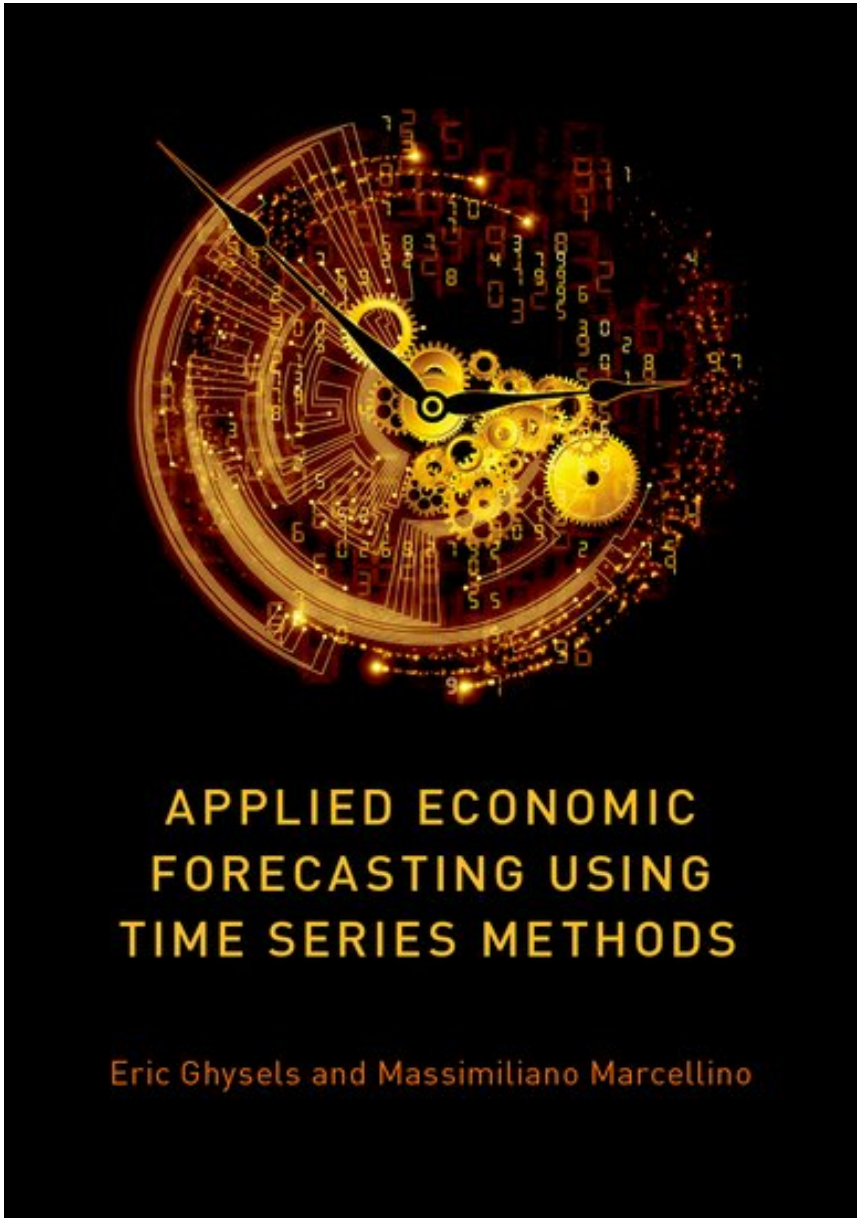

APPLIED ECONOMIC
FORECASTING USING
TIME SERIES METHODS
Eric Ghysels and Massimiliano Marcellino

Applied Economic Forecasting using Time Series Methods

Eric Ghysels and Massimiliano Marcellino

Companion Slides - Chapter 1 The Baseline Linear Regression Model

The Baseline Linear Regression Model - Outline

Overview

- We are interested in assessing the explanatory power of a set of k variables, grouped into X, for the variable y.
- From a statistical point of view, both the elements of X and y are collections of random variables, for which we have a set of realizations, X_t and y_t, $t = 1, \ldots, T$.
- For example, we can collect data on inflation and the short term interest rate over a given period at a given frequency t, e.g. each month or each quarter, or even each day or even perhaps each minute for the interest rate.

The Baseline Linear Regression Model - Outline

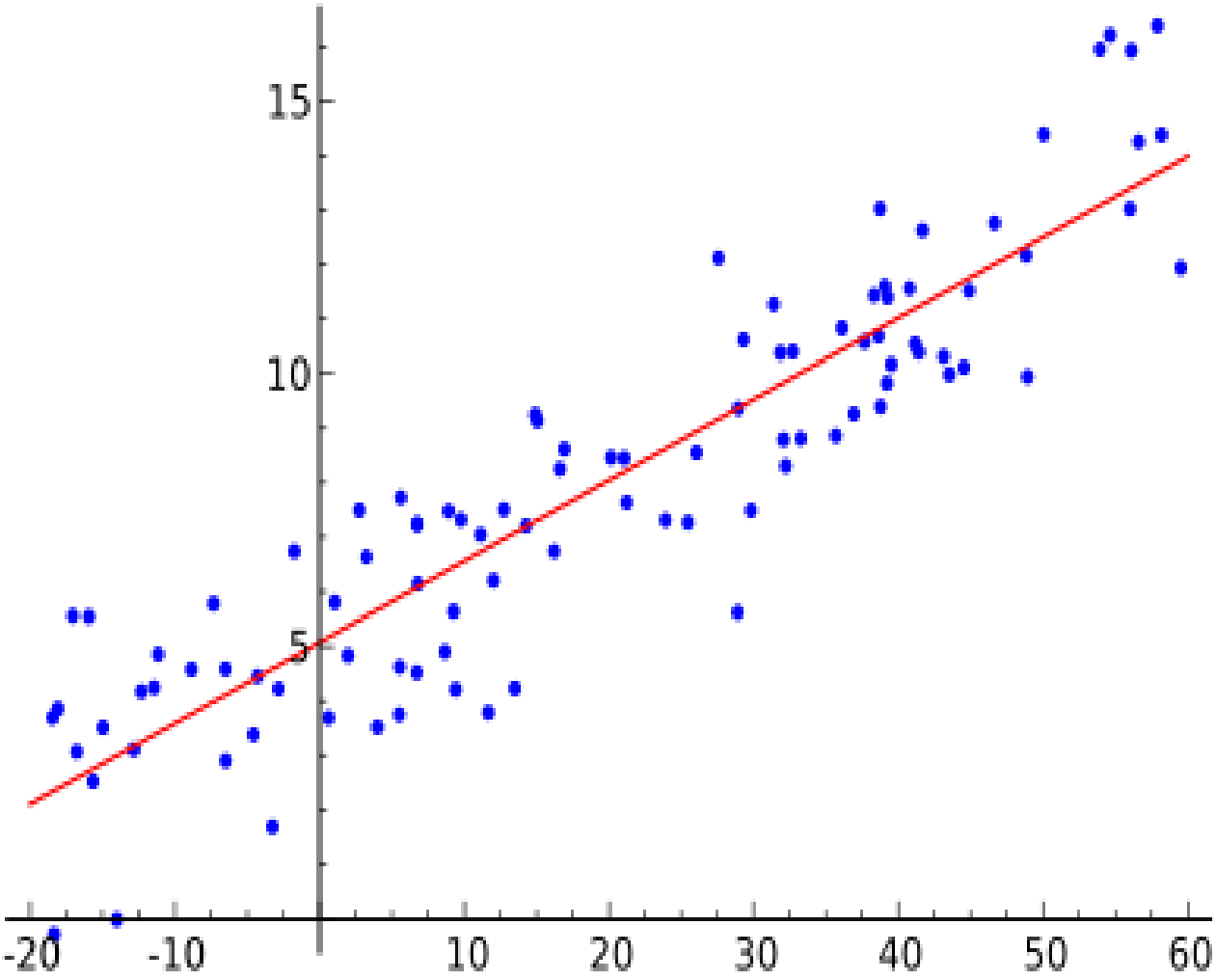
Basic Specification of Linear Regression Model

