Measuring Uncertainty: Supplementary Material*

Kyle Jurado
Columbia University

Sydney C. Ludvigson
NYU and NBER

Serena Ng
Columbia University

September 10, 2013

Abstract

This document contains supplementary material for the paper entitled "Measuring Uncertainty" and has two parts. The first part provides the results of robustness exercises based on (i) alternative weights used to aggregate individual uncertainty series; (ii) alternative location statistics of stochastic volatility to construct individual uncertainty series; (iii) alternative conditioning information using recursive forecasts to construct diffusion index forecasts; (iv) alternative ordering of the VARs used to assess the importance of uncertainty shocks. The second part is a data appendix that contains details on the construction of all data used in this study, including data sources.

*Jurado: Department of Economics, Columbia University, 1019 International Affairs Building, MC 3308, 420 West 118th Street, New York, NY 10027; Email: kej2108@columbia.edu.
Ludvigson: Department of Economics, New York University, 19 W.4th Street, 6th Floor, New York, NY 10012; Email: sydney.ludvigson@nyu.edu; `http://www.econ.nyu.edu/user/ludvigsons/`.
Ng: Department of Economics, Columbia University, 1019 International Affairs Building, MC 3308, 420 West 118th Street, New York, NY 10027; Email: serena.ng@columbia.edu; `http://www.columbia.edu/~sn2294/`.

1 Robustness

Our baseline estimate of macro uncertainty $\overline{\mathcal{U}}_t^y(h)$ is constructed as the cross-sectional average of the individual uncertainties $\mathcal{U}_{jt}^y(h)$, and each of these is based on evaluating (**??**) at the posterior mean, over the full sample, of the state and parameters of the stochastic volatility model (i.e.,$\{\log(\sigma_{jt}^y)^2\}$,$\alpha_j^y$, β_j^y, and τ_j^y) and the OLS parameter estimates from the forecasting model (i.e., ϕ_j^y, $\gamma_j^F(L)$, and $\gamma_j^W(L)$). This section assesses robustness of the results to these assumptions.

1.1 Macro Uncertainty Factor

We first entertain the possibility that uncertainty has a factor structure. In such a case, macro uncertainty at each t is a vector given by the common factor $F_t^{\mathcal{U}}(h)$ in

$$\log \mathcal{U}_{jt}^y(h) = c_j^{\mathcal{U}}(h) + \Lambda_{hj}^{\mathcal{U}\prime} F_t^{\mathcal{U}}(h) + e_{jt}^{\mathcal{U}}(h). \tag{1}$$

Macro uncertainty is then summarized by $F_t^{\mathcal{U}}(h)$ while idiosyncratic uncertainty is $e_{hjt}^{\mathcal{U}}$. Although $\mathcal{U}_{jt}^y(h)$ is always positive, the principal components estimates do not constrain the (normalized) estimated factors themselves to be positive. The log specification is therefore used to insure that both the domain and the range of the function (1) take on values on the entire real line $\mathbb{R}$. As a consequence of this log specification, our PCA estimate of macro uncertainty $\mathcal{U}_t^y(h)$ is the exponential of the PCA estimate $\widehat{F}_t^{\mathcal{U}}(h)$. Let $\widehat{\mathcal{U}}_t^y(h) \equiv \exp\left(\widehat{F}_t^{\mathcal{U}}(h)\right)$. To obtain such an estimate, we first need an estimate of the the common (log) uncertainty factor $F_t^{\mathcal{U}}(h)$. As many uncertainty series appear non-stationary, this estimate is defined by $\widehat{F}_t^{\mathcal{U}}(h) = \sum_{k=2}^t \widehat{f}_k^{\mathcal{U}}(h)$, where $f_t^{\mathcal{U}}(h)$ is an $r_{\mathcal{U}} \times 1$ vector comprised of the $r_{\mathcal{U}}$ principal components of $\Delta \log \mathcal{U}_{jt}^y(h)$.[1] As discussed in Bai and Ng (2004), this differencing-recumulating approach ensures that the factors are consistently estimated when the idiosyncratic errors are potentially non-stationary. Because of the differencing, the initial value in the sample of the common uncertainty factor, $\hat{F}_1^{\mathcal{U}}(h)$, is not identified. We initialize $\hat{F}_1^{\mathcal{U}}(h)$ to the average level of (log) uncertainty across all N series; mathematically, $\frac{1}{N}\sum_{j=1}^N \log \mathcal{U}_{j1}^y(h)$.

The problem of determining $r_{\mathcal{U}}$, the number of common uncertainty factors $f^{\mathcal{U}}(h)$, is nonstandard because the individual uncertainty measures are themselves estimated. Existing criteria for determining the number of factors do not take the first step estimation error into account and will likely overestimate the number of factors. However, there is strong evidence of a factor structure as the largest eigenvalue of forecast error variance is distinctly large. In particular,

[1] We observe $\log \mathcal{U}_{jt}^y(h)$, a data matrix with T time-series observations and N cross-section observations. The first differenced data yield a $(T-1) \times N$ vector of stationary variables. Let $f^{\mathcal{U}}(h) \equiv \left(f_1^{\mathcal{U}}(h), f_2^{\mathcal{U}}(h), ..., f_T^{\mathcal{U}}(h)\right)$ and $\Lambda_{\mathcal{U}} = (\Lambda_{\mathcal{U}\infty}, \Lambda_{\mathcal{U}\in}, ...\Lambda_{\mathcal{U}\mathcal{N}})'$. The principal component estimator of $f^{\mathcal{U}}(h)$ is the $T-1$ times the $r_{\mathcal{U}}$ eigenvectors corresponding to the first $r_{\mathcal{U}}$ largest eigenvalues of the $(T-1)(T-1)$ matrix $\left(\Delta \log \mathcal{U}_{jt}^y(h)\right)\left(\Delta \log \mathcal{U}_{jt}^y(h)\right)'$.

the first principal component of $\mathcal{U}^y_{jt}(h)$ explains 11% of the variance of the forecast errors for $h = 1$, 14% for $h = 3$, and 22% for $h = 12$. We take $r_{\mathcal{U}}$ to be one, which facilitates comparison with the base-case estimate $\overline{\mathcal{U}}^y_t(h)$ that is based on simple averaging. We also calibrate the uncertainty factor $\widehat{\mathcal{U}}^y_t(h)$ to have the same mean and standard deviation as $\overline{\mathcal{U}}_t(h)$ over the sample.

The right panel of Table 1 shows that the results using $\widehat{\mathcal{U}}^y_t(h)$ are qualitatively and quantitatively similar to the base-case. The relative importance of the uncertainty factor and idiosyncratic uncertainty is summarized in a $R^2_{jt}(h)$ statistic analogous to (**??**). The main finding continues to be that variations in macro uncertainty constitute a larger fraction of variations in individual uncertainty measures at longer horizons, and during recessions. Table 4 (second column) also reports results for the eight variable VAR, but with $\overline{\mathcal{U}}_t(h)$ replaced by recursive PCA estimates of uncertainty, $\widehat{\mathcal{U}}^y_t(h)$. The uncertainty factor has very similar dynamic effects on production, employment, and hours as $\overline{\mathcal{U}}_t(h)$. If anything, the effects due to the uncertainty are somewhat larger than the base-case of equal weighting.

1.2 Alternative Estimates of Uncertainty

We next consider alternative estimates of individual uncertainty, and alternative ways of aggregating these estimates to get macro uncertainty. The base-case implementation only requires one evaluation of uncertainty for each series j since the posterior mean of each parameter is one dimensional. Specifically, for $h = 1$, uncertainty in the variable j evaluated at the sth Monte Carlo draw is

$$\mathcal{U}_{jst}(h)\left(\theta_{js}, x_{jst}\right) = \sqrt{\exp\left(\alpha_{js} + \tau^2_{js}/2 + \beta_{js}x_{jst}\right)},$$

where $x_{jst} \equiv \ln(\sigma^y_{jst})^2$. When the function above is evaluated at the posterior mean (over all $s = 1, ..., S$ draws) of the parameters, we denote that $\mathcal{U}_{jt}(h)\left(\overline{\theta}_j, \overline{x}_{jt}\right)$. In this notation, our base case uncertainty estimate for the series j is $\mathcal{U}_{jt}(h)\left(\overline{\theta}_j, \overline{x}_{jt}\right)$. But an uncertainty estimate can also be obtained for each draw of the hyperparameters in the model for series j. Thus one can also estimate $\mathcal{U}^y_{jt}(h)$ by the posterior mean of the draws of uncertainty for series j. In this case we define individual uncertainty as $\mathcal{U}^S_{jt}(h) = \frac{1}{S}\sum_{s=1}^S \mathcal{U}_{jst}(h)\left(\theta_{js}, x_{jst}\right)$, where the superscript S denotes all S draws are used in the computation.[2] Instead of the posterior mean, it is also possible to consider other location statistics. Let $\mathcal{U}^{[s]}_{jt}(h)$ be the s-th percentile draw in the sorted sequence of $\{\mathcal{U}_{jst}(h)\}_{s=1}^S$. If $[s]$ is 50, the median obtains. We use the 90th and the 10th percentiles of the posterior distribution of $\mathcal{U}_{jst}(h)\left(\theta_{js}, x_{jst}\right)$ to assess how extreme values of individual uncertainty affect aggregate uncertainty. These are denoted $\overline{\mathcal{U}}^{10}_t(h)$ and $\overline{\mathcal{U}}^{90}_t(h)$, respectively.

