Research Guide for Datastream and Worldscope at Wharton Research Data Services

Rui Dai and Qingyi (Freda) Drechsler*

Current Version: August 19, 2021

*The views expressed in this guide are personal of the authors and do not necessarily reflect those of WRDS. All errors are our own. Rui Dai, and Qingyi (Freda) Drechsler are from Wharton Research Data Services (WRDS), University of Pennsylvania. Contact information: Dai: rdai@wharton.upenn.edu, and Drechsler: qsong@wharton.upenn.edu.

Research Guide for Datastream and Worldscope at Wharton Research Data Services

ABSTRACT

The integration of Datastream and Worldscope databases at the Wharton Research Data Services (WRDS) offers researchers a simple solution to bypass the cumbersome data collection and preparation process for global empirical research. This study guide proposes a clean approach to link these two databases through Refinitiv's Quantitative Mapping facilities. We then discuss various details about constructing stock returns through Datastream Return Index (RI). Finally, we demonstrate through an empirical case study on how to combine Datastream's pricing data and Worldscope's fundamental data to compare market characteristics across countries.

Keywords: Keywords: International Finance, Mapping, Datastream, Worldscope
JEL Classification Number: C89, G15

1 Introduction

Datastream by Refinitiv provides comprehensive historical pricing and economics data for over 65 years across 170+ countries globally. It is the most widely used database for research on international asset pricing. Based on the textual analysis result of Dai et al. (2021), 5,707 out of 53,451 SSRN Financial Economics Networks (FEN) papers mention Datastream database as a data source. More specifically, researchers from the field of finance and accounting turn to Datastream for pricing data on global equity and equity indices, macroeconomics data as well as bond and bond indices (bond related data currently not on WRDS platform). If CRSP is considered golden standard among academia for North America equity pricing data, Datastream would be arguably its counterpart in the global market arena.

For researchers that conduct empirical study on global markets, it's also crucial to have access to quality fundamental database, such as Worldscope. Worldscope provides survivorship bias-free fundamental financial data on public companies globally, covering both developed and emerging markets, and representing approximately 95% of global market capitalization. Its historical data started in 1980, though significant data coverage started from 1985 onwards. Using the same sample of study mentioned above, Worldscope is mentioned by 1,482 papers on SSRN FEN as source for fundamental data, though researchers often refer to Worldscope as Datastream fundamental data.

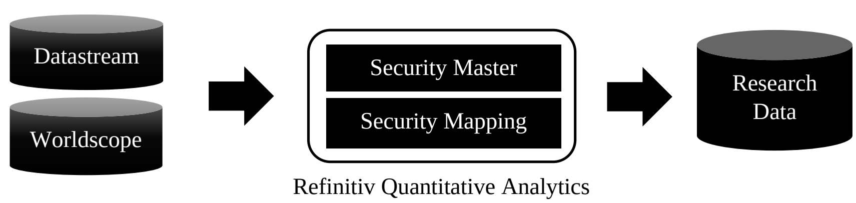
One reason for the popularity of Datastream and Worldscope among academic researchers is the many built-in features on Datastream terminals, allowing researchers to obtain pre-mapped pricing and fundamental data together. The downside of terminal data retrieval is the size limit: researchers have faced a long-lasting challenge of collecting sizeable global panel data from Datastream terminals, as the terminals are not designed for large data retrieval. The integration of Datastream and Worldscope data on WRDS platform offers researchers an optimal solution to bypass the cumbersome data collection and preparation process. As a result, it allows researchers to focus on the empirical tests using global pricing and fundamental data the same way they handle most U.S. data.

Although Datastream and Worldscope are both under the umbrella of the Refinitiv data packages, the two databases actually have different underlying data structure and identifiers. Due to its nature of international coverage, Datastream often needs additional cleaning to handle various data issues such as asynchronous trading holidays across different countries, inconsistency between return index (RI) and prices and etc. Besides, linking these two data sources together requires careful execution and supplementary data sources such as Quantitative Analytics (hereafter, QA) linking facilities. This guide serves the purpose as a white paper detailing on how to clean, link and prepare Datastream and Worldscope data for empirical study.

The rest of the paper is organized as the followings: Section 2 describes Refinitiv QA tools and how to use them to map various databases; Section 3 discusses Datastream's pricing data as well as some critical data issues; Section 4 illustrates how to merge Datastream's pricing data with the Worldscope's fundamental data through a case study; and conclusions are provided in Section 5.

2 Refinitiv Quantitative Analytics Tools

Quantitative Analytics is a data platform hosting a wide range of content from both Refinitiv and third-party data vendors along with tools to integrate data from various sources. Two useful QA tools are the *security master table*, carrying public identifiers such as CUSIP and SEDOL, and *security mapping table*, containing proprietary QA identifiers referring to entities from different data sources hosted by QA. While Datastream and Worldscope each carries a built-in identification linking table, it is our understanding through communication with Refinitiv that using QA's *security master table* and *security mapping table* provides a superior outcome when linking Datastream and Worldscope. This is due to the fact that QA's security master and mapping tables automatically align all types of identification changes among different data products.[1]



Both Datastream and Worldscope on WRDS are pulled from the QA platform. When a new database is integrated into QA, QA platform redesigns a proprietary identifier system and organizes the data pieces accordingly to facilitate seamless data integration of the new database with the existing data sources. The identifiers assigned to Datastream and Worldscope are the following:

- **Infocode**: the QA primary mapping code across all Datastream tables
- **Code**: the QA primary mapping code across all Worldscope tables

These QA proprietary identifiers are then mapped to the native permanent identifiers of both databases:

- **DsCode**: Datastream permanent identifier
- **Item6105**: Worldscope Permanent ID (referred to as **WsSctyPPI** in Datastream)

To integrate the QA proprietary identifiers and database's native identifiers, QA provides a one-to-one mapping relationship. For the case of Datastream, the QA's Datastream primary identifier

[1]This is based on the information provided in an email correspondence with Refinitiv (Ticket#: 09796827).

InfoCode is one-to-one mapped to Datastream's native identifier *DsCode*. Same exercise is done on the Worldscope's front, where QA's Worldscope primary identifier *Code* is one-to-one mapped to Worldscope's primary identifier *Item6105*. The result of this identifier translation is that *InfoCode* and *Code* can be used as unique and permanent IDs for Datastream and Worldscope, instead of their native identifiers.

Previous international finance research has focused on the country level differences manifested in stock prices and operational performance of firms (such as Griffin, et al., 2010 and Bartram, et al., 2012). One standard data preparation of this strand of research is to limit the universe of firms to those with common stocks primarily listed in their domicile/major operating countries. This treatment would essentially eliminate the following security types: preferred stocks, warrants, units or investment trusts, secondary shares, some ADRs, GDRs and other types of cross-listings. In Section 3 of this paper we provide a sample code that demonstrates a relatively strict implementation to prepare a historical linking table between Datastream and Worldscope through QA mapping tables.[2] Researchers are welcome to modify the code to generate their own link table that can handle cross-listing firms or non-common equity shares.

Using a Refinitiv datafeed on May 1, 2010, the QA mapping tableallows us to generate 65,449 pairwise historical mappings for 63,878 unique primary stocks.[3] In contrast, a linking table built through mapping *WsSctyPPI* in Datastream and *Item6105* in Worldscope would generate only 50,072 pairwise header mapping, a 20% reduction from the QA mapping table. The lower mapping number is caused by missing historical records and effective date ranges. Although most pairwise mapping is also one-to-one mapping, the QA mapping tools allow researchers to trace specific identifier changes through time. Table 1 provides two examples for demonstration.