- Assuming *explanatory variables* X have a linear impact on the *dependent variable* y, we can write the model as follows:

$$y_t = X_{1t}\beta_1 + X_{2t}\beta_2 + ...X_{kt}\beta_k + \epsilon_t,$$

  - $t = 1, \ldots, T$.
  - β_i is a parameter related to the explanatory power of X_{it}, $i = 1, \ldots, k$. It measures the change in the expected value of y_t when there is a marginal change in X_{it}, and the other Xs are kept constant.
  - ϵ_t is an error term which captures the part of y_t that is not explained by the variables in X_t.

Linear Regression Fit



Matrix Representation

If we group y_t and ϵ_t, $t = 1, \ldots, T$, into the $T \times 1$ vectors y and ϵ respectively, $\beta_1, \ldots, \beta_k$ into the $k \times 1$ vector β, and $X_{1t}, \ldots, X_{kt}$, $t = 1, \ldots, T$, into the $T \times k$ matrix X.

$$\begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} X_{11} & \ldots & \ldots & X_{k1} \\ \vdots & & & \vdots \\ X_{1T} & \ldots & \ldots & X_{kT} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} e_1 \\ \vdots \\ e_T \end{pmatrix}$$

We can rewrite the *linear regression model* as

$$\underset{(T\times 1)}{y} = \underset{(T\times k)(k\times 1)}{X \times \beta} + \underset{(T\times 1)}{\epsilon} \tag{1}$$

Linear Regression Assumptions

Additional assumptions on the model in (1), most of which will be relaxed later on:

Assumption (Linear Regression Assumptions)

The linear regression model defined in equation (1) satisfies the following assumptions:

- *LR1:* $E(\epsilon) = 0$,
- *LR2:* $E(\epsilon\epsilon') = \sigma^2 I_T$,
- *LR3:* X *is distributed independently of* ϵ,
- *LR4:* $X'X$ *is non singular.*
- *LR5:* X *is weakly stationary.*

The assumptions are typically known as "weak assumptions."

Linear Regression Assumptions

- LR1: the expected value of the error term should be equal to zero, namely, on average the model should provide a correct explanation for y, which is a basic ingredient of a reasonable model.
- LR2: the matrix of second moments of the errors (which is also equal to their variance covariance matrix given LR1) is diagonal.
  - variance of the error term is stable over time (*homoskedasticity*)
  - errors are uncorrelated over time (*no autocorrelation*)
- LR3: needed to guarantee good properties for the simplest parameter estimator.

Linear Regression Assumptions

- LR4: typically referred to as lack of perfect *multicollinearity*, and it is concerned with indicator redundancy and the identifiability of all the β parameters.

- LR5: concerned with the amount of temporal persistence in the explanatory variables.
  - if each element of X is affected by a shock, the resulting effects on that element (and on y) are not permanent.

Linear Regression Assumptions

- Three additional implicit assumptions on the model in (1):
  - All the relevant explanatory variables are included in the model (no *omitted* variables and no *redundant* variables).
  - Linearity: the model is *linear* in the parameters $beta$.
  - *Stability* of the β parameters over time – the relationship between y and X does not change over the sample under analysis.

The Baseline Linear Regression Model - Outline

OLS Estimator

Several estimators for the parameters of the linear regression model are feasible but, if assuming for the moment that X is deterministic, from the Gauss Markov theorem, the best linear unbiased estimator (*BLUE*) for β is

$$\underset{k\times 1}{\hat{\beta}} = (X'X)^{-1}X'y. \tag{2}$$

- This is *ordinary least squares* (OLS) estimator, which can be derived as the minimizer with respect to β of:

$$\epsilon'\epsilon = \sum_{t=1}^{T} \epsilon_t^2 = \sum_{t=1}^{T} (y_t - X_t\beta)^2$$

OLS – Unbiased

The OLS estimator is *unbiased* because

$$E(\hat{\beta}) = \beta, \tag{3}$$

- If we could draw a very large number of samples from y, each of dimension T, and for each sample j we computed $\hat{\beta}_j$ using the formula in (2), the average across the many samples of $\hat{\beta}_j$ would be equal to β.

- It is the best linear unbiased estimator since it is the most precise in the class of the linear unbiased estimators for β.

OLS – Most Precise Estimator

$\hat{\beta}$ has *minimal variance* in the class of estimators that are obtained as a linear combination of the $y_1, \ldots, y_T$.

- It implies that we are using optimally the available information, in the sense of obtaining the *most precise estimator* for the parameter β:

$$var(\hat{\beta}) = \sigma^2 (X'X)^{-1}, \tag{4}$$

- We can rewrite $var(\hat{\beta})$ as:

$$Var(\hat{\beta}) = \frac{\sigma^2}{T} \left(\frac{X'X}{T}\right)^{-1}$$

OLS – Consistency

Since in general $X'X/T$ converges to a matrix when T diverges while σ^2/T goes to zero, it is:

$$\lim_{T\to\infty} Var(\hat{\beta}) = 0$$

- This finding, combined with the unbiasedness of $\hat{\beta}$, implied that the OLS estimator is *consistent* for β.
  - When the size of the sample size T gets very large $\hat{\beta}$ gets very close to the *true value* β ($\hat{\beta}$ converges in probability to β).

Residuals

Use $\hat{\beta}$ to construct the residules:

$$\hat{\epsilon}_t = y_t - X_t\hat{\beta} \tag{5}$$

and collect $\hat{\epsilon}_t$, $t = 1, \ldots, T$, into the $T \times 1$ vector $\hat{\epsilon}$.

