MIDAS Estimation: Applications in Finance and Macroeconomics

Eric Ghysels

16^{th} $(EC)^2$ Conference

Istanbul 2005

Introduction

- The idea to construct regressions combining data with different sampling frequencies is explored. Think of combining annual and quarterly/monthly data, monthly/daily, daily/intradaily, etc.

- We call the regression framework a MI*xed* DA*ta* S*ampling* regression (MIDAS regression).

- MIDAS goes beyond estimation of linear regressions. Examples of filtering, panel data models, etc. will be given.

- Talk is structured around five examples. Each covering a different topic, each showing where MIDAS *makes a difference*.

- Lots of theoretical issue, but time constraint permits only focus on applications.

Motivating Examples

Example I: Risk-Return Trade-off

- The risk-return tradeoff involves the following regression:

$$R_{t+1} = \mu + \gamma \widehat{\sigma}_t^2 + \epsilon_{t+1}$$

where R_{t+1} is the excess return on the market in month $t+1$, and $\widehat{\sigma}_t^2$ is the forecasted variance of returns for the same month $t+1$, based on information known at time t.

- French et al. (1987) use within-month daily returns to estimate the realized variance in the period from $t-1$ to t (where typically $D = 22$, 44, 66, etc.):

$$\widehat{\sigma}_t^2 = \sum_{j=1}^{D} r_{t-j/22}^2$$

Results with CRSP VW excess returns - Jan. 1946 to Dec. 2000

Window (Months)	Conditional Mean Equation μ	γ	R^2
1	0.0107	-0.3422	0.0004
	(5.6932)	(-0.5365)	
2	0.0085	1.2330	0.0034
	(4.2150)	(1.5041)	
3	0.0073	2.0328	0.0072
	(3.4309)	(2.1725)	
12	0.0085	1.4310	0.0015
	(3.1820)	(0.9704)	

Example II: Predicting Realized Volatility

- Andersen, Bollerslev, Diebold and co-authors (2001 a,b, 2002), Andreou and Ghysels (2002), Barndorff-Nielsen and Shephard (2001, 2002 a,b, 2003), Taylor and Xu (1997), model realized volatility based on m intradaily returns.

Example III: Forecasting professional forecasters

- MIDAS ideally suited for exploiting high frequency financial data to predict low frequency macro data.
- One example is to 'forecast professional forecasters' (quarterly or monthly) using financial market data (daily or intra-daily).

Example IV: Macroeconomic sources of stock market volatility

- At least since Schwert (1989) the link between stock market volatility and macroeconomic fundamental risk is found to be weak.
- We revisit this with a GARCH/MIDAS framework.

Example V: News impact on the stock market

- The impact of macro and corporate news (low frequency event) on the entire cross section of individual stock returns (high frequency).
- This involves projecting low frequency data onto high frequency data.
- MIDAS regression allows us to explore this further for the *entire cross-section* (panel MIDAS approach).

Material is taken from papers downloadable at:

`http://www.unc.edu/~eghysels/`

- Example I (Risk-return) is taken from: *There is a Risk-Return Tradeoff After All* (*Journal of Financial Economics*, 2005), with P. Santa-Clara and R. Valkanov

- Example II (Predicting volatility) is taken from *Predicting Volatility: Getting the Most out of Data Sampled at Different Frequencies*(*Journal of Econometrics,* forthcoming), with P. Santa-Clara and R. Valkanov and *Why Is Realized Absolute Value Such A Good Predictor Of Volatility?* with Lars Forsberg

- Example III (Forecasting) is taken from: *Forecasting Professional Forecasters*, with J. Wright

- Example IV (Macro/Volatility) is taken from *Economic Sources of Volatility*, with R. Engle and B. Sohn

- Example V (News impact) is taken from *The Cross Section of Firm Stock Returns and Economic Announcements: A Bird's Eye View*, with A. Sinko and R. Valkanov

Related Theoretical Material

- *The MIDAS Touch: Mixed Data Sampling Regression Models*, with P. Santa-Clara and R. Valkanov

- *MIDAS Regressions: Further Results and New Directions*, with A. Sinko and R. Valkanov

- *Linear Time Series Processes with Mixed Data Sampling and MIDAS Regression Models* with R. Valkanov

- *Semiparametric MIDAS Regression*, with E. Renault.

Parameterizations of Polynomial Lag Structures

- Let us start with a simple linear regression framework.
- Suppose y_t is sampled at some fixed, say annual, quarterly, monthly or daily, frequency called interval of reference.
- Denote by $x_t^{(m)}$ a process sampled m times during interval of reference.
- We can write a simple linear MIDAS regression:

$$y_t = \beta_0 + \sum_{j=1}^{j^{max}} b(j,\theta) x_{t-j/m}^{(m)} + \varepsilon_t = \beta_0 + B(L^{1/m}) x_t^{(m)} + \varepsilon_t$$

Where $B(L^{1/m}) = b(0,\theta) + b(1,\theta)L^{1/m} + \ldots + b(j^{max},\theta)L^{j^{max}/m}$ is a polynomial of length j^{max} governed by small set of hyperparameters θ, and $L^{j/m}x_t^{(m)} = x_{t-j/m}^{(m)}$, such that $(L^{1/m})^m = L$

- Relates to distributed lag models

$$y_{t+1} = \beta_0 + \sum_{j=0}^{j^{max}} b(j,\theta) x_{t-j} + \varepsilon_{t+1} = \beta_0 + B(L) x_t + \varepsilon_{t+1}$$

where $B(L)$ is some finite or infinite lag polynomial operator, usually parameterized by a small set of hyperparameters θ.

- See e.g. Dhrymes (1971) and Sims (1974) for surveys on distributed lag models. Many econometrics textbooks also cover the topic, see e.g. Greene (2000, chap. 17), Judge et al. (1985, chap. 9 - 10), Stock and Watson (2003, chap. 13) Wooldridge (2000, chap. 18), among others.

Exponential Almon and Beta Polynomials

- We propose two parameterizations of $b(k;\theta)$.

- The first one is:

$$b(k;\theta) = \frac{e^{\theta_1 k + \ldots + \theta_Q k^Q}}{\sum_{k=1}^{k^{max}} e^{\theta_1 k + \ldots + \theta_Q k^Q}}$$

which we call the "Exponential Almon lag," since it is related to "Almon lags" (see e.g. Judge et al. 1985).

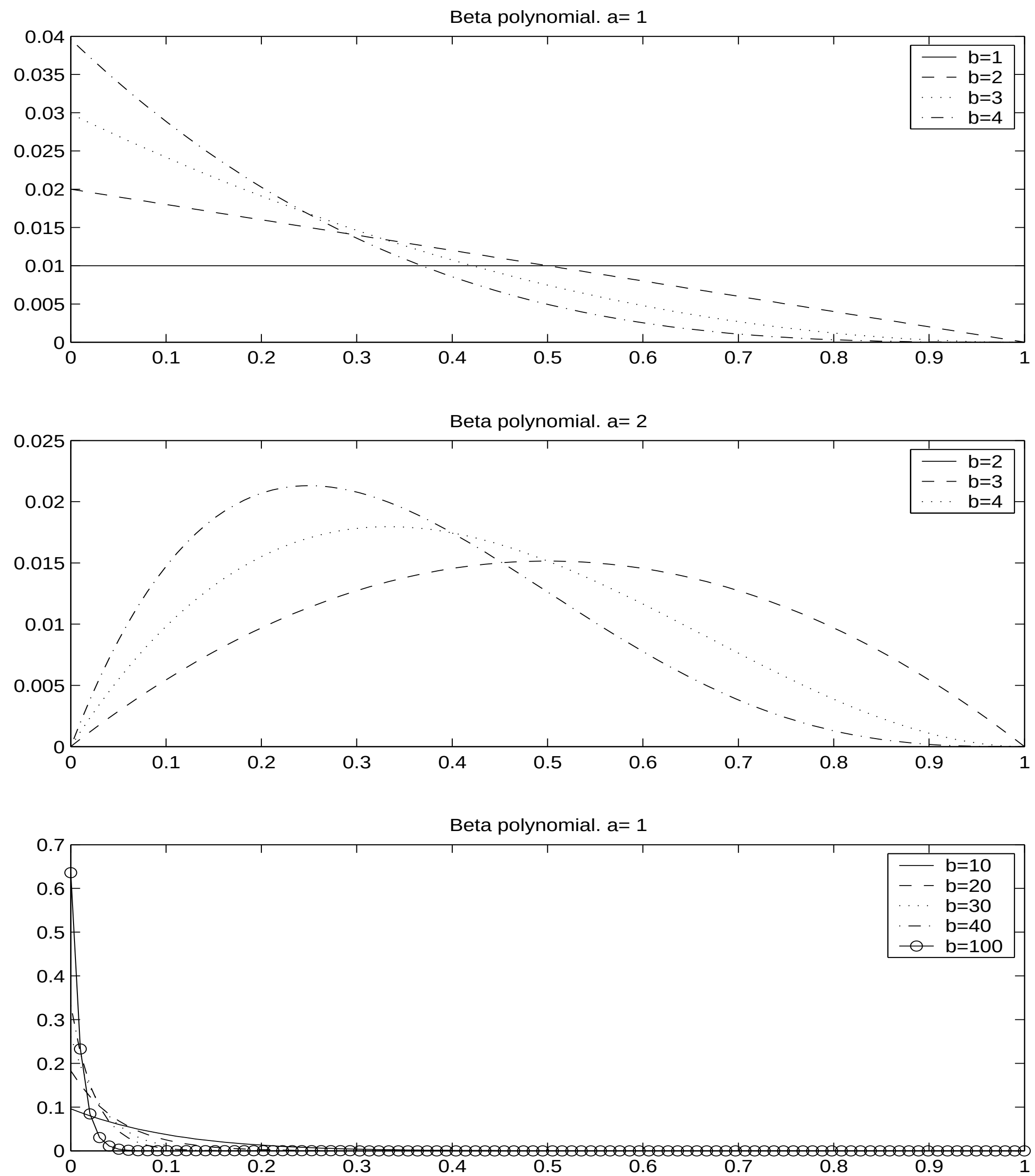
- The second parameterization has only two parameters, or $\theta = [\theta_1; \theta_2]$:

$$b(k; \theta_1, \theta_2) = \frac{f(\frac{j}{jmax}, \theta_1; \theta_2)}{\sum_{j=1}^{jmax} f(\frac{j}{jmax}, \theta_1; \theta_2)}$$

where:

$$\begin{aligned} f(x, a, b) &= \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} \\ B(a, b) &= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \\ \Gamma(a) &= \int_0^\infty e^{-x} x^{a-1} dx \end{aligned}$$

- Specification () has, to the best of our knowledge, not been used in the literature. It is based on the beta function and we refer to it as the "Beta lag."


Beta polynomial. a= 1
b=1
b=2
b=3
b=4
Beta polynomial. a= 2
b=2
b=3
b=4
Beta polynomial. a= 1
b=10
b=20
b=30
b=40
b=100

MIDAS with stepfunctions

- MIDAS with stepfunctions (special case is HAR (Heterogenous Autoregressive) Model (Corsi(2003)))

$$Y_{t+H} = \beta_0 + \beta_D X_{t-1,t} + \beta_W X_{t-5,t} + \beta_M X_{t-20,t} + \varepsilon_{t+H}$$

for X various regressors discussed later. The advantage of using stepfunctions is that one can use OLS, the disadvantage, is that parsimony may be gone.

Example I Revisited: Risk-Return Trade-off

- We estimate via QMLE the parameters θ_i jointly with μ and γ using MIDAS regression:

$$R_{t+1} \sim N\left(\mu + \gamma V_t^{\text{MIDAS}}, V_t^{\text{MIDAS}}\right)$$

where

$$V_t^{\text{MIDAS}} = 22 \sum_{d=1}^{\infty} w(d, \theta_1, \theta_2) r_{t-d}^2$$

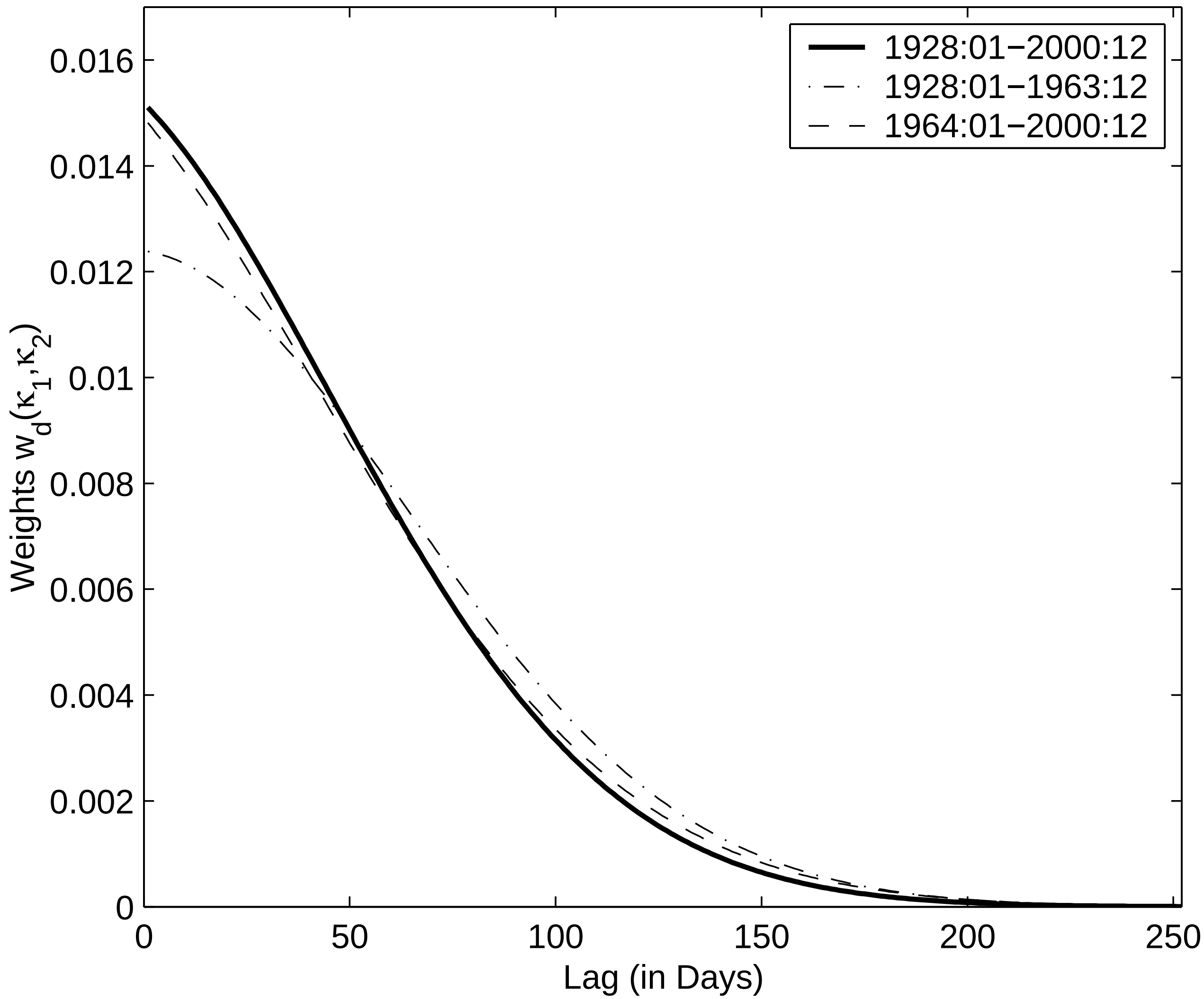
and

$$w(d, \theta_1, \theta_2) = \frac{\exp\{\theta_1 d + \theta_2 d^2\}}{\sum_{i=1}^{\infty} \exp\{\theta_1 i + \theta_2 i^2\}}$$

- In all the results that follow, we use the past 260 days as the maximum lag length (results are not sensitive to increasing the lag length beyond one year, nor to alternative polynomial specifications).

