2 Data

The task of setting up an appropriate dataset for my purposes involves decisions on the commodities to include, whether to use short-term future contracts or spot price data, what time period to analyse, and what data frequency to use.

Following Marshall, Nguyen, and Visaltanachoti (2012), I focus on the components of the S&P Goldman Sachs Commodity Index (S&P GSCI). As this index comprises 24 *important* commodities spanning energy, industrial metal, precious metal, agriculture, and livestock, it suits my desire to obtain conclusions of preferably broad, cross-sectoral applicability and to account for the heterogeneity of different commodities (Kat & Oomen, 2007).

With respect to the type of price quote to use, I would generally prefer to use spot instead of future prices for my purposes, however, Fama and French (1987) suggest that "good spot-price data are not available for most commodities" (p. 57) and Schwartz (1997) raises similar concerns about their liquidity. As a result, these authors propose the use of short-term futures data instead. Since this view is not shared by Brooks and Prokopczuk (2013) and would generally complicate the analysis and interpretation of results throughout the paper, I conjecture that in the wake of "tremendous growth in commodity [...] markets" (Tsekrekos et al., 2012, p. 543), these earlier findings are not necessarily reflective of spot price data quality in the more recent past.[9] Since the use of justifiable data can be pivotal for the results of any empirical study (Gujarati, 2003), I feel it is necessary to assess the relative data quality in spot and futures markets based on a short, back-of-the-envelope liquidity calculation.

For this task, the simple but computationally feasible[10] *Zeros* measure turns out useful (Lesmond, Ogden, & Trzcinka, 1999). For a given underlying, this met-

[9] Fama and French (1987) and Schwartz (1997) use datasets for the periods 1966 - 1984 and 1985 - 1995, respectively.

[10] Marshall et al. (2012) advise the Amihud (2002) or Effective Tick measure developed by (Goyenko, Holden, & Trzcinka, 2009) to calculate the liquidity of a security, however, these measures rely on trading volume data or price spreads, which are not consistently available from Thomson Reuters for spot price data of S&P GSCI components.

M. Schöne, *Real Options Valuation,* BestMasters,
DOI 10.1007/978-3-658-07493-7_2, © Springer Fachmedien Wiesbaden 2015

Table 1:

Overview of commodities in the dataset

BBL: Barrel, BSH: Bushel, FED: U.S. Federal Reserve, ICCO: International Cocoa Organization, ISO: International Sugar Organization, LBM: London Bullion Market, LME: London Metal Exchange, MT: Metric Ton, OZT: Troy Ounce, TR: Thomson Reuters, USDA: U.S. Department of Agriculture. Note that the data is presented following the logic of four distinctive areas: Energy, industrial metals, precious metals, and agriculture. The sample period for all items is from 02/08/1993 to 30/12/2013. All data are in US$.

Name	Code	Full description	Unit	Source
Crude oil	CRUDOIL	Crude Oil-WTI Spot Cushing	BBL	TR
Aluminum	LAHCASH	LME-Aluminium 99.7% Cash	MT	LME
Copper	LCPCASH	LME-Copper Grade A Cash	”	”
Lead	LEDCASH	LME-Lead Cash	”	”
Nickel	LNICASH	LME-Nickel Cash	”	”
Zinc	LZZCASH	LME-SHG Zinc 99.995% Cash	”	”
Gold	GOLDBLN	Gold Bullion LBM	OZT	LBM
Silver	SLVCASH	Silver Fix LBM Cash	”	”
Wheat	WHEATHD	Wheat, No.2 Hard (Kansas)	BSH	USDA
Corn	COTSCIL	Corn US No.2 South Central IL	”	”
Soybeans	SOYADSC	Yellow Soybn US NO.1 Sth Dvprt	”	”
Sugar	WSUGDLY	Raw Sugar-ISA	LB	ISO
Coffee	COFBRAZ	Coffee-Brazilian (NY)	”	TR
Cocoa	COCINUS	Cocoa-ICCO	MT	ICCO

ric relies on the proven negative relationship between the number of trading days with zero returns and liquidity in a given time interval. Hence, we can gain a rudimentary understanding of the relative liquidity and data quality of spot and future prices by simply comparing the average number of zero-return trading days per interval in both markets. As directly comparable spot and 3-months future prices are only available from the London Metal Exchange (LME) for the industrial metals part of the S&P GSCI index, this short liquidity analysis is somewhat limited in scope, but, nonetheless, the exercise reveals that over the more recent past LME data quality is, in fact, higher for spot prices than future prices. By investigating the development of *Zeros* over time, this conclusion can be reconciled with previous research as it shows unsatisfactory spot price data quality prior to July 1993, but a structural improvement thereafter. To conserve space, the graphical results of this analysis are provided in appendix A1.

Following the previous argument, I use spot price data for all commodities with

satisfactory data availability for the period 02/08/1993 - 30/12/2013. With respect to the data frequency, I use weekly quotes in the majority of calculations, as I suspect these to be more reflective of decision making horizons in capital investment decisions. However, as the analysis of empirical time series properties and calibration tasks can sometimes benefit from the information contained in more frequent daily returns, I follow Kat and Oomen (2007) and Brooks and Prokopczuk (2013) and resort to daily quotes in some cases.

An overview of the 14 commodities in the dataset is given in table 1. A complementary graph showing the historical price evolution for the four commodity sectors and the S&P 500 equity index is given in appendix A2. A table with all non-trading dates that have been erased from the dataset for the previous and all subsequent analyses is given in appendix A3. Please note that it is due to limited availability of high quality spot price data for energy commodities and live stock that only 14 out of the 24 S&P GSCI components are used.

http://www.springer.com/978-3-658-07492-0

Real Options Valuation
The Importance of Stochastic Process Choice in Commodity Price Modelling
Schöne, M.
2015, XIV, 104 p. 15 illus., Softcover
ISBN: 978-3-658-07492-0