- The residuals are related to the errors but different:

$$\hat{\epsilon} = (I - X(X'X)^{-1}X')\epsilon$$

- Variance of the errors σ^2:

$$\hat{\sigma}^2 = \hat{\epsilon}'\hat{\epsilon}/(T-k), \tag{6}$$

where we normalize by the degrees of freedom to achieve *unbiasedness* of $\hat{\sigma}^2$:

$$E(\hat{\sigma}^2) = \sigma^2.$$

Strong Assumptions

- If in addition to the assumptions stated above, we are also willing to maintain that the *errors are normally distributed*, so that

$$\epsilon \sim N(0, \sigma^2 I_k),$$

- The *distribution of the OLS estimator* is:

$$\sqrt{T}(\hat{\beta} - \beta) \sim N\left(0, \sigma^2 \left(\frac{X'X}{T}\right)^{-1}\right),$$

$$(T-k)\hat{\sigma}^2/\sigma^2 \sim \chi^2(T-k),$$

  where $\hat{\beta}$ and $\hat{\sigma}^2$ are independent

- Adding normality to the assumption LR1 - LR5 gives "*strong assumptions.*"

Large Sample Properties

- Similar results regarding the distribution of the parameter estimators hold when (some of) the elements of X are stochastic but uncorrelated with ϵ and the sample size T is *very large*
  - The results hold asymptotically

- The size of the estimated coefficients depends on the unit of measurement of the variables.

The Baseline Linear Regression Model - Outline

R^2

- It is convenient to examine measures of in sample model fit prior to forecasting, though this should not be the unique forecast selection criterion.
- Coefficient of determination, R^2 compares the *variability* in the dependent variable y with that in the model (OLS) residuals $\hat{\epsilon}$.

$$R^2 = 1 - \frac{\hat{\epsilon}'\hat{\epsilon}}{y'y} = \frac{\widehat{y}'\widehat{y}}{y'y},$$

where

$$\widehat{y} = X\widehat{\beta}$$

is the *model fit*. Hence, it is $0 \leq R^2 \leq 1$, with $R^2 = 0$ when $\hat{\epsilon}'\hat{\epsilon} = y'y$, namely the model has no explanatory power, and $R^2 = 1$ if the fit is perfect.

- It is possible to show that it is also

$$R^2 = corr(y, \widehat{y}),$$

Adjusted R^2

- A problem with the coefficient of determination is that it is monotonically increasing with the number of explanatory variables
- The adjusted R^2 is defined as

$$\overline{R}^2 = 1 - \frac{\hat{\epsilon}'\hat{\epsilon}/(T-k)}{y'y/(T-1)},$$

  - The overall impact of adding explanatory variables to the model on $\overline{R}^2$ will be positive only if the added variables have sufficiently good explanatory power for y.

The Baseline Linear Regression Model - Outline

Constructing Point Forecasts

We are interested in forecasting y_{T+h}, the value the dependent variable will take in period $T+h$, given the linear regression model in (1), with parameters estimated as in (2).

- We can show that the *best linear unbiased forecast* is

$$\underset{1\times k}{\hat{y}_{T+h}} = \underset{1\times k}{X_{T+h}}\ \underset{k\times 1}{\hat{\beta}}\ ,$$

in the sense of producing minimum *forecast error variance* and zero mean forecast error (*forecast unbiasedness*), where the forecast error is

$$e_{t+h} = y_{T+h} - \hat{y}_{T+h}.$$

Forecast Uncertainty

- The variance of forecast error is computed as

$$E(y_{T+h} - \hat{y_{T+h}})^2 = E(X_{T+h}(\beta - \hat{\beta}) + \epsilon_{T+h})^2 = \sigma^2(1 + \underset{1\times k}{X_{T+h}}(\underset{k\times k}{X'X})^{-1}\underset{k\times 1}{X_{T+h}})$$

- Other important elements that increase forecast uncertainty
  - Future values of the explanatory variables X_{T+h} are not known.
  - Explanatory variables to be inserted in the model for y are unknown.
  - The relationship between the explanatory variables and y could be non-linear and/or could change over time.

The Baseline Linear Regression Model - Outline

Constructing Density Forecasts

The density forecast can be used to assign probabilities to specific events of interest concerning the future behavior of the variable y.

- Assume that Assumptions 1 LR1-LR5 are valid, the sample size T is large and in addition the error term in (1) has a normal distribution.

- From the definition of the forecast error $e_{t+h} = y_{T+h} - \hat{y}_{T+h}$, it follows that

$$\left(\frac{y_{T+h} - \hat{y}_{T+h}}{\sqrt{V(e_{T+h})}} \right) \sim N(0,1),$$

which implies

$$y_{T+h} \sim N(\hat{y}_{T+h}, V(e_{T+h})).$$

Constructing Density Forecasts

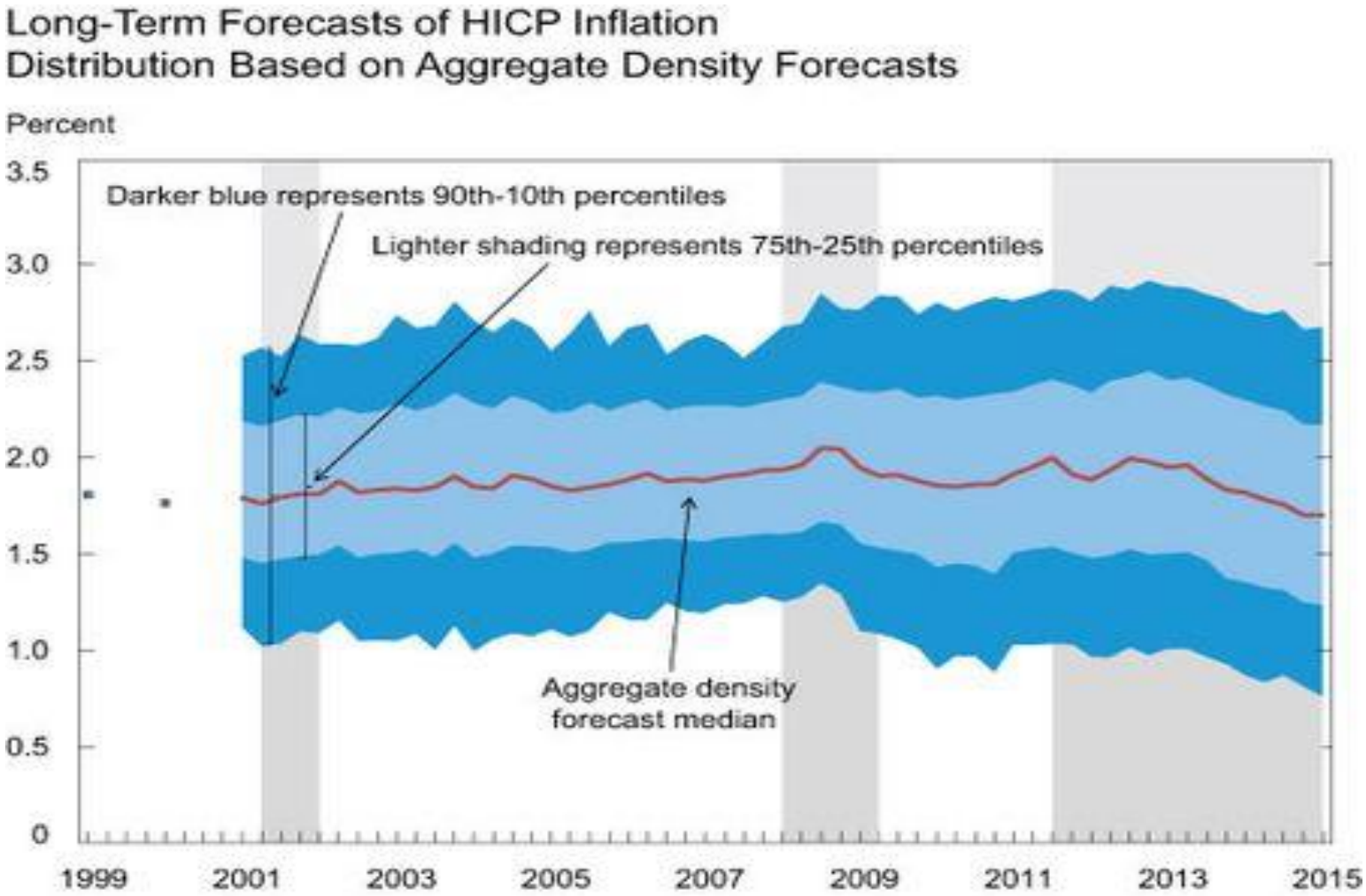
- Another use of the forecast density is to construct *interval forecasts* for y_{T+h}. A $[1-\alpha]$ % forecast interval is represented as

