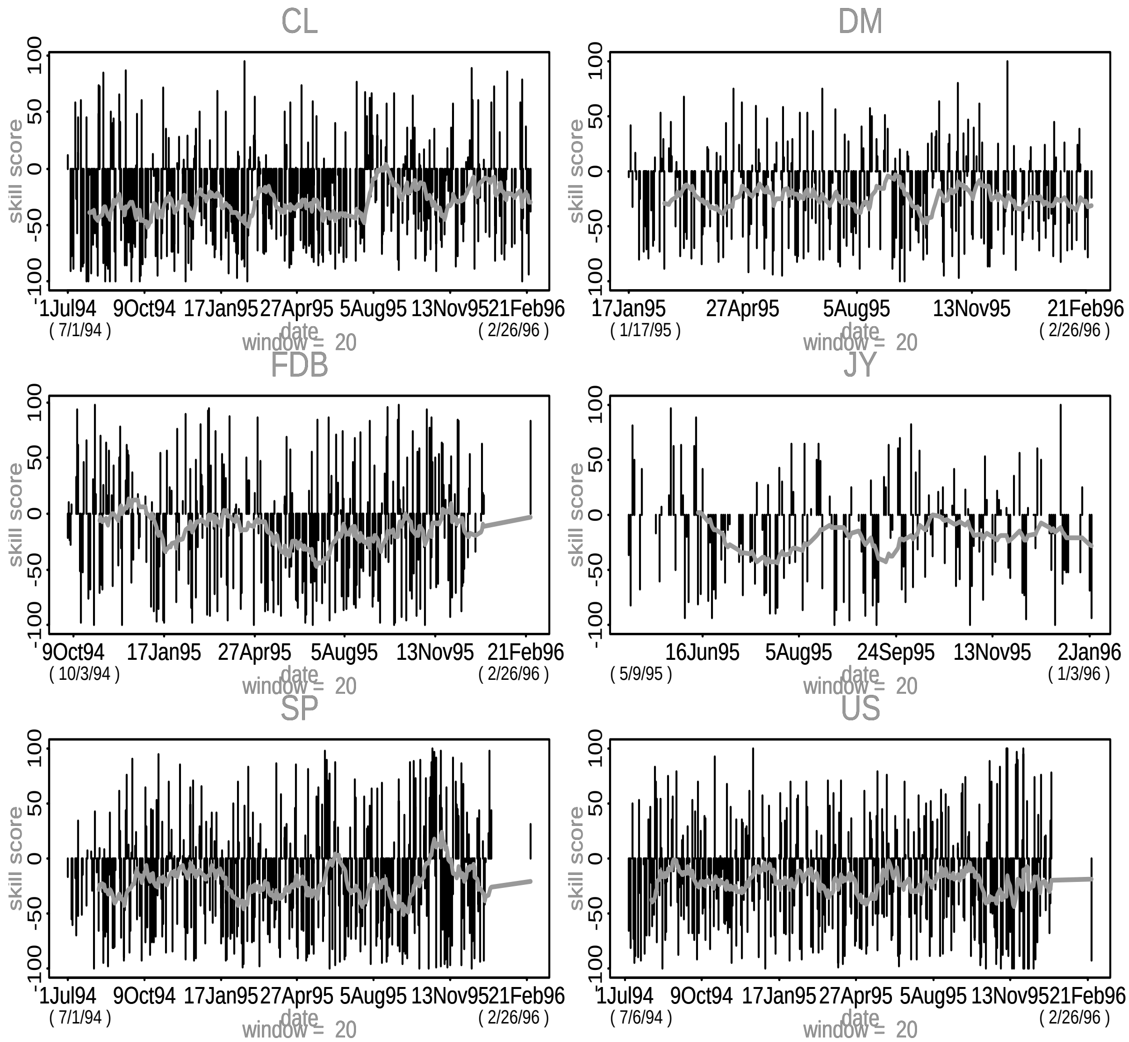
Once again, US has the worst skill score.

3.4 Variation with time

Plots showing the skill scores vs time were made using

- AllRunningSkillPlot(slip.1.stop3, mfrow=c(3,2))

Since slip.1 contains ticks 5 minutes before the order time, setting start=-10 just takes these ticks. The result is shown in the following figure:



There are some glitches in the last two months of FDB, SP, and US caused by missing ticks. A few features that are noteworthy:

- The skill level on oil seems to have generally risen since we move to a “not held” order in July 1995.
- The skill level for the SP went positive for a period after we moved to the “not held” order.
- The skill level for the US has dropped or held even since we moved to the “not held” order.

This analysis is somewhat complicated by the fact that as we change our tactics, e.g. by moving to a “not held” order, the most appropriate tick interval to measure the skill score changes. In the analysis above we are using a fixed time interval to select the ticks. This may obscure variations in time.

4. Dependence of slippage on tick behavior

Is the slippage correlated to any simple measures based on the ticks? Skill score is one such possibility. Two others are:

- **trend,** defined as (last tick price - first tick price)*sign(order). Thus, this is positive when the market is trending away from our order, e.g. buying into a rising market, and negative in the opposite case, e.g. selling into a rising market.
- **vol,** defined as the standard deviation of the ticks corresponding to a given order.

