Consistent Return and Risk Forecasting for Portfolio Optimization using Kernel Regressions

Jan Hildebrandt (Universität Bremen)
Thorsten Poddig (Universität Bremen)
Talk Outline

1. Motivation
2. Consistent Return and Risk Forecasts
3. Kernel Regression
4. Consistent Return and Risk Forecasts using Kernel Regression
5. Empirical Studies
6. Summary and Outlook
1. Motivation

- Modern asset allocation is based on the theory of portfolio selection.
- The theory was founded by Markowitz and Tobin in the 1950s.
- The theory and its application evolved over time.
- Today, many highly sophisticated asset allocation procedures are at hand.
- Nevertheless, any kind of (active) asset allocation is based on assumptions about future returns and risks from possible investments.
- Thus, return and risk forecasts are still the basis for any asset allocation procedure.
Portfolio Optimization

Objective function (usual Markowitz optimization):

\[
OF(w_1, w_2, \ldots, w_N) = \sum_{i=1}^{N} w_i \mu_i - \lambda \cdot \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{ij} \rightarrow \text{max!}
\]

Free parameters:
- asset weights

Forecast:
- Expected asset returns
- Future variances and covariances of asset returns

\( \lambda \): risk aversion parameter
Remarks

- Additionally, several constraints must be regarded, like budget and asset weights constraints.

- Nevertheless, the solution of the optimization problem is easy today (elaborated optimization procedures, powerful computers).

- **But**, it is still unclear and open to debate whether financial forecasts can be provided.

- According to the efficient market hypothesis, this is impossible and no investor can beat the market.
Problems

- If we believe in financial forecasts, there are still several practical problems in asset allocation.
- One of these problems is information aggregation.
- Financial forecasts of return and risk for many different assets are usually provided by different analysts, research institutes, models or other experts.
- How can we aggregate them?
Information Aggregation

- The problem of information aggregation even arises in very simple settings.
- For example, assume we have a return forecasting model A and a risk model B.
- Are these return and risk forecasts consistent?
- If not, does it matter?
2. Consistent Return and Risk Forecasts

- If return and risk forecasts are not consistent, does it matter?
- Several studies (e.g. Chopra/Ziemba (1993), Kallberg/Ziemba (1984)) suggest that errors in variance and covariance forecasts are of low relevance.
- Based on these insights, most effort is spend on return forecasts.
- If risk forecasts are of low relevance, inconsistent return and risk forecasts should not matter. True?
Consistent forecasts do matter

- Study by Petersmeier (2003).
- Sector allocation model for the German stock market:
  - DAX 100 stocks, represented by 9 sector indices.
  - Markowitz portfolio optimization based on the out-of-sample forecasts.
  - Equally weighted portfolio (9 sectors) as benchmark.
Consistent forecasts do matter

- Different types of forecasts were compared:
  - return forecast from kernel regression, empirical variance and co-variances
  - return and variance forecasts from kernel regression, empirical co-variance
  - return, variance and co-variance forecasts from kernel regression

- Different forecasting models (e.g., mean model, linear regression models) as additional benchmarks.
Consistent forecasts do matter

- Kernel regression showed better results than any other benchmark applied.
- Return and variance forecasts from kernel regression improved portfolio performance.
- Co-variance forecasts from kernel regression did not improve portfolio performance.
- Results were confirmed by a similar simulation study based on artificial data.
How to provide consistent forecasts?

- One possible approach: Using methods which already provide return and risk forecasts.
- Well-known class of such models: ARCH/GARCH models, especially multivariate extensions (providing variances and co-variances).
- Another interesting model class: Kernel regressions.
3. Kernel regression

Objective: Estimate an unknown functional relationship

\[ y_t = f(x_t) + \varepsilon_t \]

- \( y_t \): dependent variable (can be a vector)
- \( x_t \): vector of independent variables
- \( \varepsilon_t \): residual, i.i.d. random variable
- \( f(.) \): unknown functional relationship, to be estimated from data
Kernel regression