[2] To estimate the latter requires saving every posterior draw of $\mathcal{U}_{jst}(h)\left(\theta_{js}, x_{jst}\right)$ and is considerably more computationally demanding than the base-case where uncertainty is evaluated once at the mean of the parameters.

Since we have three ways of estimating individual uncertainties two ways of aggregating them, we have six measures of macro uncertainty summarized as follows:

$\mathcal{U}_t(h)$	Aggregator	$\mathcal{U}_{jt}(h)$
Baseline CSA: $\overline{\mathcal{U}}_t(h)$	CSA	$\mathcal{U}_{jt}(h)\left(\overline{\theta}_j, \overline{x}_{jt}\right)$
Baseline PCA: $\widehat{\mathcal{U}}_t(h)$	PCA	$\mathcal{U}_{jt}(h)\left(\overline{\theta}_j, \overline{x}_{jt}\right)$
Posterior Mean CSA: $\overline{\mathcal{U}}_t^S(h)$	CSA	$\frac{1}{S}\sum_{s=1}^{S}\mathcal{U}_{jst}(h)\left(\theta_{js}, x_{jst}\right)$
Posterior Mean PCA: $\widehat{\mathcal{U}}_t^S(h)$	PCA	$\frac{1}{S}\sum_{s=1}^{S}\mathcal{U}_{jst}(h)\left(\theta_{js}, x_{jst}\right)$
Posterior s-Percentile CSA: $\overline{\mathcal{U}}_t^{[s]}(h)$	CSA	$\mathcal{U}_{jt}^{[s]}(h)$
Posterior s-Percentile PCA: $\widehat{\mathcal{U}}_t^{[s]}(h)$	PCA	$\mathcal{U}_{jt}^{[s]}(h)$

where CSA stands for simple averaging over N_y series, and PCA stands for for the principal component of the N_y individual uncertainties constructed using the methodology as discussed above.

Figure (1) shows the baseline and posterior mean estimates of aggregate uncertainty when $h = 1$. Each of these measures are highly correlated with one another. Indeed, the estimates based on the average across draws of the parameters versus the posterior mean of the uncertainty draws are virtually indistinguishable. The estimates based on cross-section averaging are also very highly correlated with those based on the principal component estimates. Given the similarity between the CSA and PCA estimates, Figure (2) shows our base-case estimate of uncertainty $\overline{\mathcal{U}}_t(h)$, the CSA variant of $\overline{\mathcal{U}}_t^S(h)$, along with the CSA variant of $\overline{\mathcal{U}}_t^{10}(h)$ and $\overline{\mathcal{U}}_t^{90}(h)$. As for the above variations, different percentiles of the distribution have the effect of shifting our estimate of uncertainty by a constant amount only but do not much affect the dynamics of our uncertainty estimates. The 90th and 10th percentiles of the distribution have a correlation with our baseline estimate each in excess of 0.998. We conclude that results regarding the number of large uncertainty episodes, their timing, or their dynamic relation with economic activity are robust to using more extreme estimates of individual uncertainty. Overall, the results suggest that the findings reported above are not sensitive to using these alterative estimates of aggregate uncertainty.

1.3 Recursive Estimation

We next consider the sensitivity of the forecasting parameters ϕ_j^y, $\gamma_j^F(L)$, and $\gamma_j^W(L)$ to the estimation sample. Instead of full sample estimation (and hence in-sample forecasts), we also form out-of-sample forecasts for the monthly macro dataset.[3] This procedure involves fully recursive factor estimation and parameter estimation using data only through time t for forecasting at time $t+1$. Notice that, since the forecasting parameters evolve over time as new data

[3]This procedure closely follows the real-time simulation procedure of Stock and Watson (2002).

becomes available, such recursive forecasts are informative about the extent to which parameter instability in the conditional mean forecasting relation influences the uncertainty estimates. We use the first 10 years of data (t = 1,2,...,120, 1959:01-1969:01) as an initial estimation period to estimate both the factors and the parameters of the conditional mean (forecasting) regression, and to perform model selection. Next, the forecasting regressions are run over the period t =1959:01,...,1969:01, and the values of the regressors at t =1969:01 are used to forecast $y_{j1969:02}$. All parameters, factors and model selection criteria are then re-estimated from 1959:01 through 1969:02, and forecasts are recomputed for $y_{j1969:03}$, and so on, until the final out-of-sample forecast is made for $y_{j2011:12}$. Since our dataset has 622 months total, this leaves 502=622-120 forecast errors. The forecast error variances are used to compute $\overline{\mathcal{U}}_{jt}^{y}(h)$, and averaging over j gives macro uncertainty. The resulting uncertainty estimate is plotted in Figure 3 along with the original estimate. The measure is extremely highly correlated with that based on in-sample forecasts.[4] Although use of the full sample slightly under-states the level of uncertainty, it does an excellent job of capturing its time-series variation, only influencing the estimates by a constant amount. This shows that the use of in-sample versus out-of-sample forecasts has little bearing on the number of uncertainty episodes, their timing, or their dynamic relationship with economic activity. The variance decompositions for the 8 variable VAR using the recursive-based estimate of $\overline{\mathcal{U}}_{t}^{y}(1)$ are given in the first column of Table 4. The results are very similar to the base-case in Table 2.

[4]Note that this measure is feasible to compute only for $h = 1$. The multi-step ahead forecasts that are needed for uncertainty with $h > 1$ are computed once by rolling forward one-step ahead forecasts from the VAR. Recomputing the VAR in every time period would require recomputing uncertainty in every time period, which is not possible in reasonable time.

Variance Decompositions from VAR(12)			
$\mathcal{U}_{jt}(h)$ Recursively Estimated			
k	$\overline{\mathcal{U}}(1)$	$\widehat{\mathcal{U}}(1)$	VXO
Production			
1	3.78	4.05	0.31
3	6.61	7.41	0.52
12	17.53	20.45	4.92
∞	8.33	9.63	3.59
max	18.74	21.57	5.97
max k	10	10	9
Employment			
1	0.54	0.17	0.31
3	2.40	2.14	0.45
12	12.56	14.86	7.73
∞	6.38	8.21	4.43
max	12.66	14.86	8.41
max k	11	13	9
Hours			
1	6.00	5.35	0.01
3	8.75	8.75	1.50
12	15.43	16.12	7.05
∞	9.57	9.60	5.54
max	16.26	17.01	7.70
max k	10	9	9

Table 4: Eight-variable VAR(12) using the VXO Index, or a real-time version of $\overline{\mathcal{U}}_t^y(1)$ or $\widehat{\mathcal{U}}_t^y(1)$ as measure of uncertainty, estimated from the monthly macro dataset. Each VAR(12) contains, in the following order: log(S&P 500 Index), *uncertainty*, federal funds rate, log(wages), log(CPI), hours, log(employment), and log(industrial production). All shocks are a 4 standard deviation impulse, which is the same magnitude considered in Bloom (2009) Figure A.1. As in Bloom (2009), all variables are HP filtered, except for the uncertainty measures, which enter in raw levels. The data are monthly and span the period 1970:07-2011:12.

1.4 Alternative VAR Ordering

The VAR results thus far have used an ordering that puts uncertainty second in a list of eight variables, following Bloom (2009). Table 5 reports VAR variance decomposition results with uncertainty ordered last to allow uncertainty to respond contemporaneously to the five variables ordered after it. Figure 4 reports the impulse responses to orthogonal shocks created from a Cholesky decomposition of the VAR with this alternative ordering. Some variations previously attributed to uncertainty are now allocated to the orthogonalized innovations in the fed funds rate, wages, CPI, hours, employment, and industrial production. This is not surprising because our measure of uncertainty is contemporaneously correlated with these measures of economic activity, thus once we remove the variation in uncertainty that is attributable to these correlations, the effect is smaller. We again caution, however, that these results as well as the previous ones tell us only about dynamic correlations (not true causality) and differ only because of a change in the assumption about the timing of shocks. For the sake of comparison, the last column of Table 5 reports results with VXO ordered last. As documented earlier, stock market volatility and uncertainty are correlated but have significant independent variations. As expected, because our measures of uncertainty are more highly contemporaneously correlated with real activity than is VXO, the effect on production, employment, and hours attributed to uncertainty shocks is smaller compared to the results in Table 2 when uncertainty is ordered second. By contrast, the decomposition of forecast error variances to VXO shocks is not greatly affected by the ordering of VXO in the VAR, implying that VXO shocks are not as strongly contemporaneously correlated with the five real activity variables in the system as are our uncertainty estimates. These results reinforce the conclusion that the stock market can move significantly in the absence changes in fundamentals in the economy. It is thus not a good proxy for macroeconomic uncertainty, which we have found does move with these fundamentals.