The function call

- cors.slip.3 _ AllSlipCors(slip.3)

creates a list that for each instrument gives 4x4 linear and rank correlation matrices for slippage, -skill score, trend, and vol. (The sign of skill score is flipped so that it will have positive correlations to slippage, which in this older data structure has positive sign when it is unfavorable). The linear correlations to slippage using slip.3, for example, are:

	skill	trend	vol
CL	**7**	**42**	**17**
DM	**12**	**24**	**4**
FDB	**25**	**64**	**-15**
JY	**-4**	**38**	**3**
SP	**19**	**39**	**1**
US	**18**	**38**	**3**

The factor with the largest correlation to the slippage is clearly the trend. This is not surprising; the ticks in slip.3 contain values for times after the last fill, and we expect slippage to be worse buying into a rising market or selling into a falling market. The linear correlation between slippage and skill is generally higher than the rank correlation, which suggests that the correlation is larger for large slippage values (there are more outliers for slippage than for skill). It is interesting that the skill and trend factors are also highly negatively correlated. The rank correlations are:

CL	**-57**
DM	**-60**
FDB	**-31**
JY	**-64**
SP	**-54**
US	**-45**

These are all strongly negative. The signs are confusing: -skill has positive sign when we get bad slippage. Trend has a positive sign when the market is moving away from us. This result is thus quite surprising: It says that our brokers tend to do a better job of getting good fills in comparison with other market participants when the market is trending away.

I modified AllSlipCors so that it creates summary files containing the linear and rank correlations. The results from running this on slip.1 (which had data from 5 minutes before the open to the last fill time) are:

	slip-skl	slip-trnd	slip-vol	skl-trnd	skl-vol	trnd-vol	std err
CL	**39**	**47**	**34**	**1**	**-5**	**28**	**5**
DM	**41**	**25**	**19**	**6**	**-3**	**2**	**6**
FDB	**69**	**46**	**-3**	**46**	**-8**	**-1**	**6**
JY	**31**	**45**	**2**	**-1**	**-17**	**13**	**8**
SP	**60**	**31**	**3**	**46**	**-14**	**-2**	**5**
US	**66**	**46**	**9**	**51**	**-8**	**1**	**5**

As we saw before with this data set, the skill factor has a strong correlation with the slippage. Trend also has a strong correlation. What is surprising is that in this case the correlation between skill and trend is much less reliable and in general the correlations are positive rather than negative.

To test whether the relationship observed between slippage and trend is predictive, the ticks were filtered so they contained only values before the order is placed, by running

- slip.1.fut _ AllSelectTicks(slip.1, -10, 0)

The correlations are:

	slip-skl	slip-trnd	slip-vol	skl-trnd	skl-vol	trnd-vol	std.err
CL	**70**	**-8**	**4**	**27**	**7**	**6**	**5**
DM	**59**	**-25**	**14**	**38**	**-8**	**-11**	**6**
FDB	**75**	**-10**	**-1**	**30**	**-4**	**-0**	**6**
JY	**67**	**-7**	**-10**	**39**	**-13**	**15**	**8**
SP	**67**	**-3**	**6**	**50**	**-6**	**-7**	**5**
US	**71**	**-3**	**8**	**43**	**-5**	**-3**	**5**

- The most striking result here is the large correlation between the skill measure and the slippage. These correlations are higher than they are when the later ticks are included (see the previous table). A possible explanation is that the skill measure uses future data, i.e. the fill prices. Without current data, the fill prices will tend to either be above or below the ticks. This gives information about the market conditions between when the order was originated and when it was filled. For example, if the fill is above and we are buying, then this indicates we bought into a rising market, and should expect large slippage.

- The correlation between trend and slip are now statistically insignificant except for the DM, where the correlation is negative. Thus, the trend before the order is generated seems to have little predictive value for the slippage.
- The skill-trend correlation is high. This is a puzzling, considering the low correlation between slippage and trend.
- All the other correlations are fairly small.

