DATA DRIVEN RISK-RETURN COMPUTATION FOR REAL ESTATE

CapitalBAY

July 30th 2020

Technische Universität München

AGENDA

Introduction

Data

SDE Model

Risk Model

Business Plan

Results

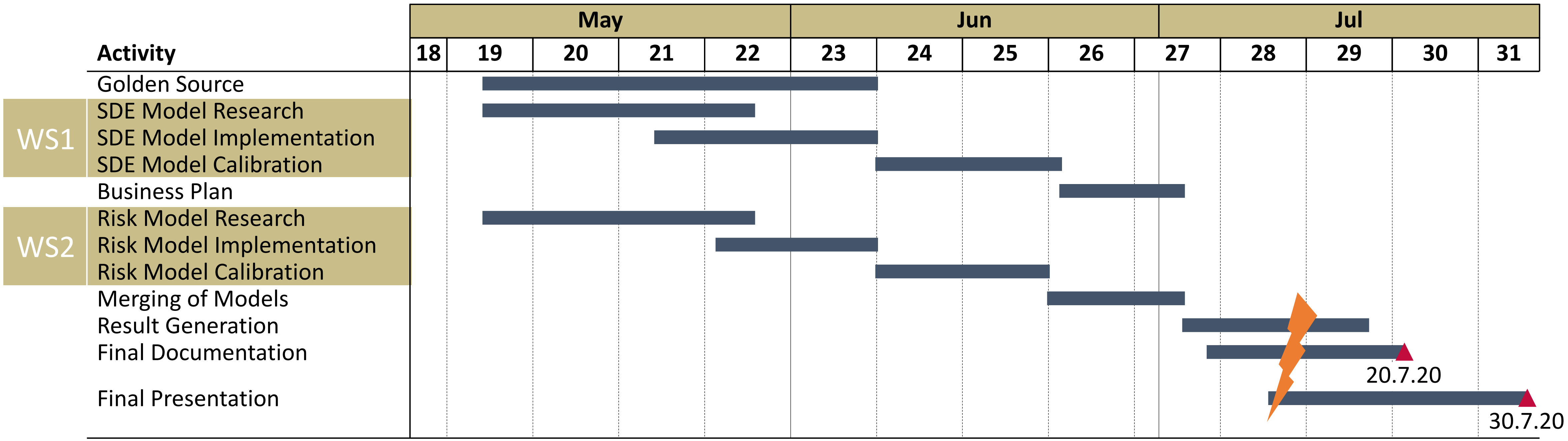
Outlook

GOALS AND PROJECT PLAN

Real estate objects such as apartments or flats can be lucrative investments. Nevertheless, as with any other investment possibility **risk** and **return** must be considered. Our goal is both,

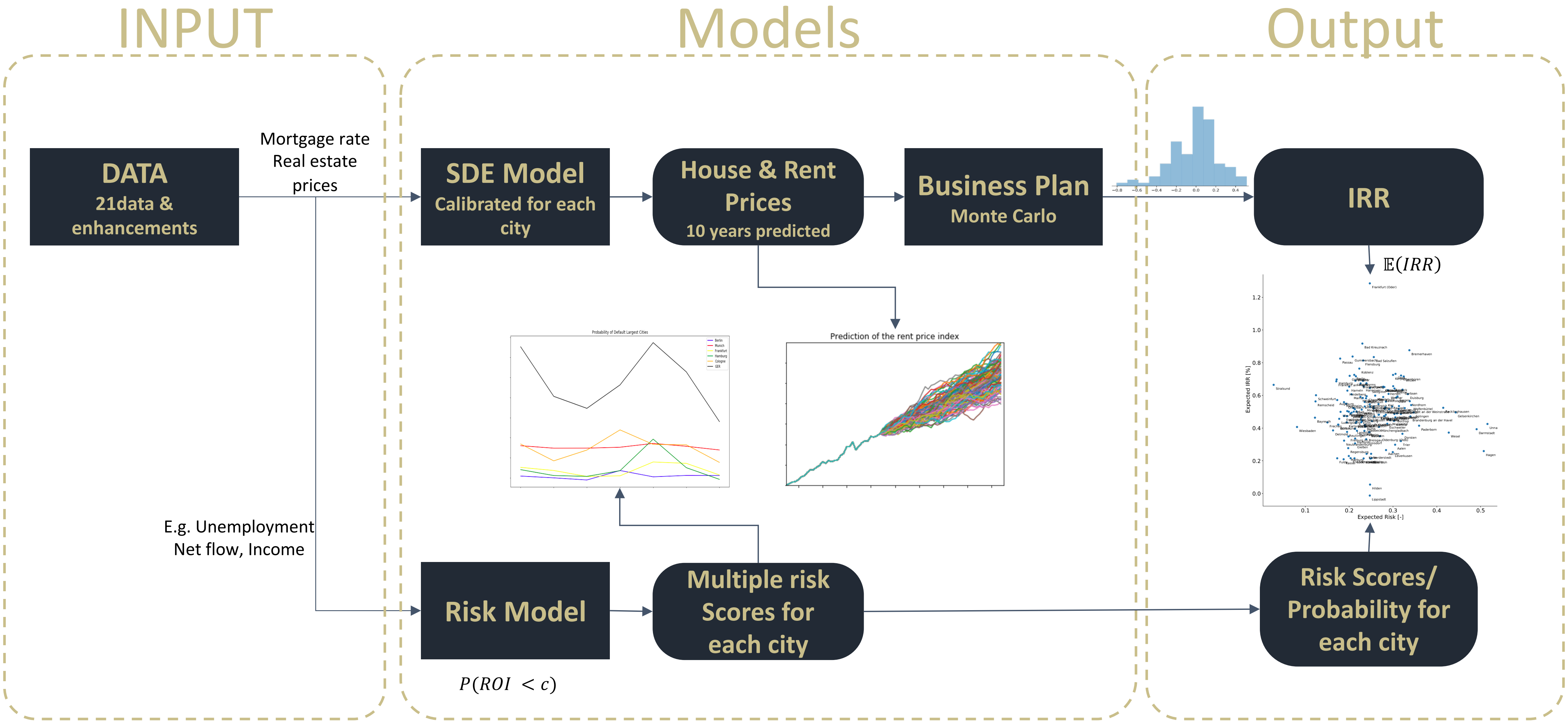
1. Compute the **expected return on investment (Internal Rate of Return - IRR)** for 149 German cities
2. Compute a compound **risk score** for these cities as well

These 2 KPIs shall help CapitalBay to choose profitable real estate investment decisions.