Table 1: A Demonstration of Historical Mapping

SecCode	Typ	ID	Code	CodeRank	InfoCode	InfoRank	StartDate	EndDate	Currency	ISO3
10987445	1	60936L10	108688	1	170026	3	20-Oct-08	5-Feb-13	CAD	CAN
10987445	1	60936L10	108688	1	265023	2	6-Feb-13	6-Jun-79	CAD	CAN
607706	6	@AACLH1	105153	1	230405	2	9-Apr-10	26-Nov-13	AUD	AUS
607706	6	@AACLH1	105153	1	230405	1	8-Dec-13	6-Jun-79	AUD	AUS

In this table, *ID* variable is a unique QA security identifier, equivalent to the combination of *SecCode* and *Typ*, where *Typ* is an indicator for whether a stock is listed in North America (Typ=1) or the rest of the world (Typ=6).[4] The variable *SecCode* is also a unique identifier in either North America or the rest of the world, suggesting *SecCode* and *Typ* together form a unique identifier

[2]One caveat in this code is that, even though the mapping table tracks historical records through security identifiers, such as CUSIP and SEDOL, the country and currency information in both Datastream and Worldscope is largely most recent header record, which may lead to certain undesirable elimination of firms before they changed country of listing or domicile.

[3]For coverage comparison purpose, this QA mapping table contains 10,949 primary shares in U.S. market, in comparison to the CRSP and Compustat merged databases (CCM), which contains 21,822, 20,139, and 9,764 primary common shares since 1962, 1982, and 2002 respectively. This suggests that Datastream, Worldscope and QA mapping tables provide a comparable coverage of the US equities to that of CRSP and Compustat no earlier than 2002.

[4]In QA platform, North America refers US and Canada, which is similar to Compustat North America.

in the combined universe. Variables *Code* and *InfoCode* are the previously mentioned primary identifiers for Datastream and Worldscope respectively. The rank variables, which are renamed as *CodeRank* for *Code* and *InfoRank* for *InfoCode*, indicate the precedence of a *Code* or *InfoCode* mapped to *ID*. A lower rank value indicates the mapping between *Code* (or *InfoCode*) and the *ID* variable is valid in a more recent period: for example, InfoRank=1 or CodeRank=1 in Table 1 indicates the most recent mapped pair. Within North America markets, the cross-listing stocks are always assigned with lower Rank for their US shares. As we explicitly restrict the country variable from Datastream and Worldscope to be the same, we would drop the *InfoCode* with Rank=1, i.e., a company's US shares when the Worldscope suggests it is Canada-based (e.g., the second row in above example). *StartDate* and *EndDate* form an effective date range for an identifier linked to *ID*. However, this effective range is only applicable to Datastream but not Worldscope in our case.[5]

Datastream and Worldscope mapping table generated by the sample code in Appendix 1 is available through a web query under WRDS Linking Suite. In addition, in the sample codes of next two sections (in Appendix 2 and 3), we provide some case studies for utilizing this mapping table and more data details about Datastream and Worldscope data from the QA platform.

3 Datastream Return and Price Data

The primary application of Datastream data in empirical finance research is to obtain security returns. There are two ways to calculate returns using Datastream data. The first option is through incorporating closing price, dividends, and cumulative adjusted factors that reflect stock dividends, splits, spin-offs, and other shares outstanding related corporate events. The second approach relies on the return index (RI), which reflects a theoretical growth in the value of an equity held over a specified period, assuming that dividends are reinvested to purchase additional equity units at the closing price on the distribution date. For a few reasons discussed in this article, such as unknown dividend payment schedules and varying institutional details of share-related corporate events across countries, we focus on the RI approach and leave the other approach to the advanced researchers for further exploration.[6]

When gross cash dividend and ex-dividend date are available, Datastream calculate RI_t as following:

$$RI_t = RI_{t-1} \times (P_t^{Adj} + D_t)/P_{t-1}^{Adj},$$

where P_t^{Adj} is split adjusted stock price at day t, D_t is the amount of gross cash dividend per share if day t is an ex-dividend day, zero otherwise. RI_t is set to be 100 (%) at the base date for a stock, the date when the price information of the stock becomes first available in Datastream. The gross

[5] See *QA Direct Training Guide and FAQs Version 1.1*, Issued June 2019.

[6] Following Ince and Porter (2006), most of recent researches have treated returns generated through changes of return index RI as the default return of Datastream.

return of a stock over a holding period can be calculated as

$$Ret_{t,t+n} = (RI_{t+n}/RI_t) - 1,$$

where n is number of calendar days in a holding period.

The information on dividend payment details received by Datastream is limited for US and Canadian firms before 1973 and firms from the rest of the world before 1988. In addition, Datastream cannot recover gross dividends when the dividend payment data it receives contains a mixture of net and gross dividends due to complex tax disclosure requirements for dividends in certain countries (e.g., Australia and Hong Kong.) To circumvent these technical challenges and provide a fair return estimation, Datastream adopts an incremental approach in which annualized dividend yield is allocated into stock price changes uniformly across 260 trading/working days of a year. In particular, RI_t in such cases is defined as

$$RI_t = RI_{t-1} \times (P_t^{Adj} + D^{Ann}/260)/P_{t-1}^{Adj},$$

where D^{Ann} is cash dividend per share over a year. In this approach, Datastream ignores market holidays other than weekends and assumes 260 trading/working days a year.[7]

While this "incremental dividend approach" appears to be the best way for return estimate, it is worth mentioning two drawbacks associated with this approach. This estimated return would not lead to price change at the ex-dividend dates due to dividend payment, ceteris paribus, causing the returns based on RI outperforming the real return with dividend over short period of time. Furthermore, this incremental treatment would mechanically cause a small monotonic increase in returns over the days in which a stock is not traded.

3.1 Trading Holidays

Trading days are an essential element for the research related to information releases, regulatory shocks, and other corporate events since they allow researchers to precisely measure the market reactions and calibrate the trading behavior of various market participants as a result of different events. However, due to the complicated nature of cross-country or cross-exchange of global stock pricing data, Datastream does not offer the trading day information at the stock exchange level. It has been a challenge for researchers to handle trading days with Datastream data. Many researchers construct their filters to exclude non-trading days. For example, Karolyi et al. (2012) define non-trading days as periods when 90% or more of the stocks listed on a given exchange have a zero return. However, as discussed above, it is well known that Datastream contains a lot of non-zero returns on non-trading days, especially in the early years. Furthermore, it is also a well-known fact that Datastream contains a lot of stale prices for the stocks listed in illiquid markets. Both

[7] For more detail information, please see the official definition of RI at: Datastream Official Support

circumstances create additional difficulty for researchers to back out trading days from market-wide stock price movement.

Presumably due to a long-lasting demand from practitioners and academia, Refinitiv provides an official list of trading holidays at exchange level. In particular, one set of tables under the QA platform offers researchers a tool to determine whether a particular day is a trading holiday at a specific exchange. A part of the code in Appendix 2 generates trading holidays for each exchange and provides a sample program to remove those trading holidays from the daily stock return data.

3.2 Inconsistency between *RI* and Transactions

Theoretically speaking, the *RI* value would not be missing over the trading days, as it reflects growth in the value of an equity holding. In fact, *RI* should have a positive value even when the stock trading is suspended. Consistent with this belief, there are around 12.93% of daily observations with positive *RI* but without closing price or trading volume among the common stocks with annual fundamentals from Worldscope, even after removing trading holidays. However, we also found that 0.23% (0.16%) of daily observations showing a missing *RI* but non-missing closing price (and trading volume) records.