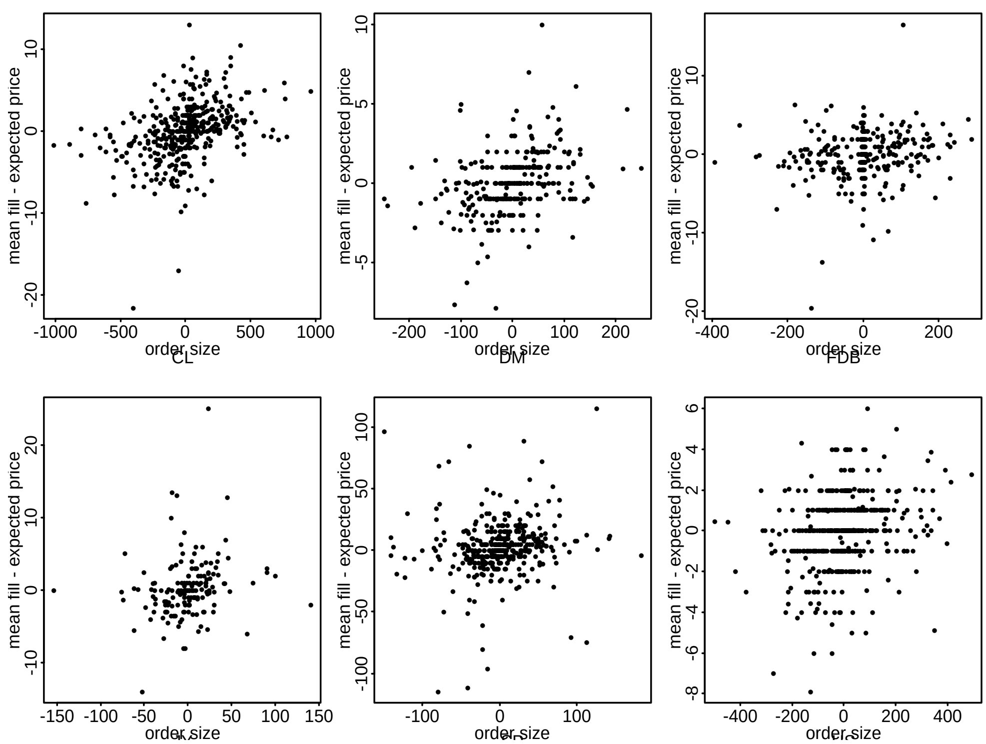
5. Market impact and dependence on order size

5.1 Scatter plots

To see whether there is any obvious dependence of slippage on order size we begin with scatter plots.

- x.3 _ AllSlipVsOrder(slip.3, slip=F)
- AllSlipScatterPlot(x.3, mfrow=c(2,3), “fill price shift vs. order”)

Figure 2. Scatter plots of fill price offset, vs. order size in dollars, for six futures



The results are pretty noisy with no obvious structure. This should be born in mind throughout the analysis that follows: Any fit to the data is not going to be very good.

5.2 Binning

To get a clearer view of whether there is any structure, we can divide the data into bins based on the order size, with roughly equal number of points in each bin. (This is done with the program ScoreVsFac.)

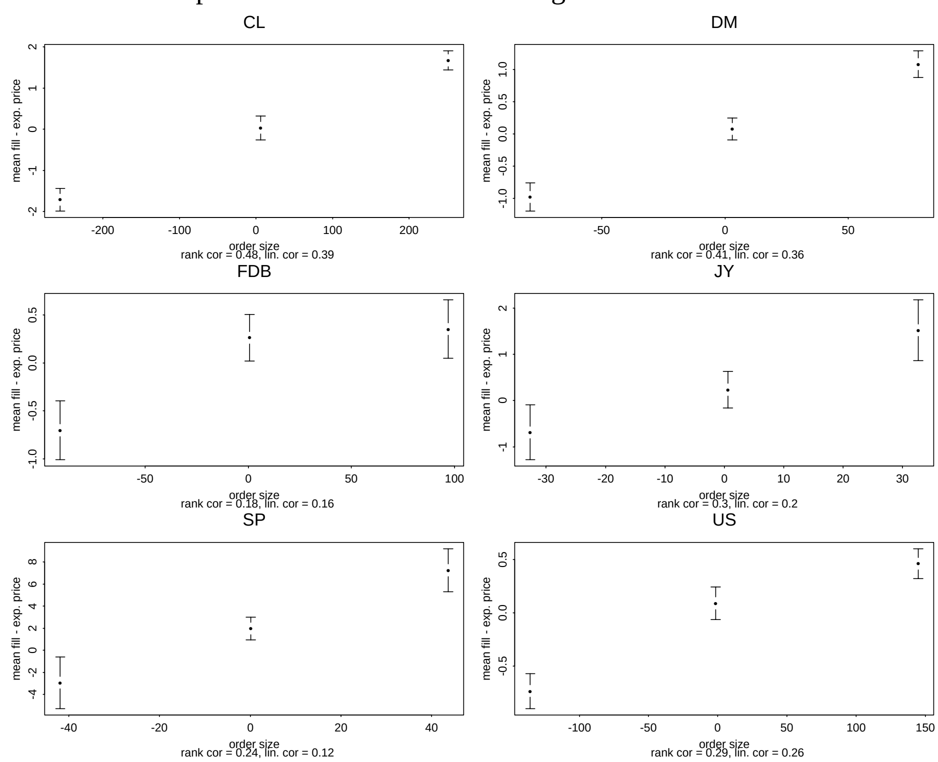
Running

- s.vs.o3 _ AllSlipVsOrder(slip.3, slip=F)
- AllPlotSlipVsOrder(s.vs.o3, mfrow=c(3,2), 3)

plots the mean fill price shift for each bin, and gives the result shown in the following figure.