Risk Return Trade-off: MIDAS regression results

Sample	μ $(\times 10^3)$	γ	θ_1 $(\times 10^2)$	θ_2 $(\times 10^9)$	R^2_R	$R^2_{\sigma^2}$	LLF
1946.01-2000.12	4.800	4.007	-1.353	-3.984	0.024	0.082	1221.837
	[2.419]	[2.647]	[-1.903]	[-0.092]			
1946.01-2000.12	4.809	4.254	-1.402	-3.293	0.041	0.251	1239.100
(No 1987 Crash)	[2.515]	[2.950]	[-1.959]	[-0.011]			


1928:01−2000:12
1928:01−1963:12
1964:01−2000:12
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
Weights $w_d(\kappa_1,\kappa_2)$
0
50
100
150
200
250
Lag (in Days)

We estimate via QMLE the GARCH-M:

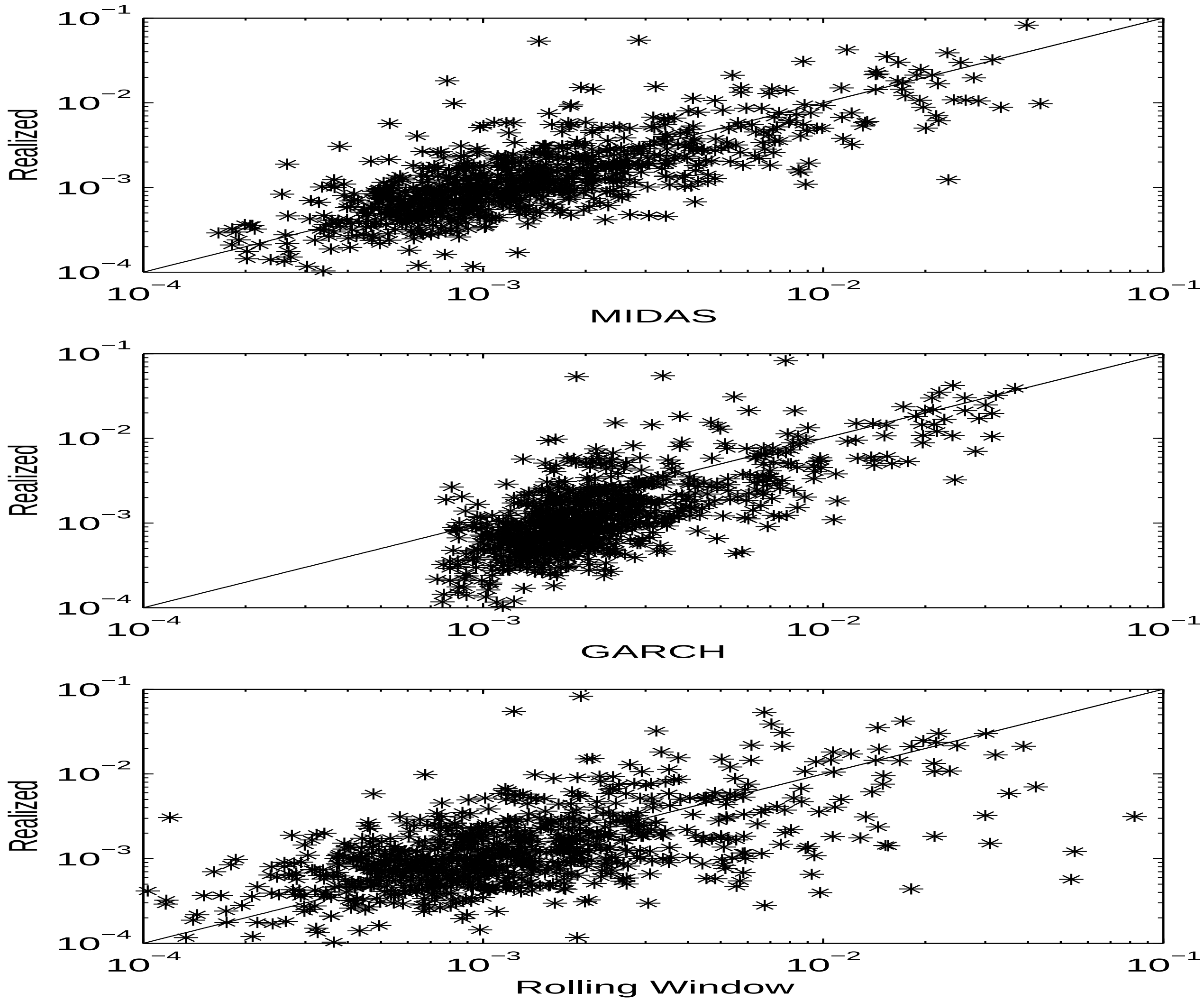
$$R_{t+1} \sim N\left(\mu + \gamma V_t^{\text{GARCH}}, V_t^{\text{GARCH}}\right)$$

where

$$V_t^{\text{GARCH}} = \frac{\omega}{1-\beta} + \alpha \sum_{i=0}^{\infty} \beta^i \epsilon_{t-i}^2$$

using past *monthly* squared returns. Results are consistent with Glosten, Jagannathan, and Runkle (1993) and others in the literature, namely:

Model	μ $(\times 10^3)$	γ	ω $(\times 10^3)$	α	β	R^2_R	$R^2_{\sigma^2}$	LLF
GARCH(1,1)-M	-0.740	6.968	0.125	0.069	0.860	0.010	0.070	1152.545
	[-0.370]	[0.901]	[0.244]	[1.398]	[18.323]			
ABS-GARCH(1,1)-M	1.727	6.013	2.751	0.099	0.858	0.010	0.071	1156.142
	[0.424]	[0.873]	[0.947]	[1.764]	[17.323]			


Realized
MIDAS
Realized
GARCH
Realized
Rolling Window

Asymmetries

To examine whether the risk-return tradeoff is robust to the inclusion of asymmetric effects in the conditional variance, we introduce the asymmetric MIDAS estimator:

$$V_t^{\text{ASYMIDAS}} = 22[\phi \textstyle\sum_{d=1}^{\infty} w(d, \theta_1^-, \theta_2^-) \mathbf{1}_{t-d}^- r_{t-d}^2 + \\ (2-\phi) \textstyle\sum_{d=1}^{\infty} w(d, \theta_1^+, \theta_2^+) \mathbf{1}_{t-d}^+ r_{t-d}^2]$$

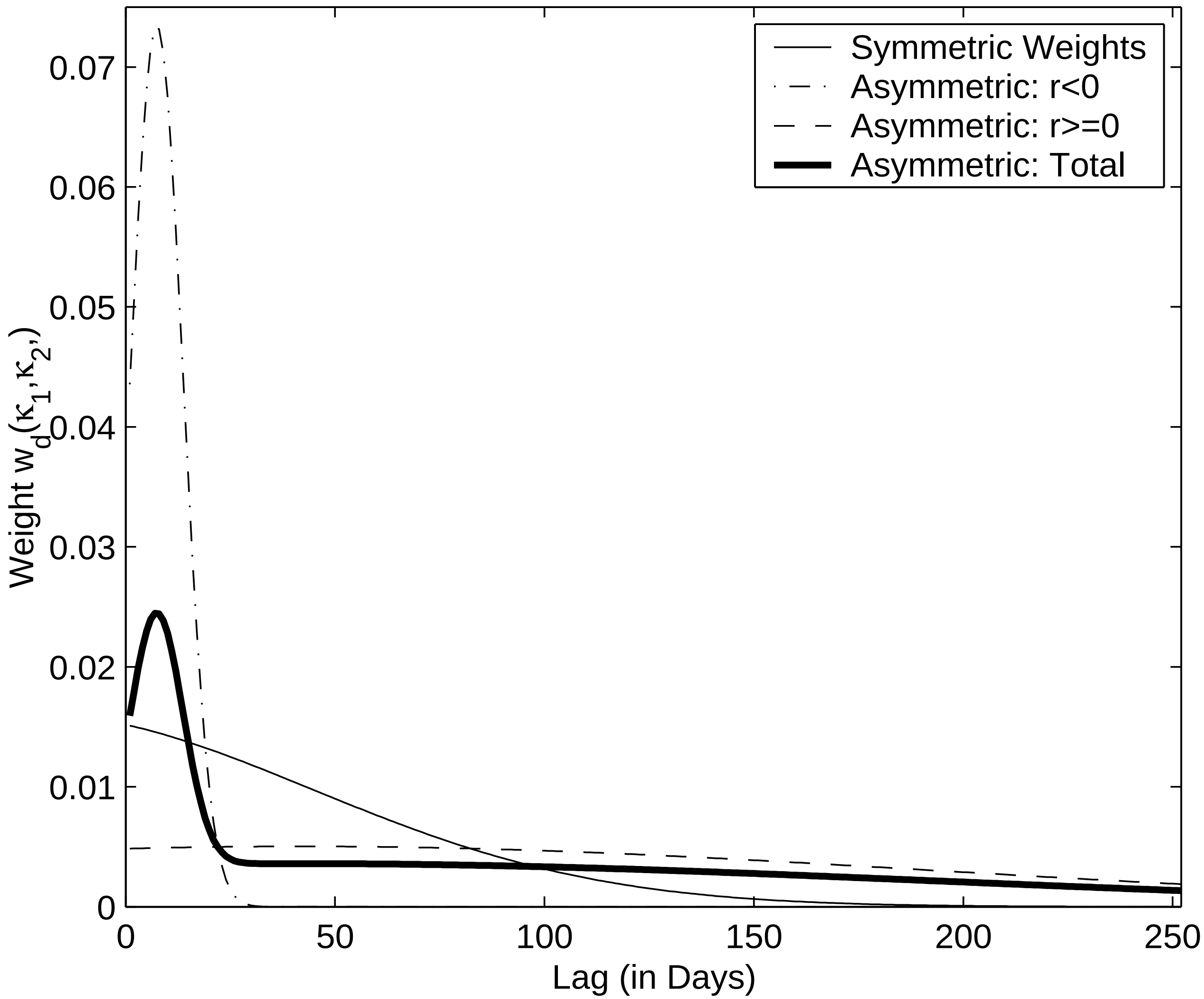
where $\mathbf{1}_{t-d}^+$ denotes the indicator function for $\{r_{t-d} \geq 0\}$, $\mathbf{1}_{t-d}^-$ denotes the indicator function for $\{r_{t-d} < 0\}$, and ϕ is in the interval $(0, 2)$.

Results with Asymmetries

Sample	μ $(\times 10^3)$	γ	+/-	θ_1 $(\times 10^2)$	θ_2 $(\times 10^3)$	ϕ	R^2_R	$R^2_{\sigma^2}$
1946:01-	5.766	3.314	(-)	9.573	-7.640	0.606	0.025	0.085
2000:12	[2.057]	[2.695]		[0.507]	[-0.929]	[3.381]		
			(+)	0.073	-0.210			
				[0.133]	[-0.539]			
1946:01-	5.550	3.735	(-)	-7.541	-2.340	0.716	0.043	0.363
2000:12	[1.989]	[2.868]		[-1.775]	[-0.342]	[5.010]		
(No Crash)			(+)	-0.214	-0.910			
				[-0.613]	[-0.498]			

GARCH-M Comparison - results are consistent with Glosten, Jagannathan, and Runkle (1993) and others in the literature.

Model	μ $(\times 10^3)$	γ	ω $(\times 10^2)$	α	β	λ $(\times 10^2)$	R^2_R	$R^2_{\sigma^2}$	LLF
EGARCH(1,1)-M	14.978	-2.521	-640.708	-0.325	0.497	-3.339	0.011	0.071	1159.102
	[6.277]	[-1.285]	[-1.790]	[-2.977]	[5.938]	[-2.206]			
ASYGARCH(1,1)-M	1.117	-3.248	0.056	0.018	0.609	-28.723	0.010	0.077	1164.023
	[0.913]	[-1.811]	[0.202]	[1.980]	[7.842]	[-2.131]			
QGARCH(1,1)-M	13.970	-1.994	0.060	0.086	0.145	-9.320	0.010	0.072	1161.173
	[2.378]	[-0.171]	[0.356]	[3.565]	[3.269]	[-7.188]			


Symmetric Weights
Asymmetric: r<0
Asymmetric: r>=0
Asymmetric: Total
Weight w_d(κ_1',κ_2')
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
50
100
150
200
250
Lag (in Days)

Summary and robustness of empirical results

- Risk-return relation has been further explored using MIDAS approach by Brown and Ferreira (2004), Jiang and Lee (2004), León, Nave and Rubio (2004), Wang (2004), Charoenrook and Conrad (2005), among others.

Example II Revisited: Predicting Realized Volatility

- We start with *daily* regressors and consider the following:

$$RV_{t+H,t} = \mu_H + \phi_H \sum_{k=0}^{k^{max}} b_H(k,\theta) X_{t-k} + \varepsilon_{Ht}^X$$

where the following regressors are considered:

  – Past daily Realized Volatility, Past daily squared returns, Past daily absolute returns, Past daily high minus low

and

  – Past *Realized Absolut Value*, i.e. sum of (say) 5-minute intra-daily **absolute** returns, see in particular Barndorff-Nielsen and Shephard (2002c) and Woerner (2002).

- All of the MIDAS regressions come in various 'flavors', i.e. using log transformations, or square root etc.

- Within Barndorff-Nielsen diffusion framework Forsberg and Ghysels (2004) show that RAV *is expected to be* the best predictor. We expect: $RAV \succ BPV\ (C) \succ RV.$

Out-of-Sample MSE Comparisons of MIDAS Models with Daily Regressors - DJ Index

	$RV_{t,t+H}$ MIDAS					$\log(RV_{t,t+H})$ MIDAS			
	RV	r^2	$\|r\|$	$[hi - lo]$	RAV	$\log RV$	$\log r^2$	$\log[hi - lo]$	$\log RAV$
1 wk	0.802	1.048	1.005	0.819	0.778	0.746	1.457	0.815	0.729
2 wks	0.920	1.168	1.039	0.855	0.726	0.794	1.296	0.882	0.731
3 wks	0.814	0.995	0.874	0.794	0.724	0.751	1.089	0.797	0.714
4 wks	0.872	1.089	0.985	0.857	0.774	0.893	1.162	0.981	0.854

Models - MIDAS with step functions

We predict the normalized Realized Variance in-sample and out-of-sample using

- MIDAS regressions with polynomials (Ghysels, Santa-Clara and Valkanov(2003a *JFE* forthcoming, 2003b *JoE* forthcoming))

$$RV_{t,t+H}^{1/2} = \mu_H + B(L) X_t + \varepsilon_{Ht}$$

- MIDAS with stepfunctions Forsberg and Ghysels (2004) (special case is HAR (Heterogenous Autoregressive) Model (Corsi(2003)))

$$RV_{t,t+H}^{1/2} = \beta_0 + \beta_D X_{t-1,t} + \beta_W X_{t-5,t} + \beta_M X_{t-20,t} + \varepsilon_{t+H}$$

In-Sample Results modeling $RV^{1/2}$ of S&P 500 1985-2003 - Sign level of the Bipower test $=$ 0.999

	Stepfunction MIDAS-RV$^{1/2}$					Beta polynomial MIDAS-RV$^{1/2}$				
Horizon	RV$^{1/2}$	BPV$^{1/2}$	C$^{1/2}$	$(CJ)^{1/2}$	RAV	RV$^{1/2}$	BPV$^{1/2}$	C$^{1/2}$	$(CJ)^{1/2}$	RAV
R^2										
1 day	0.591	0.592	0.592	0.594	0.623	0.596	0.595	0.598	0.599	0.617
1 wk	0.656	0.649	0.654	0.658	0.690	0.671	0.667	0.671	0.672	0.687
2 wks	0.648	0.639	0.646	0.651	0.690	0.670	0.665	0.670	0.671	0.689
3 wks	0.633	0.623	0.631	0.636	0.680	0.656	0.651	0.656	0.657	0.678
4 wks	0.615	0.603	0.613	0.617	0.665	0.638	0.632	0.638	0.639	0.662
MSE										
1 day	1.344	1.333	1.343	1.339	1.265	1.314	1.314	1.311	1.309	1.269
1 wk	0.513	0.519	0.517	0.510	0.458	0.482	0.486	0.481	0.480	0.463
2 wks	0.382	0.391	0.385	0.378	0.330	0.350	0.355	0.349	0.348	0.331
3 wks	0.339	0.349	0.342	0.335	0.290	0.310	0.316	0.309	0.308	0.292
4 wks	0.319	0.330	0.323	0.316	0.274	0.293	0.300	0.294	0.292	0.277

Summary and robustness of empirical results

- RAV is the best predictor/ regressor for RV (as predicted by theory)
- Robustness of empirical results:
  - Result holds when modeling Realized Variance in levels
  - Result is the same regardless of the significance level of the bipower test, we have investigated $\alpha = 0.5$, 0.95, 0.99 and 0.999
  - Same results when a dummy for the jump-days is included
  - Robust to subsample period: S&P 1990-2002
  - Results are robust to using other evaluation measures such as Heterscedasticity Adjusted Error (Bollerslev and Ghysels (1996))

- While using HF has some clear advantages, there are some costs. HF sampling may be plagued by microstructure noise. Several papers have tried to shed light on this: Aït-Sahalia, Mykland and Zhang (2003), Bandi and Russell (2003 a,b), Hansen and Lunde (2004 a,b), Zhang, Mykland and Aït-Sahalia (2003), ...

- Ghysels and Sinko (2005), Sinko (2005) show that Aït-Sahalia, Mykland and Zhang (2003) approach *does make a difference* for forecasting future RV, while other approaches don't.

- But at 5-min sampling intervals *uncorrected* RAV still predicts better. However, at 1-min interval Aït-Sahalia et al. corrections of RV dominate uncorrected uncorrected RAV.