Variance Decompositions from VAR(12)
Uncertainty Ordered Last

k	$\overline{\mathcal{U}}(1)$	$\overline{\mathcal{U}}(3)$	$\overline{\mathcal{U}}(12)$	VXO
Production				
1	0.00	0.00	0.00	0.00
3	1.16	1.31	1.03	1.04
12	6.18	8.95	6.11	5.84
∞	5.51	7.26	6.33	4.14
max	6.78	9.45	6.62	7.19
max k	10	10	10	8
Employment				
1	0.00	0.00	0.00	0.00
3	0.60	0.59	0.43	1.11
12	5.97	9.20	6.58	8.88
∞	4.99	7.03	6.18	5.18
max	6.05	9.20	6.58	9.61
max k	11	12	12	9
Hours				
1	0.00	0.00	0.00	0.00
3	1.42	1.57	0.89	1.70
12	5.82	8.00	5.56	7.12
∞	5.94	7.97	6.81	5.98
max	6.21	8.40	6.81	7.86
max k	8	10	38	8

Table 5: Eight-variable VAR(12) using the VXO Index or $\overline{\mathcal{U}}_t^y(h)$ for $h = 1, 3, 12$ as a measure of uncertainty, estimated from the monthly macro dataset. Each VAR(12) contains, in the following order: log(S&P 500 Index), federal funds rate, log(wages), log(CPI), hours, log(employment), log(industrial production), and *uncertainty*. As in Bloom (2009), all variables are HP filtered, except for the uncertainty measures, which enter in raw levels. The data are monthly and span the period 1960:07-2011:12.

2 Data Appendix

The first dataset, denoted X^m, is an updated version of the of the 132 mostly macroeconomic series used in Ludvigson and Ng (2010). The 132 macro series in X^m are selected to represent broad categories of macroeconomic time series: real output and income, employment and hours, real retail, manufacturing and trade sales, consumer spending, housing starts, inventories and inventory sales ratios, orders and unfilled orders, compensation and labor costs, capacity utilization measures, price indexes, bond and stock market indexes, and foreign exchange measures.

The 147 financial series in X^f consists of a number of indicators measuring the behavior of a broad cross-section of asset returns, as well as some aggregate financial indicators not included in the macro dataset. These data include valuation ratios such as the dividend-price ratio and earnings-price ratio, growth rates of aggregate dividends and prices, default and term spreads, yields on corporate bonds of different ratings grades, yields on Treasuries and yield spreads, and a broad cross-section of industry equity returns. Following Fama and French (1992), returns on 100 portfolios of equities sorted into 10 size and 10 book-market categories. The dataset X^f also includes a group of variables we call "risk-factors," since they have been used in cross-sectional or time-series studies to uncover variation in the market risk-premium. These risk-factors include the three Fama and French (1993) risk factors, namely the excess return on the market MKT_t, the "small-minus-big" (SMB_t) and "high-minus-low" (HML_t) portfolio returns, the momentum factor UMD_t, the bond risk premia factor of Cochrane and Piazzesi (2005), and the small stock value spread $R15 - R11$.

The raw data used to form factors are always transformed to achieve stationarity. In addition, when forming forecasting factors from the large macro and financial datasets, the raw data (which are in different units) are standardized before performing PCA. When forming common uncertainty from estimates of individual uncertainty, the raw data (which are in this case in the same units) are demeaned, but we do not divide by the observation's standard deviation before performing PCA.

Throughout, the factors are estimated by the method of static principal components (PCA). Specifically, the $T \times r_F$ matrix $\hat{F}_t$ is $\sqrt{T}$ times the r_F eigenvectors corresponding to the r_F largest eigenvalues of the $T \times T$ matrix $xx'/(TN)$ in decreasing order. In large samples (when $\sqrt{T}/N \to \infty$), Bai and Ng (2006) show that the estimates $\hat{F}_t$ can be treated as though they were observed in the subsequent forecasting regression. There is no need to correct standard errors for uncertainty in this estimate, unlike the generated regressor case analyzed in Pagan (1984) when N is fixed. This asymptotic result allows for time variation in the volatility of the forecast error.

Macro Dataset

This appendix lists the short name of each series in the macro dataset, its code in the source database, the transformation applied to the series, and a brief data description. All series are from the IHS Global Insights database, unless the source is listed (in parentheses) as FRED (St. Louis Federal Reserve Economic Data), BLS (Bureau of Labor Statistics), S (R. J. Shiller website), BEA (Bureau of Economic Analysis), IMF (IMF International Financial Statistics database), B (R Barnichon website), UM (Thomson Reuters/University of Michigan Surveys of Consumers) or AC (author's calculation). The data are available from 1959:01-2011:12.

Let X_{it} denote variable i observed at time t after e.g., logarithm and differencing transformation, and let X_{it}^A be the actual (untransformed) series. Let $\Delta = (1 - L)$ with $LX_{it} = X_{it-1}$. There are six possible transformations with the following codes:

1 Code lv: $X_{it} = X_{it}^A$.

2 Code Δlv: $X_{it} = X_{it}^A - X_{it-1}^A$.

3 Code $\Delta^2 lv$: $X_{it} = \Delta^2 X_{it}^A$.

4 Code ln: $X_{it} = ln(X_{it}^A)$.

5 Code Δln: $X_{it} = ln(X_{it}^A) - ln(X_{it-1}^A)$.

6 Code $\Delta^2 ln$: $X_{it} = \Delta^2 ln X_{it}^A$.

Group 1: Output and Income

No.	Gp	Short Name	Code	Tran	Descripton
1	1	PI	M_14386177	Δln	Personal Income
6	1	IP: total	M_116460980	Δln	Industrial Production Index - Total Index
7	1	IP: products	M_116460981	Δln	Industrial Production Index - Products, Total
8	1	IP: final prod	M_116461268	Δln	Industrial Production Index - Final Products
9	1	IP: cons gds	M_116460982	Δln	Industrial Production Index - Consumer Goods
10	1	IP: cons dble	M_116460983	Δln	Industrial Production Index - Durable Consumer Goods
11	1	IP: cons nondble	M_116460988	Δln	Industrial Production Index - Nondurable Consumer Goods
12	1	IP: bus eqpt	M_116460995	Δln	Industrial Production Index - Business Equipment
13	1	IP: matls	M_116461002	Δln	Industrial Production Index - Materials
14	1	IP: dble matls	M_116461004	Δln	Industrial Production Index - Durable Goods Materials
15	1	IP: nondble matls	M_116461008	Δln	Industrial Production Index - Nondurable Goods Materials
16	1	IP: mfg	M_116461013	Δln	Industrial Production Index - Manufacturing
17	1	IP: res util	M_116461276	Δln	Industrial Production Index - Residential Utilities
18	1	IP: fuels	M_116461275	Δln	Industrial Production Index - Fuels
19	1	NAPM prodn	M_110157212	lv	Napm Production Index
20	1	Cap util	M_116461602	Δlv	Capacity Utilization

Group 2: Labor Market

No.	Gp	Short Name	Code	Tran	Descripton
21	2	Help wanted indx	-	Δlv	Index Of Help-Wanted Advertising (B)
22	2	Help wanted/unemp	M_110156531	Δlv	Ratio of Help-Wanted Ads/No. Unemployed (AC)
23	2	Emp CPS total	M_110156467	Δln	Civilian Labor Force: Employed, Total
24	2	Emp CPS nonag	M_110156498	Δln	Civilian Labor Force: Employed, Nonagric.Industries
25	2	U: all	M_110156541	Δlv	Unemployment Rate: All Workers, 16 Years & Over
26	2	U: mean duration	M_110156528	Δlv	Unemp By Duration: Average Duration In Weeks
27	2	U < 5 wks	M_110156527	Δln	Unemploy By Duration: Persons Unempl Less Than 5 Wks
28	2	U 5-14 wks	M_110156523	Δln	Unemploy By Duration: Persons Unempl 5 To 14 Wks
29	2	U 15+ wks	M_110156524	Δln	Unemploy By Duration: Persons Unempl 15 Wks +
30	2	U 15-26 wks	M_110156525	Δln	Unemploy By Duration: Persons Unempl 15 To 26 Wks
31	2	U 27+ wks	M_110156526	Δln	Unemploy By Duration: Persons Unempl 27 Wks +
32	2	UI claims	M_15186204	Δln	Initial Claims for Unemployement Insurance
33	2	Emp: total	M_123109146	Δln	Employees On Nonfarm Payrolls: Total Private
34	2	Emp: gds prod	M_123109172	Δln	Employees On Nonfarm Payrolls - Goods-Producing
35	2	Emp: mining	M_123109244	Δln	Employees On Nonfarm Payrolls - Mining
36	2	Emp: const	M_123109331	Δln	Employees On Nonfarm Payrolls - Construction
37	2	Emp: mfg	M_123109542	Δln	Employees On Nonfarm Payrolls - Manufacturing
38	2	Emp: dble gds	M_123109573	Δln	Employees On Nonfarm Payrolls - Durable Goods
39	2	Emp: nondbles	M_123110741	Δln	Employees On Nonfarm Payrolls - Nondurable Goods
40	2	Emp: services	M_123109193	Δln	Employees On Nonfarm Payrolls - Service-Providing
41	2	Emp: TTU	M_123111543	Δln	Employees On Nonfarm Payrolls - Trade, Transportation, And Utilities
42	2	Emp: wholesale	M_123111563	Δln	Employees On Nonfarm Payrolls - Wholesale Trade.
43	2	Emp: retail	M_123111867	Δln	Employees On Nonfarm Payrolls - Retail Trade
44	2	Emp: FIRE	M_123112777	Δln	Employees On Nonfarm Payrolls - Financial Activities
45	2	Emp: Govt	M_123114411	Δln	Employees On Nonfarm Payrolls - Government
*46	2	Agg wkly hours	-	Δlv	Index of Aggregate Weekly Hours (BLS)
*47	2	Avg hrs	M_140687274	Δlv	Avg Weekly Hrs of Prod or Nonsup Workers Private Nonfarm - Goods-Producing
*48	2	Overtime: mfg	M_123109554	Δlv	Avg Weekly Hrs of Prod or Nonsup Workers Private Nonfarm - Mfg Overtime
*49	2	Avg hrs: mfg	M_14386098	Δlv	Average Weekly Hours, Mfg.
50	2	NAPM empl	M_110157206	lv	NAPM Employment Index
129	2	AHE: goods	M_123109182	$\Delta^2 ln$	Avg Hourly Earnings of Prod or Nonsup Workers Private Nonfarm - Goods-Produ
130	2	AHE: const	M_123109341	$\Delta^2 ln$	Avg Hourly Earnings of Prod or Nonsup Workers Private Nonfarm - Construction
131	2	AHE: mfg	M_123109552	$\Delta^2 ln$	Avg Hourly Earnings of Prod or Nonsup Workers Private Nonfarm - Manufacturi