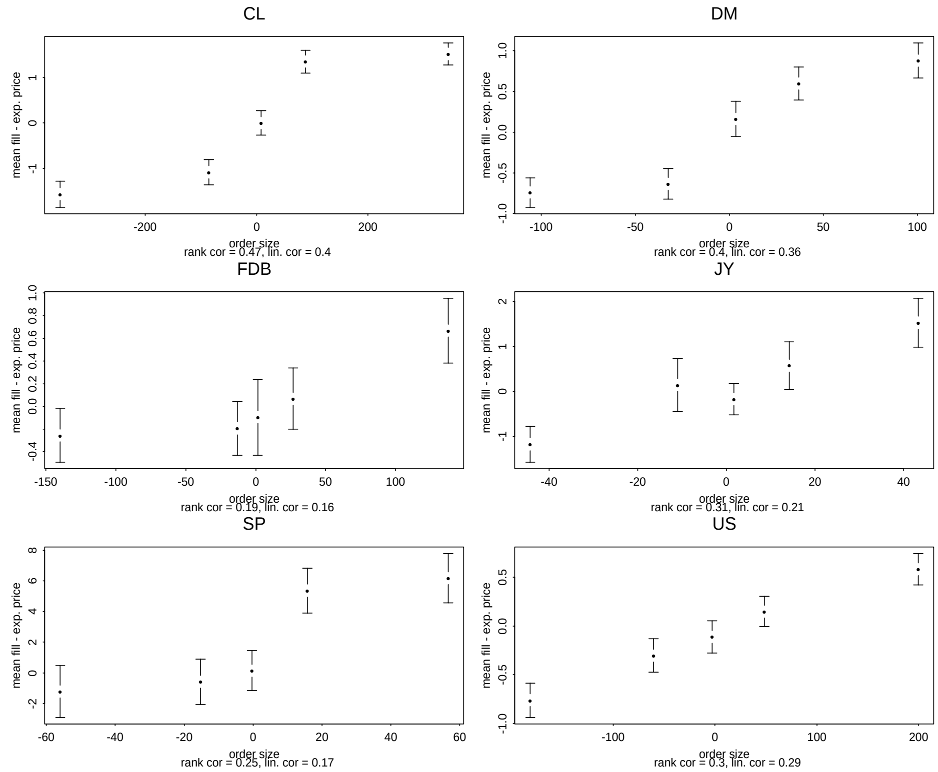
.

Figure 3. Mean of fill price shift vs. order size sorting data into three bins.



This result shows a fairly strong effect. Dividing the data into five bins shows that there are inconsistencies from the pattern seen with three bins; some of the error bars are much bigger than others. This seems to be due to outlier effects. If 5% of the outliers are removed the data becomes more consistent, as shown in the following figure.

Figure 4. Mean of fill price shift vs. order size, sorting data into five bins



5.3 Linear regression

Another view of this is obtained by performing a linear regression on each data set. This is done using:

- AllSlipVsOrderFit(s.vs.o3, intercept=T)

To reduce the dependence on tick size and contract size, which are somewhat arbitrary, the data were first scaled so that both the fill price shift and the order size are measured in dollars. This means that the slope is a non-dimensional number that can be used to characterize our slippage across different markets. The linear regression was performed both with and without a y-intercept. Since the slopes were quite similar in both cases, I give the results when the y-intercept is included.

	slope x 10^-6	**t statistic**	**intercept ($)**	**t statistic**	**R^2**	**correl.**	**rank correl**
CL	**2.65**	**8.82**	**-0.03**	**-0.02**	**0.15**	**0.39**	**0.48**
DM	**1.42**	**6.28**	**0.67**	**0.45**	**0.13**	**0.36**	**0.41**
FDB	**0.51**	**2.99**	**-0.79**	**-0.27**	**0.03**	**0.16**	**0.18**
JY	**2.06**	**2.52**	**4.39**	**1.08**	**0.04**	**0.2**	**0.3**
SP	**1.2**	**2.37**	**10.09**	**1.86**	**0.01**	**0.12**	**0.24**
US	**1.12**	**5.5**	**-2.34**	**-0.83**	**0.07**	**0.26**	**0.29**

This table shows several interesting things:

- The slopes are in a fairly narrow range, 0.51 at the low end for the Bund and 2.65 at the high end for oil (x 10^-6). As we would have expected, oil has the highest value.
- The slopes are statistically significant in every case, particularly for oil (where the p-value is effectively 0), DM, and US.
- None of the intercepts are statistically significant. The only one that approaches statistical significance is the S&P; this is not surprising, as the S&P had a large up-trend during this period. This will make the shift between the fill price and the order price be up.
- Except for oil and DM, the R^2 are generally pretty small. This emphasizes that in general the amount the linear fit "explains" the data is small.
- With the exception of the S&P and the Bund, the correlations are high. They are positive and significant in every case (the error bars vary because the lengths of the trading samples vary, but a typical value is 0.05.)
- Rank correlations are higher than the linear correlations in every case, sometimes by as much as a factor of 2. This suggests that either their are significant non-linearities, or there are outlier effects. Given the high noise level, the latter hypothesis seems more likely.