Example III: Application of MIDAS to forecasting the predictions of forecasters

Work in progress joint with Jonathan Wright

- MIDAS ideally suited to exploit high frequency financial data to predict low frequency macro data
- Low frequency macro forecasts are less noisy than *ex-post* realized macro data and so potentially easier to predict
- Once a month/quarter, observe forecasts of some future macro/finance variable (e.g. inflation four quarters hence), f_t, from
  - Survey of Professional Forecasters (quarterly)
  - Consensus Forecasts (monthly)
- Know deadline dates for completion of survey: d_t

- But surveys are infrequent (e.g. quarterly) and often stale.
- Would like survey expectations on a daily basis
- Observe daily asset price data that responds to news about likely future evolution of the economy
- Related work:
  - Evans (2005)
  - Giannone, Reichlin and Small (2005) calculate a daily coincident indicator.

Challenges

- Dealing with fuzzy timing of surveys
- Financial data are abundant, and arrive at much higher frequency than the releases of macroeconomic forecasts.
- It quickly becomes obvious that, unless a parsimonious model can be formulated, there is no practical solution for linking the quarterly forecasts to daily financial data.

Why do we care?

- Monetary policy communications such as the minutes of the Federal Open Market Committee, and the semiannual testimony of the Federal Reserve Board routinely point to survey expectations of inflation.
- Policy makers, monitoring the economy in real time, would presumably like to be able to measure these expectations at a higher frequency.
- Expectations in a multi-agent economy involve "forecasting the forecasts of others" (to paraphrase Townsend (1983)) and therefore private sector agents would likewise also wish to obtain higher frequency measures of others' expectations.
- .... beating forecasters at their own game....

Talk focuses on the SPF (results for Consensus in paper)

- SPF asks for predictions at horizons h=1,2,3,4 in the middle of each quarter (mid Feb, mid May etc.)
- Surveys of forecasters, containing respondents' predictions of future values of growth, inflation and other key macroeconomic variables, receive a lot of attention in the financial press, from investors, and from policy makers.
- In our empirical work we use the median forecasts of output growth, inflation, unemployment and certain interest rates, from the Survey of Professional Forecasters.
- The financial market data that we use to predict these forecasts are daily changes in interest rates and interest rate futures prices, and also daily stock returns.

Methods

- A priori, it is not clear how to use daily financial market data to formulate a parsimonious model.
- It is going to be 'reduced' form approaches.
- A simple linear regression will not work well as it would entail estimating a large number of parameters. For example if we use two months of data it would require estimation of 44 parameters, assuming 22 trading days a month.

We use two approaches

- Mixed Data Sampling, or MIDAS regression models, proposed in recent work of Ghysels, Santa-Clara and Valkanov *et al.* MIDAS regressions are designed to handle large high-frequency data sets with judiciously chosen parameterizations, tightly parameterized yet versatile enough to yield predictions of low frequency forecast releases with daily financial data.

- Kalman filter to estimate what forecasters would predict if they were asked to make a forecast each day, treating their forecasts as "missing data" to be interpolated (see e.g. Harvey and Pierse (1984), Harvey (1989), Bernanke, Gertler and Watson (1997)). As a by-product, this also gives forecasts of upcoming releases.

Forecasting and financial market signals

- The objective is to show, *in a very stylized setting*, how, if asset prices and forecasts both respond to news about the state of the economy, we can use asset returns to glean high-frequency information about agents' forecasts.

- The empirical specifications will be richer and more general.

- Assume that we observe forecasts of some future macroeconomic variable (e.g. inflation four quarters hence) once a quarter and suppose, for simplicity in this model, that these are observed on the last day of each quarter.

- Assume underlying macroeconomic variable y_t is AR(1): $y_{t+1} = a_0 + a_1 y_t + \varepsilon_{t+1}$
- Let f_t^{t+h} denote forecast of y_{t+h} made on last day of quarter t. If the forecaster knows the DGP, knows y_t at the end of quarter t, then the h-quarter-ahead forecast will be:

$$f_t^{t+h} = a_0 \Sigma_{j=0}^{h-1} a_1^j + a_1^h y_t$$

and, this is related to the previous h-quarter-ahead forecast and the shock ε_t as follows:

$$f_t^{t+h} = a_0 \{a_1^h - (a_1 - 1) \sum_{j=0}^{h-1} a_1^j\} + a_1 f_{t-1}^{t-1+h} + a_1^{h-1} \varepsilon_t$$

- We are interested in predicting f_t^{t+h} at some time between the end of quarter t-1 and the end of quarter t.

- Although the errors ε_t occur quarterly, we construct a *fictitious* set of daily shocks $\varepsilon_t \equiv \sum_{d=l_{t-1}+1}^{l_t} v_d$ where l_t denotes the last day of quarter t.

- We focus on a single asset and assume that its price on day τ relates to the fictitious shocks as follows:

$$p_t^\tau = p_t^0 + \sum_{d=l_{t-1}+1}^{\tau} v_d + \omega_\tau$$

where p_t^0 is the price at the beginning of the quarter t.

- Hence, daily prices provide a noisy signal of the underlying economic shocks.

- Assume that the fictitious shocks v_d are Gaussian with mean zero and variance σ_v^2 and that the noise is also Gaussian with mean zero and variance σ_w^2 and is orthogonal to the v_d process.
- There is a fundamental difference between:
  - predicting f_t^{t+h} during quarter t with concurrent and past daily financial market data via a regression model,
  - "guessing" φ_τ^h, the unobserved h-quarter-ahead expectation on day τ. Treating agents' h-quarter-ahead expectations *during* the quarter as "missing" values of a process only observed at the end of the quarter.

The regression approach

- Partial sums process $S_\tau = \sum_{d=l_{t-1}+1}^{\tau} v_d$ behaves like a random walk and therefore price p_t^τ behaves (within quarter) like r.w. + noise.

- To predict f_t^{t+h}, we would like to know ε_{t+1}, and the best predictor on day τ would be S_τ, which can be extracted from returns:

$$\mathcal{P}[f_t^{t+h} | r_d, d = l_{t-1}+1, \ldots, \tau] \approx$$

$$a_0\{a_1^h - (a_1 - 1)\sum_{j=0}^{h-1} a_1^j\} + a_1 f_{t-1}^{t-1+h} + a_1^{h-1} \sum_{d=\tau}^{l_{t-1}+1} (1 - (-\zeta)^{\tau-d+1}) r_d$$

where ζ is a function of the signal-to-noise ratio $q = \sigma_v^2/\sigma_w^2$ and r_d is the asset return on day d for days $l_{t-1}+1$, . . . , τ during quarter t.

Filtering approach

- With Gaussian distributional assumptions, we are also in the context of Kalman filter....regression approach applies to broader setting.

- The conditional expectation , combined with asset price equation holds the ingredients for a state space model to determine φ_τ^h, namely daily returns are a noisy signal of daily changes in φ_τ^h :

$$r_\tau = \phi(\varphi_\tau^h - \varphi_{\tau-1}^h) + \tilde{\varepsilon}_\tau^1$$

  where at the end of each quarter we observe $f_t^{t+h} = \varphi_{l_t}^h$. This gives us an explicit state space model.

Empirical regression model specification

- Let d_t denote the survey deadline date for quarter t: survey respondents submit their forecasts on or before this day: the survey results are released a few days later.

- We suppose that the researcher wishes to forecast f_t^{t+h} using asset return data on days up to and including day τ. Our forecasting model is:

$$f_t^{t+h} = \alpha + \rho f_{t-1}^{t-1+h} + \sum_{j=1}^{n_A} \beta_j \gamma(L) r_\tau^j + \varepsilon_t$$

where r_τ^j denotes the return on day τ for asset j, n_A denotes the number of assets and $\gamma(L)$ is a lag polynomial of order n_l so that $\gamma(L) r_\tau^j$ is a distributed lag of daily returns on asset j over the n_l days up to and including day $[d_{t-1} + \theta(d_t - d_{t-1})]$ where $0 < \theta \leq 1$.

- In empirical work we will consider $\theta = 1, 2/3$ and $1/3$ corresponding to forecasts for f_t^{t+h} made on the survey deadline date, about one month earlier, and about two months earlier, respectively.

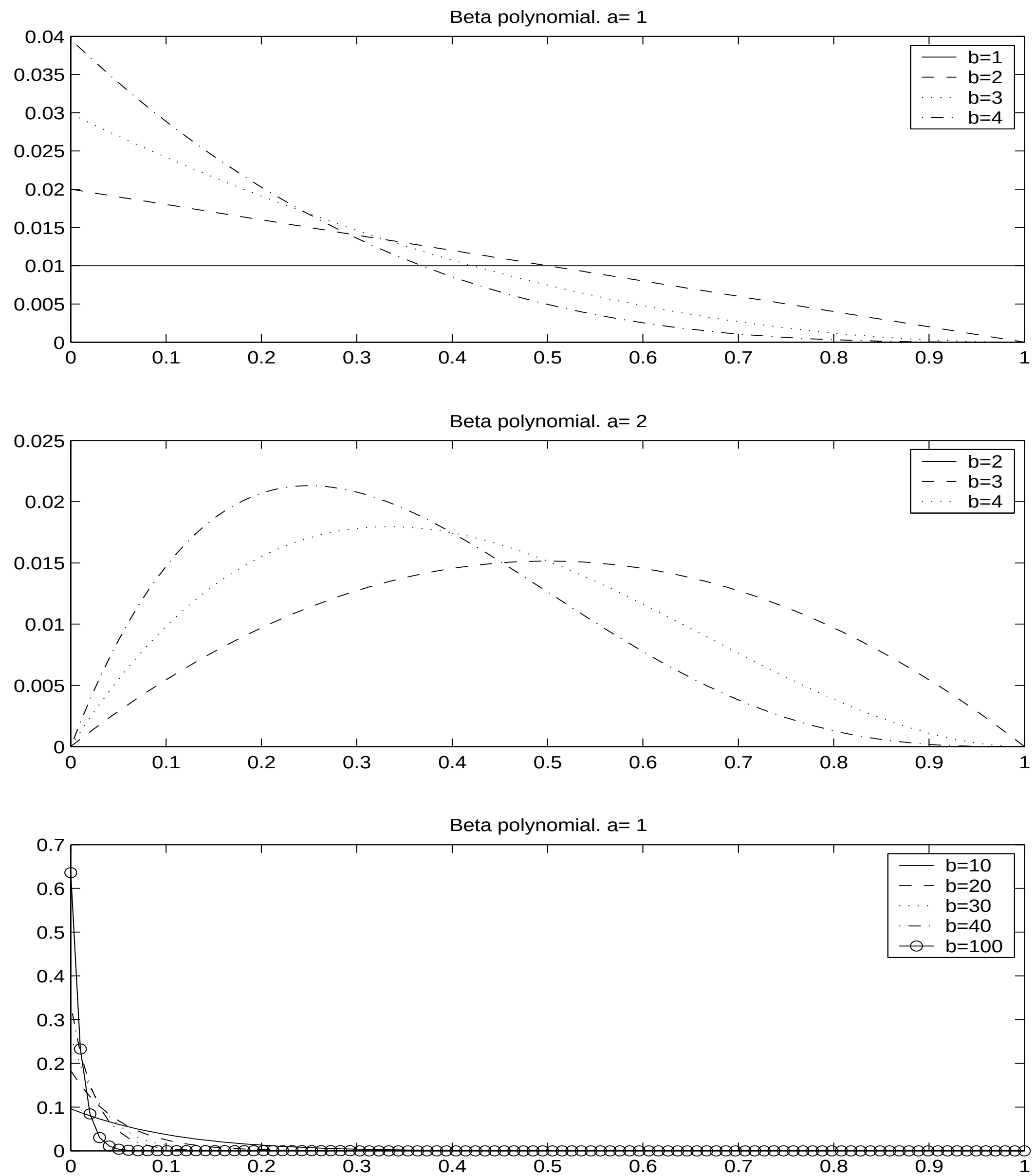
- We run a different regression for each θ and so should strictly put a θ-subscript on α,$\{\beta_j\}_{j=1}^{n_A}$, ρ, and $\gamma(L)$, but do not do so, in order to avoid excessively cumbersome notation.

- Above uses mixed frequency data: f_t^{t+h} and f_{t-1}^{t-1+h} are observed at the quarterly frequency, $t = 2, ...T$, while the returns are at the daily frequency.

- Following Ghysels, Sinko and Valkanov (2003) we use a flexible specification for $\gamma(L)$ with only two parameters, κ_1 and κ_2. In particular, the lag k coefficient is written as:

$$\gamma(k;\kappa_1,\kappa_2) = \frac{f(\frac{k}{n_l},\kappa_1;\kappa_2)}{\sum_{k=1}^{n_l} f(\frac{k}{n_l},\kappa_1;\kappa_2)}$$

where: $f(x,a,b) = x^{a-1}(1-x)^{b-1}\Gamma(a+b)/\Gamma(a)\Gamma(b)$, with $\Gamma(a) = \int_0^\infty e^{-x}x^{a-1}dx$.


Beta polynomial. a= 1
b=1
b=2
b=3
b=4
Beta polynomial. a= 2
b=2
b=3
b=4
Beta polynomial. a= 1
b=10
b=20
b=30
b=40
b=100

Three MIDAS regression models

- Model M1, where we estimate all unknown parameters α,$\{\beta_j\}_{j=1}^{n_A}$, ρ, κ_1 and κ_2.

- Model M2, imposes that $\kappa_1 = \kappa_2 = 1$, implying that the weights in $\gamma(L)$ are equal. In this case, $\gamma(L) r_\tau^j$ is simply the average return over the n_l days up until day τ and $\gamma(L)= \sum_{j=1}^{n_l} 1/n_l L^{j-1}$.

- Model M3, is an equal-weighted MIDAS regression in which the average returns from day d_{t-1} to day τ are used to predict the upcoming release, i.e. $\gamma(L)= \sum_{j=1}^{n_l} 1/n_l L^{j-1}$ and $n_l = \tau - d_{t-1}$.

- The difference between models M2 and M3 is that model M2 has a fixed lag length parameter, n_l, while model M3 will always use returns from exactly day d_{t-1} to day τ.

Kalman filter specifications

- We want the extraction of φ_τ^h for horizon h on day τ.
- We consider two thought-experiments that go beyond the simple stylized example.
  - We observe respondents' expectations on survey deadline dates, but these expectations are missing data that must be interpolated on all other days. We call this model $K1$.
  - A specification that takes into account the fuzziness about the exact timing of respondents' expectations. We call this model $K2$.

- For model $K1$, let φ_τ^h denote the respondents' h-quarter-ahead expectations on day t and write the following model

$$r_\tau = \phi(\varphi_\tau^h - \varphi_{\tau-1}^h) + \varepsilon_{1\tau}$$

$$\varphi_\tau^h = \mu_0 + \mu_1 \varphi_{\tau-1}^h + \varepsilon_{2\tau}$$
$$f_t^{t+h} = \varphi_{d_t}^h$$

where $(\varepsilon_{1\tau}^{'}, \varepsilon_{2\tau})$ is i.i.d. normal with mean zero and diagonal variance-covariance matrix.

- For model $K2$ the last equation is replaced by:

$$f_t^{t+h} = \gamma(L)\varphi_{d_t}^h$$

where $\gamma(L)$ is a MIDAS polynomial.

- The thought-experiment here is that individual respondents form their expectations each day, but that some of these get transmitted to the compilers of the survey faster than others.

- Model $K1$ is nested within $K2$ specification, as we can specify that $\kappa_1 = 1$ and $\kappa_2 = \infty$, implying that $\gamma(L) = 1$.

- We can use the Kalman filter to find maximum-likelihood estimates of the parameters, giving filtered estimates of φ_τ^h and forecasts of f_t^{t+h} (made a fraction θ of the way through the prior inter-survey period) as by-products. We can also use the Kalman smoother to obtain estimates of φ_τ^h conditional on the entire dataset.

Empirical results

- We consider the Survey of Professional Forecasters (SPF), conducted at a quarterly frequency. The respondents include Wall Street financial firms, banks, economic consulting groups, and economic forecasters at large corporations. Consesus Forecasts not report in talk.
- Hencforth, f_t^{t+h} refers to the forecast made in the quarter t SPF forecast for any one of these variables in quarter $t + h$ and our forecasting models are the MIDAS models exactly as defined earlier.
- We start with the forecasts made in 1990Q3 and end with forecasts made in 2005Q4 for a total of 62 forecasts.

- The survey deadline date is not the date that the survey results are released, but is the last day that respondents can send in their forecasts. We do not use SPF forecasts made before 1990Q3 because we do not have the associated survey deadline dates.

- The number of assets, n_A, is either 1 or 2. Our predictors are thus stock returns and changes in measures of the level and/or slope of the yield curve.

- In MIDAS models M1 and M2, the lag length n_l is a fixed parameter that we set to 90. As our data are at the business day frequency, this corresponds to substantially more than one quarter of data.

- We consider models M1, M2, M3, K1, K2 with the following daily asset returns:
  - excess stock market returns,
  - the daily change in the rate on the fourth three-month eurodollar futures contract (a futures contract on a three-month interest rate about one year hence),
  - the daily changes in the rates on the first and twelfth eurodollar futures contracts (futures contract on three-month interest rates about three months and three years hence),
  - the daily changes in three-month and ten-year Treasury yields,
  - the daily change in two-year Treasury yields.

