Some Puzzling Discrepancies between Fed and CRSP data for the US Corporate Sector

Stephen Wright*

December 2, 2014

1 Introduction

This note describes some puzzling inconsistencies between data as published in the Fed's financial accounts, and underlying figures for quoted companies, as produced by CRSP.[1]

2 Recovering implied dollar figures from CRSP data

CRSP data relate only to quoted stocks, using data from all the major exchanges, since 1925

CRSP calculate returns at the security level, dealing, as far as I can tell, very carefully with share splits etc. They calculate both price return and return including dividend

They then aggregate returns using market weights from the previous period (CRSP guide page 97). By standard aggregation results this should be equivalent to the return on the entire portfolio, hence on the entire market.

CRSP also report market value data. They do not appear to attempt any reconciliation but with correct market weighting there should be an equivalence to the market calculation, for a given universe of companies, as the following formulae show.

The capital appreciation index is constructed, effectively as a chain-weighted Laspeyres index, as

$$\begin{aligned} \frac{P_t}{P_{t-1}} &= \sum_i \frac{p_{it-1}q_{it-1}}{MV_{t-1}} \frac{p_{it}}{p_{it-1}} = \sum_i \frac{p_{it}q_{it-1}}{MV_{t-1}} \\ &= \sum_i \frac{p_{it}q_{it} - p_{it}\left(q_{it} - q_{it-1}\right)}{MV_{t-1}} = \frac{MV_t - NI_t}{MV_{t-1}} \end{aligned} \tag{1}$$

*Birkbeck College, University of London, s.wright@bbk.ac.uk

[1] All data sources and definitions are provided in the Appendix.

where p_{it} and q_{it} are prices (adjusted for share splits etc) and quantities of individual shares, and MV_t is market value. Hence we can derive aggregate new issues as

$$NI_t = MV_t - \frac{P_t}{P_{t-1}} MV_{t-1} \tag{2}$$

The total return is constructed, using the Miller Modiglani timing convention (which appears to be what CRSP do although this is a little opaque)

$$\begin{aligned} 1 + R_t &= \sum_i \frac{p_{it-1} q_{it-1}}{MV_{t-1}} \left(\frac{p_{it} + d_{it}/q_{it-1}}{p_{it-1}} \right) \\ &= \sum_i \frac{p_{it} q_{it-1}}{MV_{t-1}} + \frac{\sum_i d_{it}}{MV_{t-1}} \\ &= \frac{P_t}{P_{t-1}} + \frac{D_t}{MV_{t-1}} = \frac{MV_t - NI_t + D_t}{MV_{t-1}} \end{aligned} \tag{3}$$

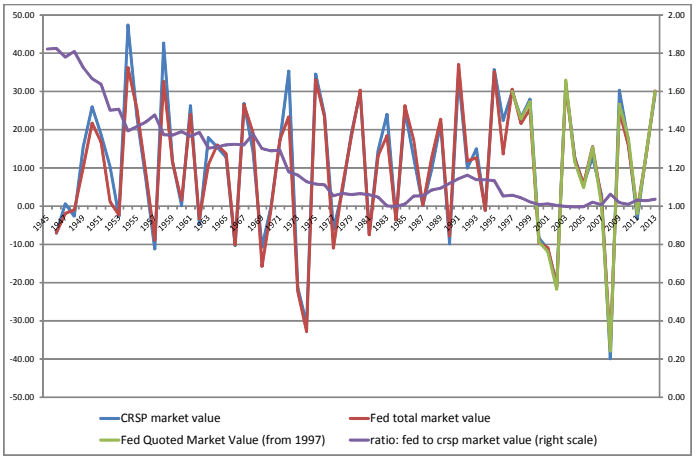
hence we should be able to reconstruct aggregate dividends as

$$D_t = \left(R_t - \frac{P_t}{P_{t-1}} - 1 \right) MV_{t-1} \tag{4}$$

3 Fed Data

The Fed flow of funds accounts take market value data, with some modifications, from CRSP. The primary modifications are a) for cross-holdings, and b) for unquoted stocks (around 14% of the net total in recent years).