Estimate \( y \) dependent on observation \( x \):

\[
\hat{y} = \sum_{t=1}^{T} \left[ \omega(x, x_t) \cdot y_t \right]
\]

\( \hat{y} \): estimate of \( y \) given \( x \)

\( \omega(x, x_t) \): weight of observation \( x_t \) given \( x \)

\( y_t \): observation \( y_t \)

\( T \): number of observations
Weighting scheme

- Estimate of $y$ given $x$ is constructed from previous observations $(y_t, x_t)$.
- Estimate of $y$ is a weighted average of previous observations of $y_t$.
- The weight $\omega(x, x_t)$ depends on the actual observation $x$ and previous observations $x_t$.
- Obviously, the estimate depends on the weighting scheme.
- The weighting scheme depends on kernel functions.
Weighting scheme

Usual weighting scheme (different variants possible):

\[ \omega(x, x_t) = \frac{K\left(\frac{d(x, x_t)}{h}\right)}{\sum_{t=1}^{T} K\left(\frac{d(x, x_t)}{h}\right)} \]

- **K(.)**: Kernel function (see next slides)
- **d(x, x_t)**: distance of observation \( x_t \) and \( x \) (e.g., city-block, Euclidean distance)
- **h**: smoothing parameter
x-Axis: distance $(x, x_t)$, usually positive values
Remarks

- Two specifications are necessary: kernel function and distance measure.
- Gaussian kernel function is often used: $\exp(-z^2)$.
- Epanechnikov kernel is easy and fast to compute.
- Euclidean distance is most often a natural measure.
- City-block distance might be more robust in the presence of outliers.
Example of kernel regression
Smoothing parameter $h$

- The estimate depends on a free, unknown parameter $h$, often called bandwidth or smoothing parameter.
- Usually, $h$ is estimated by using a one-hold-out-procedure.
- Each observation $(y_t, x_t)$ is hold out once and estimated by using all remaining observations.
- Choose $h$ such that the sum of squared residuals (errors) $SSR$ is minimized!
Smoothing parameter

$$SSR(h) = \sum_{t=1}^{T} (y_t - \hat{y}_t^-)^2 \rightarrow \min!$$

where

$$\hat{y}_t^- = \sum_{l=1}^{T} \left[ \omega(x, x_l) \cdot y_l \right]$$

$$\omega(x, x_l) = \frac{K \left( \frac{d(x, x_l)}{h} \right)}{\sum_{t=1}^{T} \sum_{l \neq t} K \left( \frac{d(x, x_l)}{h} \right)}$$
Optimal vs. suboptimal $h$
Kernel function and smoothing parameter

Schätzfehler als Funktion der Bandbreite bei verschiedenen Kernfunktionen

- Epanechnikov
- Gauss
Remarks

- The choice of the kernel function is not arbitrary.
- A kernel function might be easy to compute, but may impose difficult problems estimating $h$.
- We prefer the Gaussian kernel because it is fairly easy and fast to compute and do not cause severe problems estimating $h$.
- Accordingly, we often use the Euclidean distance and standardized $x$-values.
Variable selection

- While causing some problems in detail, the specification of the kernel function, the choice of the distance measure and the estimation of $h$ is not “really” critical.

- Most critical is the identification of the relevant independent variables ($x$).

- Unfortunately, powerful selection algorithms do not exist up to now.

- Existing tests of significance still show up some problems.
Tests of significance

- Fan and Li (1996): very slow convergence, unsuitable for medium and small sample sizes.
- Lavergne and Vuong (2000): some practical advantages (smaller bias, more robust against estimation errors in $h$).
- Ait-Sahalia, Bickel and Stoker (2001): more general, quite difficult to compute.
Alternatives

- **Bootstrapping:**
  - Easy to implement, simple algorithms.
  - Unfeasible due to computational effort.