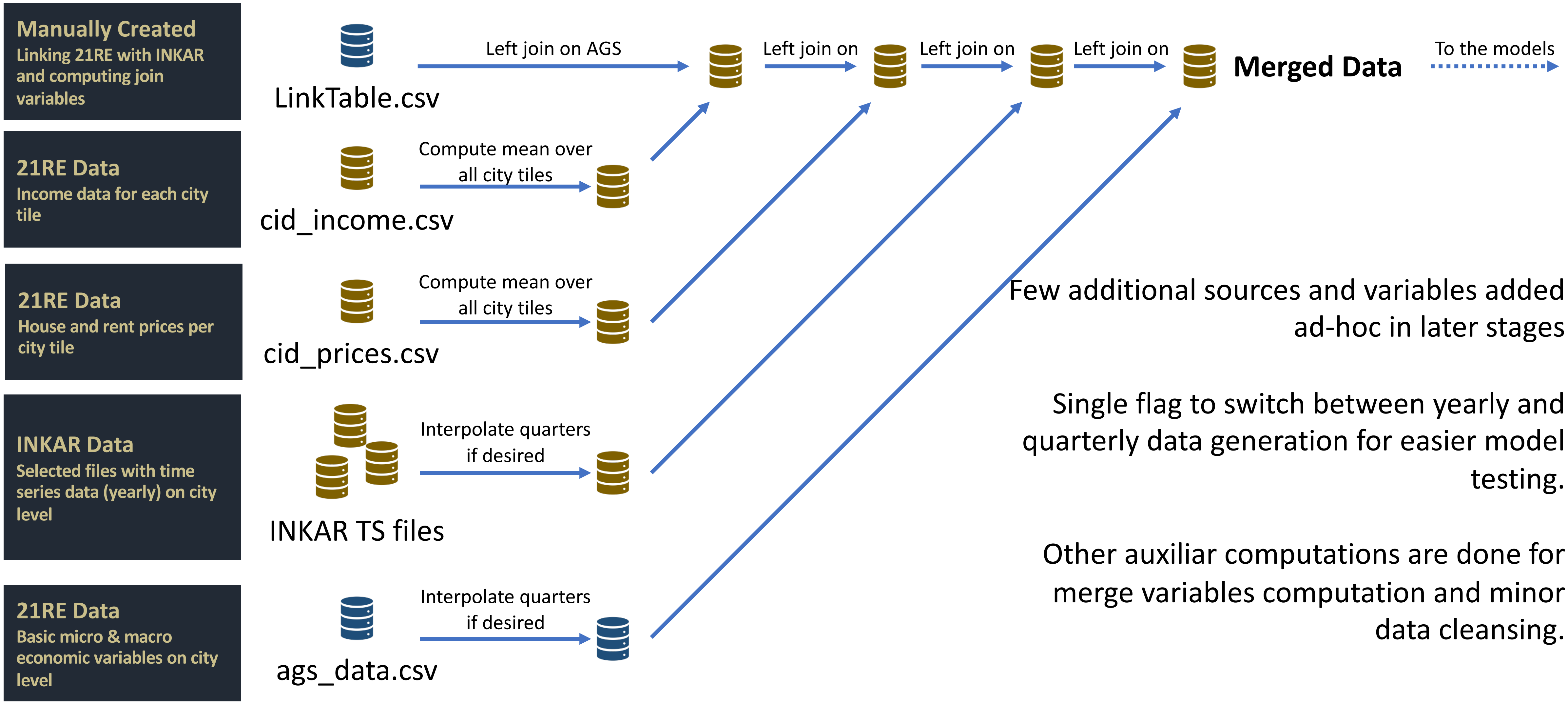


SOFTWAREARCHITECTURE

2 ½ STAND-ALONE MODELS LEAD TO RISK/RETURN



21REAL ESTATE AND INKAR AS BASELINE



Few additional sources and variables added ad-hoc in later stages

Single flag to switch between yearly and quarterly data generation for easier model testing.

Other auxiliar computations are done for merge variables computation and minor data cleansing.

SDE MODEL - WORKFLOW

Finding a theoretical model: stochastic differential equation system that describes housing and rent price dynamics

↓

Adjusting the theoretical model by using real data

↓

Discretization of the model regarding the implementation

↓

Implementation via SDE pipeline:
Parameter Estimation & Numeric Solution

↓

Prediction of Housing and Rent Price Index for the next 10 years

SDE MODEL –THEORETICAL MODEL

We consider a new SDE model derived from the Bates-Hull-White model applied to RE:

Original

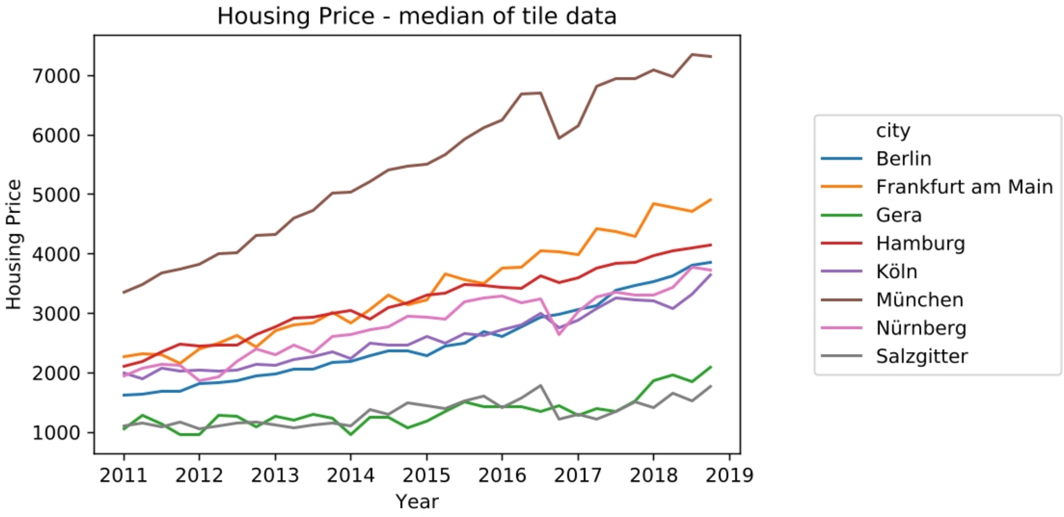
$$\begin{cases} \dfrac{\mathrm{d}h_t}{h_t} = \lambda(\mu_h - r_t)\,\mathrm{d}t + \sqrt{v_t}\,\mathrm{d}Z_t^h + dH_t^h \\ \mathrm{d}v_t = k_v(\theta_v - v_t)\,\mathrm{d}t + \sigma_v\sqrt{v_t}\,\mathrm{d}Z_t^v \\ \mathrm{d}r_t = k_r(\mu_r - r_t)\,\mathrm{d}t + \sigma_r\,\mathrm{d}Z_t^r \end{cases}$$

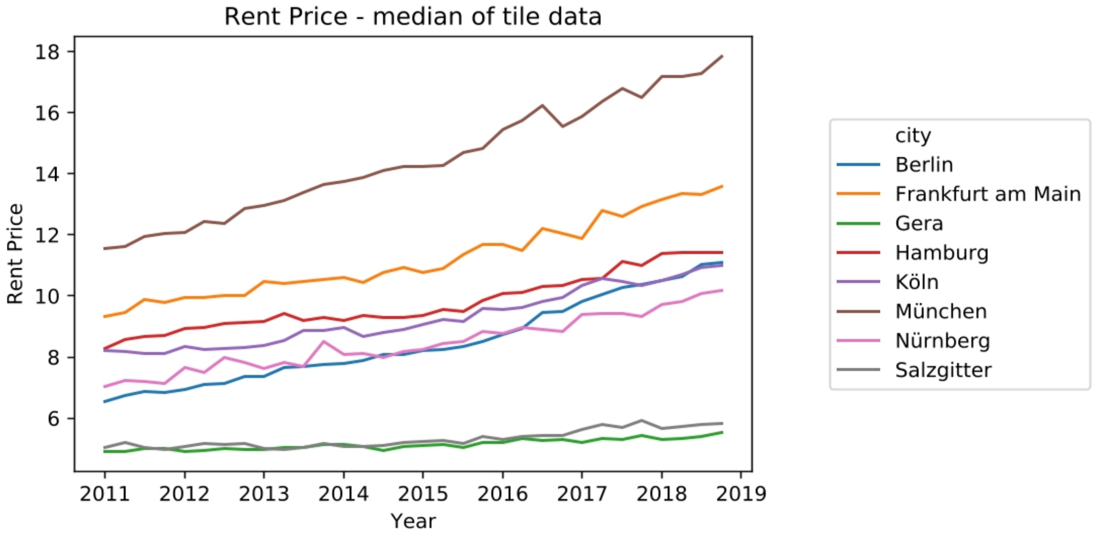
Adding Rent

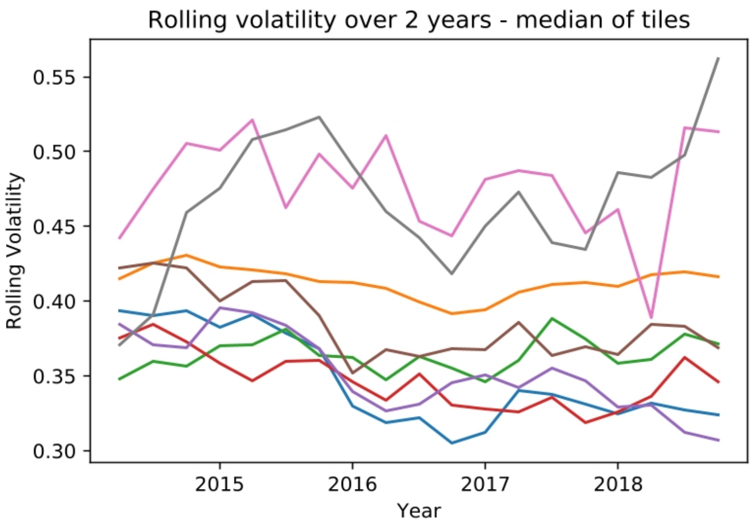
$$\begin{cases} \dfrac{\mathrm{d}m_t}{m_t} = \left(\mu_m + k_{m_1}\dfrac{\mathrm{d}h_{t-l}}{h_{t-l}} + k_{m_2}\dfrac{\mathrm{d}f_{t-l_2}}{f_{t-l_2}}\right)\mathrm{d}t + \sigma_m\,\mathrm{d}Z_t^m + dH_t^m \\ \mathrm{d}f_t = k_m(\mu_m - f_t)\,\mathrm{d}t + \sigma_m\,\mathrm{d}Z_t^m \end{cases}$$

