

```

** THIS PROGRAM CALCULATES HIGH-LOW SPREADS BASED ON HIGH-LOW PRICE DATA FROM CRSP. THE ESTIMATES
** ARE BASED ON THE METHODOLOGY IN CORWIN AND SCHULTZ (2011), AND THE CLOSED FORM SOLUTION PRESENTED
** IN THEIR EQUATIONS (14) AND (18).
*****
** FIRMS ARE IDENTIFIED BY CRSP PERM NUMBER (PERMNO). THE NECESSARY INPUT DATA INCLUDE THE DATE AND
** MONTH, AND THE DAILY SPLIT-ADJUSTED HIGH PRICE (HIPRC), LOW PRICE (LOPRC), AND CLOSE PRICE (PRC).
** IF AVAILABLE, VOLUME CAN BE USED TO IDENTIFY NON-TRADING DAYS.
*****
** OUTPUT INCLUDES DAILY AND MONTHLY ESTIMATES OF THE HIGH-LOW SPREAD.
*****
** REFERENCE:
** Corwin, Shane A., and Paul Schultz, 2011, A Simple Way to Estimate Bid-Ask Spreads from Daily
** High and Low Prices,= forthcoming, Journal of Finance.
*****
LIBNAME SAMPLIB 'Y:/crsp_spreads/samplib';

*****
** READ IN CRSP PRICE DATA
** ASSIGN VARIABLE NAMES AS FOLLOWS:
** DATE = DATE IN YYYYMMDD FORMAT
** PRC = DAILY CLOSING PRICE
** LOPRC = DAILY LOW PRICE
** HIPRC = DAILY HIGH PRICE
*****
DATA SAMPLE; SET SAMPLIB.ALLCRSP_26_09(KEEP = PERMNO SHRCD EXCHCD DATE PRICEIN_NEW LOPRC_NEW HIPRC_NEW
MONTH=INT(DATE/100);
RENAME PRICEIN_NEW=PRC LOPRC_NEW=LOPRC HIPRC_NEW=HIPRC;
PROC MEANS DATA=SAMPLE; TITLE 'SUMMARY STATISTICS FOR DAILY INPUT DATA FROM CRSP';
PROC SORT; BY PERMNO DATE;
RUN;

*****
** RETAIN GOOD HIGH-LOW PRICES AND REPLACE IN CASES WHERE HIGH=LOW
** REPLACE WITH MISSING VALUES WHEN BEGINNING OF SERIES HAS HIGH=LOW
*****
DATA SAMPLE2 (DROP = LOPRCR HIPRCR); RETAIN LOPRCR HIPRCR; SET SAMPLE; *SET SAMPLEX; BY PERMNO MONTH D
LOPRCIN=LOPRC; HIPRCIN=HIPRC;
HLRESET=0;
/* INITIAL DATA SCREENS - PRIOR TO H/L RESET */
*IF LOPRC=HIPRC OR LOPRC<=0 OR HIPRC<=0 THEN DO; *DROP BAD PRICES ONLY (USE BID
IF LOPRC=HIPRC OR LOPRC<=0 OR HIPRC<=0 OR PRC<=0 OR VOLUME=0 THEN DO; *DROP BAD PRICES AND ZERO VOLU
ISAMEPRC=0; IF LOPRC=HIPRC THEN ISAMEPRC=1;
INOTRADE=0; IF PRC<0 OR VOLUME=0 THEN INOTRADE=1;
LOPRC=.; HIPRC=.;
END;
PRC=ABS(PRC);
IF FIRST.PERMNO THEN DO;
LOPRCR=.; HIPRCR=.;
END;
*RESET RETAINED HIGH AND LOW;
IF 0<LOPRC<HIPRC THEN DO;
LOPRCR=LOPRC; HIPRCR=HIPRC;
END;
*REPLACE MISSING/BAD HIGH AND LOW PRICES WITH RETAINED VALUES;
ELSE DO;

```

```

*REPLACE IF WITHIN PRIOR DAY'S RANGE;
IF LOPRCR<=PRC<=HIPRCR THEN DO;
  LOPRC=LOPRCR; HIPRC=HIPRCR; HLRESET=1;
END;
*REPLACE IF BELOW PRIOR DAY'S RANGE;
IF PRC<LOPRCR THEN DO;
  LOPRC=PRC; HIPRC=HIPRCR-(LOPRCR-PRC); HLRESET=2;
END;
*REPLACE IF ABOVE PRIOR DAY'S RANGE;
IF PRC>HIPRCR THEN DO;
  LOPRC=LOPRCR+(PRC-HIPRCR); HIPRC=PRC; HLRESET=3;
END;
END;
/* FINAL DATA SCREENS - AFTER H/L RESET */
*DROP OBS IF HIGH/LOW>8;
IF LOPRC NE 0 AND HIPRC/LOPRC>8 THEN DO; LOPRC=.; HIPRC=.; END;
RUN;