$$\hat{y}_{T+h} - c_{\alpha/2}\sqrt{V(e_{T+h})}; \hat{y}_{T+h} + c_{\alpha/2}\sqrt{V(e_{T+h})}, \tag{7}$$

  where $c_{\alpha/2}$ is the $(\alpha/2)$ % critical value for the standard normal density (or from the Student t density in the case of estimated parameters and small samples).

- Interpretation of the interval forecast
  - In $[1-\alpha]$% of the samples the realization of y_{T+h} will fall in the interval in (7).

Density Forecasting – A practical example



Source: European Central Bank.

Notes: Shading indicates Centre for Economic Policy Research euro area recessions. HICP is harmonised index of consumer prices.

The Baseline Linear Regression Model - Outline

Parameter Testing

- We need statistical tools to decide whether a specific variable should be included or excluded from the model.

- Four main ingredients of *hypothesis testing*:
  - The hypothesis of interest, often labeled *null hypothesis* (H_0).
  - The *alternative hypothesis* (H_1), which is what we believe holds true when H_0 is rejected.
  - The *test statistics*, which is a random variable that must be informative about the null hypothesis, and whose distribution under H_0 is known.
  - A *decision rule* that tells us whether to accept or reject the null hypothesis.

Hypothesis Testing

Let us now consider the problem of testing the null hypothesis $\beta_i = c$ versus the alternative hypothesis $\beta_i < c$, where c is a generic value and $i = 1, \ldots, k$.

- Use the *t-statistic*, defined as

$$t - test = \frac{\widehat{\beta}_i - c}{\sqrt{\widehat{var}(\widehat{\beta}_i)}},$$

where $\widehat{var}(\widehat{\beta}_i)$ is the i^{th} element on the diagonal of

$$\widehat{var}(\widehat{\beta}) = \widehat{\sigma}^2 \left(\frac{X'X}{T} \right)^{-1}$$

- It can be shown that, under H_0,

$$t - test \sim t(T - k),$$

where $t(T - k)$ is a t-distribution with $T - k$ degrees of freedom.

Hypothesis Testing – One-sided and Two-sided Tests

- *Significance level* of the test: α, which is the maximum probability of rejecting the null hypothesis when it is true (type I error).

- For a *one-sided* test with $H_1 : \beta_i < c$, we then find t_α, the α % critical value from the left tail of the distribution under H_0 of the test-statistic (in our case the $t(T-k)$ distribution), and we reject if the test is lower than t_α.

- For a *two-sided* test with $H_1 : \beta_i \neq c$, the t-test can be either positive or negative, depending on the value of c. we reject when $|t-test| > t_{\alpha/2}$, where $t_{\alpha/2}$ is the $(\alpha/2)$ % critical value from the right (or left) tail of the $t(T-k)$ distribution.

Hypothesis Testing – P-value

- *P-value*: the probability, under the null hypothesis, that the absolute value of the test statistic is larger than the realized value of the statistic.

$$p - value = prob_{H_0}(|t - test| > a),$$

or

$$p - value = prob_{H_0}(t - test > a) + prob_{H_0}(t - test < -a),$$

where a is the realization of the t-test.

- Reject the null hypothesis when p-value is smaller than α.

Hypothesis Testing – Composite Hypothesis

Let us now consider the composite null hypothesis $H_0 : \beta = c$, meaning $\beta_1 = c_1, \ldots, \beta_k = c_k$ jointly.

- For testing H_0 we can use the *F-statistics*, or F-stat, defined as

$$F - stat = \frac{(\widehat{\beta} - c)' X' X (\widehat{\beta} - c)}{k \widehat{\sigma}^2} = \frac{(\widehat{\beta} - c)' \hat{var}(\hat{\beta})^{-1} (\widehat{\beta} - c)}{k}$$

- It can be shown that when H_0 is true:

$$F - stat \sim F(k, T - k),$$

  where $F(k, T - k)$ indicates an F-distribution (the ratio of two independent χ^2 distributions) with k and $T - k$ degrees of freedom.

- Reject H_0 when the realized value of the F-stat is larger than the critical value F_α

The Baseline Linear Regression Model - Outline

Variable Selection

- In small samples *redundant regressors* can be more problematic to detect, and their inclusion in the model can increase estimation uncertainty.

- *Omitting relevant explanatory variables* from the model for y will cause the OLS estimators to be biased. The omitted variables will enter as part of the error term, lowering the model fit and possibly invalidating the Assumptions LR1 - LR3.
  - An exception is when either the coefficient of a variable is different from zero but close to it or the associated parameter estimator is very imprecise. Thus, the bias induced by omitting this variable can be dominated by the increased parameter estimation precision, yielding forecasts with an overall lower variance.

Variable Selection

Top-down modeling strategy: starts from a general model and tries to reduce it by deleting non-significant regressors.