These results show a much stronger effect than the analysis that was performed in October 1995. One might think this could be due to the additional data we have accumulated since the last analysis was done. However, the more important change seems to be in the method of analysis. Previously the regression was done on slippage vs. absolute order size (in contrast above I used fill price shift vs. order size). When I redo the analysis using slippage vs. absolute order size (with intercept=T) the results are shown below:

	slope x 10^6	**t.stat**	**intercept**	**t.stat**	**R2**	**lcor**	**rcor**
CL	**1.19**	**2.91**	**10.6**	**5.06**	**0.02**	**0.14**	**0.14**
DM	**0.45**	**1.32**	**8.23**	**3.7**	**0.01**	**0.08**	**0.09**
FDB	**0.44**	**1.94**	**1.77**	**0.46**	**0.01**	**0.11**	**0.11**
JY	**1.24**	**1.09**	**5.45**	**0.97**	**0.01**	**0.09**	**0.2**
SP	**-0.24**	**-0.33**	**21.82**	**2.89**	**0**	**-0.02**	**0.05**
US	**0.44**	**1.51**	**12.79**	**3.15**	**0.01**	**0.07**	**0.06**

This gives quite a different impression. The slopes are in general smaller, and the slope for the SP is negative. Except for oil the t-statistics for the slopes are not significant. In contrast, most of the slopes for the y-intercept are significant. Also, the R2 values are all quite low -- very little of the variance is explained. The correlations are much smaller.

In contrast, the following table shows the results obtained regressing slippage vs. absolute order size, but using intercept = F:

	slope x 10^6	t.stat	R2	lcor	rcor
CL	2.65	8.83	0.15	0.14	0.14
DM	1.42	6.3	0.13	0.08	0.09
FDB	0.51	2.99	0.03	0.11	0.11
JY	2	2.43	0.04	0.09	0.2
SP	1.2	2.35	0.01	-0.02	0.05
US	1.12	5.5	0.07	0.07	0.06

The slopes, their t-statistics, and R2 values are virtually identical to the ones computed by regressing fill price shifts against orders. The fact that the R2 values are higher demonstrates that more of the variance is explained using the pure linear model. (This confuses me -- I would have thought that since this computation is in-sample adding a parameter can only decrease the mean square error, which improves R2.)

I believe there is a good explanation for this. One expects the behavior of fill price shift vs. order size to be roughly as follows:

- For orders below a critical size, there is a fixed bid-ask spread. The fill price shift approaches a positive constant for buy orders and minus this constant for sell orders. There is a discontinuity at zero corresponding to the bid-ask spread.
- For orders above the critical size, to first order the slippage grows linearly. (Of course, for sufficiently large orders we may expect nonlinearities to begin to play a role).

With binning the final result might look a bit like a sigmoid, except that unlike a sigmoid for large or small orders it approaches linear growth rather than a constant. This basic shape is suggested by the plots of figure 4 for CL, DM, and SP. When an intercept is used, and when the plot is made for slippage vs. abs(order), the constant term will capture the plateau corresponding to the fixed bid-ask spread. When no intercept is used, it will average the plateau and the linear growth region, giving a slope that is larger than the correct slope in the linear growth region (due to the discontinuity at zero order size).

Note that CL, DM and SP also have the largest discrepancy in their fits with and without an intercept.

5.4 Effect of ramping

We can use the linear fits of the previous section to determine what impact ramping will have in our slippage. For example, the results of running

- AllRampEst(slip.3, ramp.list, ramp.level=c(2,4), intercept=F, wt=F)

is shown in the next table.

	mean	model	model.adj	ramp x 2	ramp x 4
CL	**1.5**	**1**	**1.3**	**2.7**	**5.4**
DM	**0.8**	**0.6**	**0.6**	**1.1**	**2.3**
FDB	**0.4**	**0.3**	**0.6**	**1.2**	**2.4**
JY	**0.8**	**0.5**	**0.6**	**1.3**	**2.5**
SP	**4**	**1.8**	**3.4**	**6.7**	**13.4**
US	**0.5**	**0.4**	**0.6**	**1.2**	**2.3**

This is based on a regression of slippage vs. absolute value of order size, with intercept=F. The first column is just the mean slippage measured in ticks/contract. The column labeled "model" is calculated by computing the mean absolute order size for the sample, and then evaluating it using the fitted linear model. The column labeled "model.adj" is calculated by ramp adjusting the orders, computing the mean absolute value of the ramp adjusted orders, and then evaluating with the linear model. One would typically expect the mean absolute ramp adjusted order size to be larger, so that with a positive slope it should give a larger estimated slippage than the column labeled "model". However, as demonstrated by CL, it is not necessarily larger than the actual mean.

The columns labeled "ramp x 2" and "ramp x 4" give the values one would expect by ramping by a factor of 2 and 4 respectively. These values are computed by multiplying the mean absolute ramp adjusted order size by the factor and then evaluating it with the linear model. As can be seen from this table, without a constant term in the regression the effect of ramping is generally severe. This is not surprising -- with a linear model the slippage goes up proportional to the mean order size, so a ramp by a factor of 4 quadruples the slippage. As mentioned before, the linear model without a constant term is probably overly pessimistic.