*****
** ADJUST FOR OVERNIGHT RETURNS BASED ON LAGGED CLOSING PRICE.
*****

DATA SAMPLE2; SET SAMPLE2;
RETADJ=0;
TLOPRC=LOPRC; *CURRENT DAY LOW PRICE;
THIPRC=HIPRC; *CURRENT DAY HIGH PRICE;
LLOPRC=LAG(LOPRC); *PRIOR DAY LOW PRICE;
LHIPRC=LAG(HIPRC); *PRIOR DAY HIGH PRICE;
LPRC=LAG(PRC);
IF LAG(PERMNO) NE PERMNO THEN DO; LLOPRC=.; LHIPRC=.; LPRC=.; END;
IF LPRC<LOPRC AND LPRC>0 THEN DO; *ADJUST WHEN PRIOR CLOSE IS BELOW CURRENT LOW;
  THIPRC=HIPRC-(LOPRC-LPRC); TLOPRC=LPRC; RETADJ=1;
END;
IF LPRC>HIPRC AND LPRC>0 THEN DO; *ADJUST WHEN PRIOR CLOSE IS ABOVE CURRENT HIGH;
  THIPRC=LPRC; TLOPRC=LOPRC+(LPRC-HIPRC); RETADJ=2;
END;
RUN;

*****
** CALCULATE DAILY HIGH-LOW SPREAD ESTIMATES
*****

DATA SAMPLE2; SET SAMPLE2;
PI=CONSTANT('PI');
K = 1/(4*LOG(2));
K1 = 4*LOG(2);
K2 = SQRT(8/PI);
CONST = 3-2*SQRT(2);
HIPRC2=MAX(THIPRC,LHIPRC);
LOPRC2=MIN(TLOPRC,LLOPRC);
IF TLOPRC>0 AND LLOPRC>0 THEN BETA = (LOG(THIPRC/TLOPRC))**2+(LOG(LHIPRC/LLOPRC))**2;
IF LOPRC2>0 THEN GAMMA = (LOG(HIPRC2/LOPRC2))**2;
ALPHA = (SQRT(2*BETA)-SQRT(BETA))/CONST - SQRT(GAMMA/CONST);
SPREAD = 2*(EXP(ALPHA)-1)/(1+EXP(ALPHA));
*SET NEGATIVE SPREAD ESTIMATES TO ZERO;
SPREAD_0 = MAX(SPREAD,0); IF SPREAD=. THEN SPREAD_0=.;
*DROP NEGATIVE SPREAD ESTIMATES;
IF SPREAD>0 THEN SPREAD_MISS=SPREAD;

```

```

SIGMA = ((SQRT(BETA/2) - SQRT(BETA)))/(K2*CONST)+SQRT(GAMMA/(K2*K2*CONST));
*SET NEGATIVE SIGMA ESTIMATES TO ZERO;
SIGMA_0 = MAX(SIGMA,0); IF SIGMA=. THEN SIGMA_0=.;
PROC MEANS DATA=SAMPLE2; TITLE 'SUMMARY STATISTICS FOR DAILY HIGH-LOW SPREAD ESTIMATES';
RUN;