- Drawbacks of top-down modeling strategy
  - Difficult to determine the overall probability of type I error.
  - If choosing a too high value of α we could keep redundant regressors, while with a too low value we could omit relevant regressors.
  - If we delete non-significant variables in groups, we might end up with a different model.
  - The final model could possibly not satisfy the Assumptions LR1 - LR5.
  - Manually the procedure may require a lot of time.

Variable Selection

Model specification based on *information criteria*: need to select a specific criterion, compute it for each possible model, and select the one that optimizes the criterion.

- Different criteria can yield different outcomes in terms of preferred models.
- It can be very time demanding, since we need to consider a total of 2^k models, where k is the number of explanatory variables in the general model.

Automated Variable Selection

- *Hard-thresholding* rules: a regressor is selected according to the significance of its correlation coefficient with the target.
  - tends to select highly collinear targeted predictors and this can complicate parameter estimation.

- *Soft-thresholding* rules, order and select the (N) indicators on the basis of a minimization problem of the following form:

$$\min_{\beta} \Phi(RSS) + \lambda \Psi(\beta_1, \ldots, \beta_j, ..., \beta_N),$$

- The parameter λ governs the shrinkage (the higher λ the higher is the penalty for having extra regressors in the model), while Φ and Ψ are functions of, respectively, the RSS and the parameters (β) associated with the N regressors.

Automated Variable Selection

- Depending on the functional form of Φ and Ψ, different thresholding rules are obtained.

- We now discuss some of the most common ones,
  - *Forward selection regressions* (FWD)
  - *Least angle regressions* (LARS)
  - *Least absolute shrinkage selection operator* (LASSO)
  - *Elastic net estimator* (NET)

Automated Variable Selection – FWD

- *Forward Selection* (FWD) consists of regressing y on x_1, storing the residuals ($\hat{\epsilon}_1$) and then looking for the covariate in the X information set with the highest correlation with this residual, say x_2.

- The residual $\hat{\epsilon}_1$ is projected onto x_2, a new residual $\hat{\epsilon}_2$ is stored, and the covariate mostly correlated with $\hat{\epsilon}_2$ is next identified.

- The procedure continues until all the variables in the information set have been ranked, or it can be stopped when a given criterion is satisfied, e.g., the adjusted R^2 in a regression of y on the selected regressors is above a given threshold.

Automated Variable Selection – LARS

- The Least Angle Regression algorithm, devised by Efron, Hastie, Johnstone, and Tibshirani (2004), starts as FWD, by identifying the covariate that has the highest correlation with the target.
- LARS then proceeds equiangularly between x_1 and x_2.
- After k steps, there are k variables in the active set. If the algorithm is stopped here, the coefficients of the remaining $N\ K$ regressors are all set to zero. The desired shrinkage can therefore be seen as a stopping rule for k.
- Efron, Hastie, Johnstone, and Tibshirani (2004) show that the LARS algorithm encompasses other popular shrinkage methods.

Automated Variable Selection – LASSO

- Least Absolute Shrinkage Selection Operator (LASSO) can be obtained in the LARS algorithm by imposing at each step pf the algorithm a restriction on the sign of the correlation between the new candidate regressor and the projection along the equiangular direction in the previous step.

- LASSO can also be related to the *RIDGE estimator*, which is a constrained OLS estimator that penalizes overfitting. Given M regressors, RIDGE coefficients are obtained by solving the following minimization problem

$$\min_{\beta} RSS + \lambda \sum_{j=1}^{M} \beta_j^2,$$

The Lagrange multiplier λ governs the shrinkage: the higher λ the higher is the penalty for having extra regressors in the model.

Automated Variable Selection – LASSO

- LASSO introduces a slight modification of the penalty function of the RIDGE regressor, which, rather than being a quadratic function shows a kink at zero:

$$\min_{\beta} RSS + \lambda \sum_{j=1}^{M} |\beta_j|.$$

- This modification implies that, unlike in the RIDGE setup, in the LASSO some regression coefficients are set exactly at zero.

Automated Variable Selection – NET

- The Elastic Net (NET) estimator is a refinement of LASSO, and it is the solution to the follow minimization problem:

$$\min_{\beta} RSS + \lambda_1 \sum_{j=1}^{M} |\beta_j| + \lambda_2 \sum_{j=1}^{M} \beta_j^2$$

- Shrinkage under NET depends on two tuning parameters, λ_1 and λ_2.

- Bai and Ng (2008) show that it suffices to apply a variable transformation to reformulate the NET as a LASSO problem, which can be therefore obtained through the LARS algorithm.

The Baseline Linear Regression Model - Outline

Multicollinearity

– An extreme example of *multicollinearity*: consider the case where $k = 2$ and $X_1 = X_2$.

- The matrix $X'X$ in this case has dimension 2×2 but its rank is equal to one. Since the model becomes

$$\begin{aligned} y_t &= \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_t \\ &= (\beta_1 + \beta_2) x_{1t} + \epsilon_t, \end{aligned}$$

  we can only identify $\beta_1 + \beta_2$ but not β_1 and β_2 separately.

  - In empirical applications, multicollinearity makes it difficult to separately identify all the β parameters, and it is reflected in large standard errors for the associated OLS estimators.
  - Hypothesis testing can also be affected by multicollinearity.

Multicollinearity

- It is in principle possible to check for perfect multicollinearity by testing whether the rank of $X'X$ is full.
- In practice it is more common to assess the extent of cross correlation among pairs of regressors.
- When high correlation among regressors is a temporary feature, extending the sample under analysis can alleviate the problem.
- When instead high correlation is not sample specific but related to the characteristics of the explanatory variables, wither re-parameterizing the model or summarizing the highly similar regressors can work.

Multicollinearity

Multicollinearity is problematic for structural analysis: identification of each single component of the β parameter vector is relevant to assess the contribution of each explanatory variable.

- For forecasting, it is less of a problem as long as the main goal is to produce the best, in the mean square forecast error sense, forecast for the y variable.

- The optimal forecast for y_{T+h} can be written either as

$$\widehat{y}_{T+h} = \widehat{\beta}_1 x_{1T+h} + \widehat{\beta}_2 x_{2T+h},$$

or as

$$\widehat{y}_{T+h} = \widehat{\left(\beta_1 + \beta_2\right)} x_{1T+h},$$

and in this sense multicollinearity is not an issue.