- **Multiple Cross-Validation (delete-d jackknife):**
  - Easy to implement, simple algorithms.
  - Computational effort acceptable, but much slower than the former methods.
Remarks

- We found variable selection to be the most critical step using kernel regression.
- We use two alternative methods:
  - Test according to Lavergne and Vuong (2000).
  - Multiple cross-validation (delete-d jackknife).
- Both showed good performance on artificial data.
4. Consistent Return and Risk Forecasts using Kernel Regression

- Section 3 shows how to model the conditional mean of $y$ given $x$ using kernel regressions.
- In portfolio optimization, we are also interested in the conditional variances and co-variances.
- Can we predict variances (and co-variances) within the same framework?
- Yes, but we should distinguish two different approaches:
  - Direct (or explicit) risk modeling.
  - Indirect (or implicit) risk modeling.
Explicit risk modeling

Recall: Estimate $y$ is the conditional mean given $x$, so we can use this as the return model (return forecast):

$$\hat{y} = E(y \mid x) = \sum_{t=1}^{T} [\omega(x, x_t) \cdot y_t]$$

Define the residual $r$ as: $r = y - \hat{y}$

$$\hat{\sigma}^2 = E(r^2 \mid x) = \sum_{t=1}^{T} \omega(x, x_t) \cdot r_t^2$$
Remarks

- There are some critical assumptions:
  - The uncertainty is fully reflected by the residuals $r$ of the return model.
  - The unobservable variance can be properly estimated by the squared residuals.
  - The conditional variance of $r$ can be estimated using a set of independent variables $x$.
- The set of independent variables for the conditional mean and variance model can be different.
Advantages and disadvantages of explicit risk modeling

■ Advantages:
  □ The conditional variance can be explained by different independent variables.
  □ We obtain an explicit risk forecasting model.

■ Disadvantages:
  □ The whole variable selection procedure must be conducted again.
  □ The conditional variance model assumes that the underlying conditional mean model is correctly specified.
Implicit risk modeling

- The weights $\omega(x,x_t)$ in the kernel regression mean model can be interpreted as conditional probabilities.
- We can regard each observation $t$ as a possible (future) state $t$ which might occur.
- The weight $\omega(x,x_t)$ is interpreted as an estimate of the probability that state $t$ will occur, given the actual observation $x$.
- Interpreting weights $\omega(x,x_t)$ as conditional probabilities of state $t$, we can calculate the expected value, variance (and co-variances) of $y$. 
Implicit risk modeling

Data base, observations

<table>
<thead>
<tr>
<th>Observation</th>
<th>Vector of independent variables</th>
<th>Dependent variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$x_1$</td>
<td>$y_1$</td>
</tr>
<tr>
<td>2</td>
<td>$x_2$</td>
<td>$y_2$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>T</td>
<td>$x_T$</td>
<td>$y_T$</td>
</tr>
</tbody>
</table>

Actual observation $x$

Kernel regression

„Conditional“ probability

$p_1(x)$

$p_2(x)$

... 

$p_T(x)$
Implicit risk modeling

Conditional mean model \( (p_t(x) = \omega(x,x_t)) \):

\[
\hat{y} = E(y \mid x) = \sum_{t=1}^{T} [p_t(x) \cdot y_t] 
\]

Conditional variance model:

\[
Var(y \mid x) = \sum_{t=1}^{T} \left[ p_t(x) \cdot (y_t - E(y \mid x))^2 \right] 
\]
Implicit risk modeling

Conditional mean and variance (see Petersmeier (2003), p. 312):

\[
\hat{E}(y \mid x) = \sum_{t=1}^{T} [\omega(x, x_t) \cdot y_t]
\]

\[
\hat{Var}(y \mid x) = \sum_{t=1}^{T} [\omega(x, x_t) \cdot y_t^2] - \hat{E}(y \mid x)^2
\]
Advantages and disadvantages of implicit risk modeling

- **Advantages:**
  - Conditional mean, variances (and co-variances) can be estimated with the same model.
  - The variable selection procedure has to be done only once.
  - There is no “joint-hypothesis” problem.

- **Disadvantages:**
  - Conditional variances (and co-variance) can not be explained by different independent variables, even if suitable.
Explicit or implicit risk modeling?