*****
** OUTPUT DAILY HIGH-LOW SPREAD ESTIMATES.
** NOTE: SPREAD_0 IS THE PRIMARY DAILY HIGH-LOW SPREAD ESTIMATOR IN CORWIN AND SCHULTZ (2011)
*****
** OUTPUT VARIABLES:
** SPREAD = DAILY H-L SPREAD ESTIMATES WITH NEG ESTIMATES INCLUDED
** SPREAD_0 = DAILY H-L SPREAD ESTIMATES WITH NEG ESTIMATES SET TO ZERO
** SPREAD_MISS = DAILY H-L SPREAD ESTIMATES WITH NEG ESTIMATES SET TO MISSING
** SIGMA = DAILY STD. DEV. ESTIMATE WITH NEG ESTIMATES INCLUDED
** SIGMA_0 = DAILY STD. DEV. ESTIMATE WITH NEG ESTIMATES SET TO ZERO
*****
DATA SAMPLIB.HLSPRD_DAY_SAMPLE(REPLACE=YES);
  SET SAMPLE2(KEEP = PERMNO DATE MONTH SHRCD EXCHCD SPREAD SPREAD_0 SPREAD_MISS SIGMA SIGMA_0);
  IF SHRCD IN (10,11) AND EXCHCD IN (1,2,3); *EXCHANGE-LISTED U.S. COMMON STOCK ONLY;
RUN;

*****
** CALCULATE MONTHLY HIGH-LOW SPREAD ESTIMATES.
*****
PROC UNIVARIATE NOPRINT DATA=SAMPLIB.HLSPRD_DAY_SAMPLE; BY PERMNO MONTH; VAR SPREAD SPREAD_0 SPREAD_MISS SPREAD_MISS_0;
  OUTPUT OUT=SUMSPRD N=N1-N5 MEAN=MSPREAD MSPREAD_0 MSPREAD_MISS MSIGMA MSIGMA_0;
DATA SUMSPRD2; SET SUMSPRD; IF N1>=12;
  *SET NEGATIVE MONTHLY SPREAD TO ZERO WHEN NEGATIVE DAILY VALUES ARE RETAINED;
  XSPREAD_0 = MAX(MSPREAD,0); IF MSPREAD=. THEN XSPREAD_0=.;
PROC MEANS DATA=SUMSPRD2; TITLE 'SUMMARY STATISTICS FOR MONTHLY HIGH-LOW SPREAD ESTIMATES';
RUN;

*****
** OUTPUT MONTHLY HIGH-LOW SPREAD ESTIMATES.
** NOTE: MSPREAD_0 IS THE PRIMARY MONTHLY HIGH-LOW SPREAD ESTIMATOR IN CORWIN AND SCHULTZ (2011)
*****
** OUTPUT VARIABLES:
** MSPREAD = MONTHLY AVERAGE OF DAILY H-L SPREAD ESTIMATES WITH NEG DAILY ESTIMATES INCLUDED
** MSPREAD_0 = MONTHLY AVERAGE OF DAILY H-L SPREAD ESTIMATES WITH NEG DAILY ESTIMATES SET TO ZERO
** MSPREAD_MISS = MONTHLY AVERAGE OF DAILY H-L SPREAD ESTIMATES WITH NEG DAILY VALUES SET TO MISSING
** XSPREAD_0 = MAX(0,MSPREAD) - HERE NEGATIVES ARE SET TO ZERO AFTER TAKING THE MONTHLY AVERAGE
** MSIGMA = MONTHLY AVERAGE OF DAILY STD. DEV. ESTIMATE WITH NEG DAILY ESTIMATES INCLUDED
** MSIGMA_0 = MONTHLY AVERAGE OF DAILY STD. DEV. ESTIMATE WITH NEG DAILY ESTIMATES SET TO ZERO
*****
DATA SAMPLIB.HLSPRD_MO_SAMPLE(REPLACE=YES);
  SET SUMSPRD2(KEEP = PERMNO MONTH MSPREAD MSPREAD_0 MSPREAD_MISS XSPREAD_0 MSIGMA MSIGMA_0);
RUN;

```