Table 2 reports the number of observations with non-missing closing price but missing *RI* by year and country. We only keep the countries that have more than 20 distinct firms with this missing *RI* problem. The developed countries show up only five times in this table: Poland in 2016, Finland in 2017, Belgium in 2019, and Australia in 2020, and these four markets experience only 1 to 3.06 daily missing *RI* records per firm in a given year. One noticeable exception is Canada in 2018, where 32 firms miss roughly four months or more of daily RI on average. The Datastream content team has been aware of some of those missing observations while consistently working on data consistency checking and corrections. For example, some Datastream content specialist comments that *RIs* for Brazilian stocks before July 1994 are not available due to currency changes.[8]

3.3 Datastream Return Problems

It has been a conventional belief among academic researchers that data from Datastream can be prone to errors. By comparing the return data from Datastream and CRSP, Ince and Porter (2006) find that abnormal return reversals from Datastream *RI* may reveal certain problematic stock prices. After removing those erroneous prices of common stocks, the authors find this correction approach allows them to obtain similar summary statistics on return using Datastream data to those from CRSP data. Griffin et al. (2010) propose a similar correction on daily returns. In particular, the authors propose if Ret_t or $Ret_{t-1} > 100\%$ and $(1 + Ret_t) \times (1 + Ret_{t-1}) - 1 < 20\%$,

[8] In the same email correspondence, the Refinitiv content team informs us they are investigating and correcting the issues, particularly in Canadian and US markets, found through the development of this *Guide*. (ticket #: 09797654).

then both Ret_t and Ret_{t-1} are set equal to a missing value. Additionally, they also suggest setting any daily return greater than 200% to be missing. Likewise, it is also common among researchers to winsorize the top and bottom 0.1% of stock returns in their return sample to achieve the same goal of weeding out return outliers caused by erroneous closing prices.

Like many other successful data vendors, Datastream keeps correcting errors and adding content to meet the needs of practitioners and academia. By 2017, Refinitiv revamped old Datastream and released a newer version of Datastream.[9] In addition to fixing some known errors, the latest version of Datastream expands to cover more markets and provides historical public records and additional data fields. Despite those changes, we found that daily return filters of Griffin et al. (2010) are still able to identify a small proportion of potential errors. In particular, with a data vintage from April 2021, we find close to 23,000 rows failed to pass the reversal filter and around 10,000 extreme daily returns ($> 200\%$) out of roughly 235 million daily observations.

The sample code in Appendix 2 provides a way to generate a daily stock file from Datastream and a QA mapping table between Datastream and Worldscope (see Appendix 1). In addition, we add cumulative adjusted factors and stock pricing records in US Dollars for all primary shares through a web query at WRDS Datastream Daily Stock File.

4 Merging Worldscope and Datastream: A Case Study

Academic research in the finance and accounting field usually complements international pricing data from Datastream with fundamental data from Worldscope. WRDS presents Worldscope data from the QA platform in an intuitive pre-merged format similar to other fundamental databases, such as Compustat, in addition to the complete set of raw data. Comparing to Compustat, Worldscope offers more ratio variables and a set of selected key metrics in USD across all the firms to facilitate cross-country comparison. One relevant and subtle difference between Worldscope and Compustat is its cutoff days for fiscal year-end. In particular, the fiscal year of Worldscope is determined by a cutoff date of February 10th for US firms and January 15th for non-US firms. Data for a fiscal year ending on or before a January cutoff date is classified as the previous year's data. Furthermore, for firms with multiple common shares, Worldscope cumulates firm level (Item6100 = 'C'), security level (Item6100 = 'S'), and ADR/Foreign Filer (Item6100 = 'A') data together, when available. A firm-level record contains all general and fundamental data along with the combined number of shares. In contrast, security-level (ADR-level) records contain detailed market and stock performance data for each share type.

In terms of coverage, many North American firms and larger firms in the other developed

[9] According to an email correspondence with Refinitiv (ticket#: 09856259), Datastream version 1 is no longer supported effective December 31, 2017. Datastream version 2 has all of the content available in version 1 with added benefits, including global VWAP, consolidated volume, consolidated market cap, market coverage, additional trading price&volume, historical CUSIP, SEDOL and primary exchange records.

markets have history back to 1980. Annual history for most medium-sized (small-sized) firms in the developed market is added from the mid-1980s (1990s). In addition, Worldscope gradually expands coverage of firms from the emerging market from the early 1990s. Interim financial data are added for US companies (non-US firms) starting in 1998 (2001). Since 2004, Worldscope stacks restated/revised data with the original data, which is not overwritten. The restated data items can be identified through a few data fields (such as freq). Worldscope occasionally reports pro forma data items. However, these pro forma items are restricted to the amount of history in the registration document for the companies that underwent IPO recently. The currency used for data items is the local currency of the firms' home nation specified by Worldscope *Item6027*. When a firm changes domicile or listing country, all monetary items are retrospectively adjusted according to the nation code in *Item6027*.[10]

Appendix 3 provides a sample code to merge Datastream and Worldscope into one panel data and generate summary statistics similar to Panel A of Table I in Bartram et al. (2012). The results are reported in Table 3 and are broadly consistent with the original work. The current vintage tends to cover more firm-year observations in most markets. The two noticeable countries in which Bartram et al. (2012) report more firm-year observations are Brazil and Germany. There are only 440 Brazilian firm-year observations in our sample, but the original work reports 1,142 observations. This relatively large discrepancy could be attributed to the unavailability of RI over trading days for Brazilian firms before 1994. Similarly, there are 4,451 year-firm observations for the German market in our sample but 6,156 observations reported in the original work. It is less clear what could be the reasons for this difference without access to the original data. Furthermore, the firms from Venezuela (123 firm-year observations) reported in the original work do not pass the filters to be included in our final table. In contrast, our final table adds 531 firm-year observations from Bulgaria, Jordan, Kuwait, and Saudi Arabia, which are not reported in the original work.

Using a subset of major markets over a more extended history, Figure 1 reveals some well-known data patterns of global stock market development, consistent with previous literature. For example, the post-2000 downtrend in the number of US stocks confirms the reduction in the number of US public firms, which is first documented in Doidge et al. (2017). Figure 2 plots the number of primary US shares covered by CRSP and Datastream side-by-side. Datastream covers the US market back to 1973 but includes significantly fewer firms in the early years, and the coverage difference then becomes narrower and reverse around 2011. By 2012, Datastream covers more primary shares in the US exchanges than CRSP, suggesting two data sources may define common stocks or/and USA firms differently.

[10] This is based on the information provided in an email correspondence with Refinitiv (ticket#: 09864168).

5 Conclusion

This article intends to provide researchers a guide on how to use Datastream and Worldscope data through WRDS. We start with some essential knowledge of Refintiv QA tools for mapping data sources. We then discuss the definition of Datastream Return Index and some problems in using this variable to calculate holding period returns. Finally, we demonstrate through an empirical case on how to prepare data for an international study by combining records from Datastream and Worldscope.

Through this article, we provide some crucial technical details for researchers to bypass some otherwise non-trivial coding problems, using tools already built into the system. In particular, we discuss the exchange-level trading holiday records directly available from the QA platform and the newly added historical primary listing exchange records from Datastream. In addition, we fill the void on some technical details using the answers from the correspondence with Refintiv content specialists and comparison with other data sources. We provide SAS codes to generate the key data pieces and summary statics in the appendix.

References

Bartram, Söhnke M., Gregory Brown, and René M. Stulz, 2012, Why are US stocks more volatile?, *The Journal of Finance* 67, 1329–1370.

Dai, Rui, Lawrence Donohue, Qingyi Freda Drechsler, and Wei Jiang, 2021, Dissemination, publication, and impact of finance research: When novelty meets conventionality, *SSRN Working Paper* .