- h_t is the House Price Index
- v_t is the volatility of the index and is a **Cox–Ingersoll–Ross (CIR)** process
- r_t is the Mortgage rate defined by a generalized Ornstein-Uhlenbeck (OU)
- m_t is the Market Rent Index
- f_t is an additional factor (ie. vacancy) defined by a generalized Ornstein-Uhlenbeck (OU)
- $Z_t^S, Z_t^v, Z_t^r, Z_t^m$ are correlated Brownian Motions
- H_t^h, H_t^m are compound Poisson processes

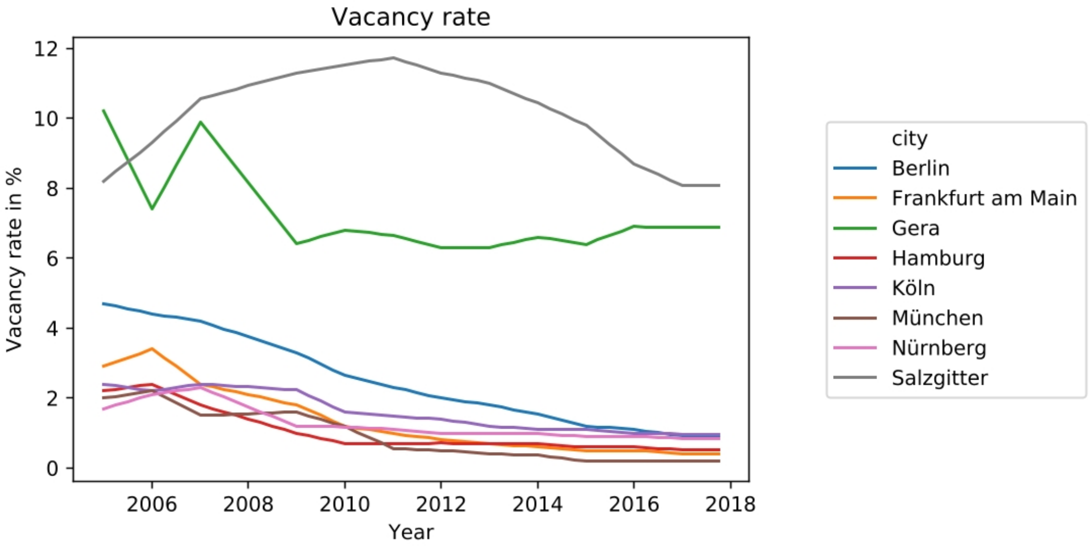
SDE MODEL – INCORPORATING THE DATA







Note: We consider the volatility of the house price index constant since we noticed from the data observations that there is no relevant variation of the volatility.



SDE MODEL – DISCRETIZED MODEL

We compute the numerical solution of the house price index using the implicit Euler method :

$$h_{t+1} = h_t + (1-\theta)\lambda(\mu_h - r_t)h_t\Delta t + \theta\lambda(\mu_h - r_t)h_{t+1}\Delta t + \sigma_h h_t \Delta Z_t^h + v h_t \Delta H_t^h$$
$$= \{h_t + (1-\theta)\lambda(\mu_h - r_t)h_t\Delta t + \sigma_h h_t \Delta Z_t^h + v h_t \Delta H_t^h\}/\{1 - \theta\lambda(\mu_h - r_t)\Delta t\}$$

$$r_{t+1} = r_t + k_r(\mu_r - r_t)\Delta t + \sigma_r \Delta Z_t^r$$

Where:

- Δt quarter
- $\Delta Z_t^h = Z_{t+1}^h - Z_t^h$
- $\Delta H_t^h = H_{t+1}^h - H_t^h$
- $\theta \in [0,1]$, NOTE: if $\theta = 0$ Euler-Maruyama

Analogously will be implemented :

$$\frac{\Delta m_t}{m_t} = \left(\mu_m + k_{m_1}\frac{\Delta h_{t-l}}{h_{t-l}} + k_{m_2}\frac{\Delta f_{t-l_2}}{f_{t-l_2}}\right)(1-\theta)\Delta t + \left(\mu_m + k_{m_1}\frac{\Delta h_{t-l}}{h_{t-l}} + k_{m_2}\frac{\Delta f_{t-l_2}}{f_{t-l_2}}\right)\theta\Delta t + \sigma_m \Delta Z_t^m + \Delta H_t^m$$

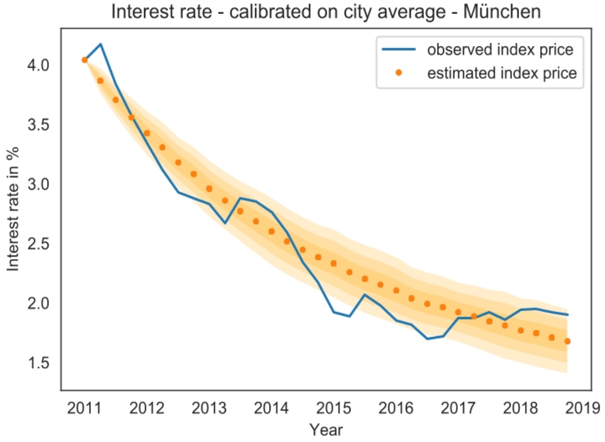
$$\Delta f_t = k_m(\mu_m - f_t)\Delta t + \sigma_m \Delta Z_t^m$$

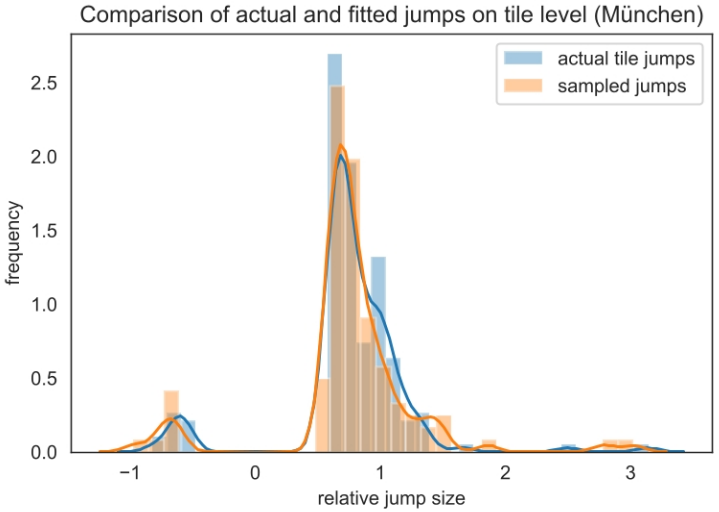
SDE MODEL – IMPLEMENTATION PIPELINE (1/2)

Pipeline SDE

Data	**Interest Rate** Parameter estimation and numeric solution	**Housing Price Index Analytical Parameter estimation** - of diffusion part - of starting values HPI	**Housing Price Index Optimization:** Parameter estimation (of drift part)	**Prediction** Interest Rate & Housing Price Index

$$(\hat{k}_r, \hat{\mu}_r) = argmin \sum_{i=1}^{N-1} (r_{i+1} - r_i - k_r(\mu_r - r_i))$$





$$\theta^* = arginf \left\| Y^{obs} - \hat{Y}^{\theta} \right\|^2$$

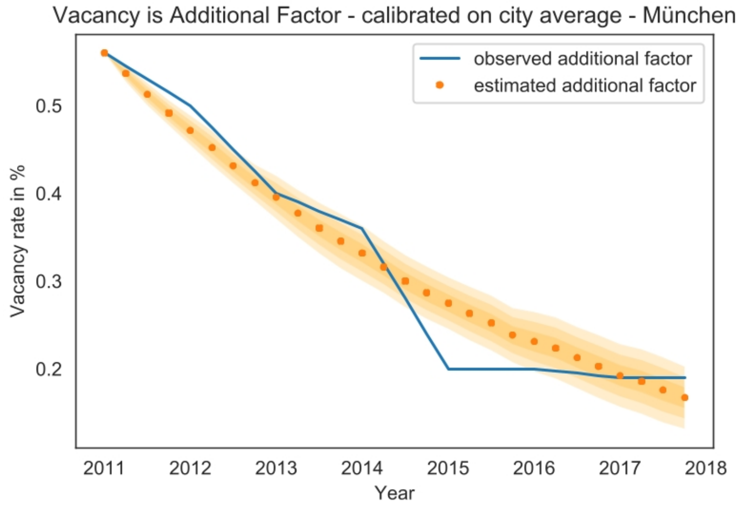
$$\equiv arginf \sum_{t=0}^{T} (Y_t^{obs} - \hat{Y}_t^{\theta})^2$$