The Baseline Linear Regression Model - Outline

Data Simulation Procedure

- The data generating process (DGP) is the following:

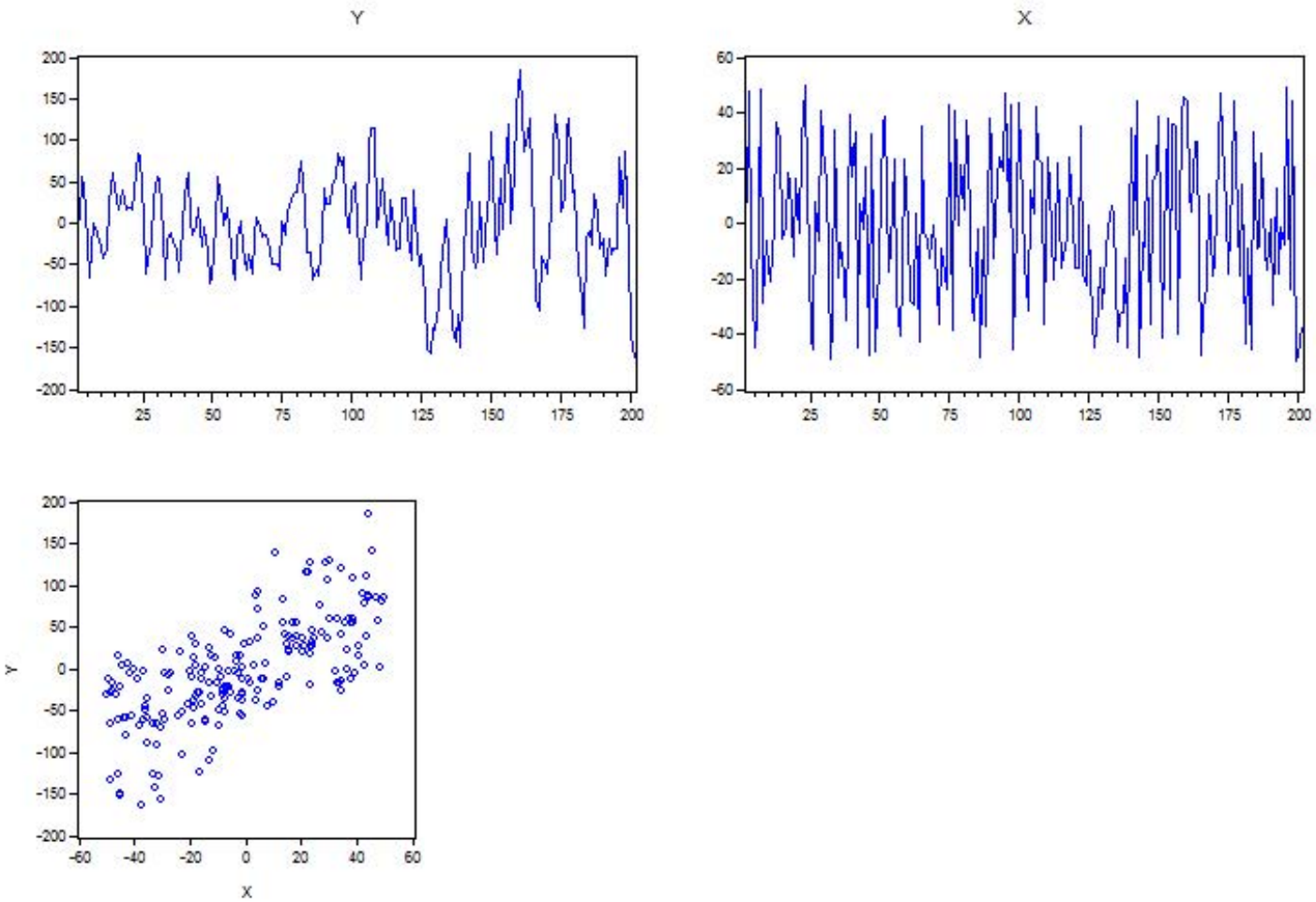
$$y_t = 1 + D_t + D_t x_t + x_t + 0.5x_{t-1} + 0.5y_{t-1} + \varepsilon_t$$

  - $y_t = \ln(Y_t)$ is the dependent variable
  - x_t is the independent variable
  - lags of $\ln(Y_t)$ and x appear also as explanatory variables
  - D_t is a dummy variable that determines a change in the intercept and in the coefficient of x_t
  - ε_t is i.i.d. distributed as $N(0, 1)$.

  - Estimation sample: observations 102-301
  - Forecasting sample: observations 302-501
  - The dummy variable is equal to 1 for the observations from 202-501 and 0 otherwise.

Linear Regression Model

Simulated dependent variable y, regressor x, and scatter plot y and X.



Linear Regression Model – Baseline Model

- The baseline linear regression model we consider is the following:

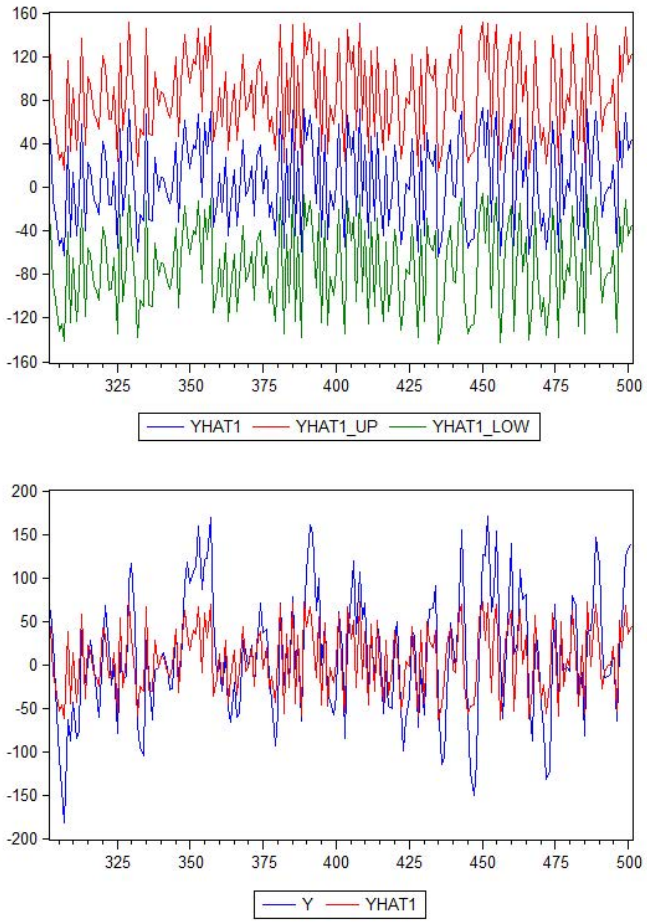
$$y = \beta_0 + \beta_1 x_1 + \epsilon_t,$$

Baseline linear regression model for the sample: 102-301

	Coefficient	Std. Error	t-Statistic	Prob.
BETA(0)	5.380	0.667	8.072	0.000
BETA(1)	1.491	0.117	12.801	0.0000
R-squared	0.453	Mean dep var		6.695
Adjusted R-squared	0.450	S.D. dep var		12.560
S.E. of regression	9.314	Akaike IC		7.311
Sum squared resid	17176.480	Schwarz IC		7.344
Loglikelihood	-729.086	Hannan-Quinn criter.		7.324
F-statistic	163.861	DW stat		0.955
Prob(F-statistic)	0.000			