Doidge, Craig, G Andrew Karolyi, and René M Stulz, 2017, The us listing gap, *Journal of Financial Economics* 123, 464–487.

Griffin, John M., Patrick J. Kelly, and Federico Nardari, 2010, Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets, *The Review of Financial Studies* 23, 3225–3277.

Ince, Ozgur S., and R. Burt Porter, 2006, Individual equity return data from Thomson Datastream: Handle with care!, *Journal of Financial Research* 29, 463–479.

Karolyi, G. Andrew, Kuan-Hui Lee, and Mathijs A. Van Dijk, 2012, Understanding commonality in liquidity around the world, *Journal of Financial Economics* 105, 82–112.

Figure 1: Number of Common Stocks in The Top 20 Markets by Maturity

This figure reports time-series of the numbers of stocks from top 20 markets in term of firm-year data availability. The data construction follows Table I in Bartram et al. (2012) and the number of primary shares are derived from data for 1981–2019 excluding firm-years with more than 30% of weekly stock returns equal to zero in the prior year. We also eliminate firms with missing data for total assets, market-to-book, and firm age in the prior year.

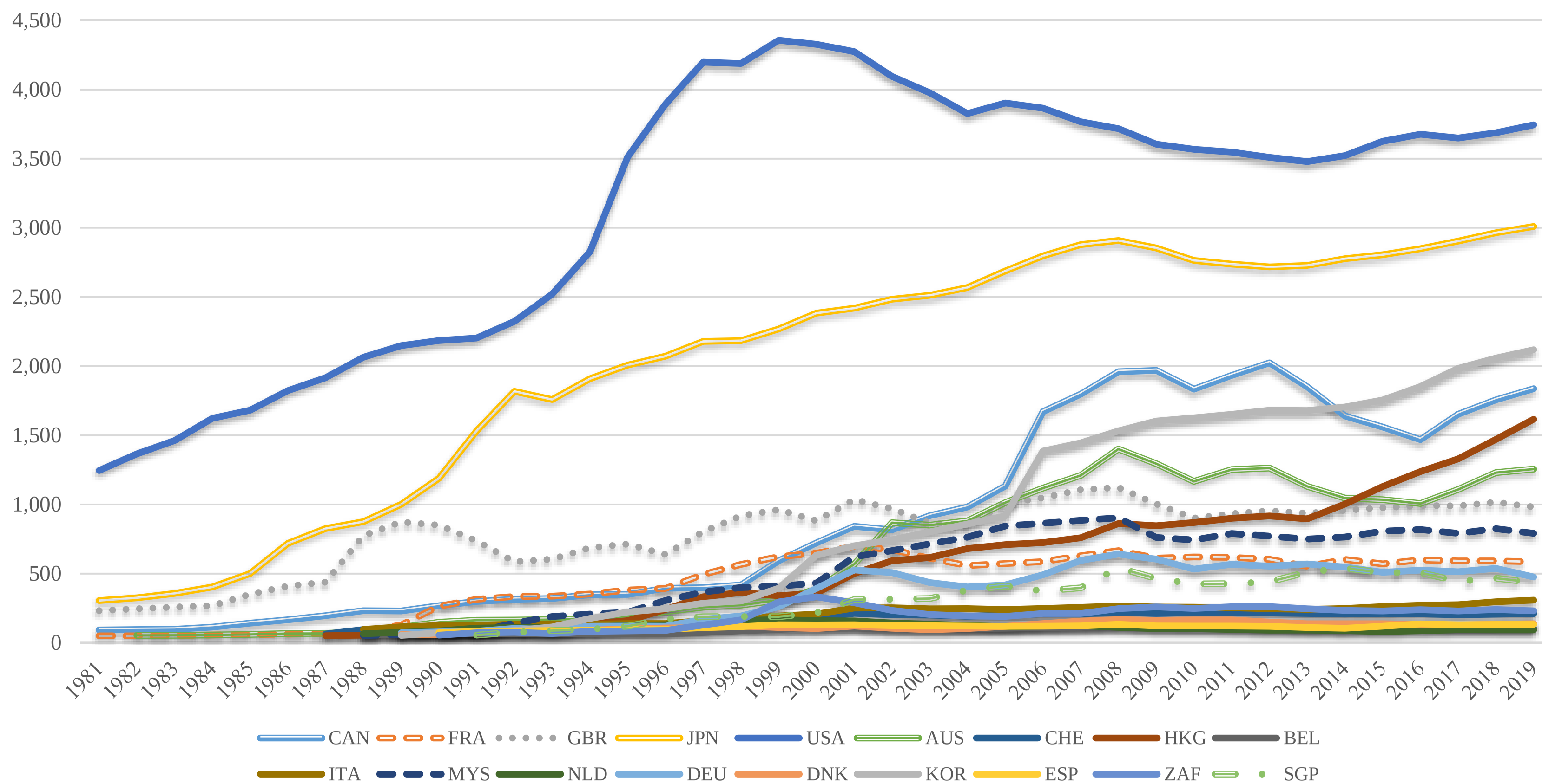


Figure 2: Number of US Common Stocks Covered by CRSP vs. Datastream

This figure reports a time series of the numbers of primary common stock covered by CRSP and Datastream. The common primary US stocks from CRSP are determined by the CRSP share code (10 and 11), as well as the reporting currency (USD) and domicile country (USA). The common primary US stock from Datastream is determined by QA's security information, trading currency in Datastream, and domicile country information from Worldscope.