Group 3: Housing

No.	Gp	Short Name	Code	Tran	Descripton
*51	3	Starts: nonfarm	M_110155536	Δln	Housing Starts:Nonfarm(1947-58);Total Farm&Nonfarm(1959-)
*52	3	Starts: NE	M_110155538	Δln	Housing Starts:Northeast
*53	3	Starts: MW	M_110155537	Δln	Housing Starts:Midwest
*54	3	Starts: South	M_110155543	Δln	Housing Starts:South
*55	3	Starts: West	M_110155544	Δln	Housing Starts:West
*56	3	BP: total	M_110155532	Δln	Housing Authorized: Total New Priv Housing Units
*57	3	BP: NE	M_110155531	Δln	Houses Authorized By Build. Permits:Northeast
*58	3	BP: MW	M_110155530	Δln	Houses Authorized By Build. Permits:Midwest
*59	3	BP: South	M_110155533	Δln	Houses Authorized By Build. Permits:South
*60	3	BP: West	M_110155534	Δln	Houses Authorized By Build. Permits:West

Group 4: Consumption, Orders, and Inventories

No.	Gp	Short Name	Code	Tran	Descripton
61	4	PMI	M_110157208	lv	Purchasing Managers' Index
62	4	NAPM new ordrs	M_110157210	lv	Napm New Orders Index
63	4	NAPM vendor del	M_110157205	lv	Napm Vendor Deliveries Index
64	4	NAPM Invent	M_110157211	lv	Napm Inventories Index
65	4	Orders: cons gds	M_14385863	Δln	Mfrs' New Orders, Consumer Goods And Materials
66	4	Orders: dble gds	M_14386110	Δln	Mfrs' New Orders, Durable Goods Industries
67	4	Orders: cap gds	M_178554409	Δln	Mfrs' New Orders, Nondefense Capital Goods
68	4	Unf orders: dble	M_14385946	Δln	Mfrs' Unfilled Orders, Durable Goods Indus.
69	4	M&T invent	M_15192014	Δln	Manufacturing And Trade Inventories
70	4	M&T invent/sales	M_15191529	Δlv	Ratio, Mfg. And Trade Inventories To Sales
3	4	Consumption	M_123008274	Δln	Real Personal Consumption Expenditures (AC)
4	4	M&T sales	M_110156998	Δln	Manufacturing And Trade Sales
5	4	Retail sales	M_130439509	Δln	Sales Of Retail Stores
132	4	Consumer expect	hhsntn	Δlv	U. Of Mich. Index Of Consumer Expectations (UM)

Group 5: Money and Credit

No.	Gp	Short Name	Code	Tran	Descripton
71	5	M1	M_110154984	$\Delta^2 ln$	Money Stock: M1
72	5	M2	M_110154985	$\Delta^2 ln$	Money Stock: M2
73	5	Currency	M_110155013	$\Delta^2 ln$	Money Stock: Currency held by the public
74	5	M2 (real)	M_110154985	Δln	Money Supply: Real M2 (AC)
75	5	MB	M_110154995	$\Delta^2 ln$	Monetary Base, Adj For Reserve Requirement Changes
76	5	Reserves tot	M_110155011	$\Delta^2 ln$	Depository Inst Reserves:Total, Adj For Reserve Req Chgs
77	5	Reserves nonbor	M_110155009	$\Delta^2 ln$	Depository Inst Reserves:Nonborrowed,Adj Res Req Chgs
78	5	C&I loans	BUSLOANS	$\Delta^2 ln$	Commercial and Industrial Loans at All Commercial Banks (FRED)
79	5	C&I loans	BUSLOANS	lv	Change in Commercial and Industrial Loans at All Commercial Banks (FRED)
80	5	Cons credit	M_110155009	$\Delta^2 ln$	Consumer Credit Outstanding - Nonrevolving
81	5	Inst cred/PI	M_110154569	Δlv	Ratio, Consumer Installment Credit To Personal Income

Group 6: Bond and Exchange Rates

No.	Gp	Short Name	Code	Tran	Descripton
86	6	Fed Funds	M_110155157	Δlv	Interest Rate: Federal Funds
87	6	Comm paper	CPF3M	Δlv	3-Month AA Financial Commercial Paper Rate (FRED)
88	6	3 mo T-bill	M_110155165	Δlv	Interest Rate: U.S.Treasury Bills,Sec Mkt,3-Mo.
89	6	6 mo T-bill	M_110155165	Δlv	Interest Rate: U.S.Treasury Bills,Sec Mkt,6-Mo.
90	6	1 yr T-bond	M_110155165	Δlv	Interest Rate: U.S.Treasury Const Maturities,1-Yr.
91	6	5 yr T-bond	M_110155174	Δlv	Interest Rate: U.S.Treasury Const Maturities,5-Yr.
92	6	10 yr T-bond	M_110155169	Δlv	Interest Rate: U.S.Treasury Const Maturities,10-Yr.
93	6	Aaa bond	M_14386682	Δlv	Bond Yield: Moody's Aaa Corporate
94	6	Baa bond	M_14386683	Δlv	Bond Yield: Moody's Baa Corporate
95	6	CP-FF spread	-	lv	CP-FF spread (AC)
96	6	3 mo-FF spread	-	lv	3 mo-FF spread (AC)
97	6	6 mo-FF spread	-	lv	6 mo-FF spread (AC)
98	6	1 yr-FF spread	-	lv	1 yr-FF spread (AC)
99	6	5 yr-FF spread	-	lv	5 yr-FF spread (AC)
100	6	10 yr-FF spread	-	lv	10 yr-FF spread (AC)
101	6	Aaa-FF spread	-	lv	Aaa-FF spread (AC)
102	6	Baa-FF spread	-	lv	Baa-FF spread (AC)
103	6	Ex rate: avg	-	Δln	Nominal Effective Exchange Rate, Unit Labor Costs (IMF)
104	6	Ex rate: Switz	M_110154768	Δln	Foreign Exchange Rate: Switzerland - Swiss Franc Per U.S.$
105	6	Ex rate: Japan	M_110154768	Δln	Foreign Exchange Rate: Japan - Yen Per U.S.$
106	6	Ex rate: UK	M_110154772	Δln	Foreign Exchange Rate: United Kingdom - Cents Per Pound
107	6	EX rate: Canada	M_110154744	Δln	Foreign Exchange Rate: Canada - Canadian $ Per U.S.$