The picture is somewhat different when the regressions are made including a constant term, as shown in the following table:

	mean	model	model.adj	ramp x 2	ramp x 4
CL	**1.5**	**1.5**	**1.6**	**2.1**	**3.2**
DM	**0.8**	**0.8**	**0.8**	**1**	**1.4**
FDB	**0.4**	**0.4**	**0.6**	**1.1**	**2.2**
JY	**0.8**	**0.8**	**0.8**	**1.2**	**2**
SP	**4**	**4**	**3.7**	**3**	**1.7**
US	**0.5**	**0.5**	**0.6**	**0.9**	**1.3**

Comparing this to the previous table shows that the match between the mean and the model is much better when an intercept is used. In addition, the effect of the ramp is generally much less severe. Nonetheless, except for the S&P (which has a negative slope when the intercept is used), at a ramp of 4 the effect is large enough to be a matter of concern. (Using fits weighted by order size gave similar results).

6. Market impact

By market impact, I mean the influence our orders have on the price, in particular the shift in price caused by our order. This is in general different from slippage. The prior is that curve for market impact vs. order size should be continuous, without the jump at order size = 0 for market friction. (A very small order should in general have a very small effect on the price). Slippage should be a stronger effect that market impact.

In principle market impact can be measured by removing our fills from the price ticks, and examining the remaining price ticks to see whether they show any trends that correlate with our orders. In practice this is difficult for several reasons:

- We do not have accurate fill times; typically the price ticks through a given level several times, so it is not clear which tick to remove. This may or may not be a significant problem. On one hand, if we remove the wrong tick it doesn't matter. On the other hand, it is useful to look in a restricted time range relative to the order, e.g. a few minutes after the order, because we expect to see the largest effect immediately following the order. This creates problems, because we may remove ticks from the wrong time block, leaving ticks that are actually ours, and confusing market impact with the stronger effect of friction.
- Multiple fills at the same price are condensed into one fill. Thus, one of our fills may correspond to many different ticks in the tick stream. We only remove one of them. What is left still contains several ticks corresponding to our fills, and so gives a distorted view of the market impact; it may more closely represent our market friction than our market impact.

To assess market impact, it is first necessary to attempt to remove the fill ticks. There are S+ functions available to this in two different ways:

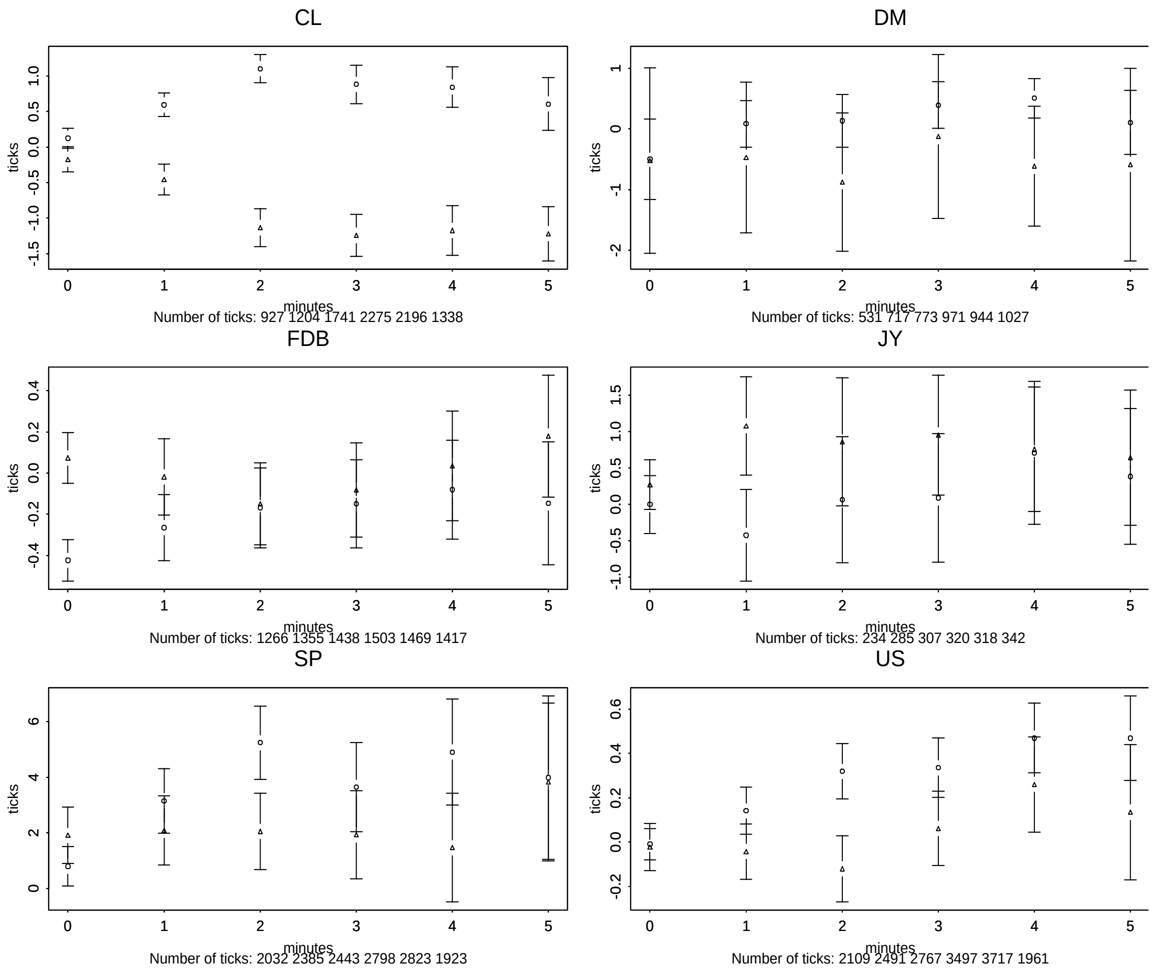
- RmFirstFillTick removes the first tick it finds in the interval of data given corresponding to the fill price. AllRmFirstFillTick does this for a list of results for different instruments.
- RmTicksInMinute removes ticks from the minute that you specify. Minutes are labeled 0, 1, 2,... (for 0-1 minutes from the order time, etc.). AllRmTicksInMinute does this for a list for different instruments.