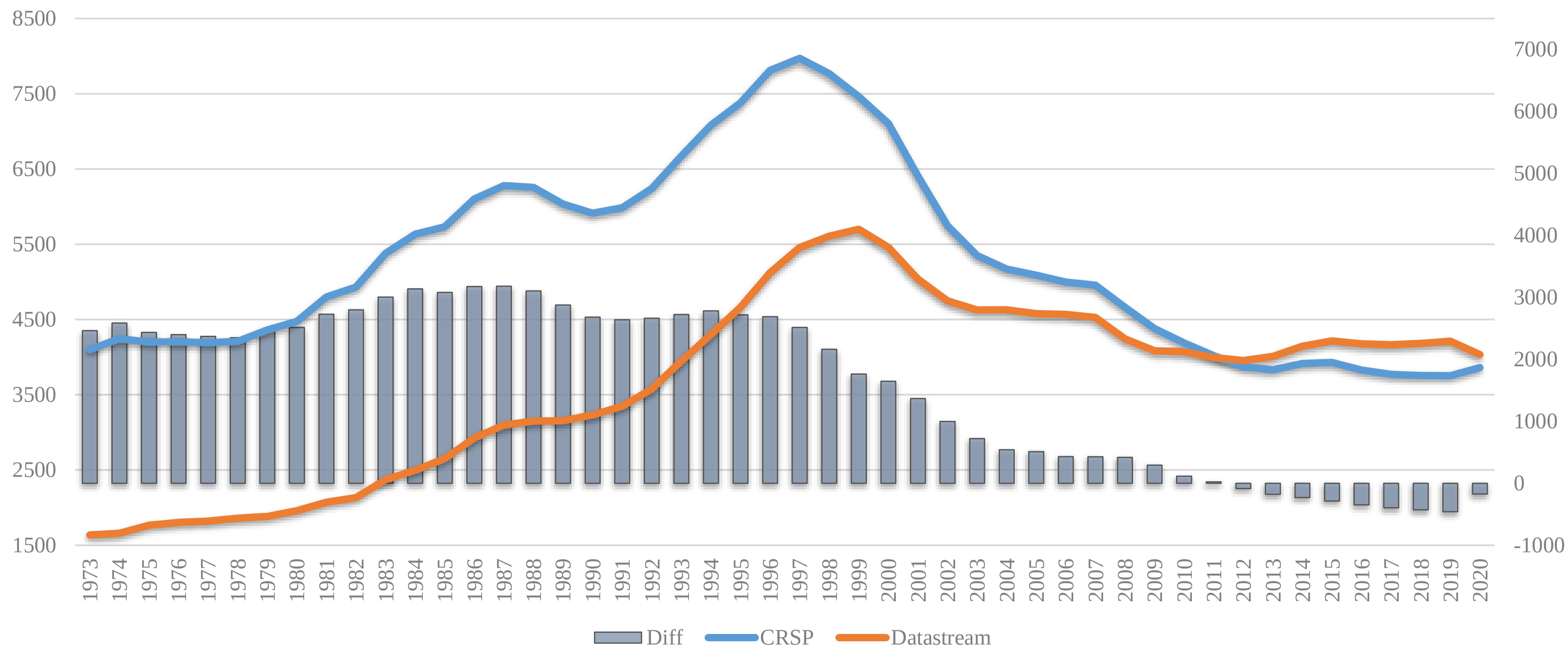


Table 2: Missing RI Trading Days

This table reports countries in the years when they have more than 20 primary shares missing Datastream Return Index but not stock pricing data in a given data. The sample is limited to the common equities that are traded in the currency of home countries and have fundamental data from Worldscope.

Year	Country	#OBS. Missing RI	# Firms	Year	Country	#OBS. Missing RI	#Firms
1990	Brazil	4,988	33	2007	Mauritius	5,828	33
1991	Brazil	10,924	50	2008	Nigeria	27,556	134
1992	Brazil	12,147	56	2008	Mauritius	3,254	33
1993	Brazil	14,486	65	2008	Ghana	1,398	25
1994	Brazil	7,291	65	2009	Nigeria	18,852	142
1998	Mauritius	4,250	33	2009	Ghana	780	24
1999	Mauritius	4,228	33	2011	Jamaica	2,860	21
2000	Mauritius	4,300	33	2012	Jamaica	2,718	21
2001	Nigeria	9,607	81	2013	Sri Lanka	120	120
2001	Kuwait	8,210	68	2013	Jamaica	3,383	29
2001	Mauritius	5,458	32	2014	Jamaica	3,451	29
2002	Nigeria	10,700	79	2015	Jamaica	2,924	28
2002	Kuwait	9,277	76	2016	Malaysia	60	60
2002	Mauritius	7,738	32	2016	India	627	44
2003	Kuwait	13,545	86	2016	Poland	21	21
2003	Nigeria	12,758	80	2017	Thailand	56	51
2003	Mauritius	7,393	33	2017	Finland	50	50
2004	Nigeria	14,944	87	2018	Thailand	40	40
2004	Kuwait	4,725	42	2018	Canada	2,746	32
2004	Mauritius	5,087	34	2018	Saudi Arabia	33	25
2004	Croatia	2,666	21	2018	India	33	24
2005	Nigeria	16,356	91	2019	India	42	39
2005	Kuwait	2,783	38	2019	Indonesia	37	37
2005	Mauritius	5,125	34	2019	Kenya	34	34
2005	Croatia	2,767	28	2019	Belgium	27	22
2006	Nigeria	17,246	102	2019	Venezuela	4,173	22
2006	Mauritius	5,545	34	2020	Australia	2,103	688
2007	Nigeria	23,382	115	2020	Venezuela	3,803	22

Table 3: Firm-Level Characteristics

This table reports country-level median values of variables for firm-year observations in our primary data set. It replicates Panel A of Table I in Bartram et al. (2012) and reports values for firm-level variables. Statistics are derived from data for 1991–2006 excluding firm-years with more than 30% of weekly stock returns equal to zero in the prior year. We also eliminate firms with missing data for total assets, market-to-book, and firm age in the prior year. Definitions of variables are provided in the Appendix of in Bartram et al. (2012)

	Firm-Year Obs.	Total Risk	Total Asset	Age	Market-to-Book	Percent Zero Returns	Leverage	Profit Margine	Cash/Total Assets	Debt Maturity
Argentina	487	0.418	662.40	8	1.007	0.038	0.355	0.276	0.045	0.590
Australia	7,618	0.424	61.06	8	1.675	0.038	0.186	0.104	0.093	0.750
Austria	838	0.258	511.74	8	1.460	0.038	0.400	0.198	0.072	0.600
Belgium	1,404	0.260	354.32	11	1.428	0.038	0.362	0.069	0.084	0.551
Brazil	440	0.469	1,022.70	6	0.917	0.038	0.325	0.323	0.089	0.549
Bulgaria	98	0.662	8.70	1	0.801	0.193	0.119	0.145	0.044	0.549
Canada	10,491	0.450	107.99	10	1.693	0.038	0.239	0.216	0.065	0.798
Chile	837	0.286	429.30	8	1.484	0.075	0.292	0.302	0.034	0.695
China	6,000	0.380	202.56	7	2.208	0.020	0.364	0.211	0.138	0.092
Colombia	163	0.388	1,276.74	7	0.934	0.077	0.213	0.337	0.065	0.584
Czech Republic	239	0.497	229.78	4	0.651	0.075	0.265	0.066	0.040	0.529
Denmark	1,665	0.272	279.49	11	1.371	0.077	0.413	0.202	0.114	0.588
Egypt	245	0.422	267.50	7	1.494	0.000	0.262	0.315	0.134	0.505
Finland	1,250	0.325	263.49	6	1.583	0.038	0.351	0.222	0.082	0.706
France	8,149	0.344	188.24	7	1.512	0.038	0.366	0.107	0.104	0.574
Germany	4,451	0.393	166.34	5	1.672	0.019	0.280	0.202	0.092	0.551
Greece	2,680	0.481	100.49	7	1.659	0.019	0.339	0.237	0.051	0.226
Hong Kong	6,094	0.468	205.08	9	0.895	0.077	0.210	0.206	0.153	0.416
Hungary	263	0.388	176.75	6	1.102	0.020	0.211	0.223	0.070	0.463
India	4,269	0.514	134.22	9	1.378	0.000	0.446	0.139	0.031	0.663
Indonesia	1,513	0.519	183.44	7	0.998	0.173	0.445	0.250	0.103	0.552
Ireland	521	0.303	389.80	15	1.607	0.000	0.409	0.282	0.096	0.746