Linear Regression Model – Forecast

Plot of actual y series, linear model forecast yhat1, and yhat1_up and yhat1_low series of forecasts' 95% approximate forecast interval upper and lower bounds respectively

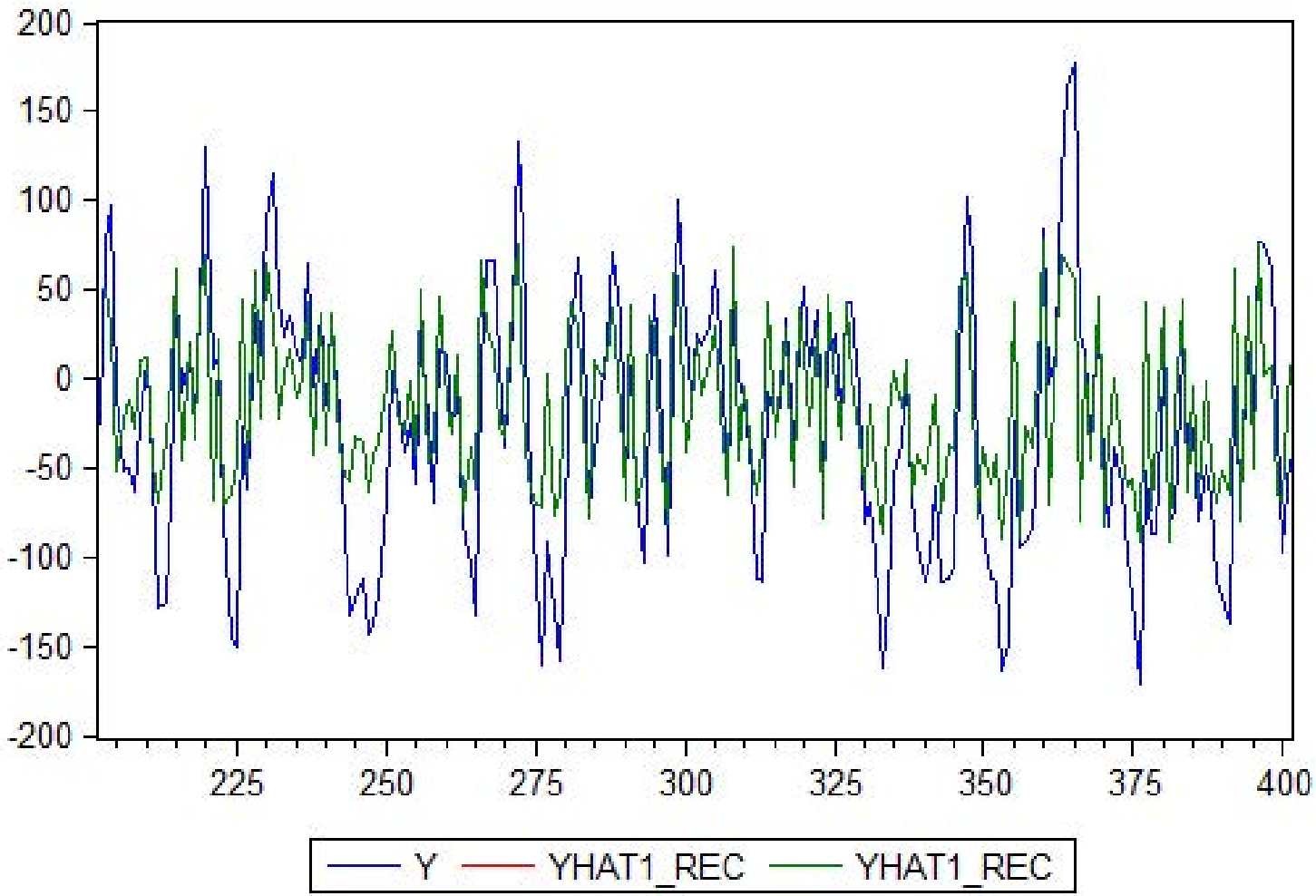


Linear Regression Model – Recursive Forecast

- We can compute a recursive forecast where the model is estimated over an expanding estimation sample, and the forecasts are computed with the recursively estimated parameters.
  - Estimate the model for a given estimation sample.
  - Compute (one-step-ahead) forecasts using the estimated coefficients from step 1.
  - Add one observation to the estimation sample and go to step 1 again.

Linear Regression Model – Recursive Forecast

Plot of actual y series and yhat1_rec which presents the forecasts obtained from a recursive forecast procedure



Linear Regression Model – Recursive Forecast

Baseline linear regression model forecasting performance

	Forecasting Method	
	Static	*Recursive*
RMSE	9.7426	9.5705
MAFE	8.0798	7.9432

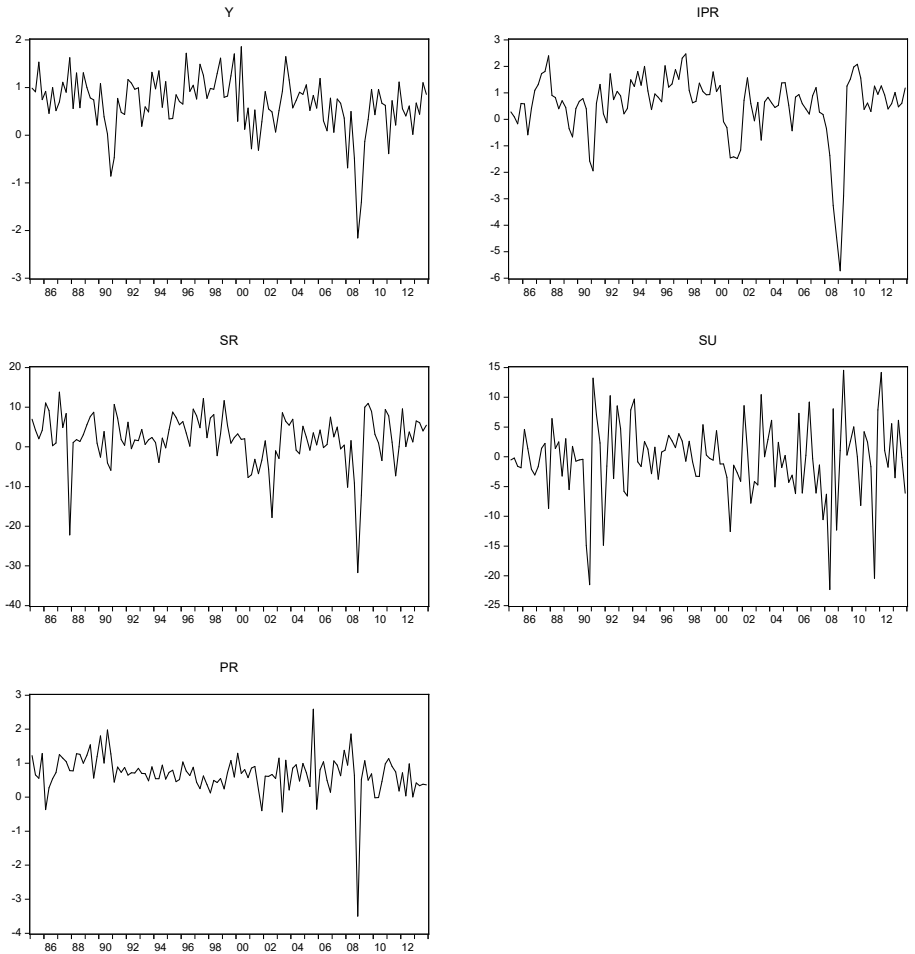
The Baseline Linear Regression Model - Outline