- We evaluate the forecasts by comparing the in-sample and pseudo-out-of-sample root mean-square prediction error (RMSPE), relative to the RMSPE from using the prior survey release as a predictor (a "random walk" forecast).

- The first observation for out-of-sample prediction is the first observation in 1998, with parameters estimated using data from 1997 and earlier, and prediction then continues from this point on in the usual recursive manner, forecasting in each period using data that were actually available at that time.

- Note that because we are working with asset price data and survey forecasts (rather than actual macroeconomic realizations), we have no issues of data revisions to contend with; the out-of-sample forecasting exercise is a fully real-time forecasting exercise.

In-Sample RMSE using Changes in three-month and ten-year Treasury yields

Horizon (Qtrs.)	M1	M2	M3 ($\theta = 2/3$)	K1	K2	M1	M2	M3 ($\theta = 1/3$)	K1	K2
					Real GDP Growth					
1	0.733	0.834	0.780	0.806	0.766	0.772	0.886	0.805	0.855	0.814
2	0.870	0.870	0.836	0.842	0.832	0.845	0.901	0.859	0.895	0.864
					CPI Inflation					
1	0.827	0.834	0.828	0.848	0.837	0.809	0.855	0.832	0.906	0.833
2	0.797	0.814	0.808	0.826	0.804	0.783	0.844	0.819	0.886	0.801
					T-Bill					
1	0.299	0.497	0.487	0.466	0.387	0.488	0.646	0.651	0.713	0.621
2	0.356	0.540	0.524	0.503	0.440	0.505	0.684	0.643	0.717	0.633
					Unemployment Rate					
1	0.625	0.735	0.739	0.739	0.672	0.694	0.818	0.777	0.846	0.759
2	0.648	0.760	0.744	0.732	0.686	0.728	0.849	0.790	0.825	0.767

Pseudo-Out-Of-Sample RMSE using Changes in three-month and ten-year Treasury yields

Horizon (Qtrs.)	M1	M2	M3 $\theta = 2/3$	K1	K2	M1	M2	M3 $\theta = 1/3$	K1	K2
Real GDP Growth										
1	0.805	0.868	0.816	0.858	0.777	0.845	0.945	0.842	0.903	0.849
2	0.927	0.942	0.925	0.968	0.917	1.049	0.992	0.929	0.946	0.951
CPI Inflation										
1	1.053	0.953	0.987	0.945	1.007	1.008	0.994	1.000	0.941	1.072
2	0.991	0.928	0.951	0.891	0.947	0.979	0.973	0.990	0.917	1.018
T-Bill										
1	0.390	0.630	0.593	0.434	0.403	0.521	0.706	0.647	0.666	0.628
2	0.422	0.664	0.621	0.472	0.472	0.530	0.732	0.649	0.682	0.672
Unemployment Rate										
1	0.752	0.865	0.805	0.711	0.682	0.754	0.879	0.701	0.763	0.711
2	0.802	0.849	0.849	0.732	0.714	0.775	0.872	0.759	0.769	0.738

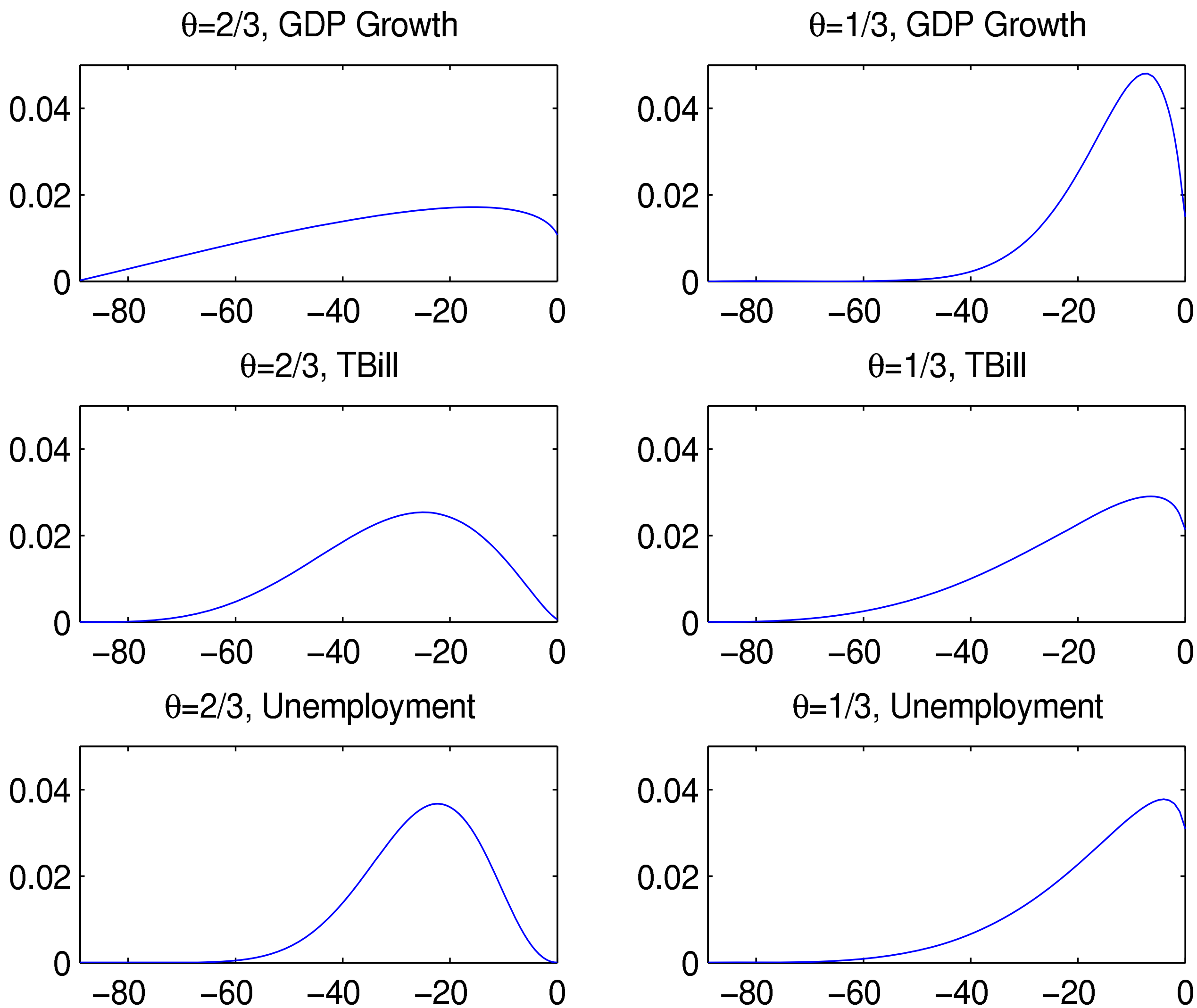
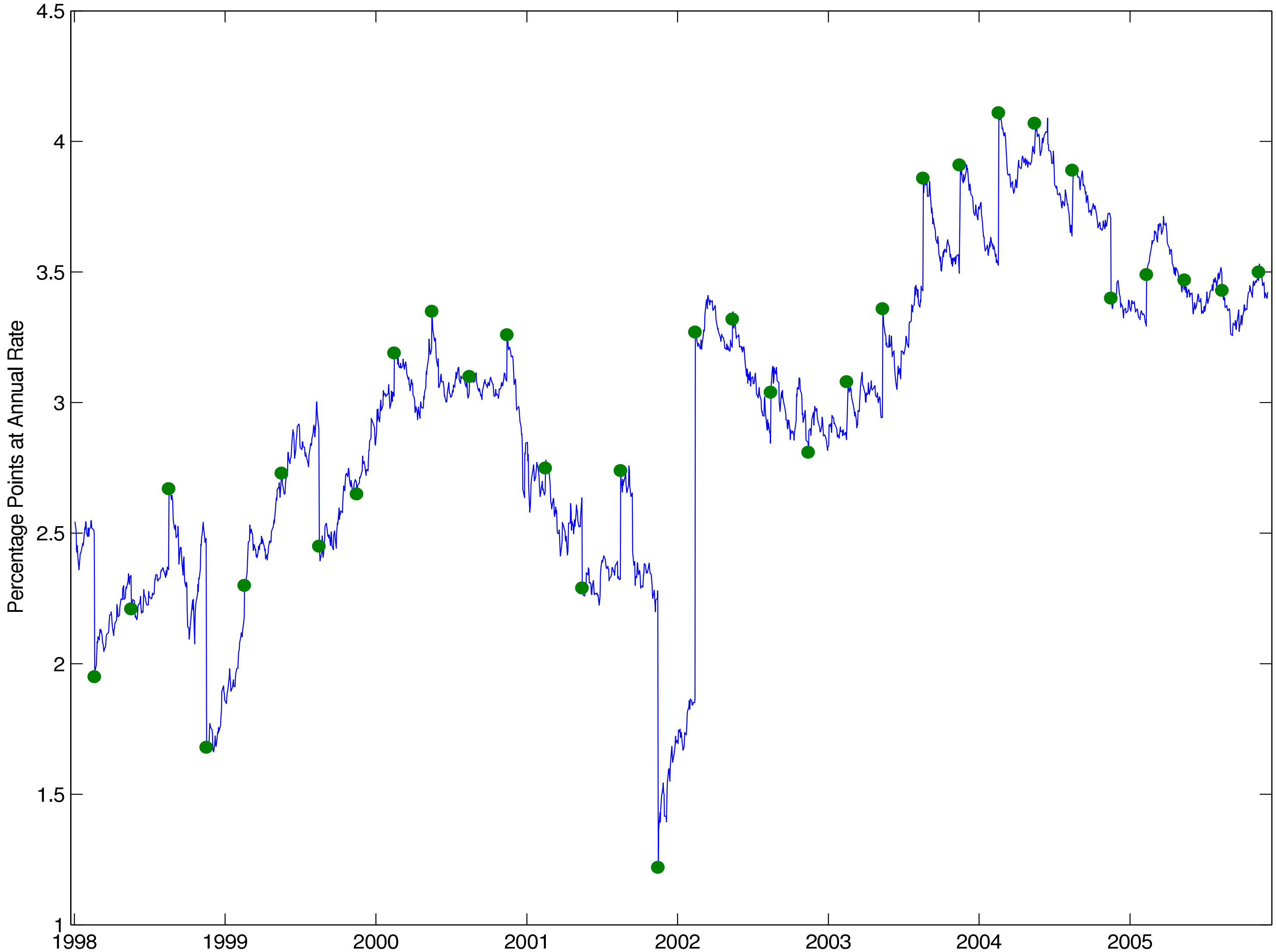

θ=2/3, GDP Growth
θ=1/3, GDP Growth
θ=2/3, TBill
θ=1/3, TBill
θ=2/3, Unemployment
θ=1/3, Unemployment
0.04
0.02
0
−80
−60
−40
−20
0

Fig. 1: Filtered Estimates of 2–Quarter GDP Growth Forecasts from Model K1 with predictors (d)
Actual SPF median forecasts are shown by dots



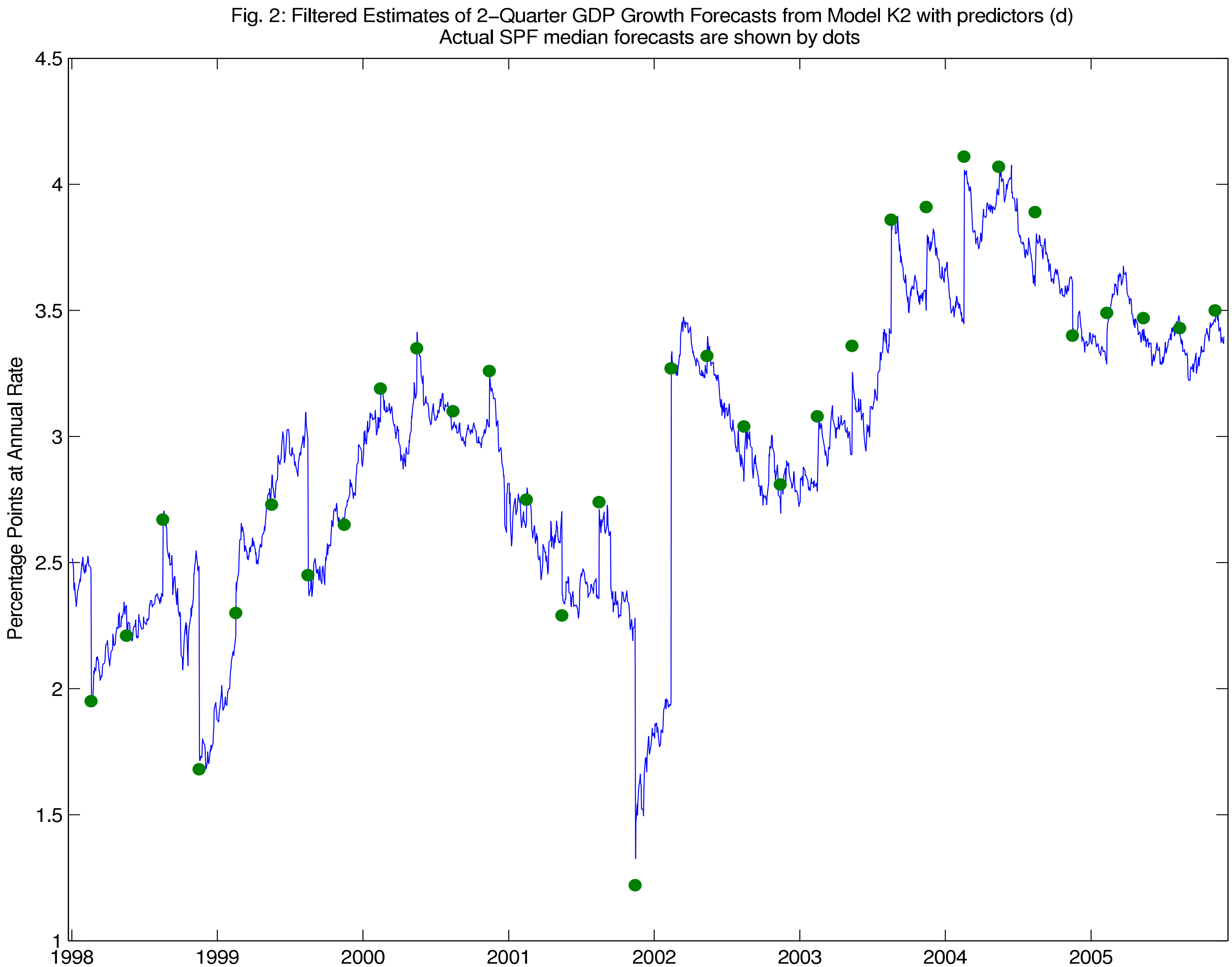


Fig. 2: Filtered Estimates of 2–Quarter GDP Growth Forecasts from Model K2 with predictors (d)
Actual SPF median forecasts are shown by dots
Percentage Points at Annual Rate
4.5
4
3.5
3
2.5
2
1.5
1
1998
1999
2000
2001
2002
2003
2004
2005

- The in-sample relative RMSPEs from MIDAS models M1, M2 and M3 are generally well below one.
- Not surprisingly, relative RMSPEs are higher in the pseudo-out-of-sample forecasting exercise.
- Survey forecasts of the unemployment rate and T-Bill yields appear to be generally the most predictable out-of-sample, but relative RMSPEs are in many cases below one out-of-sample for GDP growth and CPI inflation as well.
- Overall, the best out-of-sample results appear to obtain with the daily changes in the three-month and ten-year Treasury yields, but the other yield curve variables.

- On average, across all four variables and all four horizons, the pseudo-out-of-sample relative RMSPE from MIDAS model M1 with $\theta = 1$ and using changes in three-month and ten-year Treasury yields is 0.78.

- Even though MIDAS model M1 involves estimation of two additional parameters, it generally gives smaller out-of-sample RMSPEs than models M2 or M3.

- Model K2 generally gives smaller out-of-sample RMSPEs than model K1, even though the former involves estimation of two additional parameters. This reinforces the evidence that surveys represent agents' beliefs a considerable lag to the survey deadline date.

- Models K1 and K2 seem to generally give less good forecasts of what the upcoming survey release is going to be than the reduced form MIDAS regression models M1, M2 and M3.

Measuring the Effect of News Announcements on Agents' Expectations

- Applying the Kalman smoother to models, K1 and K2, we can estimate what forecasters' expectations were on a day-to-day basis conditional on the whole sample.

- Hence, we can, in principle, measure agents' expectations immediately before and after a specific event (e.g. macroeconomic news announcements, Federal Reserve policy shifts or major financial crises) and so measure the impact of such events on their expectations.

- As an illustration, we show how our method can be used to estimate the average effect of a nonfarm payrolls data release (one of the most important macroeconomic news announcements) coming in 100,000 stronger than expected on the expectations of respondents to the SPF.