The effect of slippage can then be assessed by taking the mean price move for the ticks in each minute before or after the order. Price movements can be separated based on the sign of the order -- orders of positive sign are lumped into one group, and orders of negative sign into another group. This is done by the program PricePush. Plots can be made either with PlotPricePush (for a single instrument) or AllPlotPricePush for multiple instruments.

The effect on the price is seen in figure 5. The first fill ticks have been removed. Mean price movements for buy orders are shown with octagons; mean price movements for sell orders are shown with triangles. This plot suggests that there is a large impact for oil, with smaller impacts for S&P and US. The price impacts for the other instruments are never statistically significant. In general the largest impact is seen in minute 2. This may not be surprising; typically 1-3 minutes elapse before the orders reach the floor; this corresponds to minutes 0-2 on these plots. It is interesting to note that there are several cases where the market impact appears to go in the opposite direction of what one would expect, e.g the Bund and the JY.

Plots made removing no ticks at all show similar results. Not surprisingly, the main effect is in minute 0; removing the fill ticks tends to decrease the apparent market impact.

Figure 5. Plots showing the impact of orders on price, with first fill ticks removed

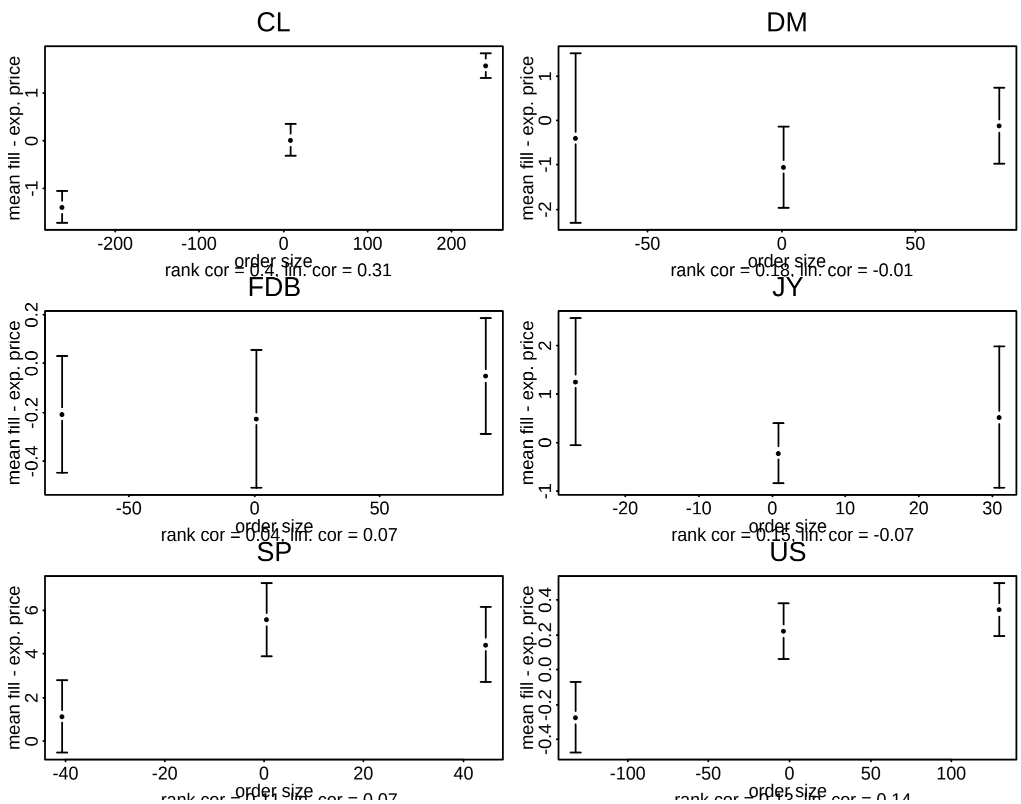


As with slippage, the dependence on order size can be assessed by binning the data based on the order size and computing the mean slippage for each bin. This is done by

- push.3 _ AllSlipVsOrder(slip.3, minute=2)
- AllPlotSlipVsOrder(push.3, mfrow=c(3,2), ncats=3)

The results are shown in figure 6.