Forecasting US GDP Growth

- Hard data: industrial production (IPR), the consumer prices index (PR), and the S&P500 index (SR).
- Soft data: the Michigan University consumer sentiment index for the United States
- Sample period: 1985Q1 - 2013Q4
  - The estimation sample should be as long as possible, but also homogeneous when working under the assumption of stable model parameters.

Forecasting US GDP Growth

Quarterly US GDP growth, IP growth, S&P500 returns, CSI growth, CPI inflation



Forecasting US GDP Growth

- To separate the role of the intercept, we denote it via the parameter α. We start with three cases in which the regressors X_t vary

$$y_t = \alpha + X_t\beta + \epsilon_t$$

  - Case 1 (Model 1): $X_t = (ipr_t, su_t, pr_t, sr_t)$
  - Case 2 (Model 2): $X_t = (ipr_t, su_t, sr_t)$
  - Case 3 (Model 3): $X_t = (ipr_t, su_t)$

Forecasting US GDP Growth - Summary Statistics

	Model 1	*Model* 2	*Model* 3
R^2	0.533456	0.522913	0.508604
$\overline{R^2}$ (adjusted R^2)	0.516644	0.510134	0.499907
AIC	1.131974	1.137080	1.149389
SC	1.250663	1.232031	1.2206003
HQ	1.180155	1.175625	1.178298

According to the R^2 and adjusted R^2, the best model is the first one. Other criteria, such as the AIC, SC and HQ are not consistent. We know that the lower is the value of the criteria the ”better” is the model. The minimum for AIC corresponds to Model 1. At the same time SC is minimized for Model 3 and HQ is lowest for Model 2. The differences are however small, so that we do not expect major differences in the forecasting performance of the models.

Forecasting US GDP Growth - Forecast

- Estimation output Model 1 for sample: 1985Q1 - 2006Q4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.579	0.088	6.586	0.000
IPR	0.310	0.049	6.328	0.000
SU	0.013	0.009	1.403	0.164
PR	-0.015	0.095	-0.153	0.879
SR	0.003	0.009	0.313	0.755
R-squared	0.387	Mean dep var		0.778
Adjusted R-squared	0.357	S.D. dep var		0.493
S.E. of regression	0.395	Akaike IC		1.036
SSR	12.956	Schwarz IC		1.176
Loglikelihood	-40.572	Hannan-Quinn		1.092
F-stat	13.096	DW stat		2.028
Prob(F-stat)	0.000			

Forecasting US GDP Growth - Forecast

- Estimation output Model 2 for sample: 1985Q1 - 2006Q4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.569	0.058	9.827	0.00
IPR	0.31	0.049	6.376	0.00
SU	0.013	0.009	1.509	0.135
SR	0.003	0.009	0.306	0.761
R-squared	0.387	Mean dep var		0.778
Adjusted R-squared	0.365	S.D. dep var		0.493
S.E. of regression	0.393	Akaike IC		1.013
SSR	12.96	Schwarz IC		1.126
Loglikelihood	-40.584	Hannan-Quinn		1.059
F-stat	17.659	DW stat		2.026
Prob(F-stat)	0.00			

Forecasting US GDP Growth - Forecast

- Estimation output Model 3 for sample: 1985Q1 - 2006Q4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.579	0.053	10.931	0
IPR	0.309	0.048	6.404	0
SU	0.015	0.008	1.912	0.059
R-squared	0.386	Mean dep var		0.778
Adjusted R-squared	0.372	S.D. dep var		0.493
S.E. of regression	0.391	Akaike IC		0.992
SSR	12.974	Schwarz IC		1.076
Loglikelihood	-40.633	Hannan-Quinn		1.026
F-stat	26.726	DW stat		2.01
Prob(F-stat)	0			

Forecasting US GDP Growth - Forecast

- According to R^2 Model 1 is "better", but $\overline{R^2}$ prefers Model 3. This conclusion is consistent with other criteria: AIC, SC, and HQ are lowest for Model 3.

- Forecast evaluation statistics of static and recursive forecasts

	Model 1	Model 2	Model 3
	Static forecasts		
RMSFE	0.547	0.541	0.549
MAFE	0.417	0.414	0.418
	Recursive forecasts		
RMSFE	0.524	0.529	0.546
MAFE	0.385	0.402	0.412

Forecasting US GDP Growth - Forecast

- Model 3 has the highest value of RMSFE and MAFE while Model 1 is best according to both criteria.
- Statistics associated with the recursive forecasts are lower, suggesting the possibility of parameter instability associated with the financial crisis.

The Baseline Linear Regression Model - Outline

Dynamic regression

- While we are entering into dynamic regression territory, we will provide a simple example of how one can still use the classical regression model with the following specification:

$$y_t = \alpha + X_{t-1}\beta + \epsilon_t$$

- Shifting the regressors one period back implies that the regressors are more likely available when performing one-step ahead forecasts.

Revisiting GDP forecasting

- Refit the GDP growth example with dynamic regression.
- Estimation sample: 1985Q1 - 2006Q4; forecast evaluation period: 2007Q1 - 2013Q4.
- Estimation sample measures of fit

	Model 1	Model 2	Model 3
R^2	0.310	0.304	0.248
$\overline{R^2}$	0.286	0.285	0.235

- Model 1 stays the best overall in the new dynamic setting.

Revisiting GDP forecasting

- Forecast evaluation statistics of static and recursive forecasts

	Model 1	Model 2	Model 3
	Static forecasts		
RMSFE	0.726	0.711	0.785
MAFE	0.515	0.492	0.528
	Recursive forecasts		
RMSFE	0.713	0.702	0.724
MAFE	0.502	0.476	0.485

- The more parsimonious Model 2, while inferior to Model 1 in-sample, is the superior one out-of-sample according to both RMSFE and MAFE criteria, whether using static or recursive forecasting schemes.

Revisiting GDP forecasting

Recursive GDP growth forecasts and 95% confidence intervals Model 2