SDE MODEL – IMPLEMENTATION PIPELINE (2/2)

Pipeline SDE

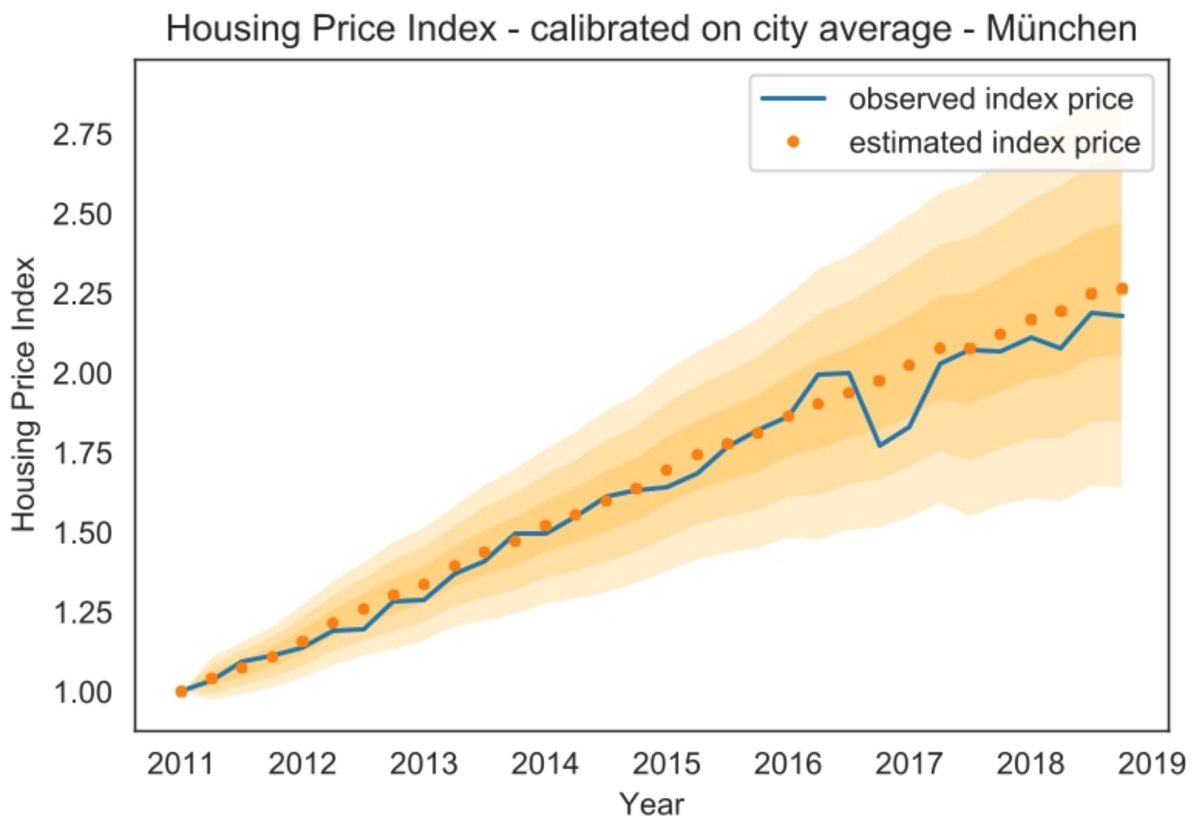
Business plan

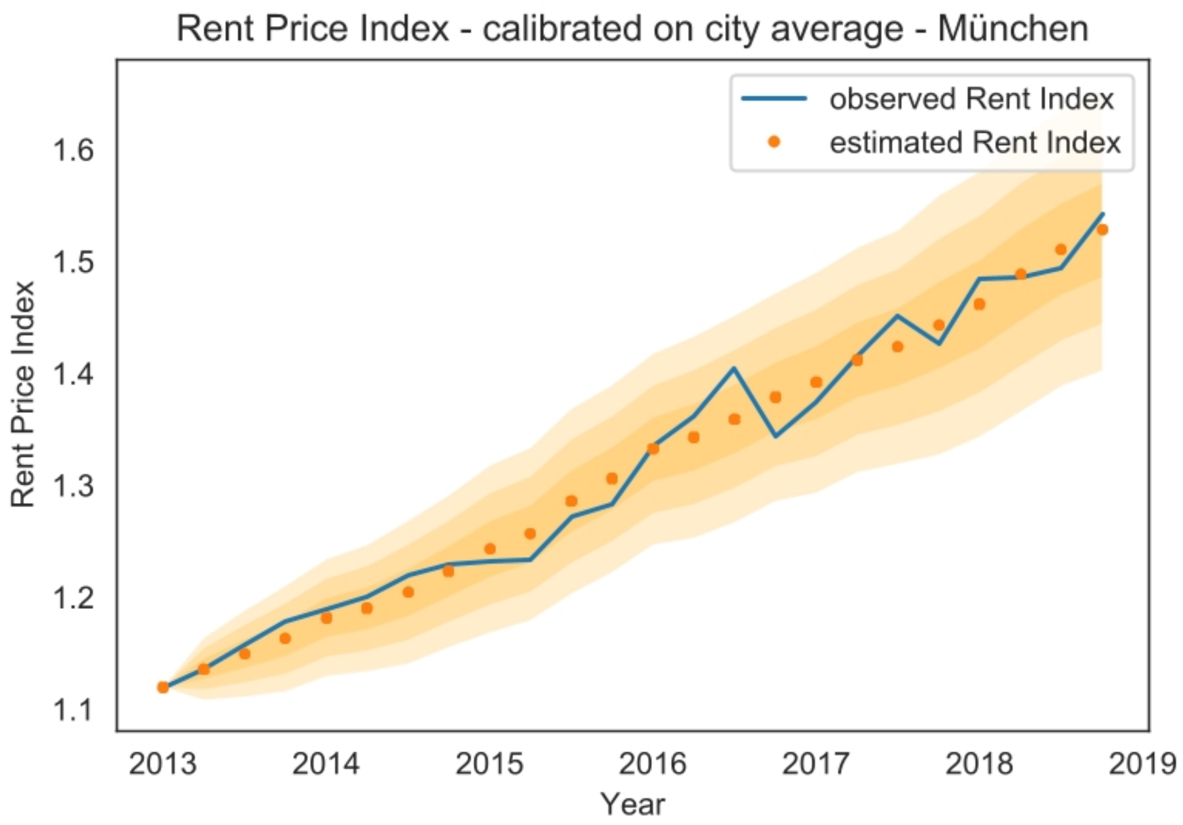
PREVIOUS	**Vacancy rate** - Parameter estimation and numeric solution	**Rent Price Index Analytical Parameter estimation** - of diffusion part	**Rent Price Index Optimization**: Parameter estimation (of drift part)	**Prediction** Additional Factor & Rent Price Index



$$\theta^* = arginf \left\| Y^{obs} - \hat{Y}^{\theta} \right\|^2 \equiv arginf \sum_{t=0}^{T} (Y_t^{obs} - \hat{Y}_t^{\theta})^2$$

SDE MODEL – OBSERVED VS ESTIMATED DATA





SDE MODEL – OUTLOOK AND LIMITATIONS

LIMITATIONS

Model derived by an historical point of view:

- *no embedding of economic point of view into the future*
- *no adjustment of predictions due to the outbreak of the Corona Virus*

OUTLOOK

Possible model extension:

- *Non-constant volatility*
- *Incorporating further additional factors*
- *Prediction on tile level if location information is available*

RISK MODEL – WORKFLOW

Define Risk:
Build model(s) that generate risk scores for available dataset + Research

↓

Data Pipeline: Pre-process Data

↓

Run and calibrate models:
1. Logistic Regression Model
2. Location Factor Model

↓

Aggregate risk score

↓

Visualize Results

RISK MODEL – LOGIT MODEL

Risk as the probability of the profit not reaching a certain level u.

$$profit_{c,y} < u$$

$$profit_{c,y} = \frac{rent_{c,y} + purchasing\ price_{c,y+1}}{purchasing\ price_{c,y}} - 1$$

Independent variables with a timeseries X from data pipeline.

$$y = \alpha + \beta * X + \varepsilon$$

Define best set of features X' out of X and best level for u.