- Nonfarm payrolls data are released by the Bureau of Labor Statistics once a month, at 8:30 AM sharp.
- We measure *ex-ante* expectations for nonfarm payrolls releases from the median forecast from Money Market Services (MMS) taken the previous Friday. The surprise component of the nonfarm payrolls release is then the released value less the MMS survey expectation.
- We regress the *change* in these expectations from the day before the nonfarm payrolls release to the day of the nonfarm payroll release on the surprise component of that release.

$$\varphi^h_{\tau|T} - \varphi^h_{\tau-1|T} = \lambda s_\tau + \eta_\tau$$

where $\varphi^h_{\tau|T}$ denotes the Kalman smoothed estimates of the h-quarter-ahead forecast for any variable being predicted in the SPF, s_τ denotes the surprise component of the nonfarm payrolls release.

- The estimated coefficients seem to be of a reasonable magnitude. For example, using changes in three-month and ten-year Treasury yields and model K2, a 100,000 positive nonfarm payrolls surprise (which is approximately a one standard deviation announcement surprise) is estimated to raise
  - one-quarter -ahead growth forecasts by 6/100ths of a percentage point
  - four-quarter -ahead growth forecasts by 1/100th of a percentage point.
  - four-quarter inflation forecasts by 1/100th of a percentage points,
  - four-quarter T-Bill yield forecasts by 3 basis points.
- These estimated effects are all small, but it seems reasonable that forecasts are not adjusted much in response to a one-standard deviation surprise in employment growth for one month. And, though small, these effects are all highly significant.

Ongoing and future work

- So far we have focused on forecasting upcoming releases of surveys of forecasters.
- What about forecasting actual macro variables?
- If our forecasts are conditional expectations of the upcoming survey forecasts, and those survey forecasts are in turn conditional expectations of actual future outcomes then, by the law of iterated expectations, our forecasts must also be conditional expectations of those actual future outcomes.
- Recall Ang et al. found surveys to be better than 'time series' models,....but these models use aggregate data (quarterly) data whereas survey forecasts have an informational advantage.

Example III: Macroeconomic sources of stock market volatility

- At least since Schwert (1989) the link between stock market volatility and macroeconomic fundamental risk is found to be weak.
- We revisit this with a GARCH/MIDAS framework.
- Example III is taken from *Economic Sources of Volatility*, with R. Engle and B. Sohn

Most of the talk based on:

- *On the Economic Sources of Stock Market Volatility, Engle, Ghysels and Sohn*
- *A Component Conditional Correlation Model for Financial Assets, Colacito, Engle and Ghysels*

Introduction

- We have made substantial progress on modeling the time variation of volatility.

- We have a better understanding of forecasting volatility over relatively short horizons, ranging from one day ahead to a couple of months. A key ingredient is volatility clustering, a feature and its wide-range implications, first explored in the seminal paper on ARCH models by Engle (1982).

- We also bridged the gap between discrete time models, such as the class of ARCH models, and continuous time models, such as the class of Stochastic Volatility (SV) models with close links to the option pricing literature.

- Despite the impressive list of areas where we made measurable and lasting progress, we are still struggling with at least two basic issues.

- Schwert (1989) wrote a paper with the pointed title, *Why Does Stock Market Volatility Change Over Time*?

- The contributions of this paper pertain to both the economic sources of volatility (and therefore risk premia) and the impact of aggregate volatility on the cross-section of returns.
- The progress of the last fifteen years allows us to revisit these basic questions with various new insights, matured during the decade and a half of research on volatility.

- In recent years, various authors have advocated the use of component models for volatility. Engle and Lee (1999) introduced a GARCH model with a long and short run component. See also Ding and Granger (1996), Gallant et al.(1999), Alizadeh et al. (2002) and Chernov et al. (2003), among many others.

- While the principle of multiple components is widely accepted, there is no clear consensus how to specify the dynamics of each of the components.

- The purpose of this paper is to suggest several new component model specifications and introduce methods to link them *directly* to economic activity.
- We study both the time series dynamics as well as the impact of volatility (and its components) on the cross section of returns

- We start with Spline-GARCH model

$$r_i = \mu + \sqrt{\tau_i g_i} \varepsilon_i$$

where $\varepsilon_i \mid \Phi_{i-1} \sim N(0,1)$ with Φ_{i-1} is the information set up to day $(i-1)$.

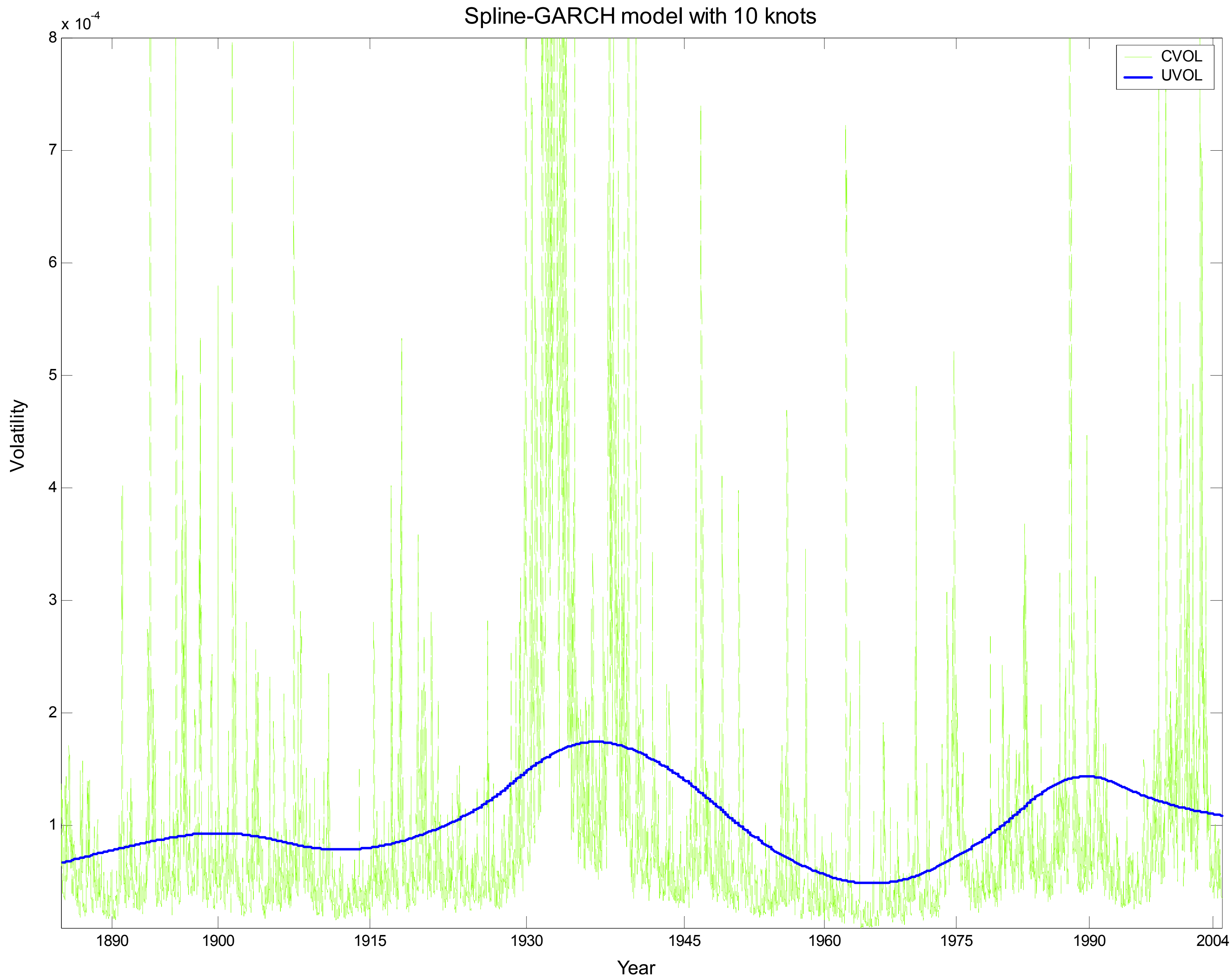
$$g_i = (1 - \alpha - \beta) + \alpha \frac{(r_{i-1} - \mu)^2}{\tau_i} + \beta g_{i-1}$$

$$\tau_i = c\exp\left(w_{(0)} i + w_{(1)} i^2 + \sum_{k=1}^{K} w_k((i - i_{k-1})_+)^2\right)$$

where $\{i_1, i_2, \ldots, i_K\}$ denotes a partition of the time horizon T in (K+1) equally-spaced intervals.

parameter space $\Theta = \{\mu, \alpha, \beta, c, w_{(0)}, w_{(1)}, w_1, \ldots, w_k\}$

The number of knots for the spline is determined by BIC.


Spline-GARCH model with 10 knots
x 10^-4
8
7
6
5
4
3
2
1
Volatility
1890
1900
1915
1930
1945
1960
1975
1990
2004
Year
CVOL
UVOL

- Our work pertains to modelling ‘tau’ and is inspired by the recent work on mixed data sampling, or MIDAS.
- We replace spline by a MIDAS polynomial that is applied to long horizon volatility or macroeconomic variables, the latter to provide a direct link between market volatility and economic activity.
- Note that in the original MIDAS filter setting of Ghysels et al. (2005) we applied a filter to ***high frequency*** data. Here we combine GARCH with a MIDAS filter applied to ***low frequency*** data.

- GARCH-MIDAS with fixed span RV

$$\tau_t = m + \theta \sum_{k=1}^{K} \varphi_k(\omega) RV_{t-k} \qquad RV_t = \sum_{i=1}^{N_t} r_{it}^2$$

- GARCH-MIDAS with rolling window RV

$$\tau_i^{(rw)} = m^{(rw)} + \theta^{(rw)} \sum_{k=1}^{K} \varphi_k(\omega) RV_{i-k}^{(rw)}$$

$$RV_i^{(rw)} = \sum_{j=1}^{N'} r_{i-j}^2$$

- GARCH-MIDAS with macroeconomic volatility

$$\begin{aligned}\tau_t &= m + kI_t^{BC} + \sum_{k=1}^{K} \varphi_k(\omega_1)\big(\theta_1 X_{t-k}^{spread} + \theta_2 X_{t-k}^{IP} + \theta_3 X_{t-k}^{GDP}\big) + \\ &\quad + \sum_{k=1}^{K} \varphi_k(\omega_2)\big(\theta_4 X_{t-k}^{PPI} + \theta_5 X_{t-k}^{MB}\big) \\ &= m + kI_t^{BC} + \tau_t^{real} + \tau_t^{nominal}\end{aligned}$$

- Relates to formulating ‘index models’ (see e.g. Ghysels and Ng (1998) for similar real/nominal index specificaiton in term structure models)

- There are various parsimoniously parameterized specifications, see Ghysels, Sinko and Valkanov (2006).

- Weighting function: Beta lag and exponential weights (like RiskMetrics). Both are single-parameter weighting schemes.

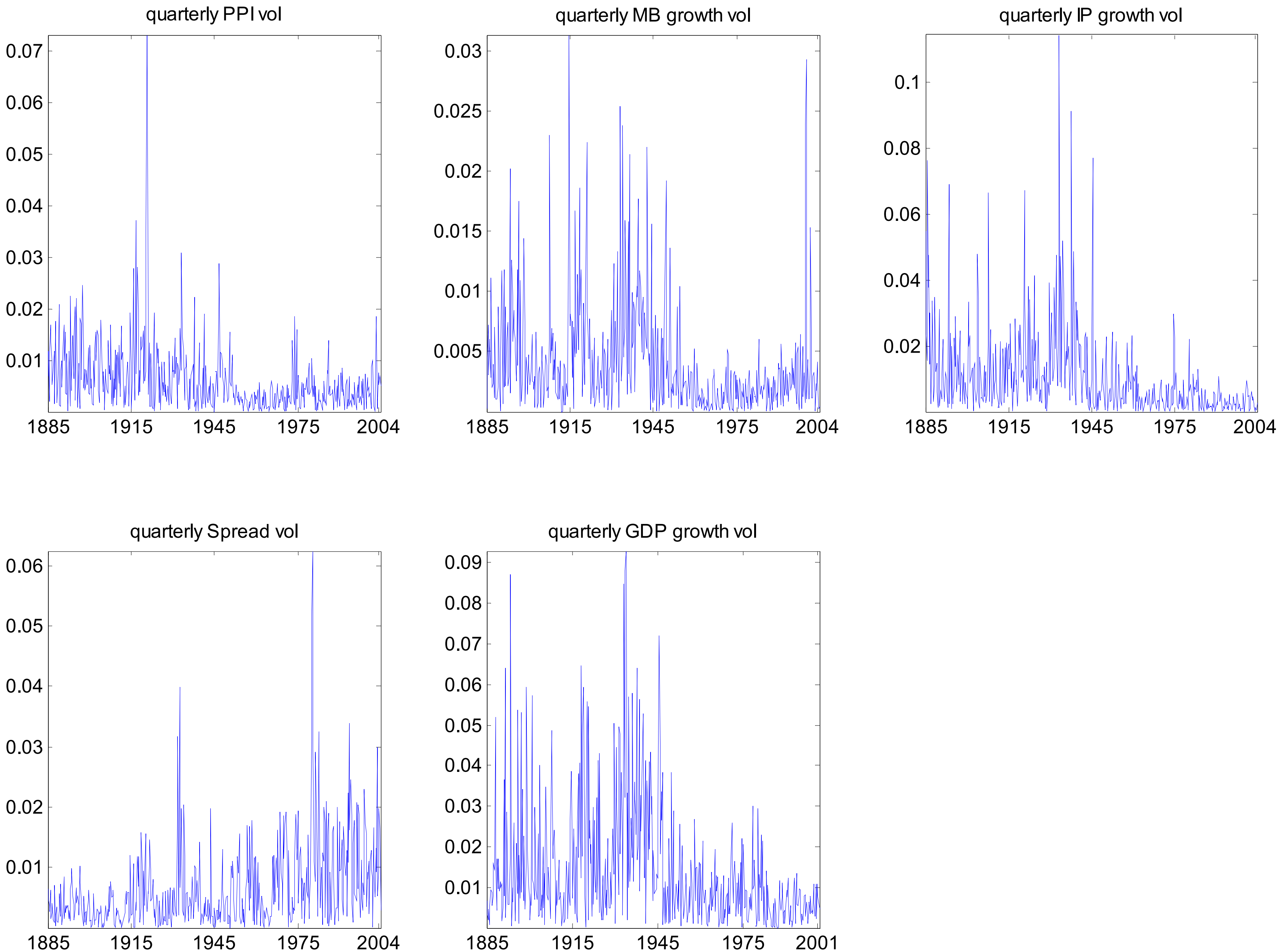
Data

- Stock market return
  - Daily U.S. stock market returns from 1885/2/16 to 1962/7/2. (Schwert’s website)
  - Completed return series by using CRSP daily stock market return series.
- Macroeconomic variables
  - Monthly data (1885-2004)
    - PPI (Producer Price Index) inflation rate
    - MB (Monetary Base) growth rate
    - IP (Industry Production) growth rate
    - LT-ST (Long term – Short term) spread
  - Quarterly data (1885-2001)
    - Real GDP (Gross Domestic Product) growth rate

- Macroeconomic volatility
  - To measure quarterly volatility of quarterly macroeconomic variables, we follow long tradition in economic literature: we run regression specified below for quarterly macroeconomic variable X_t

$$X_t = \sum_{j=1}^{4} \alpha_j D_{jt} + \sum_{i=1}^{4} \beta_i X_{t-i} + \varepsilon_t$$

then, $|\hat{\varepsilon}_t|$ is the estimate for quarterly volatility of macroeconomic variable X_t .


quarterly PPI vol
0.07
0.06
0.05
0.04
0.03
0.02
0.01
1885
1915
1945
1975
2004
quarterly MB growth vol
0.03
0.025
0.02
0.015
0.01
0.005
1885
1915
1945
1975
2004
quarterly IP growth vol
0.1
0.08
0.06
0.04
0.02
1885
1915
1945
1975
2004
quarterly Spread vol
0.06
0.05
0.04
0.03
0.02
0.01
1885
1915
1945
1975
2004
quarterly GDP growth vol
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
1885
1915
1945
1975
2001