Figure 6. Market impact vs. order size for minute 2. Vertical axis shows mean tick price.



The only two cases showing a significant market impact are CL and US. It is not surprising that we see an effect for CL. In this case we know that the broker fills our order in small fills, many of which are at the same price. Thus, for CL the problem mentioned at the beginning of this section are quite strong -- I suspect that the market impact computed for CL shows substantial corruption by our own ticks. It is difficult to say how much is market impact and how much is just friction. Making a plot using 5 bins shows a result that is almost indistinguishable from figure 4.

For US this is more surprising. It is suspicious that in figure 5 the price impact for oil persists across the whole period, whereas the price impact for US comes within the error bars by minute 3. (Note that for oil filling our order often takes more than 10 minutes). Our order is typically filled in just a few fills (it would be good to compute the mean number of fills since the February ramp, to make this more quantitative.) It is certainly possible that multiple fills cause contamination.

A linear regression can be performed on the market impact data as follows:

- push.3 _ AllSlipVsOrder(slip.3, minute=2, slip=F, tick.size=tick.size, contract.size=contract.size)
- AllSlipVsOrderFit(push.3, intercept=T)

The list push.3 was rebuilt so that the price shift and orders are both in units of dollars, to make the slopes comparable to those computed for market friction. This produces output in the file “stats.slip”. The results are shown in the following table:

	slope	t.stat	intercept	t.stat	R2	lcor	rcor
CL	**2.12**	**5.82**	**0.81**	**0.44**	**0.1**	**0.31**	**0.4**
DM	**-0.11**	**-0.07**	**-6.54**	**-0.68**	**0**	**-0.01**	**0.18**
FDB	**0.18**	**1.15**	**-2.89**	**-1.14**	**0**	**0.07**	**0.04**
JY	**-1.37**	**-0.68**	**6.97**	**0.81**	**0**	**-0.07**	**0.15**
SP	**0.63**	**1.39**	**18.4**	**3.76**	**0.01**	**0.07**	**0.11**
US	**0.65**	**2.66**	**3.15**	**0.99**	**0.02**	**0.14**	**0.13**

Running the regression with intercept=F produces similar slopes. The only significant slopes are for CL and US.

Given the problems involved in removing our fills from the tick stream, I do not believe we have conclusive evidence of market impact in any of the cases above. The most interesting result is that for four of our models, making an effort to remove our fills gives a null result (in contrast to the strong results seen with a similar analysis of fill price shift).

7. Conclusions

- Our results on slippage are now statistically significant in every case, with t-statistics ranging from 2.1 for the bund to 9.8 for oil.
- There are several indications that since June 1995 our brokers have been doing a poor job in filling the T-bond. Some of them are:
  - High slippage in dollar terms in comparison with SP and FDB. Measured in non-dimensional units, it is more than twice as high as the SP and 4.5 times as high as the FDB.
  - Low value of volatility/spread: Again the SP is more than twice US, and FDB is 2.6 times US.
  - Large value of mean/standard deviation. More than twice SP, 2.5 times FDB.
  - Large and statistically significant change in slippage beginning June 1995. The slippage since June 1995 is roughly 2.5 times that prior to June 1995, a 2.6 standard deviation result. The odds of this occurring at random are 1 in 100. Prior to June 1995 the slippage on the US was comparable to that on the other instruments.
  - Worst skill score when more realistic (non-future) ticks are used.

- There is evidence that close monitoring of the brokers gives improved slippage. During the 5 weeks where we closely monitored our oil fills our slippage actually went in our favor, a results with a t-statistic of 3.8. The odds of this occurring at random are 1 in 10,000.
- The measure of skill correlates strongly with our actual slippage as long as the right interval of ticks is used. Using ticks after the order time degrades the correlation. It is puzzling that the highest correlation is obtained when the ticks are included from before the order is made.
- Trend across the tick interval also correlates well, but only if it is contemporaneous to the period of the fill. If only ticks prior to the order are used to compute the trend, the correlation is small. There is also some correlation to volatility, but the correlation is low.
- The fill price shift vs. the order size shows a strong correlation, with rank correlation values ranging from 48% for oil to 18% for the bund. Since the standard error is roughly 5%, all of these are highly statistically significant. Linear regression gives highly statistically significant slopes, and the results are quite similar with and without inclusion of a constant term in the regression.
- To estimate expected increase in slippage it is more convenient to examine slippage vs. absolute order size. Regression gives similar results to those above when the constant term is omitted. The slopes are much smaller when it is included; several of the constant terms are statistically significant. My belief is that the inclusion of the constant term is a better model of the slippage, although I do not understand why the R^2 values are much smaller. Even in this case the model suggests that we will see a mean increase in slippage of 30% for a factor of 2 ramp and 100% for a factor of 4 ramp.
- The only clear suggestions of market impact are in oil and the T-bond. In the other cases any dependence of price shift on order size disappears when our fills are removed. For both oil and the T-bond there are concerns that the price shifts that are seen may be caused by lack of total removal of our fills from the tick stream.