Run regression pipeline for ~40 different values for u. Tested city-, time-specific and constant levels for u.

1. *Drop missing values*
2. *Remove correlated features*
3. *Include Interactions in X*
4. *Include Lags*
5. *Scale Features*
6. *Remove correlated features again*
7. *Recursive Feature Elimination*

RISK MODEL – LOGIT MODEL

Highest measure of certainty R^2 for u=5%.

```
                                                          Results: Logit
======================================================================================================================================
Model:                                               Logit                                         Pseudo R-squared:                                0.210
Dependent Variable:                                  y                                             AIC:                                             936.4231
Date:                                                2020-07-17 14:36                              BIC:                                             1040.3701
No. Observations:                                    1043                                          Log-Likelihood:                                  -447.21
Df Model:                                            20                                            LL-Null:                                         -566.20
Df Residuals:                                        1022                                          LLR p-value:                                     3.0117e-39
Converged:                                           1.0000                                        Scale:                                           1.0000
No. Iterations:                                      8.0000
--------------------------------------------------------------------------------------------------------------------------------------
                                                                                                    Coef.  Std.Err.    z     P>|z|   [0.025   0.975]
--------------------------------------------------------------------------------------------------------------------------------------
Einkommensteuer                                                                                   -2.0572   0.9829 -2.0930 0.0363  -3.9837 -0.1308
Population                                                                                        -4.9122   2.5345 -1.9381 0.0526  -9.8797  0.0553
growth_Anteil Schutzsuchender an Bevölkerung                                                       2.1385   1.5746  1.3581 0.1744  -0.9477  5.2248
growth_Population                                                                                 -6.2201   2.2120 -2.8120 0.0049 -10.5556 -1.8847
mult_Income_x_Pendlersaldo                                                                        -2.8226   1.0901 -2.5892 0.0096  -4.9592 -0.6859
mult_Studierende_x_growth_app_licence_residential                                                  4.2318   1.2072  3.5055 0.0005   1.8658  6.5979
mult_Einkommensteuer_x_Growth_Shrink_Ratio                                                        -2.0190   1.6854 -1.1979 0.2310  -5.3223  1.2844
mult_Verhältnis junge zu alten Erwerbsfähigen_x_Beschäftigte Tertiärer Sektor                     -1.4476   0.7644 -1.8939 0.0582  -2.9458  0.0505
mult_growth_Einkommensteuer_x_growth_Population                                                   -0.6167   1.6642 -0.3706 0.7110  -3.8784  2.6451
mult_growth_Anteil Schutzsuchender an Bevölkerung_x_growth_Bruttowertschöpfung                     6.2480   2.8407  2.1995 0.0278   0.6804 11.8157
mult_growth_Empfänger von Grundsicherung im Alter (Altersarmut)_x_growth_Growth_Shrink_Ratio     -3.7728   1.4599 -2.5844 0.0098  -6.6341 -0.9115
growth_Schuldnerquote_lagged_1                                                                     1.7541   0.7536  2.3276 0.0199   0.2770  3.2312
growth_Existenzgründungen_x_lagged_2                                                               2.3993   1.0993  2.1826 0.0291   0.2447  4.5539
growth_Anteil Teilzeitbeschäftigte_lagged_2                                                       -2.1682   0.8031 -2.6999 0.0069  -3.7421 -0.5942
growth_Beschäftigte am Wohnort mit akademischem Abschluss_lagged_2                                 1.9486   0.8279  2.3536 0.0186   0.3259  3.5713
growth_SGB II - Quote_lagged_1                                                                     1.6386   0.9479  1.7286 0.0839  -0.2193  3.4965
growth_Personen in Bedarfsgemeinschaften_lagged_2                                                  2.7145   0.8769  3.0954 0.0020   0.9957  4.4332
growth_birth_death_lagged_2                                                                       -1.7249   0.7221 -2.3888 0.0169  -3.1401 -0.3097
Growth_Shrink_Ratio_lagged_1                                                                       2.3353   1.5280  1.5283 0.1264  -0.6596  5.3301
growth_Growth_Shrink_Ratio_lagged_1                                                                2.9805   1.4196  2.0996 0.0358   0.1982  5.7627
const                                                                                              0.0958   1.2975  0.0739 0.9411  -2.4472  2.6389
======================================================================================================================================
```

Significant features with p-values<5%

feature	coef	pvalues
Einkommensteuer	-2.057223	0.036350
growth_Population	-6.220143	0.004923
mult_Income_x_Pendlersaldo	-2.822570	0.009620
mult_Studierende_x_growth_app_licence_residential	4.231842	0.000456
mult_growth_Anteil Schutzsuchender an Bevölker...	6.248036	0.027844
mult_growth_Empfänger von Grundsicherung im Al...	-3.772807	0.009756
growth_Schuldnerquote_lagged_1	1.754109	0.019934
growth_Existenzgründungen_x_lagged_2	2.399322	0.029069
growth_Anteil Teilzeitbeschäftigte_lagged_2	-2.168165	0.006936
growth_Beschäftigte am Wohnort mit akademische...	1.948616	0.018591
growth_Personen in Bedarfsgemeinschaften_lagged_2	2.714470	0.001965
growth_birth_death_lagged_2	-1.724871	0.016903
growth_Growth_Shrink_Ratio_lagged_1	2.980461	0.035766

*We end up with a **probability of default p** for every city and year.*

$$p = \frac{e^{\beta_1 * X_1 + \beta_2 * X_2 + .. + \beta_n * X_n}}{e^{\beta_1 * X_1 + \beta_2 * X_2 + .. + \beta_n * X_n} + 1}$$

RISK MODEL – LOCATION FACTOR MODEL

$$Risk \stackrel{?}{=} f(LocationFactors)$$

$$y = \alpha + \beta * X + \varepsilon$$

Independent variables without timeseries from data pipeline

Risk as stability versus instability over time

$$Sharpe\ Ratio_c = \frac{mean(profit_{c,qy})}{std(profit_{c,qy})}$$

$$Variance\ of\ profitgrowth_c = Var(\frac{profit_{c,qy} - profit_{c,qy-1}}{profit_{c,qy-1}})$$

Fig 1: Government district

Fig 2: Airport accessibility

Fig 3: Infrastructure connection (Autobahn)