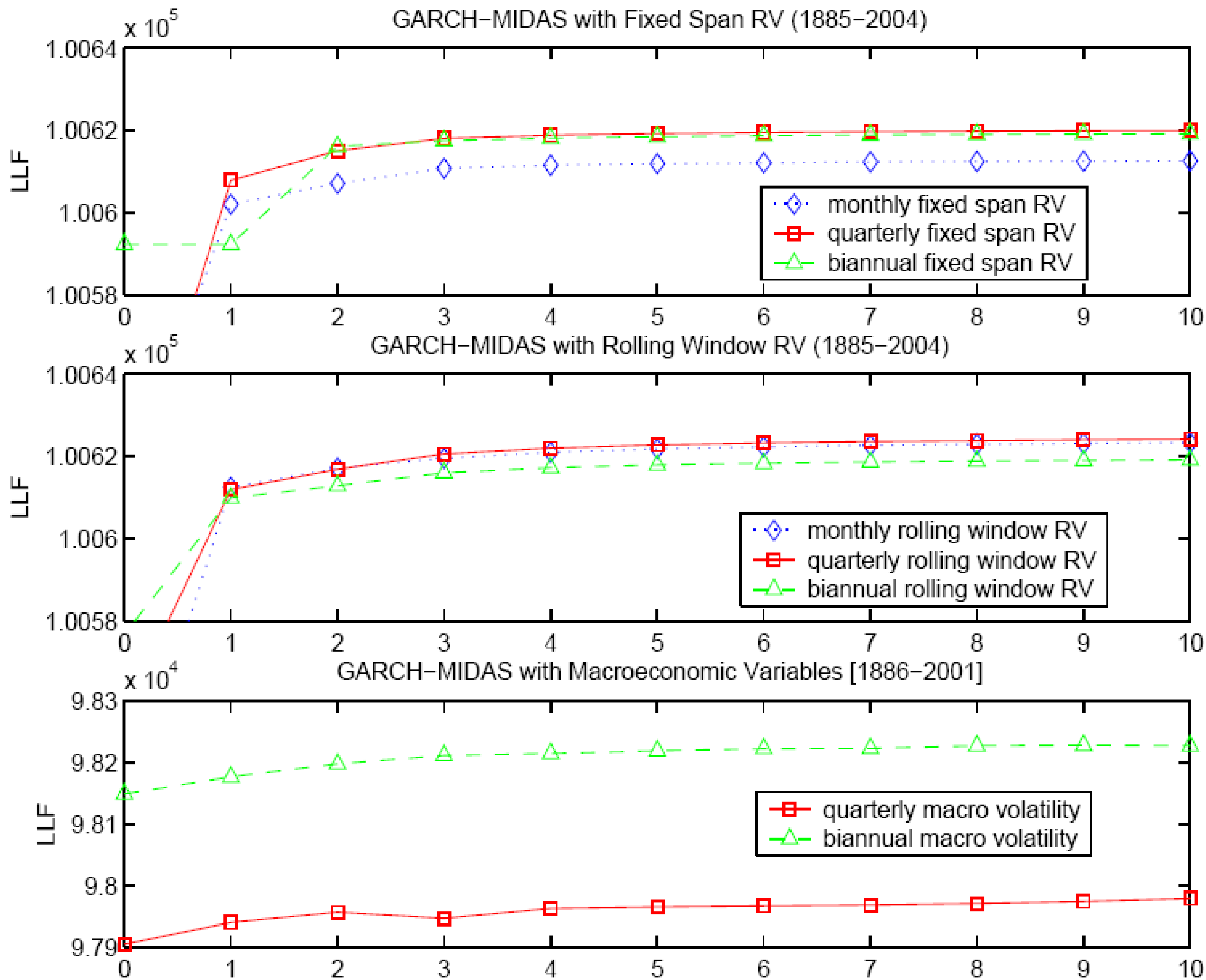
Estimation & Model Selection

- We take the conventional methodology which is used to estimate GARCH-type models, namely QMLE.

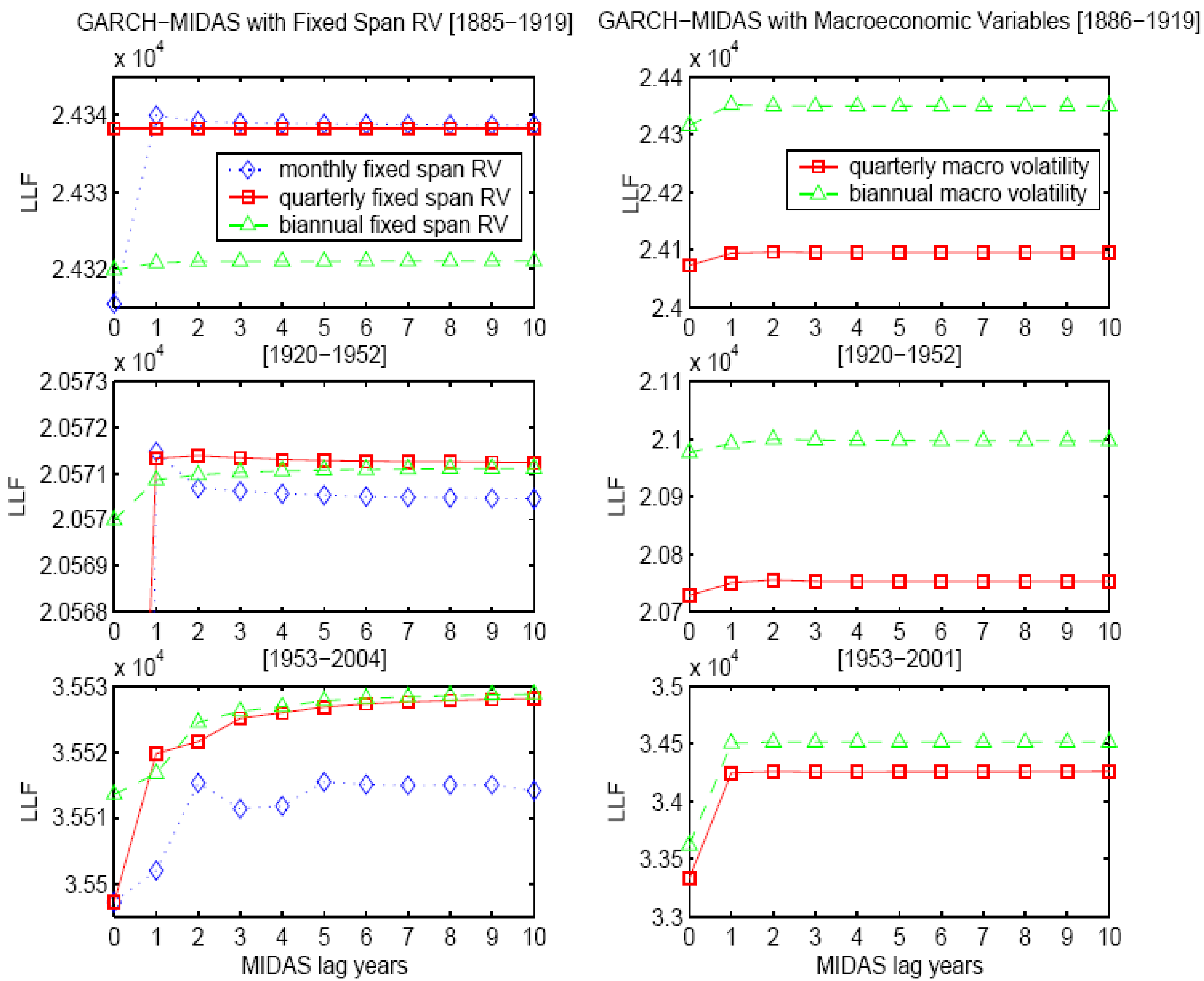
- Log Likelihood function is the following:

$$LLF = -\frac{1}{2}\sum_{t=1}^{T}\left[\log(\tau_t g_t) + \frac{(r_t - \mu)^2}{\tau_t g_t}\right]$$

- Remember that *K*, the number of lags in the MIDAS regression for ‘tau,’ and ‘t’ (period) are the choice variables. For given K and ‘t,’ the model is estimated. Moreover, recall that the number of parameters does not change by varying K and ‘t.’

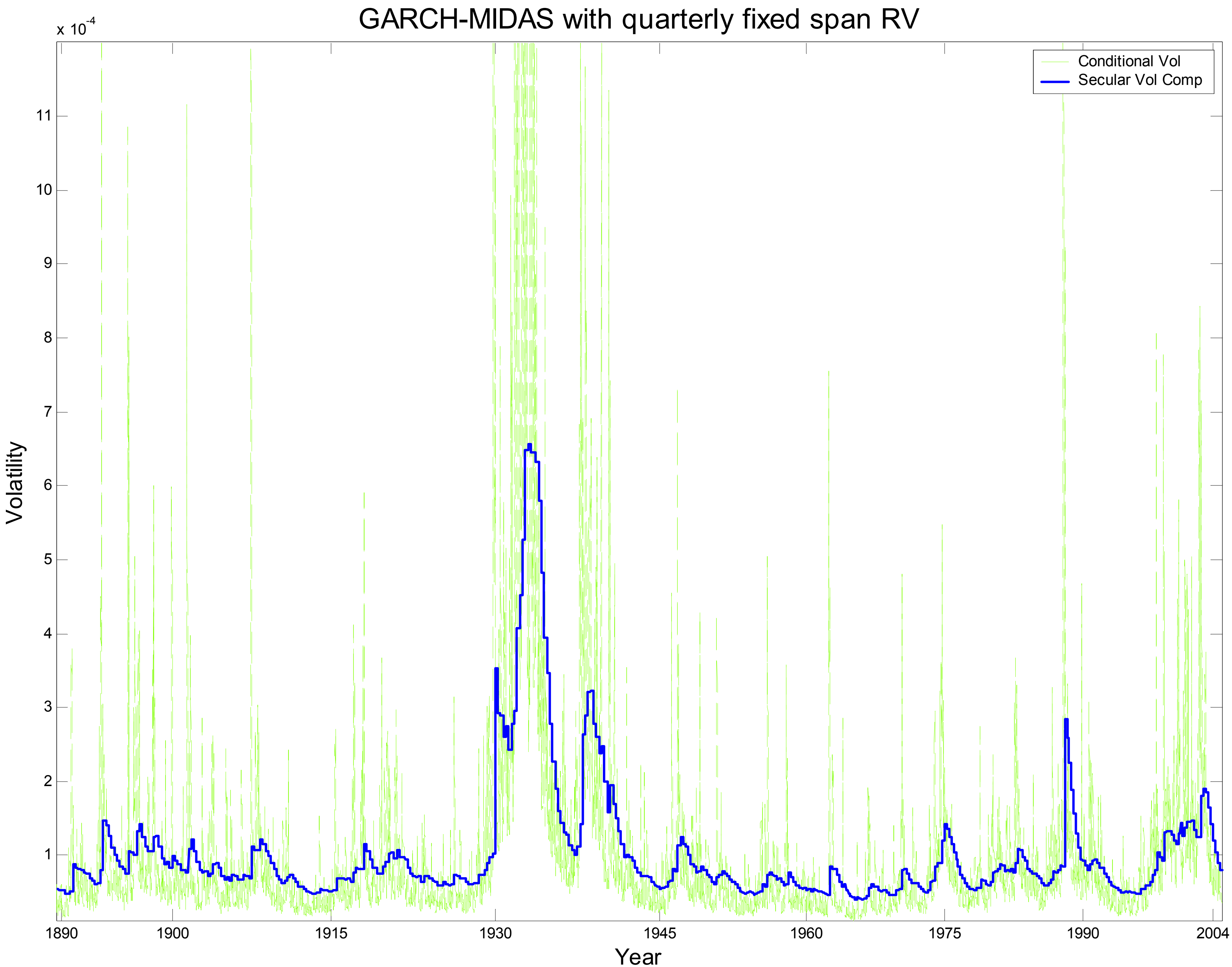


GARCH−MIDAS with Fixed Span RV (1885−2004)
x 10^5
1.0064
1.0062
1.006
1.0058
LLF
0 1 2 3 4 5 6 7 8 9 10
monthly fixed span RV
quarterly fixed span RV
biannual fixed span RV
GARCH−MIDAS with Rolling Window RV (1885−2004)
x 10^5
1.0064
1.0062
1.006
1.0058
LLF
0 1 2 3 4 5 6 7 8 9 10
monthly rolling window RV
quarterly rolling window RV
biannual rolling window RV
GARCH−MIDAS with Macroeconomic Variables [1886−2001]
x 10^4
9.83
9.82
9.81
9.8
9.79
LLF
0 1 2 3 4 5 6 7 8 9 10
quarterly macro volatility
biannual macro volatility

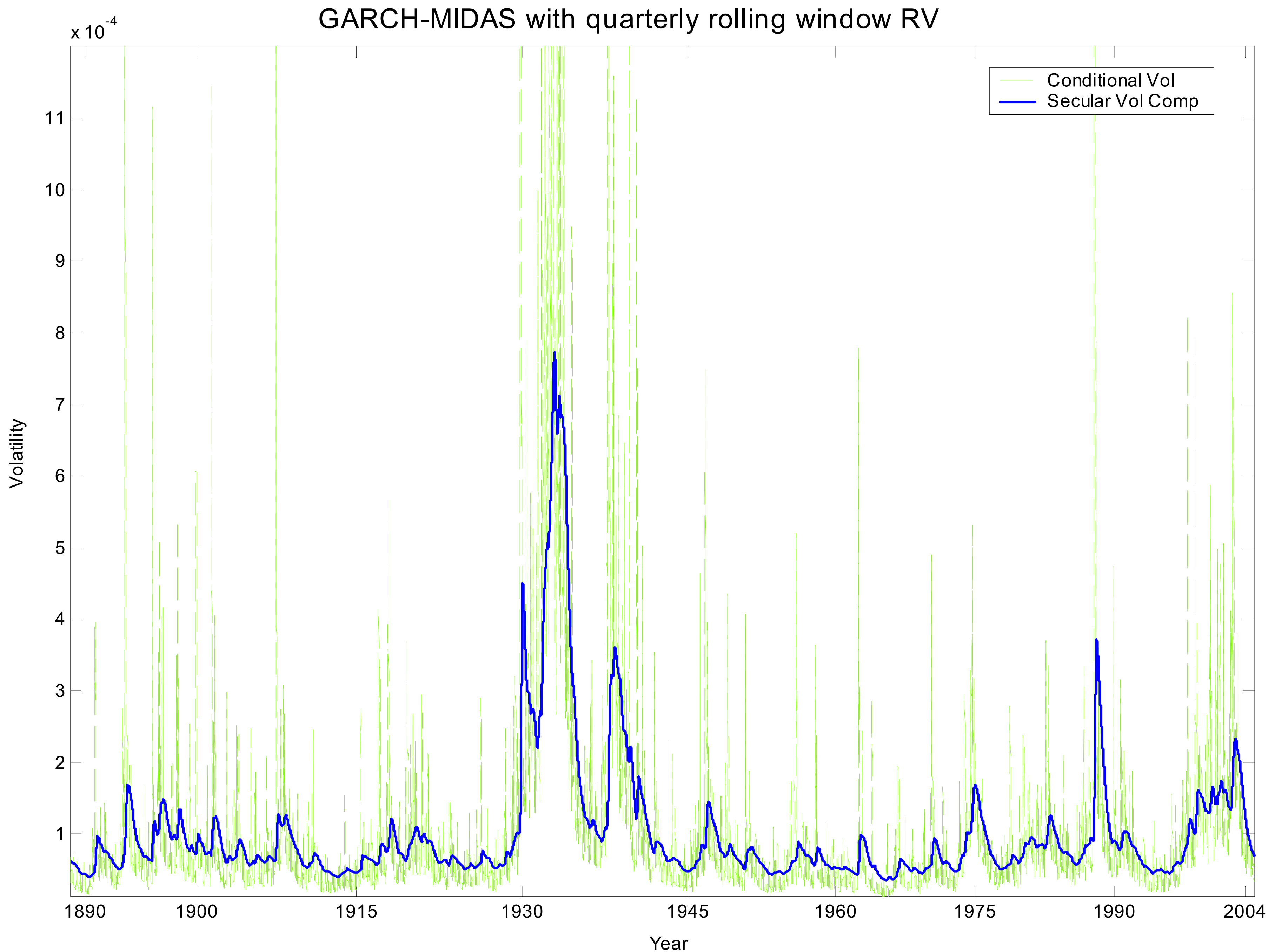

GARCH−MIDAS with Fixed Span RV [1885−1919]
GARCH−MIDAS with Macroeconomic Variables [1886−1919]
x 10^4
LLF
monthly fixed span RV
quarterly fixed span RV
biannual fixed span RV
quarterly macro volatility
biannual macro volatility
[1920−1952]
[1953−2004]
[1953−2001]
MIDAS lag years

RV-based GARCH-MIDAS model estimation

Whole Sample	Specification	μ	α	β	θ	ω	m
Fixed span RV	Qtr/4yr	0.00058	0.10690	0.86032	0.00964	4.83019	0.00003
		(14.37)	(13.28)	(77.22)	(17.32)	(3.46)	(11.71)
Rolling window RV	Qtr/4yr	0.00061	0.10640	0.85481	0.01125	5.99019	0.00002
		(13.89)	(12.75)	(54.55)	(18.20)	(1.41)	(10.61)
Subsample (w/ fixed span RV)							
1885 - 1919	Mth/1yr	0.00062	0.15460	0.75121	0.02021	4.86451	0.00003
		(7.98)	(11.14)	(28.99)	(10.63)	(5.26)	(10.48)
	Qtr/2yr	0.00051	0.14737	0.78901	0.00485	22.65451	0.00004
		(7.27)	(9.31)	(36.14)	(6.99)	(1.34)	(15.33)
1920 - 1952	Mth/1yr	0.00068	0.09793	0.84742	0.03177	1.89311	0.00002
		(6.39)	(7.34)	(35.94)	(15.34)	(2.55)	(5.19)
	Qtr/2yr	0.00077	0.10399	0.85701	0.00943	2.66315	0.00003
		(7.29)	(10.09)	(61.62)	(12.52)	(4.14)	(6.87)
1953 - 2004	Bia/6yr	0.00049	0.08344	0.90340	0.00806	3.61519	0.00001
		(7.66)	(6.51)	(60.83)	(6.12)	(0.94)	(1.38)
	Qtr/6yr	0.00052	0.08924	0.89537	0.01603	2.54990	0.00001
		(7.18)	(6.70)	(63.00)	(7.65)	(2.61)	(1.87)



GARCH-MIDAS with quarterly fixed span RV
x 10^-4
Conditional Vol
Secular Vol Comp
Volatility
11
10
9
8
7
6
5
4
3
2
1
1890
1900
1915
1930
1945
1960
1975
1990
2004
Year



GARCH-MIDAS with quarterly rolling window RV
x 10^-4
11
10
9
8
7
6
5
4
3
2
1
Volatility
1890
1900
1915
1930
1945
1960
1975
1990
2004
Year
Conditional Vol
Secular Vol Comp

I. Economic Sources of Volatility

- We start with GARCH-MIDAS with fixed span RV

$$\tau_t = m + \theta \sum_{k=1}^{K} \varphi_k(\omega) RV_{t-k}$$

- We follow the two-step procedure of Schwert – later we look at the direct MIDAS specification with macro variables. The difference with Schwert is that we replace the 'noisy' RV measure with tau.