Table 3: Firm-Level Characteristics, Cont

	Firm-Year Obs.	Total Risk	Total Asset	Age	Market-to-Book	Percent Zero Returns	Leverage	Profit Margine	Cash/Total Assets	Debt Maturity
Israel	865	0.359	347.25	9	1.573	0.000	0.462	0.284	0.130	0.630
Italy	3,038	0.310	805.67	9	1.417	0.019	0.446	0.340	0.091	0.463
Japan	35,567	0.364	502.95	13	1.346	0.019	0.383	0.211	0.158	0.433
Jordan	146	0.396	28.01	1	1.609	0.058	0.129	0.262	0.056	0.053
Korea (Republic of)	7,956	0.541	202.19	11	0.714	0.019	0.407	0.151	0.094	0.385
Kuwait	184	0.364	481.91	2	1.881	0.115	0.298	0.323	0.243	0.310
Luxembourg	89	0.279	442.37	8	1.377	0.000	0.130	0.215	0.094	0.506
Malaysia	7,233	0.413	114.37	9	1.006	0.058	0.282	0.178	0.070	0.322
Mexico	933	0.372	1,247.53	6	1.361	0.000	0.322	0.316	0.061	0.715
Morocco	153	0.259	440.19	9	1.793	0.058	0.166	0.196	0.068	0.032
Netherlands	2,144	0.286	372.85	15	1.810	0.019	0.384	0.206	0.054	0.599
New Zealand	844	0.266	134.59	8	1.437	0.077	0.318	0.133	0.020	0.874
Norway	1,484	0.399	294.59	6	1.617	0.058	0.456	0.202	0.132	0.824
Pakistan	740	0.449	125.30	9	1.400	0.038	0.476	0.154	0.071	0.433
Peru	378	0.367	182.03	8	1.040	0.000	0.251	0.267	0.037	0.456
Philippines	963	0.520	226.30	8	0.887	0.135	0.312	0.267	0.065	0.519
Poland	1,003	0.424	93.72	6	1.515	0.038	0.231	0.189	0.072	0.483
Portugal	736	0.293	301.05	8	1.313	0.038	0.484	0.065	0.033	0.526
Russia	114	0.405	2,298.33	3	2.086	0.019	0.212	0.354	0.055	0.577
Saudi Arabia	103	0.697	661.14	2	4.149	0.019	0.128	0.222	0.098	0.339
Singapore	3,513	0.389	140.46	8	1.135	0.077	0.245	0.161	0.138	0.397
South Africa	2,726	0.394	261.68	9	1.623	0.075	0.186	0.192	0.098	0.630
Spain	1,772	0.279	862.99	8	1.648	0.019	0.381	0.232	0.049	0.489
Sri Lanka	322	0.419	46.74	12	1.073	0.096	0.393	0.211	0.054	0.372
Sweden	3,036	0.377	139.63	6	1.877	0.038	0.284	0.146	0.105	0.797
Switzerland	2,514	0.268	679.72	12	1.471	0.038	0.361	0.248	0.128	0.700
Taiwan	5,178	0.426	243.94	7	1.288	0.020	0.313	0.170	0.123	0.315
Thailand	3,399	0.433	107.65	8	1.116	0.058	0.437	0.225	0.049	0.373
Turkey	1,783	0.563	108.13	9	1.481	0.058	0.261	0.256	0.061	0.264
United Kingdom	13,305	0.335	227.63	12	1.993	0.077	0.288	0.291	0.091	0.655
United States	58,129	0.387	392.74	10	1.932	0.019	0.330	0.338	0.083	0.828
All Country	220,102	0.387	278.11	9	1.565	0.038	0.329	0.231	0.103	0.578

Appendix 1: QA Mapping between Datastream and Worldscope

```
/* ************************************************************************************* */
/* ****************** W R D S   R E S E A R C H   A P P L I C A T I O N S ****************** */
/* ************************************************************************************* */
/* Summary   : Construct a mapping table between Datastream and Worldscope through Refinitiv   */
/*             Quantitative Analytics Tables.                                              */
/*                                                                                         */
/* Date      : May 2021                                                                    */
/* Author    : Rui Dai, WRDS and Qingyi (Freda) Drechsler, WRDS                            */
/*                                                                                         */
/* Variables : - Id: Refintiv QA Security ID                                               */
/*             - SecCode: Refintiv QA Regional Security ID                                 */
/*             - typ: Refintiv QA Regional ID                                              */
/*             - Code: Worldscope QA ID                                                    */
/*             - Coderank: Worldscope QA ID Rank                                           */
/*             - Infocode: Datastream QA ID                                                */
/*             - Inforank: Datastream QA ID Rank                                           */
/*             - Startdate: Datastream Price Start Date                                    */
/*             - Enddate: Datastream Price End Date                                        */
/*             - PrimISOCurrCode: Latest Common Currency                                   */
/*             - PrimISO: Latest Common Country                                            */
/*             - DStype: Datastream Security Type                                          */
/*             - WStype: Worldscope Entity Type                                            */
/*             - ExchIntCode: Datastream Exchange ID                                       */
/*             - exEndDate: Latest non-OTC Exchange End Date                               */
/* To run the program, a user should have access to to Datastream, Worldscope, and QA Tables   */
/* ************************************************************************************* */

/*NOTE: [SecCode]: A unique security ID in either North American or Rest of World (Global).   */
/*NOTE: [Typ]=1: Equities traded in North American (USA and Canada) and =6 for Global Equities.*/
/*NOTE: [ID] is equivalent to [SecCode + Typ], i.e. [seccode] is not a universal id per se.    */
/*NOTE: [VenType] Vendor Type 10 is for Worldscope and 33 is for Datastream.                   */
/*NOTE: [VenCode] Refinitiv QA Vendor Code (Not Native IDs from Vendors)                       */
/*NOTE: [Rank] is the id precedence for each [ID]. Smaller [Rank] value indicates more recent  */
/*     mapping, through public security id, or larger listing exchanges for cross-listed.      */
/*NOTE: [startdate] to [enddate] is the effective range for a [id + rank], which is applicable */
/*      to Datastream but not Worldscope in this guide.                                        */
proc sql;
        create table secmapx as
        select distinct a.SecCode, a.Typ, Id, VenType, VenCode, Rank, Exchange,
        datepart(startdate) as startdate format=date9.,
        datepart(endDate) as enddate format=date9.
        from core.Vw_securitymasterx as a inner join core.Vw_securitymappingx as b
        on a.seccode=b.seccode and a.typ=b.typ
        where VenType in (10 33)
        group by a.SecCode, a.typ
        having count(distinct VenType)>1;
quit;
```

```
/*NOTE: Matchinig Worldscope Annual Fundamental to linking table for content validation.       */
/*NOTE: Primary Worldscope ID [code] is assigned by Refinitiv Quantitative Analytics.          */
/*NOTE: [ITEM7230]: TOTAL ASSETS (USD) and ITEM7240:NET SALES OR REVENUES (USD).               */
proc sql;
        create table Wrds_ws_funda as
        select distinct SecCode, typ, id, rank, a.code, ITEM5350 as datadate
        from ws.Wrds_ws_funda as a inner join Secmapx(where=(VenType=10)) as b
        on a.code=b.VenCode and (ITEM7230>. or ITEM7240>.)
        group by a.code, ITEM5350/*[ITEM5350]: Fiscal Year End Date*/
        having rank=min(rank) /*Keep one with highest precedence at a given [code+datadate] */
        order by SecCode, typ, id, ITEM5350;
quit;
/*NOTE: Primary Datastream ID [Infocode] is assigned by Refinitiv Quantitative Analytics.      */
/*NOTE: Construct a pairwise Datastream ID [InfoCode] to Worldscope ID [Code] mapping table.   */
/*NOTE: [Startdate + Enddate] is applicable to Datastream according on QA Training Guide v1.1  */
proc sql;
        create table secmapx_ws as
        select distinct a.*, code, b.rank as coderank
        from secmapx(where=(VenType=33)) as a inner join Wrds_ws_funda as b
        on a.SecCode=b.SecCode and a.typ=b.typ and a.id=b.id and
        startdate<=datadate<=EndDate; /*datadate is fiscal year end date*/
quit;
/*NOTE: Bring in Country and Currency info from Worldscope and Datastream for cross checking.  */
proc sql;
        create table ds2ws_id_iso as
        select distinct a.SecCode, a.typ, a.id, a.code, a.coderank, b.infocode, rank as inforank,
        startdate, EndDate, PrimISOCurrCode label="DS Currency ISO3",
        IsoUsed label="WS Currency ISO3",  region as ISO2 label="DS ISO2",
        Desc_ as PrimISO length=3 format=$3. informat=$3. label="DS ISO3",
        IsoCtry label="WS ISO3", TypeCode as DStype,
        ITEM6100 as WStype length=3  format=$3. informat=$3.
        from secmapx_ws as a inner join ds.Ds2ctryqtinfo as b
        on a.VenCode=b.infocode and TypeCode in (/*"ADR"*/ "EQ")/*[EQ]:Datastream Common Share*/
        inner join ws.Wrds_ws_company as c
        on a.code=c.code and ITEM6100 in ('C' /*'A'*/) and ITEM6027>./*[C]:Worldscope Company*/
        inner join ws.wscurr as d
        on ITEM6027=d.natcode
        inner join ws.wsinfo as e
        on a.code=e.code
        inner join ds.Ds2xref as f
        on f.type_=1006 and f.Code=region
        where Desc_ =IsoCtry and IsoUsed=PrimISOCurrCode;
quit;
/*NOTE: Bring in historical primary exchanges (OTC-alike exchanges are removed).               */
proc sql;
        create table ds2ws_id_exch(drop=iso2) as
        select distinct id.*, ex.EXCHINTCODE, ex.EndDate as exEndDate
        from ds2ws_id_iso as id inner join (
```

```
            select distinct a.*, b.EXCHCTRYCODE
            from ds.Ds2primexchqtchg as a inner join ds.Ds2exchange as b
            on a.ExchIntCode=b.ExchIntCode
            where index(b.ExchName, 'BULLETIN')=0 and index(b.ExchName, 'OTC')=0
            and index(b.ExchName, 'PINK')=0  /*Remove OTC-alike Exchanges etc.*/
        ) as ex
        on id.infocode=ex.infocode and ISO2=ExchCtryCode
        group by id.infocode
        having ex.EndDate=max(ex.EndDate) /*Match the last non-OTC exchange*/;
quit;
/*NOTE: Matchinig Datastream ID [InfoCode] to linking table for content validation.            */
proc sql;
        create table core.ds2ws_linktable as
        select distinct a.Id, a.SecCode, a.typ,
        code "Worldscope QA ID", coderank  "Worldscope QA ID Rank",
        a.infocode "Datastream QA ID", inforank  "Datastream QA ID Rank",
        startdate "Datastream Price Start Date", enddate "Datastream Price End Date",
        PrimISOCurrCode "Latest Common Currency", PrimISO "Latest Common Country",
        DStype "Datastream Security Type", WStype "Worldscope Entity Type",
        a.ExchIntCode "Datastream Exchange ID",
        exEndDate format=date9. "Latest non-OTC Exchange End Date"
        from Ds2ws_id_exch as a inner join ds.Ds2primqtri as b
        on a.infocode=b.infocode and startdate<=MarketDate<=EndDate and MarketDate<=exEndDate;
quit;
```