SIEMENS

Fig 4: DAX companies

RISK MODEL – LOCATION FACTOR MODEL

Higher measure of certainty R^2 for y=Sharpe Ratio.

```
                                        Results: Ordinary least squares
==============================================================================================================================================
Model:                                        OLS                                           Adj. R-squared:                              0.477
Dependent Variable:                           SharpeRatio_Absolut                           AIC:                                         -40.8291
Date:                                         2020-07-20 11:21                              BIC:                                         7.2341
No. Observations:                             149                                           Log-Likelihood:                              36.415
Df Model:                                     15                                            F-statistic:                                 10.00
Df Residuals:                                 133                                           Prob (F-statistic):                          1.46e-15
R-squared:                                    0.530                                         Scale:                                       0.040233
----------------------------------------------------------------------------------------------------------------------------------------------
                                                                                                    Coef.  Std.Err.    t    P>|t|   [0.025  0.975]
----------------------------------------------------------------------------------------------------------------------------------------------
Leistungen für Wohngeld_mean                                                                       -0.1817  0.5132 -0.3541 0.7238 -1.1967  0.8333
Anteil Erholungsfläche_mean                                                                         0.6042  0.1676  3.6052 0.0004  0.2727  0.9357
Ärzte je  Einwohner_mean                                                                            0.0144  0.7646  0.0189 0.9850 -1.4980  1.5268
Ein- und Zweifamilienhäuser_mean                                                                   -0.6722  0.7080 -0.9494 0.3441 -2.0727  0.7282
Großunternehmen_mean                                                                               -0.3073  0.1263 -2.4336 0.0163 -0.5571 -0.0575
mult_Erreichbarkeit von Flughäfen_x_Anteil Erholungsfläche_mean                                    -0.5965  0.1693 -3.5236 0.0006 -0.9314 -0.2617
mult_Leistungen für Wohngeld_mean_x_Ein- und Zweifamilienhäuser_mean                                0.6850  0.6656  1.0292 0.3053 -0.6315  2.0016
mult_Nahversorgung Grundschulen Durchschnittsdistanz_x_Erreichbarkeit von Oberzentren               0.0301  0.1100  0.2739 0.7846 -0.1874  0.2476
mult_Nahversorgung Supermärkte Durchschnittsdistanz_x_Erreichbarkeit von Autobahnen                 0.2633  0.1658  1.5875 0.1148 -0.0648  0.5913
mult_Anteil Erholungsfläche_mean_x_Erreichbarkeit von Autobahnen                                    0.4851  0.1780  2.7252 0.0073  0.1330  0.8372
mult_Anteil Erholungsfläche_mean_x_Erreichbarkeit von IC/EC/ICE-Bahnhöfen                          -0.5054  0.2481 -2.0365 0.0437 -0.9962 -0.0145
mult_Anteil Erholungsfläche_mean_x_Erreichbarkeit von Oberzentren                                  -0.3560  0.1275 -2.7912 0.0060 -0.6082 -0.1037
mult_Erreichbarkeit von Autobahnen_x_Nahversorgung Apotheken Durchschnittsdistanz                  -0.5366  0.2240 -2.3953 0.0180 -0.9797 -0.0935
mult_Erreichbarkeit von IC/EC/ICE-Bahnhöfen_x_Nahversorgung Apotheken Durchschnittsdistanz          0.3353  0.1592  2.1059 0.0371  0.0204  0.6502
mult_Ärzte je  Einwohner_mean_x_Ein- und Zweifamilienhäuser_mean                                    0.2848  0.8441  0.3374 0.7363 -1.3849  1.9545
const                                                                                               0.7101  0.2951  2.4061 0.0175  0.1264  1.2939
----------------------------------------------------------------------------------------------------------------------------------------------
Omnibus:                                        49.134                                  Durbin-Watson:                                   1.731
Prob(Omnibus):                                  0.000                                   Jarque-Bera (JB):                                266.930
Skew:                                           1.024                                   Prob(JB):                                        0.000
Kurtosis:                                       9.229                                   Condition No.:                                   132
==============================================================================================================================================
```

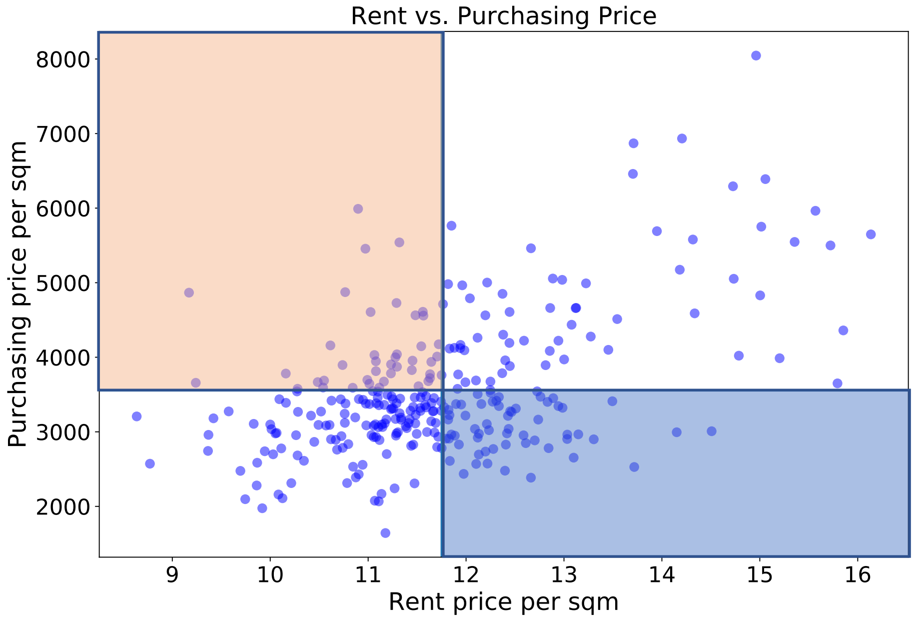
Significant features with p-values<5%

feature	coef	pvalues
Anteil Erholungsfläche_mean	0.604224	0.000440
Großunternehmen_mean	-0.307319	0.016276
mult_Erreichbarkeit von Flughäfen_x_Anteil Erh...	-0.596523	0.000584
mult_Anteil Erholungsfläche_mean_x_Erreichbark...	0.485129	0.007291
mult_Anteil Erholungsfläche_mean_x_Erreichbark...	-0.505358	0.043680
mult_Anteil Erholungsfläche_mean_x_Erreichbark...	-0.355964	0.006026
mult_Erreichbarkeit von Autobahnen_x_Nahversor...	-0.536618	0.018001
mult_Erreichbarkeit von IC/EC/ICE-Bahnhöfen_x_...	0.335290	0.037094

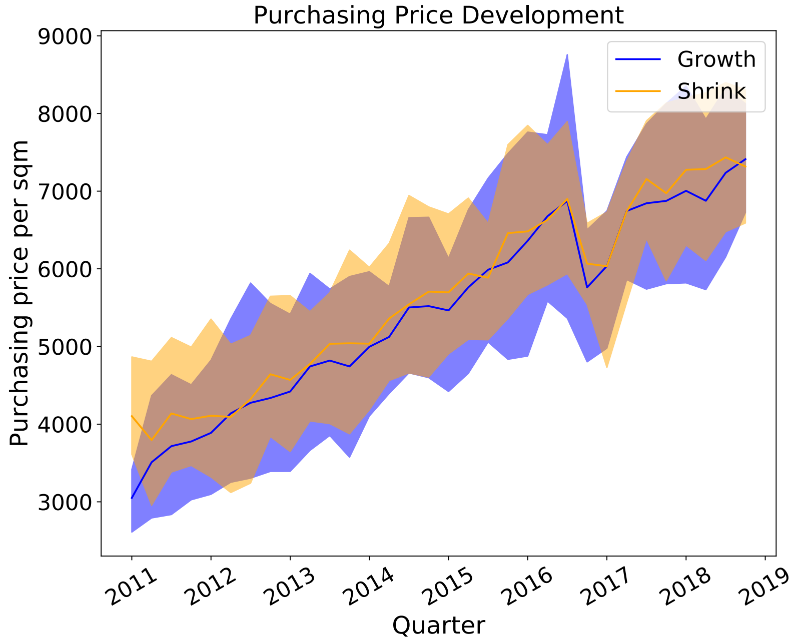
We end up with normalized risk score for every city and year q.

$$q = \beta_1 * X_1 + \beta_2 * X_2 + .. + \beta_n * X_n$$

RISK MODEL – QUADRANTS



Purchasing vs. Rent price for first quarter of 2011 for Munich



Development of purchasing prices of growth and shrink tiles from 2011.00 for Munich

Include observations into Logit Model using two additional features

$$Growth\ Shrink\ Ratio_{y,c} = \frac{\#tilesgrowth_{y,c}}{\#tilesshrink_{y,c}}$$

$$Price\ Rent\ Ratio\ Variance_{y,c}$$

RISK MODEL – OUTLOOK AND LIMITATIONS

LIMITATIONS

Different way of calculating the profit

$$grossprofit_{c,y} = \frac{rent_{c,y}}{purchasing\ price_{c,y}}$$

vs.

$$profit_{c,y} = \frac{rent_{c,y} + purchasing\ price_{c,y+1}}{purchasing\ price_{c,y}} - 1$$

Location Factor Model

-> aggregating the data leaves us with only 149 observations, one could think of just incorporating the historic Sharpe Ratio

OUTLOOK

City specific level of u

-> with more granular data, one could run city specific regressions

Quadrants

-> came up with this concept, further research might be interesting

TRANSFORMING SCENARIOS TO IRR

SDE Model
Calibrated for each city

→

House & Rent Prices
10 years predicted

→

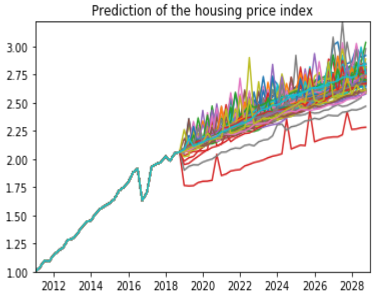
Business Plan
Monte Carlo

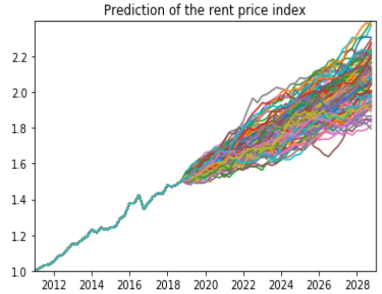
→

IRR

Scenarios

100 rent and sale price scenarios give us cash flows for an artificial real estate object portfolio