- Simulation studies on artificial data with variances conditional on different independent variables than in the mean model showed:
  - In the absence of noise, explicit modeling leads to more accurate variance forecasts.
  - But, even very low noise levels distort explicit risk modeling.
  - Implicit risk modeling does not lead to inferior forecast in the presence of even low noise levels.
  - Implicit risk modeling is easier and much faster.
5. Empirical studies

- Kernel regressions have been applied to financial forecasting for many years.
- They became very popular under the name „General Regression Neural Network (GRNN)“ in the neural network community.
- We give a short overview over several studies.
- Since the objective is to deal with consistent return and risk forecasts, we will focus on such studies.
Literature overview:

- **Return forecasts:**
  - Dichtl (2001): different asset classes (stocks, bonds, exchange rates).
  - Richter/Poddig/Hildebrandt (2007): different asset classes (stocks, bonds, exchange rates).
Literature overview

- Risk forecasts:

- Return and risk forecasts:
Hildebrandt/Poddig (2008)

- Stock (Europe, Germany, Emerging markets) and bond market (Europe, USA) indices as asset classes.
- Independent variables: financial and macro-economic indicators, sentiment data.
- Forecast of compounded returns, variances and co-variances, Markowitz portfolio optimization.
Return/Risk-Profile

Consistent return and risk forecasting using kernel regressions

- Lin.Regr./GARCH
- Kernregr./emp.Kov.
- AR(1)
- emp.MW/Kov.
- Lin.Regr./GARCH
First results

- Study confirmed results obtained by Petersmeier (2003).
- Seemingly, portfolio performance was improved by consistent return and risk forecasts.
- Kernel regression outperformed several benchmark models.
- Will results hold in the long run?
Hildebrandt (2009)

- Extended study, monthly data, 1975-2008, 397 observations (almost 33 years).
- Different lengths of in-sample window were compared (18, 12, 6 years).
- Forecast of the next monthly return.
- Markowitz portfolio optimization.
Hildebrandt (2009)

- Asset classes: DJIA (USA), MSCI World, MSCI Europe, DAX, GSCI (Commodities), USD/EUR exchange rate.
- Several benchmark models:
  - Naive and mean model (return forecasts).
  - Empirical co-variance matrix (variance and co-variance forecasts).
  - ARIMA and linear regression models (return forecasts).
  - GARCH (1,1) model (variance forecasts).
  - Ensemble of 3-layer perceptrons (return and variance forecasts).
- Use of variable selection procedures (except 3-layer perceptrons due to computational effort).
Results

■ Kernel regression can not outperform simple forecasting models (mean model and empirical co-variance matrix).
■ Ensemble of 3-layer perceptrons show best portfolio performance.
■ However, results are very instable, no general conclusions can be drawn.
■ Variable selection procedure shows very instable selection of independent variables which appears to be rather randomly.
Shortcomings

■ Asset class universe is not well balanced (4 stock market indices, highly correlated).

■ Two target time series are reconstructed from different data sources (EUR/USD, DAX).

■ Other important asset classes, e.g. bonds, are not part of the asset class universe due to availability.

■ Many important time series can not be tracked back to the 1970s or early 1980s.
6. Summary and Outlook

- Results from Petersmeier (2003), Hildebrandt/Poddig (2008) and Hildebrandt (2009) are part of an ongoing research.
- There is some evidence that consistent return and risk forecasts might improve portfolio performance.
- Perhaps, the predictability of financial markets might be time dependent.
- Kernel regression shows encouraging results, but it should be used as an complementary – not single – forecasting tool.
- More complex and complicated models are not automatically better models.
- Model ensembles might be superior to single forecasting models which is already well known from literature.
Outlook

- The study by Hildebrandt (2009) should be redesigned addressing the shortcomings mentioned above.
- Kernel regression might be improved:
  - More complex: Local-linear kernel regression.
  - Less complex: Probabilistic Neural Network.
- Instable variable selection might be “worked around” using ensembles of kernel regression models.
- Model ensembles provide a third way to generate consistent variance and co-variance forecasts.