Group 7: Prices

No.	Gp	Short Name	Code	Tran	Descripton
108	7	PPI: fin gds	M_110157517	$\Delta^2 ln$	Producer Price Index: Finished Goods
109	7	PPI: cons gds	M_110157508	$\Delta^2 ln$	Producer Price Index: Finished Consumer Goods
110	7	PPI: int materials	M_110157527	$\Delta^2 ln$	Producer Price Index:I ntermed Mat.Supplies & Components
111	7	PPI: crude matâ€™ls	M_110157500	$\Delta^2 ln$	Producer Price Index: Crude Materials
112	7	Spot market price	M_110157273	$\Delta^2 ln$	Spot market price index: bls & crb: all commodities
113	7	PPI: nonferrous materials	M_110157335	$\Delta^2 ln$	Producer Price Index: Nonferrous Materials
114	7	NAPM com price	M_110157204	lv	Napm Commodity Prices Index
115	7	CPI-U: all	M_110157323	$\Delta^2 ln$	Cpi-U: All Items
116	7	CPI-U: apparel	M_110157299	$\Delta^2 ln$	Cpi-U: Apparel & Upkeep
117	7	CPI-U: transp	M_110157302	$\Delta^2 ln$	Cpi-U: Transportation
118	7	CPI-U: medical	M_110157304	$\Delta^2 ln$	Cpi-U: Medical Care
119	7	CPI-U: comm.	M_110157314	$\Delta^2 ln$	Cpi-U: Commodities
120	7	CPI-U: dbles	M_110157315	$\Delta^2 ln$	Cpi-U: Durables
121	7	CPI-U: services	M_110157325	$\Delta^2 ln$	Cpi-U: Services
122	7	CPI-U: ex food	M_110157328	$\Delta^2 ln$	Cpi-U: All Items Less Food
123	7	CPI-U: ex shelter	M_110157329	$\Delta^2 ln$	Cpi-U: All Items Less Shelter
124	7	CPI-U: ex med	M_110157330	$\Delta^2 ln$	Cpi-U: All Items Less Midical Care
125	7	PCE defl	gmdc	$\Delta^2 ln$	Pce, Impl Pr Defl:Pce (BEA)
126	7	PCE defl: dlbes	gmdcd	$\Delta^2 ln$	Pce, Impl Pr Defl:Pce; Durables (BEA)
127	7	PCE defl: nondble	gmdcn	$\Delta^2 ln$	Pce, Impl Pr Defl:Pce; Nondurables (BEA)
128	7	PCE defl: service	gmdcs	$\Delta^2 ln$	Pce, Impl Pr Defl:Pce; Services (BEA)

Group 8: Stock Market

No.	Gp	Short Name	Code	Tran	Descripton
82	8	S&P 500	M_110155044	Δln	S&P's Common Stock Price Index: Composite
83	8	S&P: indust	M_110155047	Δln	S&P's Common Stock Price Index: & Industrials
84	8	S&P div yield	-	Δlv	S&P's Composite Common Stock: Dividend Yield Real (S)
85	8	S&P PE ratio	-	Δln	S&P's Composite Common Stock: Price-Earnings Ratio Real (S)

Notes:

1. Series # 87, 104 and 105 were spliced with the data available on the previous data set.

2. Series # 3 and 74 were calculated dividing the series by # 125.

3. Series # 21 is a vacancy posting index built by R. Barnichon by combining the print help-wanted index and the on-line help-wanted index. See Barnichon, R. , Building a composite Help-Wanted Index, Economic Letters Dec 2010, for more details.

4. Series # 22 was computed dividing series # 21 by series M_110156531 of the IHS GI database.

5. Series # 84 was computed as D_t/P_t. Both Price and Dividends are real.

6. Series # 85 was computed as $P_t/AVERAGE\,(E_{t-1},...E_{t-12})$. Both Price and Earnings are real.

7. Series 125-128 (implicit price deflators) were calculated as (Nominal Cons / Real Cons) * 100. Real consumption is computed as: $\text{RealCons}_t = \text{RealCons}_{base} * \text{Qindex}_t/\text{Qindex}_{base}$. The quantity indices are from table 2.8.3. The Base is Jan 2005, Real Consumption for the base comes from table 2.8.6. The Nominal consumption comes from table 2.8.5.

Data for VAR Analysis

Monthly VAR Data

REX 3M: Log Excess Equity return, NSA (CRSP and Board of Governors)

The log equity return is the VWRETD series obtained from CRSP. For each month, we create the quarterly return by adding over the log return for that month and the following two months.

To obtain the quarterly excess return, we subtract the 3-month log t-bill return (secondary market), obtained from the Board of Governors via FRED (series name: TB3MS).

For example, the January excess return is defined as the sum of the January, February, and March log equity returns, minus the log 3-month t-bill return for January.

Log returns are multiplied by 100 to express in percent.

REX 1Y: Log 1-year excess return.

Equity return is obtained by compounding the log of the CRSP series VWRETD over 12 consecutive months and subtracting off the 1-year log T-Bill return.

For example, a January observation is given by the sum of January through December equity returns, minus the January T-Bill return.

The 1-year T-Bill series is the constant maturity series, obtained from the Board of Governors, via FRED (series name: GS1).

REX 5Y: Log 5-year excess return.

Equity return is obtained by compounding the log of the CRSP series VWRETD over 60 consecutive months and subtracting off the 1-year log T-Bill return.

For example, a January observation is given by the sum of January through December five years hence equity returns, minus the January T-Bill return of the initial year.

The 5-year T-Bill series is the constant maturity series, obtained from the Board of Governors, via FRED (series name: GS5).

FEDFUNDS: Log Effective Federal Funds Rate, NSA (Board of Governors)

Obtained via FRED (series name: FEDFUNDS).

Log returns are multiplied by 100 to express in percent.

EARN_ALL: Average Hourly Earnings of Production and Nonsupervisory Employees: Total Private, SA (BLS)

Obtained via FRED (series name: AHETPI).

EARN_MAN: Average Hourly Earnings Of Production And Nonsupervisory Employees: Manufacturing, SA (BLS)

Obtained via FRED (series name: AHEMAN).

CPI: Consumer Price Index for All Urban Consumers: All Items (BLS)

Obtained via FRED (series name: CPIAUSCL).

HOURS_ALL: Average Weekly Hours Of Production And Nonsupervisory Employees: Total Private, SA (BLS)

Obtained via FRED (series name: AWHNONAG).

HOURS_MAN: Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing, SA (BLS)

Obtainted via FRED (series name: AWHMAN).

EMP_ALL: All Employees: Total Private Industries, SA (BLS)

Obtained via FRED (series name: USPRIV).

EMP_MAN: All Employees: Manufacturing, SA (BLS)

Obtained via FRED (series name: MANEMP)

IP_ALL: Industrial Production Index, SA (Board of Governors)

Obtained via FRED (series name: INDPRO).

IP_MAN: Industrial Production: Manufacturing (NAICS) (Board of Governors)

Obtained via FRED (series name: IPMAN)

Note: to obtain the quarterly series, run the Matlab file to_quarterly.m.

This will draw from individual csv files for each monthly series.

Quarterly dates are expressed as the month in the BEGINNING of the quarter (i.e. Jan for Q1).

Variables in QDATA.xls:

REX 3M: Log Excess Equity return, NSA (CRSP and Board of Governors)

The log equity return is the quarterly VWRETD series obtained from CRSP. For each month, we create the quarterly return by adding over the log return for that month and the following two months.

Quarterly VAR Data

To obtain the quarterly excess return, we subtract the 3-month log t-bill return (secondary market), obtained from the Board of Governors via FRED (series name: TB3MS).

For example, the Q1 log excess return is the annualized Q1 quarterly log equity return, minus the log 3-month t-bill return for January of that year.

Log returns are multiplied by 100 to express in percent.

REX 1Y: Log 1-year excess return.

Equity return is obtained by compounding the log of the quarterly CRSP series VWRETD over 12 consecutive months and subtracting off the 1-year log T-Bill return.

For example, a January observation is given by the sum of January through December equity returns, minus the January T-Bill return.

The 1-year T-Bill series is the constant maturity series, obtained from the Board of Governors, via FRED (series name: GS1).

For example, the Q1 log excess return is the compounded Q1-Q4 quarterly log equity return, minus the log 1 year t-bill return for January of that year.

Log returns are multiplied by 100 to express in percent.

REX 5Y: Log 5-year excess return.

Equity return is obtained by compounding the log of the quarterly CRSP series VWRETD over 60 consecutive months and subtracting off the 1-year log T-Bill return.

For example, a January observation is given by the sum of January through December five years hence equity returns, minus the January T-Bill return of the initial year.

The 5-year T-Bill series is the constant maturity series, obtained from the Board of Governors, via FRED (series name: GS5).

For example, the Q1 log excess return is the compounded quarterly log equity return over 5 years annualized, minus the annualized log 5 year t-bill return for January of that year.

Log returns are multiplied by 100 to express in percent.

FEDFUNDS: Log Effective Federal Funds Rate, Not Seasonally Adjusted (Board of Governors)

Obtained via FRED (series name: FEDFUNDS).

Quarterly log returns are obtained by averaging monthly log returns over the quarter.

Log returns are multiplied by 100 to express in percent.

EARN_ALL: Average Hourly Earnings of Production and Nonsupervisory Employees: Total Private, Seasonally Adjusted (BLS)

Obtained via FRED (series name: AHETPI).

Quarterly series is obtained by averaging over the quarter.

EARN_MAN: Average Hourly Earnings Of Production And Nonsupervisory Employees: Manufacturing, SA (BLS)

Obtained via FRED (series name: AHEMAN).

Quarterly series is obtained by averaging over the quarter.

CPI: Consumer Price Index for All Urban Consumers: All Items (BLS)

Obtained via FRED (series name: CPIAUSCL).

Quarterly series is obtained by averaging over the quarter.