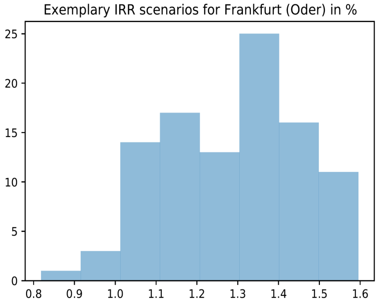
CB BP

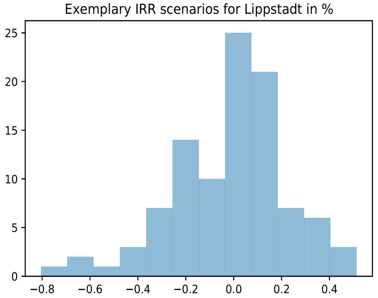
With these cash flows the IRR for every scenario and city is computed. Vacancy rate, fixed rent prices for specified periods etc. are simulated.

$$NPV = \Sigma_{t=0}^{T} \frac{CF}{IRR^t} \overset{!}{=} 0$$

IRRs

We obtain an IRR distribution for every city with 100 samples each





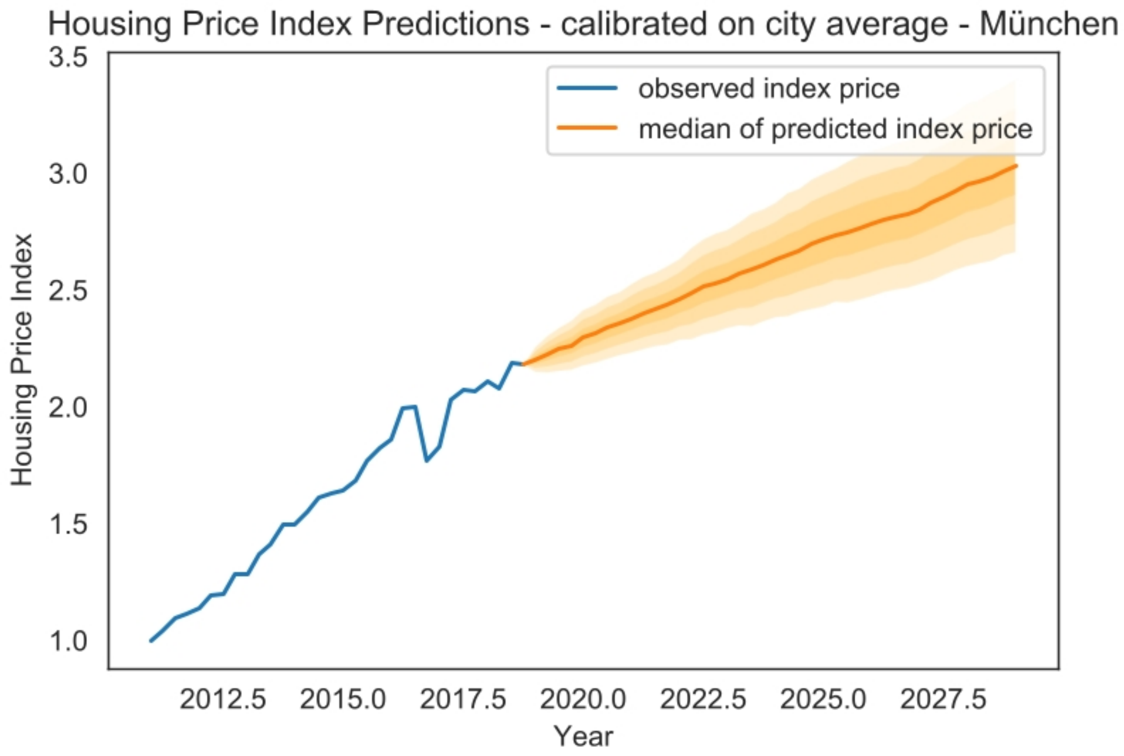
$\mathbb{E}(IRR)$

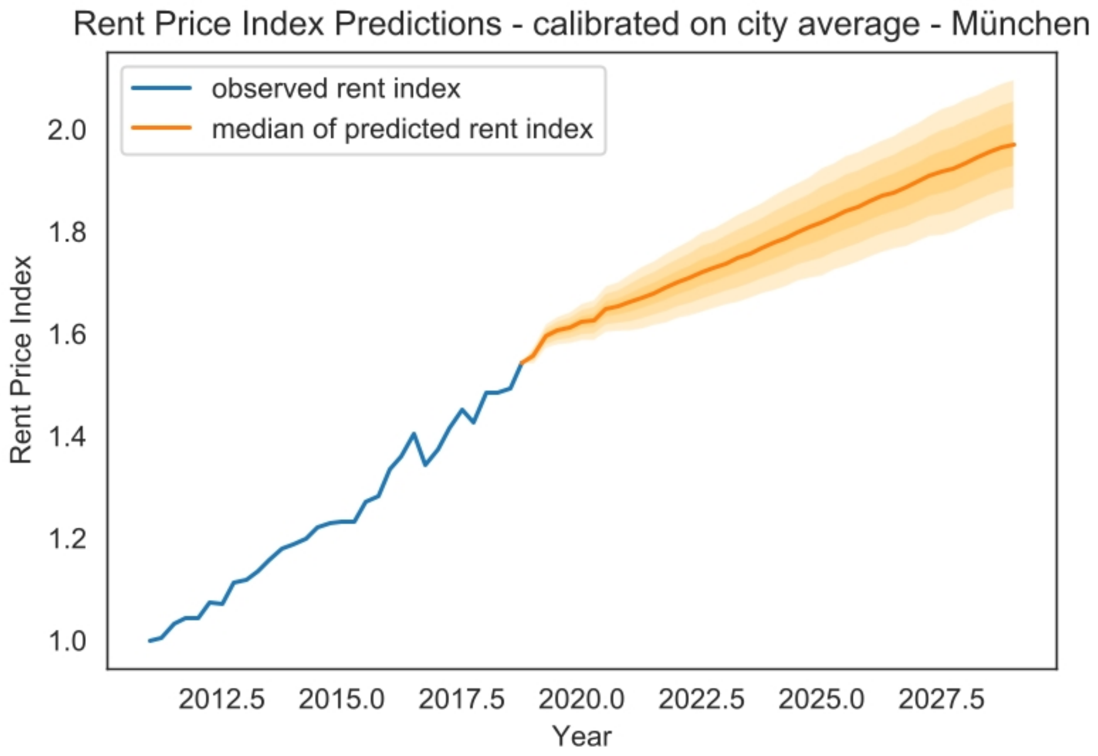
As single KPI for each city we take the mean of the IRR distribution

1.3%

0.1%

SDE MODEL – RESULTS: PREDICTIONS OVER 10 YEARS





RISK MODEL – AGGREGATED RISK MODEL

Set of possible explanatory variables X – Timeseries and $var(x) > \epsilon$?