Appendix 2: Construction of Daily Security File

```
/* ************************************************************************************ */
/* ****************** W R D S   R E S E A R C H   A P P L I C A T I O N S ****************** */
/* ************************************************************************************ */
/* Summary   : Construct a daily security file from Datastream Return Index Table, Price Table */
/*             and a mapping table between Datastream and Worldscope built by Refinitiv        */
/*             Quantitative Analytics Tables.                                                  */
/*                                                                                             */
/* Date      : May 2021                                                                        */
/* Author    : Rui Dai, WRDS and Qingyi (Freda) Drechsler, WRDS                                */
/*                                                                                             */
/* Variables : - Id: Refintiv QA Security ID                                                   */
/*             - AR_flag: Ince and Porter(2006) reversal flag Defined in Griffin et al (2010)  */
/*             - PRC: Daily Close Price                                                        */
/*             - RET: Daily Return Calculated from Return Index (RI)                           */
/*             - Other Variables: Inherited from Datastream                                    */
/* To run the program, a user should have access to to Datastream, and QA Tables               */
/* ************************************************************************************ */

/*NOTE: Obtain Trading Holidays (e.g., Christmas and Thanksgiving) from QA Special Date Tables */
proc sql;
        create view _holidays_ as
        select e.exchintcode, a.cntryname as sd_country_name, a.isoctry as sd_isoctry,
            a.holtype as sd_holding_type, datepart(b.date_) as holidays format=date9.,
                Name as holiday_name
        from core.sdexchinfo_v a inner join core.sddates_v b
                on a.exchcode = b.exchcode
        inner join core.sdinfo_v c
                on c.code = b.code
        inner join ds.ds2xref d                              /*=1017 Trading Holiday Mapping*/
                on input(d.desc_, 8.0) = a.exchcode and type_ = 1017
        inner join ds.ds2exchange e
                on e.exchintcode = input(d.code, 8.);
quit;
/*NOTE: 1) Restrit Sample into Datastream and Worldscope universe 2) Remove Trading Holidays   */
proc sql;
        create view _Ds2primqtri_ as
        select distinct id, b.infocode, inforank, marketdate, PrimISOCurrCode, a.PrimISO,
        a.exchIntCode, ri
        from core.Ds2ws_linktable as a inner join ds.Ds2primqtri as b
        on a.infocode=b.infocode and startdate<=marketdate<=enddate and marketdate<=exEndDate
        left join _Holidays_ as c
        on marketdate=Holidays and a.ExchIntCode=c.ExchIntCode
        having holidays=.
        order by infocode, marketdate;
quit;
/* NOTE: 1)Calculate Daily Total Return From RI; 2)Flag Abnormal Reversal for Day t (ar_flag)  */
data Ds2primqtri; set _Ds2primqtri_; by infocode marketdate;
```

```
        lag_ri=lag(ri);
        lag_date=lag(marketdate);
        if first.infocode^=1 then do; ret=(ri/lag_ri)-1; lag_ret=lag(ret); end;
        else do; lag_ri=.; ret=.; lag_ret=.; lag_date=.; end;
        if 0<marketdate-lag_date<=7 and (ret>1 or lag_ret>1) /*Griffin et al (2010)'s reversal*/
                                        and .<(1+ret)*(1+lag_ret)-1<.2 then ar_flag=1;
        else ar_flag=0;
        drop lag_date lag_ri lag_ret;
run;
/*NOTE: 1)Sort RI Descending; 2)Flag Abnormal Reversal for Day t+1 (ar_flag)                  */
proc sort data=Ds2primqtri out=Ds2primqtri; by infocode descending MarketDate; run;
data Ds2primqtri; set Ds2primqtri; by infocode descending MarketDate;
        lead_ret=lag(ret);
        lead_date=lag(marketdate);
        if first.infocode then do; lead_ret=.; lead_date=.; end;
        if 0<lead_date-marketdate<=7 /*only on adjacent returns in a week*/
            and (ret>1 or lead_ret>1) and .<(1+ret)*(1+lead_ret)-1<.2 then ar_flag=1;
        label ar_flag = "Ince and Porter (2006) abnormal reversal flag";
        label ret = "Return Based on RI";
        format ret percentn8.2;
        drop lead_date lead_ret;
run;
/*NOTE: Prepare Shares Outstanding for Constructing Final Data Set                             */
proc sort data=ds.Ds2numshares out=Ds2numshares; by infocode descending EventDate; run;
data Ds2numshares; set Ds2numshares; by infocode descending EventDate;
        fromdate=EventDate; todate=lag(EventDate)-1;
        if first.infocode then todate=today();
        if last.infocode then fromdate=.;
        format fromdate todate date9.;
run;
/*NOTE: Merging Price Information to Return data built on RI                                   */
proc sql;
        create table core.Ds2DSF as
        select distinct a.*, Open_ as Open, Close_ as Prc, High, Low, Bid, Volume,
        Ask, VWAP, MostTrdPrc, ConsolVol, NumShrs
        from Ds2primqtri as a left join ds.Ds2primqtprc as b /*Ignore records with missing RI*/
        on a.infocode=b.infocode and a.MarketDate=b.MarketDate
        left join Ds2numshares as c
        on a.infocode=c.infocode and fromdate<=a.MarketDate<=todate
        group by id, a.MarketDate
        having inforank=min(inforank);
quit;
```