HOURS_ALL: Average Weekly Hours Of Production And Nonsupervisory Employees: Total Private, Seasonally Adjusted (BLS)

Obtained via FRED (series name: AWHNONAG).

Quarterly series is obtained by averaging over the quarter.

HOURS_MAN: Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing, SA (BLS)

Obtainted via FRED (series name: AWHMAN).

Quarterly series is obtained by averaging over the quarter.

EMP_ALL: All Employees: Total Private Industries, Seasonally Adjusted (BLS)

Obtained via FRED (series name: USPRIV).

Quarterly series is obtained by averaging over the quarter.

EMP_MAN: All Employees: Manufacturing, SA (BLS)

Obtained via FRED (series name: MANEMP)

Quarterly series is obtained by averaging over the quarter.

GDP: Real Gross Domestic Product, 1 Decimal, Seasonally Adjusted Annual Rate (BEA)

Obtained via FRED (series name: GDPC1).

Financial Dataset

The data set is at monthly frequency, with 147 observations spanning the period 1960:01-2013:01. All returns and spreads are expressed in logs (i.e. the log of the gross return or spread), are displayed in percent (i.e. multiplied by 100), and are annualized by multiplying by 12. Federal Reserve data are annualized by default and are therefore not "re-annualized." Note: this annualization means that the annualized standard deviation (volatility) is equal to the data standard deviation divided by $\sqrt{12}$. The data series used in this dataset are listed below by data source.

Source: CRSP

Value-weighted price and dividend data were obtained from the Center for Research in Security Prices (CRSP). From the Annual Update data, we obtain monthly value-weighted returns series vwretd (with dividends) and vwretx (excluding dividends). These series have the interpretation

$$VWRETD_t = \frac{P_{t+1} + D_{t+1}}{P_t}$$
$$VWRETX_t = \frac{P_{t+1}}{P_t}$$

From these series, a normalized price series P, can be constructed using the recursion

$$P_0 = 1$$
$$P_t = P_{t-1} \cdot VWRETX_t.$$

A dividend series can then be constructed using

$$D_t = P_{t-1}(VWRETD_t - VWRETX_t).$$

In order to remove seasonality of dividend payments from the data, instead of D_t we use the series

$$D_t^* = \frac{1}{4}(D_t + D_{t-1} + D_{t-2} + D_{t-3}.$$

For the price and dividend series under "reinvestment," we calculate the price under reinvestment, P_t^{re}, as the normalized value of the market portfolio under reinvestment of dividends, using the recursion

$$P_0^{re} = 1$$
$$P_t^{re} = P_{t-1} \cdot VWRETD_t$$

Similarly, we can define dividends under reinvestment, D_t^{re}, as the total dividend payments on this portfolio (the number of "shares" of which have increased over time) using

$$D_t^{re} = P_{t-1}^{re}(VWRETD_t - VWRETX_t).$$

As before, we can remove seasonality by using

$$D_t^{re,*} = \frac{1}{4}(D_t^{re} + D_{t-1}^{re} + D_{t-2}^{re} + D_{t-3}^{re}$$

Five data series are constructed from the CRSP data as follows:

- D_log(DIV): $\Delta \log D_t^*$.

- D_log(P): $\Delta \log P_t$.
- D_DIVreinvest: $\Delta \log D_t^{re,*}$
- D_Preinvest: $\Delta \log P_t^{re,*}$
- d-p: $\log(D_t^*) - \log(P_t)$

Source: Monika Piazzesi, Stanford University

The Cochrane-Piazzesi factor (Cochrane and Piazzesi (2005)), which forms the data series CP, was obtained directly from Monika Piazzesi.

Source: Kenneth French, Tuck School of Business, Dartmouth College

The following data are obtained from the data library of Kenneth French's website:

- Fama/French Factors: From this dataset we obtain the data series RF, Mkt-RF, SMB, HML.
- 25 Portolios formed on Size and Book-to-Market (5 x 5): From this dataset we obtain the series R15-R11, which is the spread between the (small, high book-to-market) and (small, low book-to-market) portfolios.
- Momentum Factor (Mom): From this dataset we obtain the series UMD, which is equal to the momentum factor.
- 49 Industry Porfolios: From this dataset we use all value-weighted series, excluding any series that have missing observations from Jan. 1960 on, from which we obtain the series Agric through Other. The omitted series are: Soda, Hlth, FabPr, Guns, Gold, Softw.
- 100 Portfolios formed in Size and Book-to-Market: From this dataset we use all value-weighted series, excluding any series that have missing observations from Jan. 1960 on. This yields variables with the name X_Y where X stands for the index of the size variable (1, 2, ..., 10) and Y stands for the index of the book-to-market variable (Low, 2, 3, ..., 8, 9, High). The omitted series are 1_low, 1_3, 7_high, 9_9, 10_8, 10_9, 10_high.

Firm-level Dataset

Firm level observations are from COMPUSTAT Fundamentals Quarterly dataset. The unit of observation is the change in firm pre-tax profits $P_{i,t}$, normalized by a two-period moving average of sales, $S_{i,t}$, following Bloom (2009). Bloom constructs

$$dpretax_{i,t} = (P_{it} - P_{it-1}) \,/\, (0.5 \cdot S_{it} + 0.5 \cdot S_{it-1}), \tag{2}$$

for each firm i in quarter t. This is the same measure reported on in Bloom (2009), Table 1, and discussed in footnote c. We find, however, that (2) exhibits clear seasonality patterns, thus we instead use year-over-year changes for the variable (2), normalized by average sales:

$$Y_{i,t} = dpretaxy_{i,t} = (P_{it} - P_{it-4}) \,/\, (0.5 \cdot S_{it} + 0.5 \cdot S_{t-4}), \tag{3}$$

We follow the trimming procedures used by Bloom, which includes considering any observation with sales $S = 0$ a missing value, and windsorizing observations at the top and bottom 0.05% values (replacing values in the top and bottom 0.05% with the values at the 0.05th and 99.95th percentile values).[5] After converting to a balanced panel, we are left with 155 firms from 1970:Q1-2011:Q2 without missing values.

These variables are constructed from COMPUSTAT Fundamentals Quarterly dataset. It contains 155 firms observed from 1970Q1 to 2011Q2 that have non-missing observations for $P_{i,t}$ (Compustat identifier *piq*) and $S_{i,t}$ (Compustat identifier for net sales *saleq*) across the entire time period.[6]

- gvkey: firm identifier
- date: period (1 to 166)
- *dpretax*: quarterly change in pretax profits scaled by average sales in current and past quarter:
$$dpretax_{i,t} = \frac{piq_{i,t} - piq_{i,t-1}}{0.5\,(saleq_{i,t} + saleq_{i,t-1})}.$$
- *dpretaxy*: year-over-year change in quarterly pretax profits scaled by average sales:
$$dpretaxy_{i,t} = \frac{piq_{i,t} - piq_{i,t-4}}{0.5\,(saleq_{i,t} + saleq_{i,t-4})}$$

[5] A detailed description of these procedures are given in the code to Bloom (2009) http://www.stanford.edu/~nbloom/Uncertainty_shocks_code.zip.

[6] This item represents operating and nonoperating income before provisions for income taxes and minority interest. Earnings (COMPUSTAT code ibq) are measured as the income of a company after all expenses, including special items, income taxes, and minority interest, but before provisions for common and/or preferred dividends. Formally: $ibq = piq - txt$ (income taxes) $-mii$ (minority interest).

References

Bai, J., and S. Ng (2004): "A PANIC Attack on Unit Roots and Cointegration," *Econometrica*, 72(4), 1127–1177.

——— (2006): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 74(4), 1133–1150.

Bloom, N. (2009): "The Impact of Uncertainty Shocks," *Econometrica*, 77, 623–685.

Cochrane, J. H., and M. Piazzesi (2005): "Bond Risk Premia," *American Economic Review*, 95(1), 138–160.

Fama, E. F., and K. R. French (1992): "The Cross-Section of Expected Returns," *Journal of Finance*, 47, 427–465.

——— (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

Ludvigson, S. C., and S. Ng (2010): "A Factor Analysis of Bond Risk Premia," in *Handbook of Empirical Economics and Finance*, ed. by A. Ulah, and D. E. A. Giles, vol. 1, pp. 313–372. Chapman and Hall, Boca Raton, FL.

Pagan, A. R. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25(1), 221–247.

Stock, J. H., and M. W. Watson (2002): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20(2), 147–162.