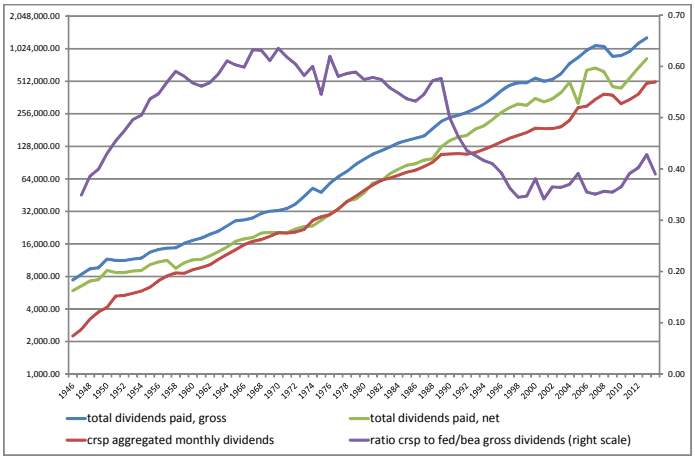
The net effect of these adjustments in recent years appears to more or less cancel out, such that the market value of equities outstanding in the Fed's figures tracks both the level and changes in the CRSP series very closely:



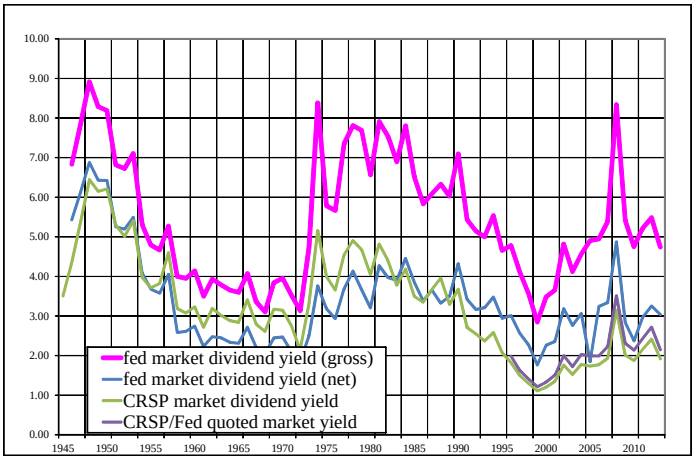
The Fed however independently measure both new issues and dividends (the latter coming from the BEA). Both these series differ significantly from the CRSP series.

4 Dividends

Since CRSP takes no account of cross-holdings we would expect that the appropriate comparator would be the Fed/BEA gross dividend series. In terms of % changes and long-term trend growth (the chart below has a log scale) this is the case (the correlation with gross dividend growth is 0.73, whereas with the net figure is only 0.48) and the share of CRSP dividends (calculated using (4) on a monthly basis, aggregated to produce an annual flow) in the total has been stable in recent years, however in level terms the CRSP figure seems surprisingly low:



If we then construct comparable estimates of "market dividend yields" (ie, the ratio of dividends paid within the year to end-year market value[2]), we get very different answers indeed:



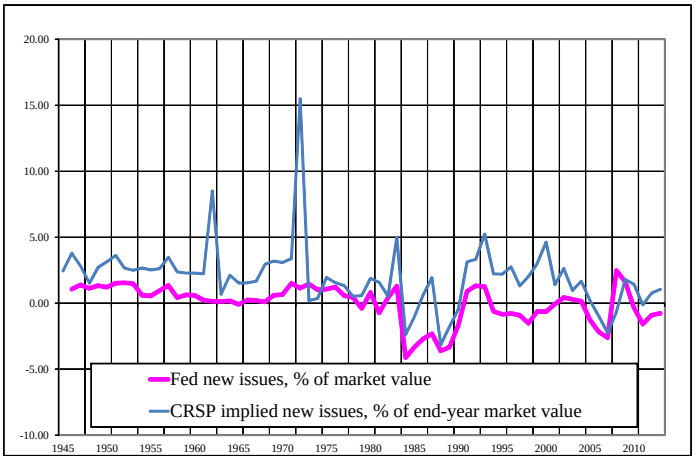
[2] Opinions seem to differ on whether the dividend yield should be measured relative to value at the end of the period, as here, or at the start of the period, in which case it is the same as the income return, but is then contaminated by inflation rates.

Over the common sample the ratio of CRSP dividends to the Fed's estimate of quoted market value looks, unsurprisingly, very like the same ratio constructed with only CRSP data, but as soon as we use Fed/BEA dividend data the figures look very different. The implied gross dividend yield is very much higher, and even the net dividend yield is higher in the last couple of decades. And of course, if we believe the CRSP dividend figures for quoted companies, the implied gross yield for unquoted companies must be fantastically high - over 20% in recent years.

I have asked the BEA if they have any breakdown of dividend data into quoted and unquoted companies; apparently they do not. This seems surprising: presumably there must be some database somewhere with a full list of companies and dividends paid.

5 New Issues

The gap between new issues as measured by the Fed and the implied figure from CRSP (using (2)) is also very significant. Fed figures imply net equity purchases by the corporate sector in most recent years (with more or less continuous equity purchases by nonfinancials offset somewhat by equity issues by financials).