Appendix 3: Global Firm-Level Characteristics Analysis

```
/* *************************************************************************************** */
/* ****************** W R D S   R E S E A R C H   S A M P L E   C O D E********************* */
/* *************************************************************************************** */
/* Summary   : Replicate Panel A of Table I of Bartram et al (2012) through a daily Datastream */
/*             Security File, Worldscope Annual File, and a mapping table between Datastream   */
/*             and Worldscope built by Refinitiv Quantitative Analytics Tables.                */
/* Date      : May 2021                                                                        */
/* Author    : Rui Dai, WRDS and Qingyi (Freda) Drechsler, WRDS                                */
/*                                                                                             */
/* Variables : -All Variables: Defined in Bartram et al (2012)                                 */
/* To run the program, a user should have access to to Datastream, Worldscope, and QA Tables   */
/* *************************************************************************************** */

/*NOTE: Worldscope ITEMS:                                                                      */
/*NOTE: ITEM7210-MARKET CAPITALIZATION (USD); ITEM7220-COMMON EQUITY (USD)                     */
/*NOTE: ITEM7230-TOTAL ASSETS (USD); ITEM8206-Preferred Stock % Total Capital                  */
/*NOTE: ITEM8221-Total Debt % Total Capital; ITEM8206-Preferred Stock % Total Capital          */
/*NOTE: ITEM8306-Gross Profit Margin; ITEM2001-Cash & Short Term Investments                   */
/*NOTE: ITEM2999-Total Assets; ITEM3255-Total Debt;                                            */
/*NOTE: ITEM3251-LONG TERM DEBT; ITEM3101-CURRENT LIABILITIES - TOTAL                          */
proc sql;
        create table _Wrds_ws_funda_ as
        select distinct id, a.code, coderank, year_ as year label='Fiscal Year', freq,
        ITEM5350 as datadate, PrimISOCurrCode, ITEM7230, ITEM7210/ITEM7220 as M2B,
        sum(item8206, item8221) as leverage, Item8306, item2001/(item2999-item2001) as Cash,
        item3251/item3255 as Maturity, ITEM3101, ITEM3251
        from ws.Wrds_ws_funda as a inner join core.Ds2ws_linktable as b
        on a.code=b.code and freq='A' and seq=1 and startdate<=ITEM5350<=enddate
        group by id, year_
        having coderank=min(coderank) /*Keep lowest rank at id level*/
        order by a.code, year;
quit;
/*NOTE: 1) Bring Earliest Year from WS or DS; 2) Calculate 3-year moving gross profit margin   */
proc sql;
        create table Wrds_ws_funda as
        select distinct a.*, a.year-min(b.minyear,c.minyear) as age,mean(d.ITEM8306) as avg_gpm
        from _Wrds_ws_funda_ as a left join
        (select distinct id, min(year) as minyear from _Wrds_ws_funda_ group by id) as b
        on a.id=b.id left join
        (select distinct id, min(year(MarketDate)) as minyear from core.Ds2dsf group by id) as c
        on a.id=c.id left join _Wrds_ws_funda_ as d
        on a.id=d.id and 0<=a.year-d.year<=2
        group by a.id, a.year;
quit;
/* NOTE: Keep only stocks that traded more than 25 weeks in a given calendar year              */
/* NOTE: Keep markets that have more than 10 (included) stocks traded in a given calendar year */
proc sql;
```

```
        create view _ds2ret_ as
        select distinct id, InfoCode, MarketDate, PrimISO, Ret
        from core.Ds2dsf(where=(ret>. and ret<=2 and ar_flag=0)) /*Drop extreme value*/
        group by id, year(MarketDate)
        having count(distinct week(marketdate))>=25;
quit;
/* NOTE: Calculate weekly return by compounding daily return                                    */
proc sql;
        create table _ds2wret_ as
        select distinct id, intnx('week', MarketDate, 0, 'e') as weekday format=date9.,
        PrimISO, exp(sum(log(sum(1,ret)))) - 1 as ret, count(distinct MarketDate) as daycnt
        from _ds2ret_
        group by id, intnx('week', MarketDate, 0, 'e');
quit;
/* NOTE: Exclude firm-years with >= 30\% of weekly stock returns equal to 0 in the prior year.  */
/* NOTE: Eliminate firms with missing total assets, market-to-book, and age in the prior year.  */
proc sql;
        create table ds2wret as
        select distinct a.*
        from _ds2wret_ as a inner join Wrds_ws_funda(where=(ITEM7230>.)) as b
        on a.id=b.id and year(weekday)=year(b.datadate)
        inner join Wrds_ws_funda(where=(ITEM7230>. and M2B>. and age>.)) as c
        on a.id=c.id and year(weekday)-1=year(c.datadate)
        left join ( select distinct year(weekday) as year, id, sum(ret=0) as cnt
                                from _ds2wret_ group by id, year(weekday)
                                having sum(ret=0)/count(distinct weekday)>=.3
                ) as d
        on a.id=d.id and year(weekday)-1=d.year
        having d.cnt=.;
quit;
/* NOTE: Winsorize by replacing two tail values with 0.1% and 99.9% values.                     */
proc univariate data=ds2wret noprint;
        var ret;
         output out=ret_w pctlpts=0.1 99.9 pctlpre=ret;
run;
proc sql;
        create table ds2wret_win as
        select distinct a.*, case when .<ret<ret0_1 then ret0_1
        when ret>ret99_9 then ret99_9 else ret end as ret_win
        from ds2wret as a, ret_w;
quit;
/*NOTE: Aggregate Weekly Date into Annual Level.                                                 */
proc sql;
        create view _ds2ws_ann_ as
        select distinct id, PrimISO, year(weekday) as year,
        sum(ret_win=0)/count(ret_win) as zeropct "Percent Zero Returns",
        STD(ret_win)*sqrt(52) as ret_std "Total Risk"
        from ds2wret_win
```

```
        group by id, year;
quit;
/*NOTE: Merge Datastream and Worldscope Data at annual level.                                   */
/*NOTE: "We drop firms that are missing data on total assets, market price at yearend, book      */
/*             value per share, shares outstanding, book value of long-term debt, and book value */
/*             of short-term debt." See Bartram et al (2012).                                    */
proc sql;
        create table ds2ws_ann as
        select distinct a.*, age, ITEM7230/1000000 as asset "Total Assets",
        M2B "Market-to-Book", leverage/100 as leverage "Leverage",
        avg_gpm/100 as avg_gpm "Profit Margin", cash "Cash/Total Assets",
        Maturity "Debt Maturity"
        from _ds2ws_ann_ as a inner join Wrds_ws_funda as b
        on a.id=b.id and a.year=b.year
        where nmiss(ITEM7230, M2B, ITEM3251, ITEM3101)<4
        group by a.PrimISO, a.year
        having count(distinct id)>=10;
quit;
/*NOTE: Generate Panel A of Table I of Bartram et al (2012).                                    */
proc tabulate data=ds2ws_ann;
        where 1991<=year<=2006;
        class PrimISO;
        var ret_std zeropct asset age M2B leverage avg_gpm cash Maturity;
        table PrimISO='' all='', n=''*ret_std*f=comma8.0 median=''*((ret_std asset)*f=8.3
        age*f=8.0 (M2B zeropct leverage avg_gpm cash Maturity)*f=8.3);
run;
```