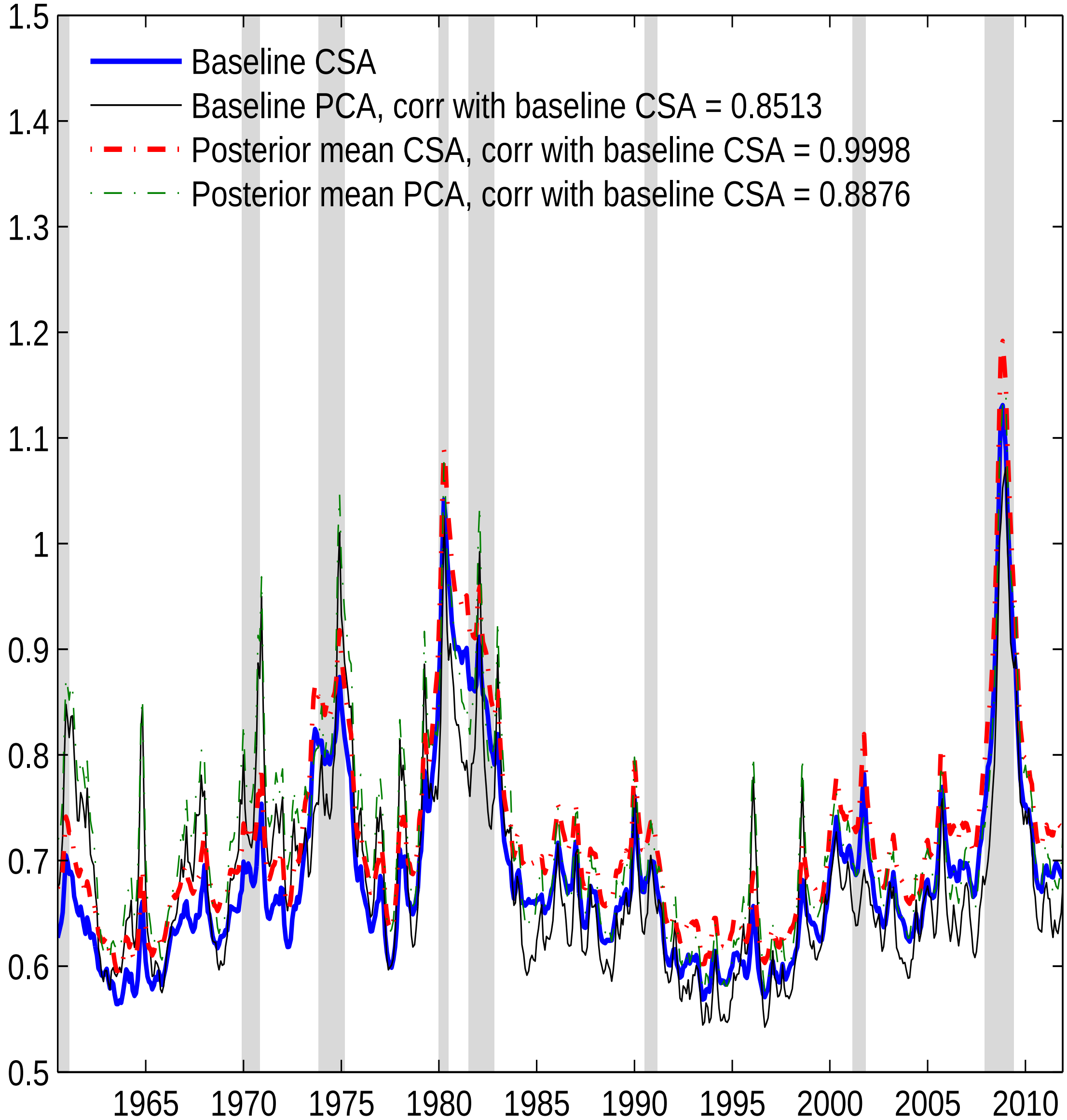

Figure 1: Different estimates of macro uncertainty when $h = 1$. Baseline CSA is $\overline{\mathcal{U}}_t(1) = \frac{1}{N_y}\sum_{j=1}^{N_y}\mathcal{U}_{jt}(1)\left(\overline{\theta}_j, \overline{x}_{jt}\right)$. Baseline PCA shows the principal component based on $\mathcal{U}_{jt}(1)\left(\overline{\theta}_j, \overline{x}_{jt}\right)$. Posterior mean CSA is the cross-section average of $\frac{1}{S}\sum_{s=1}^{S}\mathcal{U}_{jst}(1)\left(\theta_{js}, x_{jst}\right)$. Posterior mean PCA shows the first principal component based on $\frac{1}{S}\sum_{s=1}^{S}\mathcal{U}_{jst}(1)\left(\theta_{js}, x_{jst}\right)$. The full sample spans the period 1960:01-2011:12.

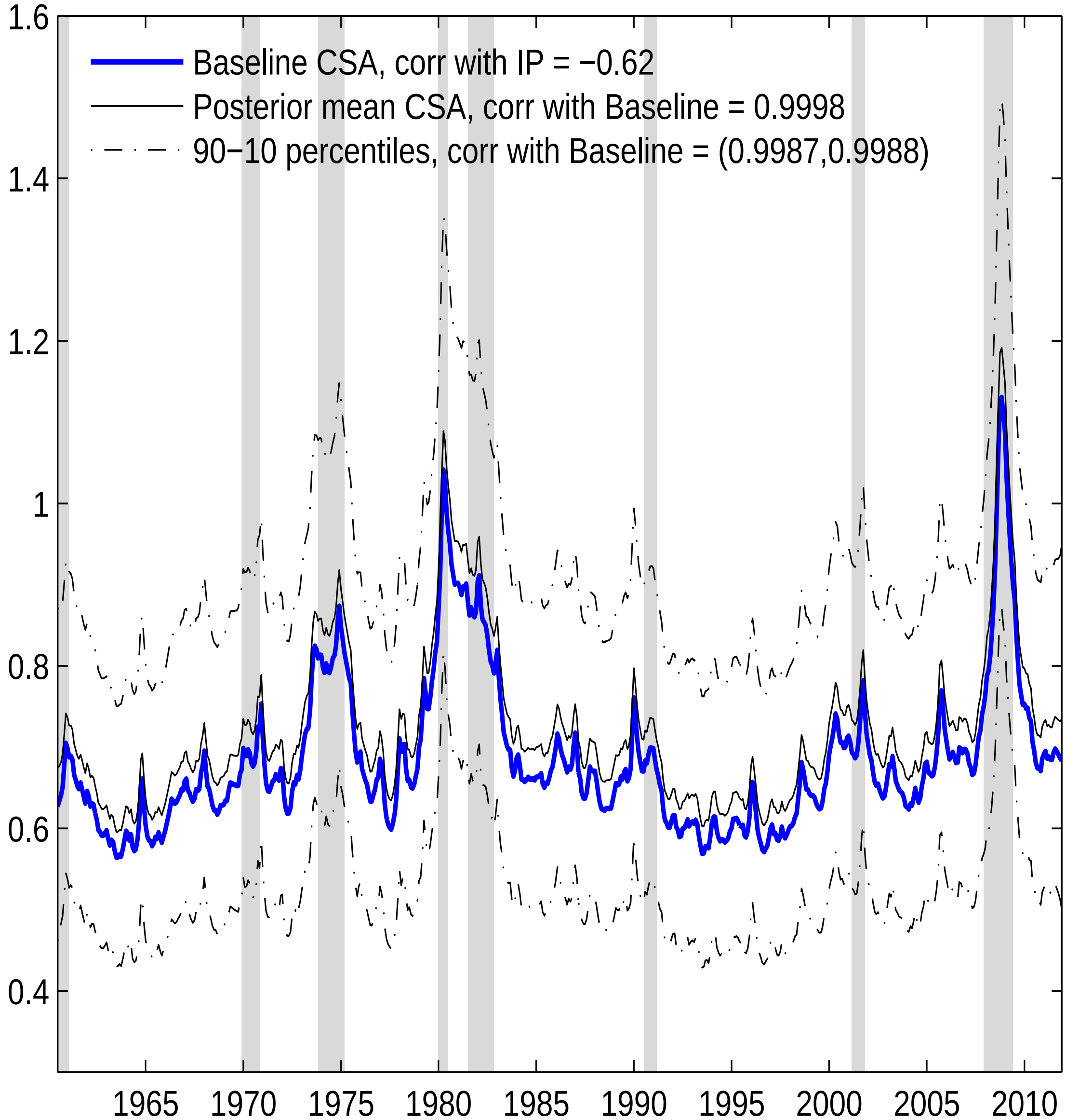

Figure 2: Percentile-based estimates of aggregate uncertainty when $h = 1$. Baseline denotes our base-case CSA estimate of macro uncertainty: $\overline{\mathcal{U}}_t(1) = \frac{1}{N_y}\sum_{j=1}^{N_y}\mathcal{U}_{jt}(1)\left(\overline{\theta}_j, \overline{x}_{jt}\right)$ and $\overline{\theta}_j$ and $\bar{x}_{jt}$ are posterior means over S draws. Posterior mean CSA is $\overline{\mathcal{U}}_t(1) = \frac{1}{N_y}\sum_{j=1}^{N_y}\frac{1}{S}\sum_{s=1}^{S}\mathcal{U}_{jst}(1)\left(\theta_{js}, x_{jst}\right)$. The posterior percentile-s CSA is $\overline{\mathcal{U}}_t(1) = \frac{1}{N_y}\sum_{j=1}^{N_y}\mathcal{U}_{jt}^{[s]}(1)$ where $\mathcal{U}_{jt}^{[s]}(1)$ is the s-th percentile draw in the ordered sequence of $\mathcal{U}_{jst}(1)(\theta_{js}, x_{jst})$, for $s = 1, \ldots, S$. The sample spans the period 1960:01-2011:12.