Economic Sources: Two-step procedure

- We take a two-step procedure like Schwert (1989) to investigate the relation between stock market volatility and macroeconomic volatility.

- Granger Causality Test:

$$y_t = c + \sum_{i=1}^{8} a_i y_{t-i} + \sum_{j=1}^{8} a'_j x_{t-j} + u_t$$

- *F*-test of VAR

$$\Psi_t = \sum_{i=1}^{8} B_i \Psi_{t-i} + \sum_{j=1}^{4} b_j D_{jt} + \eta_t$$

This table summarizes the results of Granger causality tests and F-tests from the VAR. The relations significant at 5% level are marked with $\rightarrow$ and ones significant at 1% level are marked with $\Rightarrow$. In the table, Spread is represented with Δ. Also, note that, although we use, for notational convenience, RV and τ in the table below, it is $\sqrt{RV_t}$ and $\sqrt{\tau_t}$ that are actually used to investigate relations with macroeconomic volatility.

	RV_t				τ_t			
	Granger Causality		VAR		Granger Causality		VAR	
	RV $\rightarrow$ Macro	Macro $\rightarrow$ RV	RV $\rightarrow$ Macro	Macro $\rightarrow$ RV	$\tau \rightarrow$ Macro	Macro $\rightarrow \tau$	$\tau \rightarrow$ Macro	Macro $\rightarrow \tau$
				1885-2004				
PPI								
MB	RV $\Rightarrow$ MB		RV $\Rightarrow$ MB		$\tau \Rightarrow$ MB		$\tau \Rightarrow$ MB	
IP	RV $\Rightarrow$ IP		RV $\Rightarrow$ IP		$\tau \Rightarrow$ IP	IP $\Rightarrow \tau$	$\tau \Rightarrow$ IP	IP $\Rightarrow \tau$
Spread								
GDP	RV $\Rightarrow$ GDP		RV $\Rightarrow$ GDP		$\tau \rightarrow$ GDP	GDP $\rightarrow \tau$	$\tau \Rightarrow$ GDP	
				1885-1919				
PPI								PPI $\rightarrow \tau$
MB	RV $\Rightarrow$ MB		RV $\Rightarrow$ MB		$\tau \Rightarrow$ MB		$\tau \Rightarrow$ MB	
IP						IP $\rightarrow \tau$		
Spread			RV $\rightarrow \Delta$			$\Delta \Rightarrow \tau$		$\Delta \rightarrow \tau$
GDP	RV $\rightarrow$ GDP				$\tau \rightarrow$ GDP			
				1920-1952				
PPI								
MB	RV $\rightarrow$ MB				$\tau \rightarrow$ MB			
IP	RV $\Rightarrow$ IP				$\tau \rightarrow$ IP			IP $\rightarrow \tau$
Spread	RV $\Rightarrow \Delta$		RV $\Rightarrow \Delta$		$\tau \Rightarrow \Delta$		$\tau \Rightarrow \Delta$	
GDP								
				1953-2004				
PPI						PPI $\rightarrow \tau$		PPI $\rightarrow \tau$
MB								
IP								
Spread								
GDP								

Economic Sources: Direct link

- Now, we look at the relation between stock market volatility and macroeconomic volatility by the direct link in GARCH-MIDAS with macroeconomic volatility:

$$
\begin{aligned}
\tau_t &= m + kI_t^{BC} + \sum_{k=1}^{K} \varphi_k(\omega_1)\left(\theta_1 X_{t-k}^{spread} + \theta_2 X_{t-k}^{IP} + \theta_3 X_{t-k}^{GDP}\right) + \\
&\quad + \sum_{k=1}^{K} \varphi_k(\omega_2)\left(\theta_4 X_{t-k}^{PPI} + \theta_5 X_{t-k}^{MB}\right) \\
&= m + kI_t^{BC} + \tau_t^{real} + \tau_t^{nominal}
\end{aligned}
$$

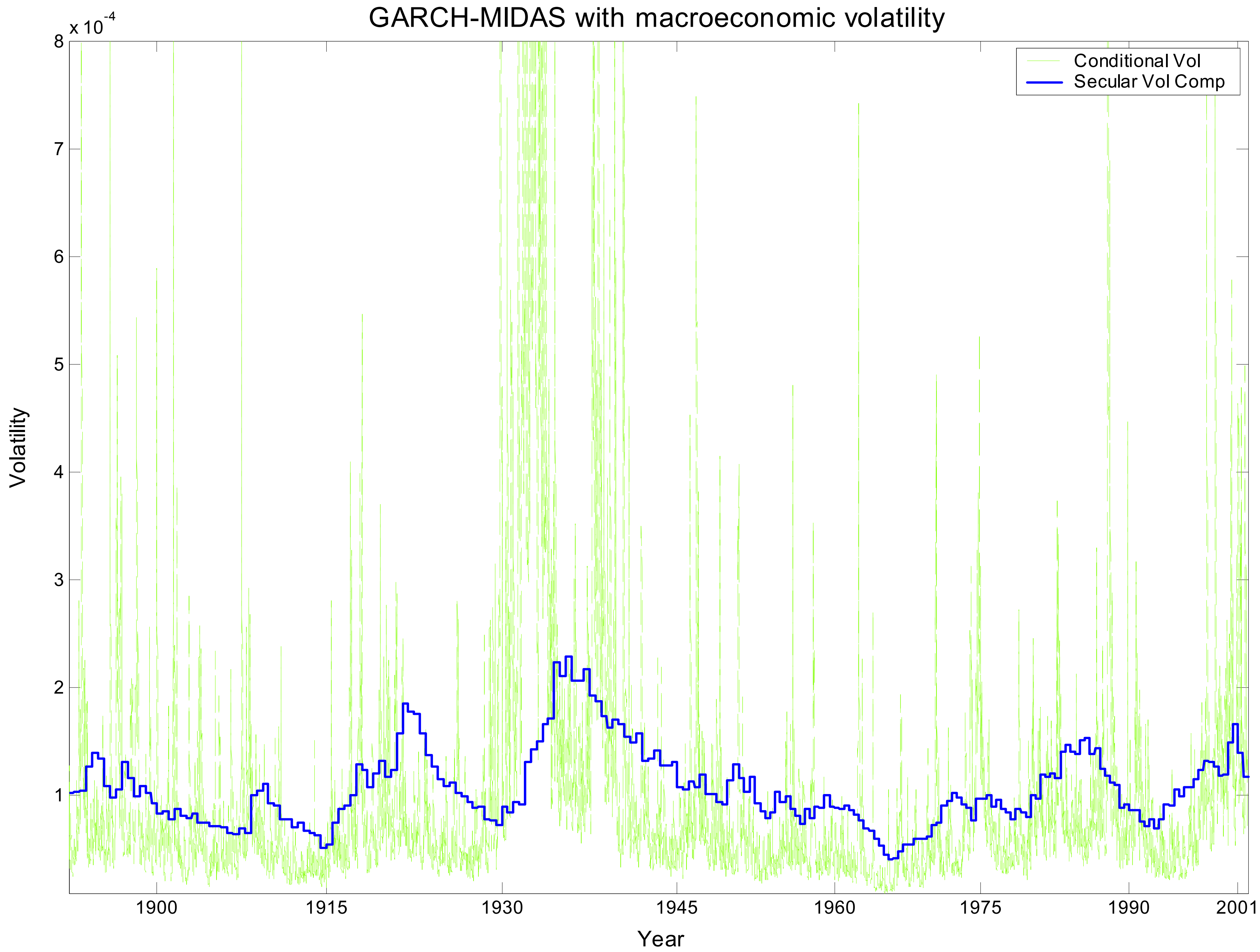
- Note that the one-sided filters imply Granger causality restrictions. We stick with one-sided filters as we want to stay with prediction models. (Two-sided filters in progress)

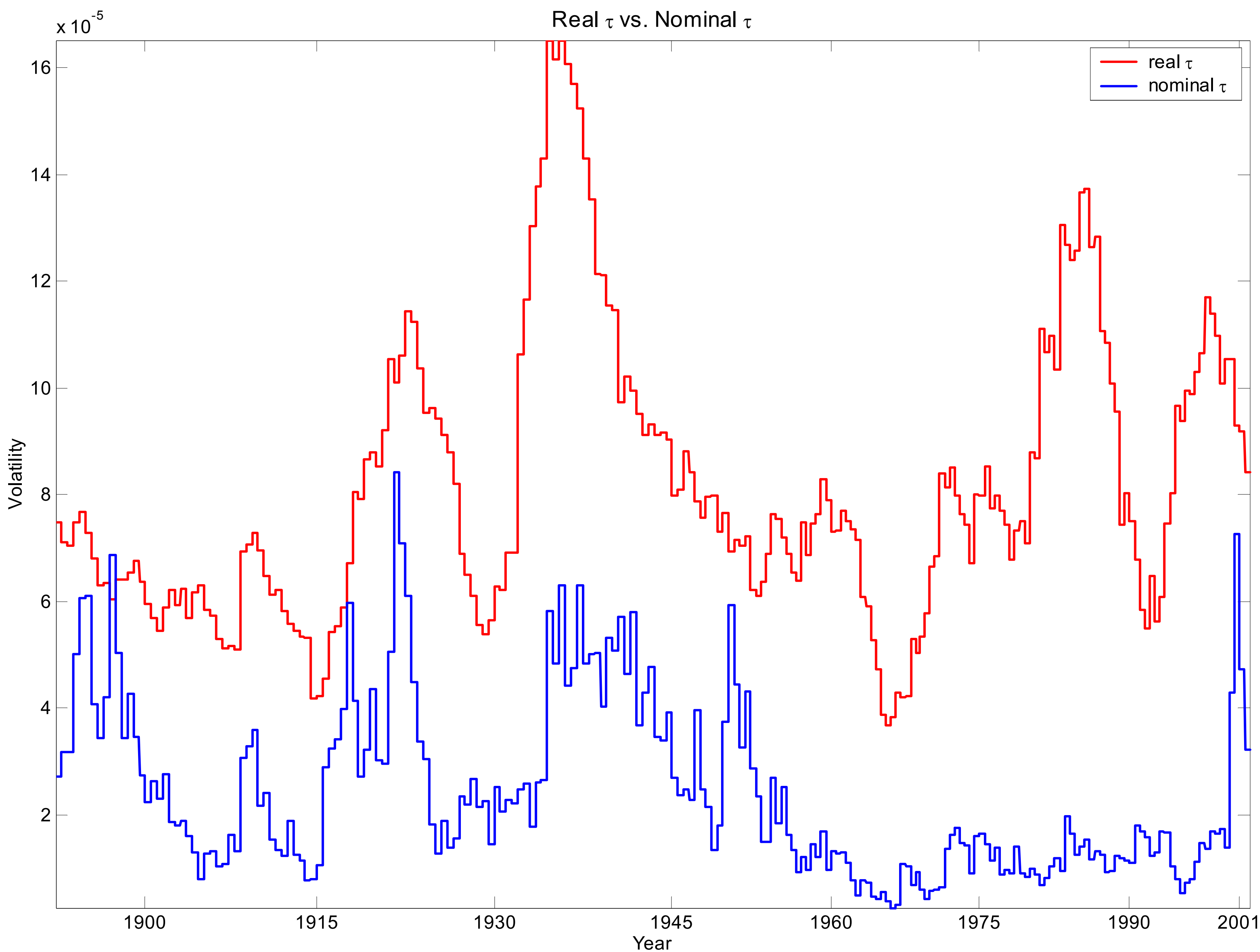
Findings

- From the previous results, we expect IP and PPI will be important in predicting future tau component. However, IP is only significant at the whole sample whereas PPI stays strong over the whole sample and various subsamples.

- Overall, GDP and IP are weak and others appear quite strong in prediction.

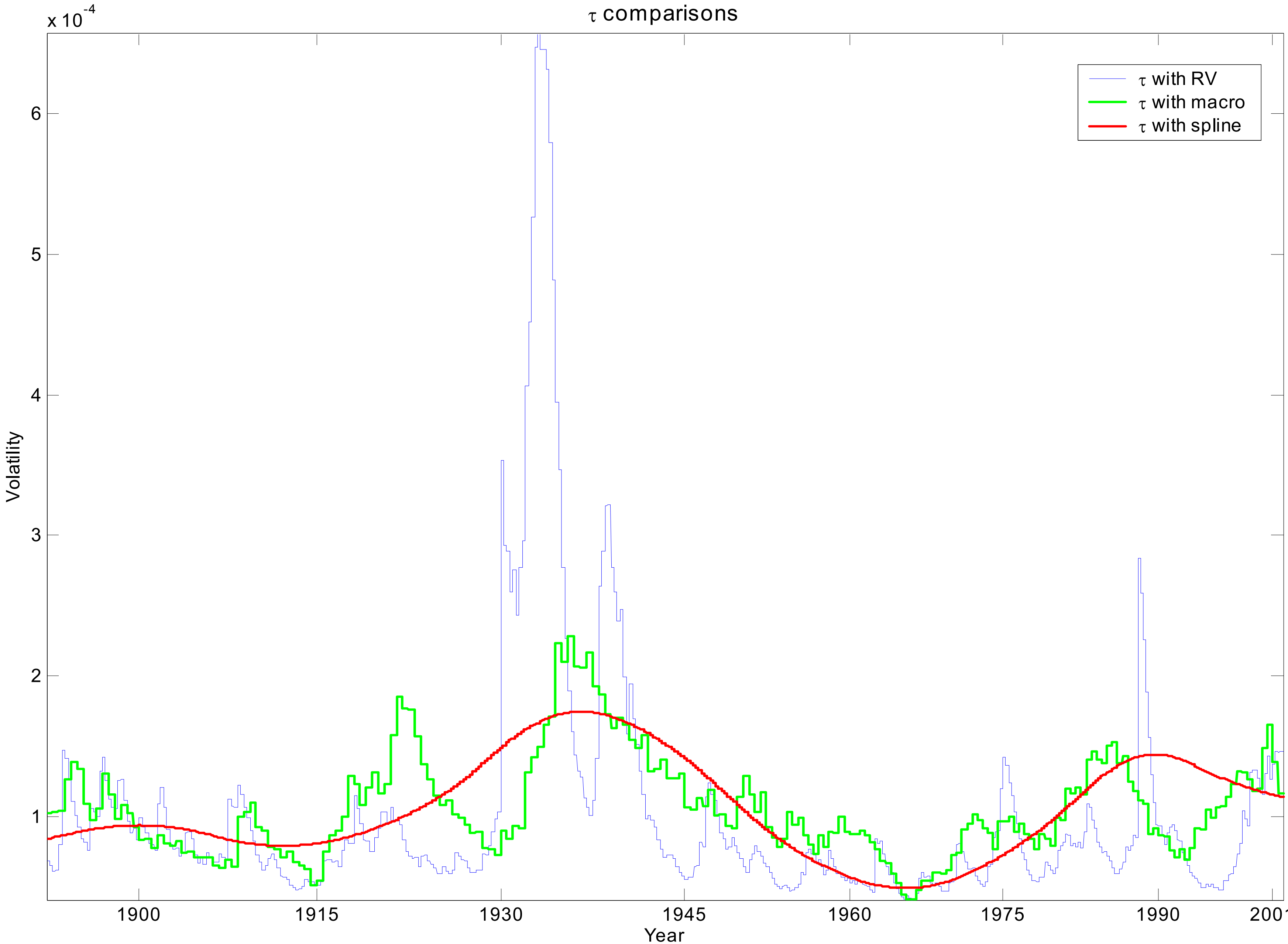
- NBER recession indicator turns out to be insignificant (because of presence of other variables).


