

Financial Data Resources for Thesis Research

Gregory Connor
Department of Economics, Finance and
Accounting
NUI Maynooth

Topics

1. Research Categories and their Data Requirements
2. A Partial List of Finance Databases
3. Strategies for Data Manipulation

1. Research Categories and their Data Requirements

Indices vs. individual securities

- Many important research projects require index-only data
- The relative ease and reliability of index-level data should not be forgotten
- Industry-level indices provide a theoretically-justified middle ground between national indices and individual securities

1. Research Categories and their Data Requirements

International and Macro-Finance

- International and macro-finance are well served at the index level
 - Currencies, equity index returns, trade account balances
- Security-specific data is *usually* not appropriate in international and macro-finance research
- Karl Whelan will discuss data sources for macro-finance and international finance later today

1. Research Categories and their Data Requirements

Estimating Equity Market Risk-Return

- 99% of empirical research on equity market risk premiums is empty of statistical content, redundant, or fallacious
 - This proportion is higher than in other subfields within financial economics
 - The very best empirical finance researchers (Fama, Roll, Campbell, et al.) cannot agree on the US equity premia within a factor of two
 - Much more difficult to estimate the risk premia in smaller countries with shorter time series and less reliable data

1. Research Categories and their Data Requirements

Corporate Finance Research

- Accounting data can be a very powerful source of scientific information, but the data is difficult to work with
- Near-impossible to create balanced panels
- Dating conventions are problematic
- Requires sometimes-detailed knowledge of accounting rules
- Event studies (security price data, plus news announcements) relatively easy from a data perspective

1. Research Categories and their Data Requirements

Fixed Income and Derivatives Research

- Individual-security fixed income models require detailed information on bond covenants and cash flow schedules
- The risk-return features of individual bonds change through time
- Bond indices are representative bonds (Federal Reserve) are a good alternative to individual bonds

1. Research Categories and their Data Requirements

Market Microstructure Research

- Expensive, massive databases are characteristic of microstructure research
- Many proprietary databases at financial institutions via automated storage of real-time feeds

1. Research Categories and their Data Requirements

Alternative Asset Classes

- Hedge fund returns and indices
- Commodities, real estate, collectibles
- Hedge fund returns are well-represented in terms of available data
- Commodity prices are often proxied by commodity futures prices
- Alternative asset classes such as real estate and collectibles are improving markedly in terms of data availability and quality

2. A Partial List of Finance Databases

Center for Research in Security Prices (CRSP)

- Preeminent source of US stock and bond prices and returns at daily and monthly frequencies
- Data history begins in 1926 and is updated annually
- Very high quality data in terms of low error frequency

2. A Partial List of Finance Databases

Compustat

- Compustat North America and Compustat Global
- Comprehensive source of accounting data and associated equity security prices
- Links with CRSP equity returns via GVKey common numerical identifier
- Both annual and quarterly accounting data

2. A Partial List of Finance Databases

Wharton Research Data Services

- Essentially a CRSP-Compustat packaging and download service (and some others)
- Subscribe to WRDS as well as CRSP and Compustat
- Describe the variables, time period, and frequencies of a desired data download using an on-line template
- WRDS builds your database and temporarily places it on its server for collection by you

2. A Partial List of Finance Databases Datastream

- Datastream
 - Very important, very comprehensive source of both security returns and accounting data, worldwide
 - Some concerns about pricing reliability for individual fixed income securities relative to Bloomberg
 - Benchmark bonds are a good data source
 - The main source of finance data for a large number of doctoral theses

2. A Partial List of Finance Databases

Ken French – CRSP-based Indices

- Available at <http://mba.tuck.dartmouth.edu/faculty/ken.french>
- Completely free source of CRSP-based market, style and industry index returns
 - Very high quality with frequent updates
 - Daily, monthly and annual
 - Provides size and value based sort portfolios as well as a wide range of industry indices
 - Invaluable both as a data source and as an error-check on parallel sources and processed series

2. A Partial List of Finance Databases

Yahoo Finance Free Downloads

- Available at www.yahoofinance.com
- Completely free source of daily closing prices for individual equities on a large number of equity markets (Ireland, UK, USA, etc)
- Very valuable free resource
- There are many other free-downloads on the web

2. A Partial List of Finance Databases

- Bloomberg
 - High quality source of fixed income and derivatives prices
- CBOT Treasury Package
 - Derivatives prices from the Chicago Board of Trade
- CISDM Hedge Fund Database
 - One of the two pre-eminent databases of hedge fund styles and returns, both for individual funds and hedge fund style indices

2. A Partial List of Finance Databases

- Lipper-Tass
 - Alternative to CISDM for hedge fund returns
- DealScan
 - Fixed income prices and transactions
- Standard and Poors Emerging Markets Database
 - Equity index returns (both complete indices and investable indices) for 35 emerging equity markets
- First Call
 - Alternative to Compustat

2. A Partial List of Finance Databases

- Thompson Research (formerly Global Access)
 - Alternative to Compustat Global
- I/B/E/S
 - Important source of earnings and dividend forecasts – reliable quality; less expensive than some other sources
- ISSM
 - Market microstructure database of tick-by-tick NYSE equity market transactions

2. A Partial List of Finance Databases

- LIFFE Tick Data
 - London International Financial Futures Exchange tick-by-tick derivatives trade prices
- Moody's Research Services
 - Corporate bond database
- OptionMetrics Ivy Database
 - Daily closing prices on virtually all exchange-traded US individual equity options

2. A Partial List of Finance Databases

- Trade and Quote (TAQ) database
 - High quality database of US equity market tick-by-tick trades and bid-ask quotes
- Reuters Foreign Exchange Transactions
 - Extremely large database of tick-by-tick currency market transactions

2. A Partial List of Finance Databases

- Proprietary databases within financial services firms
 - A huge number of databases are maintained within individual firms in the financial services industry
 - Often, temporary access for academic research can be arranged, and suitably processed findings used in doctoral research

2. A Partial List of Finance Databases

- Web pages of individual academics
 - Some academic researchers provide data access on their web pages for their completed studies
 - Not possible if the researcher uses subscription-based sources (e.g., CRSP or Compustat)

3. Strategies for Data Manipulation

- Build parallel error-control systems
 - Compare your findings to approximately parallel existing findings in reputable published research
 - Use alternative or higher-level data series to approximately replicate your findings
 - Get a sense of magnitudes for key parameters and findings, and use it to check the reasonableness of your findings
 - Keep in mind that unusual results can be valid; they do not always identify data errors.

3. Strategies for Data Manipulation

- Stale price observations are an endemic problem
- Within indices, across indices, within and across individual security series
- Be aware of quoted price versus transactions price observations within data series

3. Strategies for Data Manipulation

- Dividends and interest coupons are a necessary evil
- Where possible, download total returns rather than building them up from prices and cash flows
- In many risk-measurement contexts, dividends can be ignored
- Bond price conventions sometimes rely on coupon-accrual pricing

3. Strategies for Data Manipulation

- Ask yourself carefully: what research outcome am I seeking, both for my own career and to push the research frontier
- Many good doctoral dissertations include innovative theory, illuminating empirical analysis, and valuable new data sets
 - You might want to settle for two-out-of-three!
- US data is easiest and cleanest, but also the most exhaustively flogged