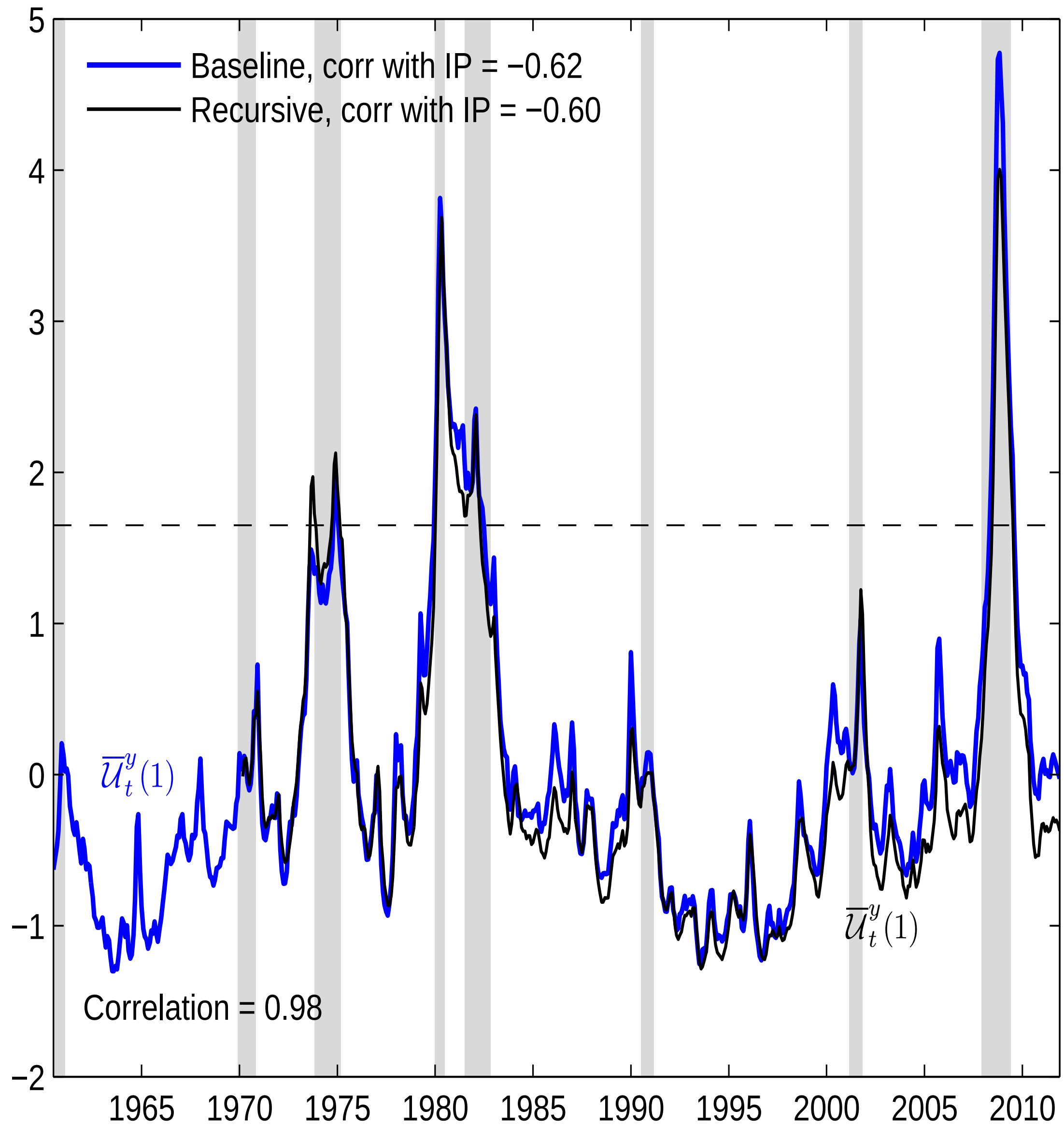

Figure 3: Uncertainty factor based on recursive forecasts. This plot displays $\overline{\mathcal{U}}_t^y(h)$ based on forecasts which use information from the full sample ("Baseline"), and based on recursively computed out-of-sample forecasts ("Real-time"), expressed in standardized units. The recursive forecasting procedure involves estimating model parameters and predictor variables only using information available up to time t. A training sample of 10 years (120 observations) is used to compute the first out-of-sample forecast, for 1970:01. The full sample spans the period 1960:01-2011:12.

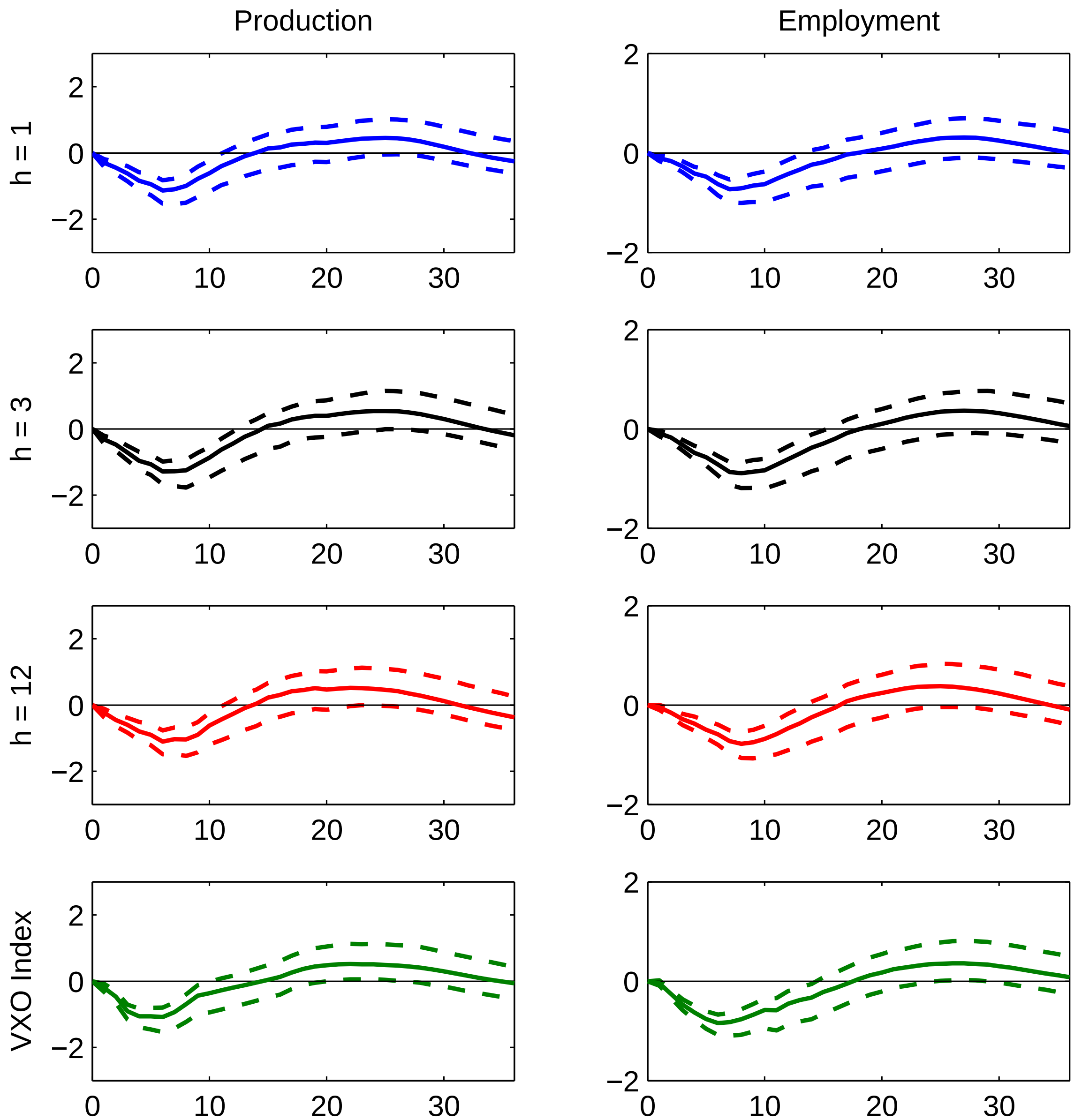

Figure 4: Eight-variable VAR(12) using the VXO Index or $\overline{\mathcal{U}}^y_t(h)$ for $h = 1, 3, 12$ as a measure of uncertainty. Each VAR(12) contains, in the following order: log(S&P 500 Index), federal funds rate, log(wages), log(CPI), hours, log(employment), log(industrial production), and *uncertainty*. All shocks are a 4 standard deviation impulse, which is the same magnitude considered in Bloom (2009) Figure A.1. As in Bloom (2009), all variables are HP filtered, except for the uncertainty measures, which enter in raw levels. The data are monthly and span the period 1960:07-2011:12.