GARCH-MIDAS with macroeconomic volatility
x 10^-4
8
7
6
5
4
3
2
1
Volatility
Conditional Vol
Secular Vol Comp
1900
1915
1930
1945
1960
1975
1990
2001
Year

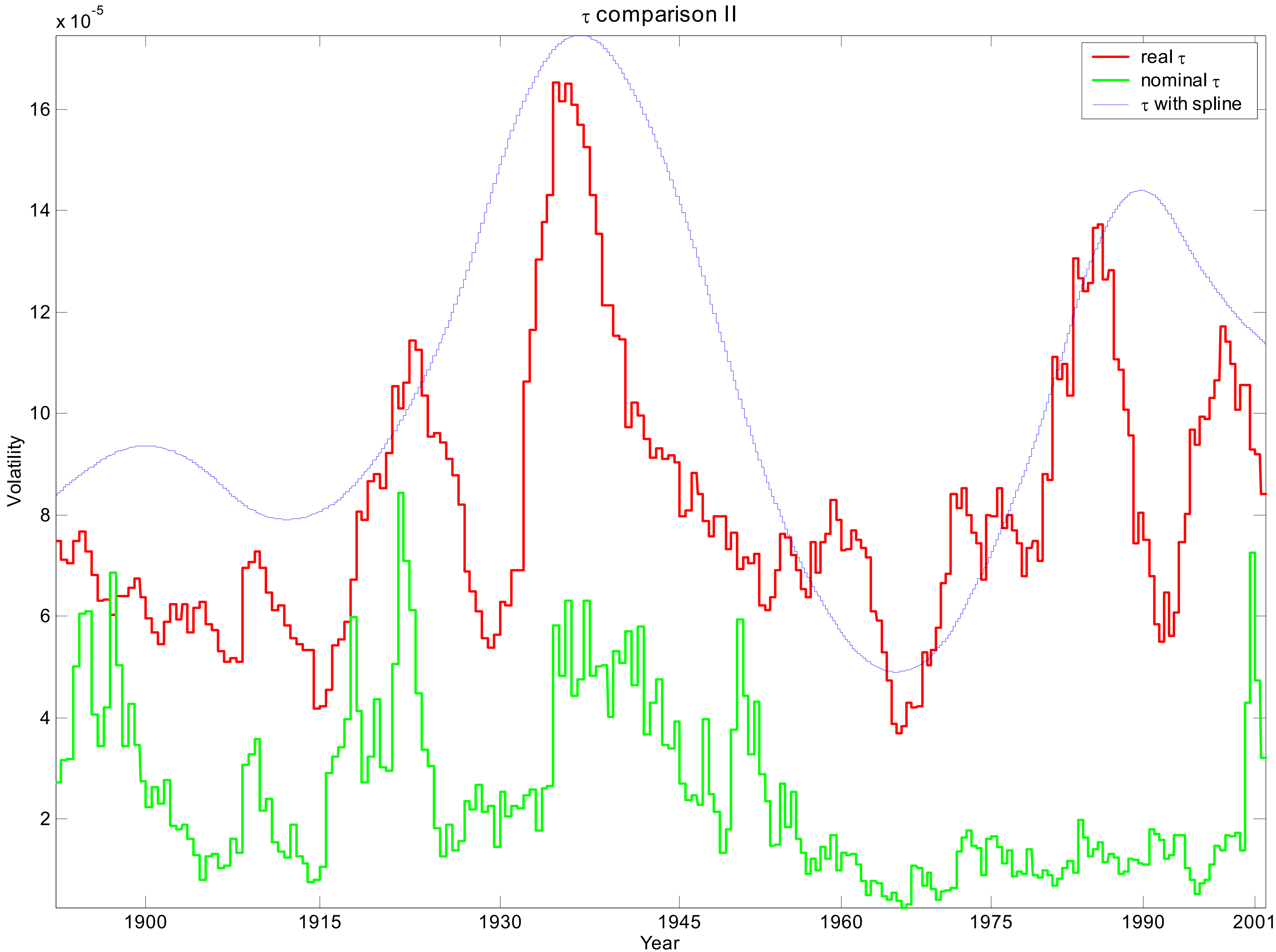

Real τ vs. Nominal τ
x 10^-5
16
14
12
10
8
6
4
2
Volatility
1900
1915
1930
1945
1960
1975
1990
2001
Year
real τ
nominal τ

Revisiting the Spline-GARCH

- We compare the ‘tau’ components identified by three different models:
  - GARCH-MIDAS with quarterly fixed span RV
  - GARCH-MIDAS with macroeconomic volatility
  - Spline-GARCH



τ comparisons
x 10^-4
Volatility
Year
τ with RV
τ with macro
τ with spline
1900
1915
1930
1945
1960
1975
1990
2001
1
2
3
4
5
6


τ comparison II
x 10^-5
real τ
nominal τ
τ with spline
16
14
12
10
8
6
4
2
Volatility
1900
1915
1930
1945
1960
1975
1990
2001
Year

II. Cross-sectional Study

Introduction

- The question of how aggregate volatility affects the cross-section of expected returns has received less attention
- We are going to study whether the two volatility components extracted by the GARCH-MIDAS model are priced across assets.

- Ang, Hodrick, Xing, and Zhang (2006)
  - Sample period: 1986-2000
  - Aggregate market volatility has significant and negative price of risks.
  - However, a two-factor model with the market return and total market risk as pricing factors only marginally reduces pricing errors compared to the CAPM.
- Adrian and Rosenberg (2006)
  - Sample period: 1963-2003
  - Both short-run component (‘*sres*’) and long-run component (‘*lres*’) of market volatility have significant and negative price of risks.
  - Three factor model with the market return and two volatility components as pricing factors reduces pricing errors as much as the Fama-French three factor model.

Conditional Linear Factor Pricing Model

- Whole sample period: 1926-2004
- We consider a factor model for the monthly frequency
- For construction of factors from volatility component innovations, we used GARCH-MIDAS model with monthly fixed span RV and monthly macroeconomic volatility. (time aggregation problem)
- We focus on pricing size and book-to-market sorted portfolio used in Fama and French (1992, 1993)

$$
\begin{aligned}
r_{t+1}^{j} &= \alpha_t^j + \beta_t^{j,M}(r_{t+1}^{M} - \gamma_t^M) + \beta_t^{j,\bar{g}}(\bar{g}_{t+1} - \gamma_t^{\bar{g}}) + \beta_t^{j,\bar{\tau}}(\bar{\tau}_{t+1} - \gamma_t^{\bar{\tau}}) + \\
&\quad + \sum_{k=1}^{K} \beta_t^{j,k}(f_{t+1}^{k} - \gamma_t^k) \\
\alpha_t^j &= E_t(r_{t+1}^j) = \beta_t^{j,M}\lambda_t^M + \beta_t^{j,\bar{g}}\lambda_t^{\bar{g}} + \beta_t^{j,\bar{\tau}}\lambda_t^{\bar{\tau}} + \sum_{k=1}^{K}\beta_t^{j,k}\lambda_t^k
\end{aligned}
$$

where γ's are conditional means of the corresponding factors.

Prices of Market Volatility Component Risk

- The following table reports from Fama and MacBeth (1973) regressions for the Size and Book-to-Market sorted portfolios of Fama and French (1993).

- In the first stage, excess portfolio returns are regressed on the pricing factors to obtain factor loadings.

- In the second stage, for each month, excess portfolio returns are regressed on the loadings with an intercept, giving an estimate of the price of risk for each factor.

Findings (RV-based GARCH-MIDAS)

- (***g-risk*** / ***τ-risk***) have persistently (negative / positive) and significant price of risk whereas price of ***τ*g-risk*** changes sign depending on the specification.

- ***g inv*** is fairly correlated with market factor and ***τ inv***. Other than these relation, these factors do not show any substantial correlation.

- ***g inv*** is highly correlated with ***liq inv*** as was expected. However, ***τ inv*** shows no substantial correlation with other factors.

- Short- and long-run component of Adrian and Rosenberg (2006) are quite closely correlated and both of them fairly correlates with ***liq inv***.

Volatility component factors from GARCH-MIDAS with monthly fixed span RV (Table 8)

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)
intercept	0.00402	0.00219	0.01975	0.01661	0.02214	0.02074	0.00673	0.01296	0.01556	0.02212
	1.18	0.64	5.17	3.28	3.55	4.87	2.00	3.38	3.10	3.55
market	0.00402	0.00553	-0.01278	-0.00968	-0.01537	-0.01373	0.00035	-0.00581	-0.00867	-0.01528
	1.06	1.44	-3.12	-1.83	-2.40	-3.06	0.09	-1.40	-1.65	-2.39
g innov		-3.04063			-3.90517	6.37459	-12.59631	-13.24991	-5.75426	
		-1.05			-1.30	1.92	-4.01	-4.11	-1.93	
τ innov		0.00014			0.00016	0.00032	-0.00006	0.00002		0.00015
		2.77			2.81	5.55	-1.07	0.33		2.74
SMB			0.00166	0.00178	0.00301	0.00326			0.00208	0.00243
			1.44	1.56	2.60	2.86			1.83	2.12
HML			0.00485	0.00432	0.00526		0.00474		0.00472	0.00465
			3.69	3.11	3.68		3.60		3.38	3.29
UMD				0.00679	-0.00862			-0.01787	0.00204	-0.00040
				1.13	-1.14			-4.22	0.34	-0.06

	(xi)	(xii)	(xiii)	(xiv)	(xv)
intercept	0.00863	0.01532	0.01579	0.00575	0.01192
	2.13	3.03	4.20	1.52	2.65
market	-0.00078	-0.00837	-0.00812	0.00166	-0.00436
	-0.18	-1.59	-2.02	0.40	-0.91
τ*g innov	0.00196	-0.00097	0.00275	-0.00151	-0.00100
	2.11	-1.39	3.27	-2.12	-1.46
SMB		0.00184	0.00133		
		1.62	1.15		
HML		0.00449		0.00475	
		3.23		3.61	
UMD		0.00540			-0.01592
		0.91			-3.87

Correlation of Factors Considered

	market	g innov	τ innov	SMB	HML	UMD
			1926 ~ 2004			
market	1.00	-0.26	-0.01	0.33	0.22	-0.34
g innov	-0.26	1.00	0.24	-0.14	0.07	-0.08
τ innov	-0.01	0.24	1.00	-0.03	0.12	-0.06
SMB	0.33	-0.14	-0.03	1.00	0.10	-0.19
HML	0.22	0.07	0.12	0.10	1.00	-0.51
UMD	-0.34	-0.08	-0.06	-0.19	-0.51	1.00

Correlations with Other Factors (1962-2003)

	vres	sres	lres	$\tau * g$ inv	g inv	τ inv	liq inv	skew inv
vres	1	0.45	0.68	0.56	0.85	0.02	-0.50	-0.08
sres	0.45	1	0.51	0.16	0.24	0.03	-0.34	-0.13
lres	0.68	0.51	1	0.37	0.49	0.14	-0.44	-0.03
$\tau * g$ inv	0.56	0.16	0.37	1	0.66	0.72	-0.30	-0.05
g inv	0.85	0.24	0.49	0.66	1	0.13	-0.47	-0.10
τ inv	0.02	0.03	0.14	0.72	0.13	1	-0.07	0.05
liq inv	-0.50	-0.34	-0.44	-0.30	-0.47	-0.07	1	0.04
skew inv	-0.08	-0.13	-0.03	-0.05	-0.10	0.05	0.04	1

Findings (Macro-based GARCH-MIDAS)

- Just like the previous results, ***g-risk*** is has negative and significant price of risk. For **τ*-risk***, **τ_n*-risk*** is positively priced whereas **τ_r*-risk*** is negatively priced. However, **τ_n-*risk*** seems to have positive price of risk only when we include Great Depression in the sample period (***i.e.*** 1926-2004 and 1926-1962).

- ***g inv*** is fairly correlated with market factor. However, ***g inv***, ***τ_n inv***, and ***τ_r inv*** have very low correlations and these factors make good orthogonal factors.

- ***g inv*** closely correlates with ***liq inv*** whereas ***τ_n inv*** and ***τ_r inv*** have virtually no correlation with any other factors.

Volatility component factors are from GARCH-MIDAS model with macroeconomic volatility (Table 9b)

	(xi)	(xii)	(xiii)	(xiv)	(xv)	(xvi)	(xvii)	(xviii)	(xix)	(xx)
intercept	0.01122	0.00749	0.01408	0.01535	0.01459	0.01855	0.01538	0.01265	0.01871	0.01498
	3.98	2.05	4.90	4.76	3.47	5.58	4.72	3.75	5.25	4.25
market	-0.00290	0.00039	-0.00608	-0.00698	-0.00669	-0.01059	-0.00715	-0.00450	-0.01069	-0.00705
	-0.90	0.10	-1.85	-1.93	-1.51	-2.86	-1.98	-1.20	-2.74	-1.81
g innov	-10.73904	-6.39250	-8.22595	-8.79356	-5.44028	-6.07756	-6.02131			
	-4.28	-2.80	-3.44	-3.59	-2.47	-2.63	-2.51			
τ innov										
τ_r innov	0.00001		0.00001	0.00001		0.00001		-0.00002		-0.00001
	2.01		1.77	0.95		0.85		-2.43		-2.21
τ_n innov		0.00004	0.00004		0.00005	0.00005			0.00003	0.00003
		3.00	2.65		3.49	3.40			2.23	1.90
SMB							0.00256	0.00235	0.00196	0.00237
							2.22	2.05	1.67	2.06
HML	0.00488	0.00512	0.00510				0.00415	0.00371	0.00403	0.00412
	3.95	4.12	4.10				3.00	2.70	2.87	2.93
UMD				-0.02552	-0.02231	-0.02359	0.00503	0.00538	0.00683	0.00303
				-4.69	-4.11	-4.40	0.90	0.99	1.13	0.53

Correlation of Factors Considered

	market	g innov	τ innov	τ_r innov	τ_n innov	SMB	HML	UMD
				1926 $\sim$ 2004				
market	1.00	-0.24	0.06	-0.02	0.08	0.38	0.09	-0.17
g innov	-0.24	1.00	0.01	0.04	0.01	-0.11	0.06	-0.11
τ innov	0.06	0.01	1.00	0.12	0.74	0.11	0.03	-0.06
τ_r innov	-0.02	0.04	0.12	1.00	0.02	0.08	-0.02	0.02
τ_n innov	0.08	0.01	0.74	0.02	1.00	0.11	0.06	-0.06
SMB	0.38	-0.11	0.11	0.08	0.11	1.00	0.10	-0.18
HML	0.09	0.06	0.03	-0.02	0.06	0.10	1.00	-0.37
UMD	-0.17	-0.11	-0.06	0.02	-0.06	-0.18	-0.37	1.00

Correlations with Other Factors (1962-2003)

	vres	sres	lres	$\tau * g$ inv	g inv	τ inv	τ_{r} inv	τ_n inv	liq inv	skew inv
vres	1	0.45	0.68	0.87	0.83	0.02	0.04	-0.01	-0.50	-0.08
sres	0.45	1	0.51	0.22	0.22	-0.02	0.01	-0.02	-0.34	-0.13
lres	0.68	0.51	1	0.49	0.45	-0.03	0.01	-0.05	-0.44	-0.03
$\tau * g$ inv	0.87	0.22	0.49	1	0.98	0.06	0.05	0.03	-0.43	-0.09
g inv	0.83	0.22	0.45	0.98	1	0.03	0.03	0.01	-0.44	-0.11
τ inv	0.02	-0.02	-0.03	0.06	0.03	1	0.55	0.77	0.00	-0.01
τ_{r} inv	0.04	0.01	0.01	0.05	0.03	0.55	1	-0.11	0.04	-0.02
τ_n inv	-0.01	-0.02	-0.05	0.03	0.01	0.77	-0.11	1	-0.04	0.00
liq inv	-0.50	-0.34	-0.44	-0.43	-0.44	0.00	0.04	-0.04	1	0.04
skew inv	-0.08	-0.13	-0.03	-0.09	-0.11	-0.01	-0.02	0.00	0.04	1

Conclusion

- With regard to relation between stock market volatility and macroeconomic volatility, it looks like stock market volatility contains fair amount of information in predicting macroeconomic uncertainty. For reverse direction, PPI appears strong.

- As was expected, short-run component $\boldsymbol{g}$ is related to the day-to-day liquidity concerns. On the other hand, $\boldsymbol{\tau}$ is also priced in various specifications, but it is not obvious what it is capturing. However, we conjecture that $\boldsymbol{\tau}$ relates to the future expected cashflows and future discount rates.

Extension (Work in Progress)

- $r_i = \mu + \sqrt{\tau_i g_i} \varepsilon_i$

- g_i ~ GARCH (1,1) (like a residual)
- τ_i ~ MIDAS (macro variables)

  × MIDAS (liquidity)

  × MIDAS (analyst's disagreement)

Extension to Correlations (work in progress)

- The component GARCH/MIDAS model can be extended to DCC/MIDAS, where the dynamic correlations have long and short run components.