The Fed figures would imply that, from (1) market value should be rising less rapidly than price, whereas the reverse is the case in CRSP data.

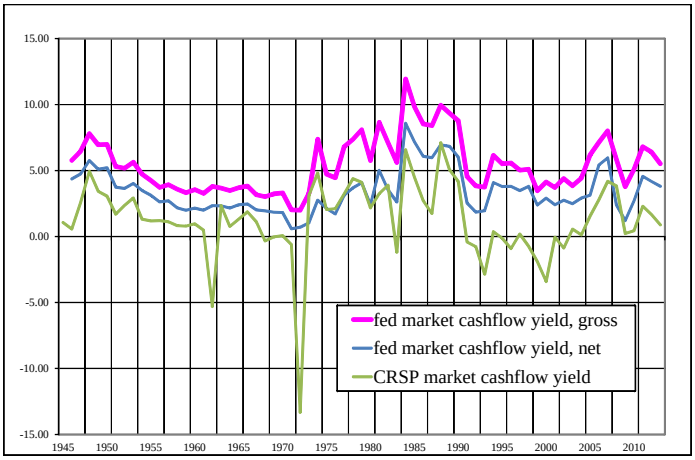
In the Fed dataset stock flow consistency is notionally maintained by a "revaluation" figure. I understand that this is a misnomer, and that implicitly this figure also includes mis-measured flows. The CRSP figure, for example, will include the face value of IPOs, while the Fed will only record that part of the issue that is actually bought.

6 Cashflow Yields and Returns

We can re-write the definition of the return in (3) as:

$$1 + R_t = \frac{MV_t}{MV_{t-1}} \left(1 + \frac{D_t - NI_t}{MV_t}\right) = \frac{MV_t}{MV_{t-1}} (1 + CY_t)$$

where CY_t, the "Cashflow Yield" captures the Miller-Modigliani consistent net cashflow out of the corporate sector. This decomposition is helpful since, as shown above, the two datasets have virtually identical growth rates of market value in recent years; however as shown above dividends and new issues data are so different that the cashflow yields, as shown below, and hence implied return figures, if we believed them, would be very significantly higher. (Note that on the basis of CRSP data we would conclude that the US corporate sector has frequently generated a negative net cashflow).



7 Conclusions

It seems inescapable that, unless I have made some major error in calculations, there must be something wrong with one, or both of the datasets, for them to have such radically different implications. If we assume that returns on quoted securities are reasonably well measured, then it seems highly unlikely, on the face of it, that returns earned by investors in unquoted stocks could be higher by anywhere near enough to explain the apparent discrepancy. The discrepancies in both dividend and new issues data would appear worthy of scrutiny; and in the absence of a persuasive reconciliation of these figures with the market value data it is hard to know how confident to be about the overall picture conveyed by any of these series.

Possible issues that might be relevant are:

- The CRSP market cap data include REITs and closed end funds (see CRSP Data Descriptions p1) as well as a quite long list of other securities

that would appear conceptually to be derivatives, rather than true equities. This appears not to contaminate CRSP return series relative to eg S&P returns, but this is perhaps unsurprising given that returns on these other securities will be expected to be so closely aligned. But to the extent that market value has been growing due to issues of these securities the implied new issue data may well include spurious flows.

- IPOs may also introduce distortions, by two routes. First, they will raise market cap, but conceptually leave the total market value of the corporate sector unchanged, since the represent a transfer of market value from the quoted to unquoted sector. Unless this is allowed for systematically (which it does not appear to be - since the figure for unquoted equities moves quite closely with the total) it will overstate growth of market value. Second, at the time of an IPO market cap appears to rise in line with the total value of all stocks in the relevant company, rather than the number actually sold. But proceeds of aggregate IPOs appear to be quite small[3], averaging only around 1/4% of market cap in recent years, so it is hard to know whether this can be the explanation unless the second factor is very significant.

[3]IPO data are available from Jay Ritter, University of Florida, http://bear.warrington.ufl.edu/ritter/ipodata.htm

Data Sources

A CRSP Data

Capital appreciation index P_t satisfies $P_t/P_{t-1}-1$ =VWRETX, Value-Weighted Return-excl. dividends

Total Return, R_t =VWRETD, Value-Weighted Return-incl. dividends

Market Value MV_t = TOTVAL, Total Market Value

All series are measured on an end-month basis, monthly flows are cumulated to derive annual flows

B Fed Data

Market Value of Corporate Equities = FL103164103.A+FL793164105.A

New Issues =FA103164103.A+FA793164105.A

Gross dividends=FA106121001.A+FA796121001.A

Net Dividends=Gross Dividends- FA106121101.A-FA796121101.A

Unquoted Equities=+ FL103164123.A+fl793164123.A