Yes

Logit model

$Logistic\ Regression$

$$y = \alpha + \beta * X + \epsilon$$

$$with\ y = \begin{cases} 0, if\ profit > u \\ 1, if\ profit < u \end{cases}$$

Probability of Default between 0 and 1

No

Location factor model

$Linear\ Regression$

$$y = \alpha + \beta * X + \epsilon$$

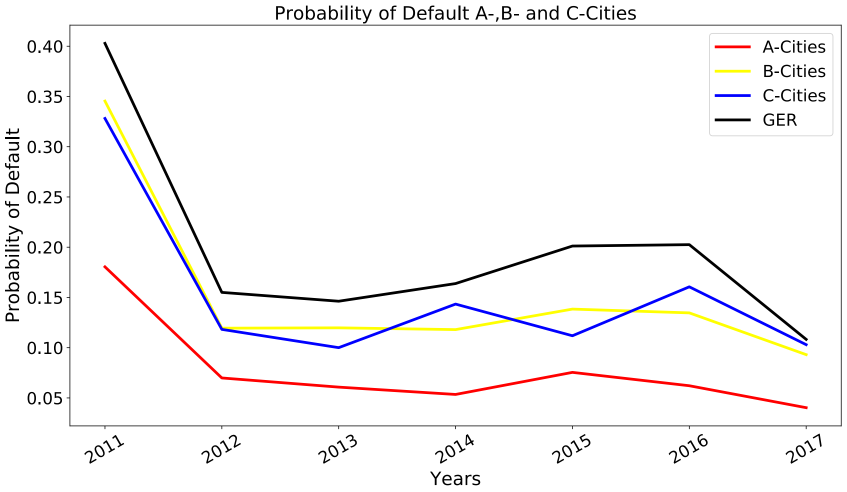
$$with\ y = \frac{mean(profit_c)}{std(profit_c)}$$

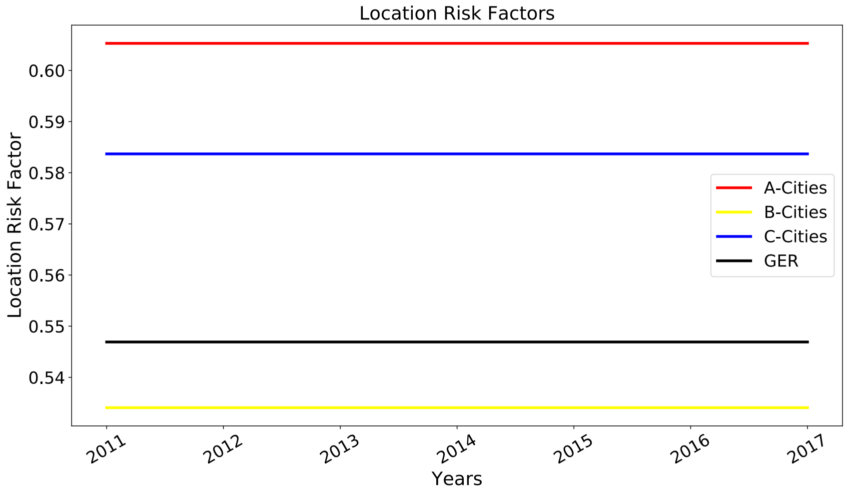
Normalized risk score between 0 and 1

$$Aggregated\ Risk\ Score = w * Normalized\ Risk\ Score + (1 - w) * Probability\ of\ Default$$

FINALRESULTS

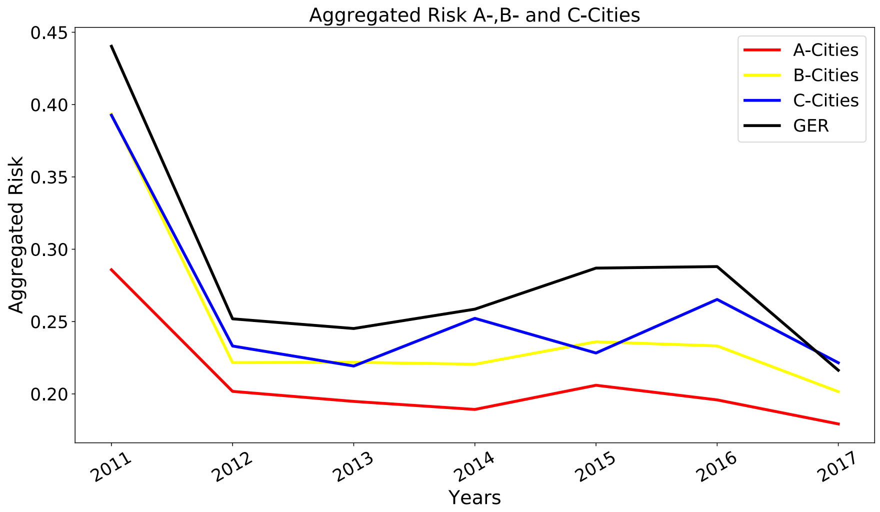
RISK MODEL - RESULTS





$* \frac{3}{4}$

$* \frac{1}{4}$



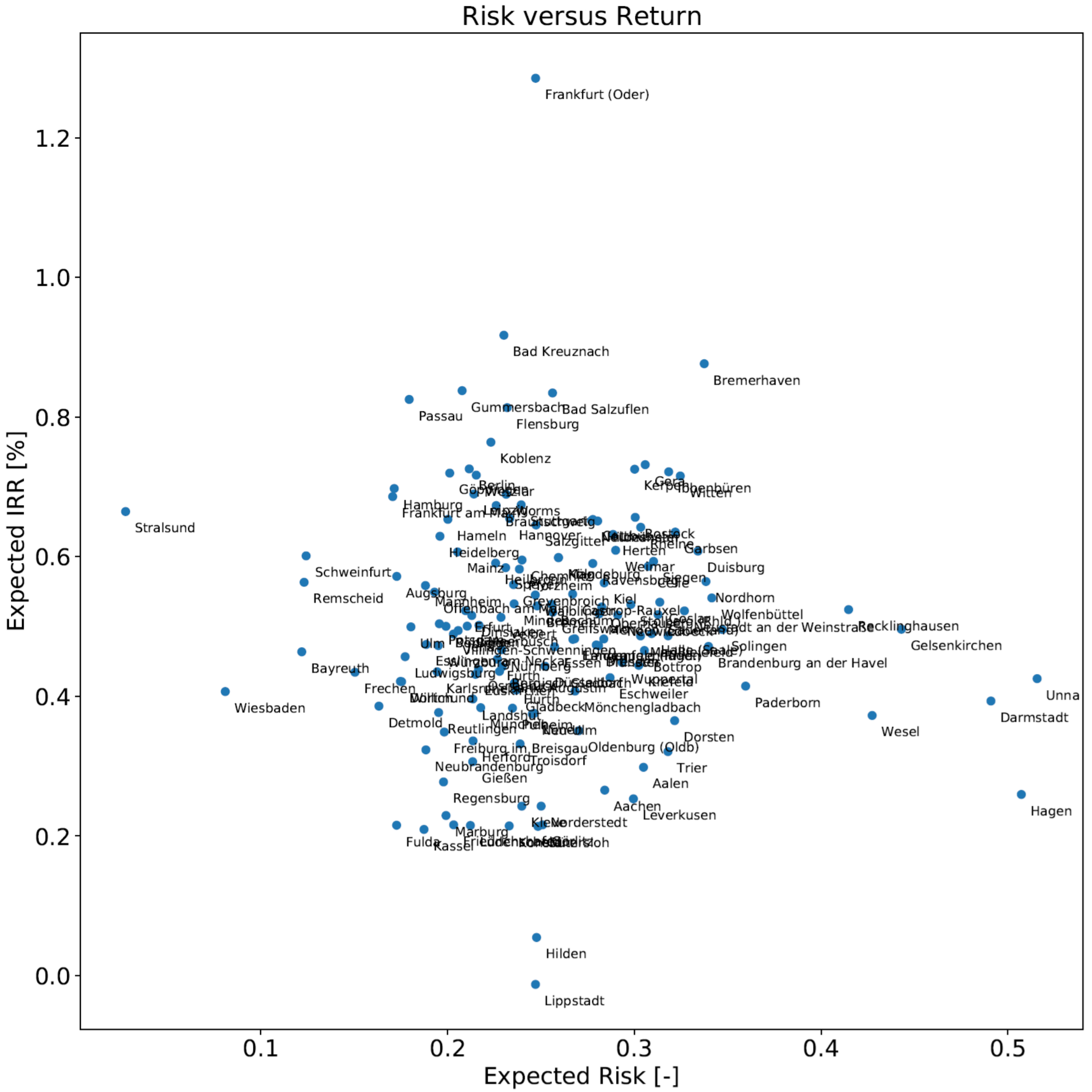
A-Cities	B-Cities	C-Cities
Berlin	Bochum	Aachen
Düsseldorf	Bonn	Augsburg
Frankfurt (Main)	Bremen	Bielefeld
Köln	Dresden	Darmstadt
München	Duisburg	Erfurt
Stuttgart	Essen	Freiburg
	Hannover	Heidelberg
	Karlsruhe	Kiel
	Leipzig	Lübeck
	Mannheim	Magdeburg
	Münster	Mainz
	Nürnberg	Mönchengladbach
	Wiesbaden	Offenbach (Main)
		Osnabrück
		Potsdam
		Regensburg
		Rostock
		Saarbrücken
		Wuppertal

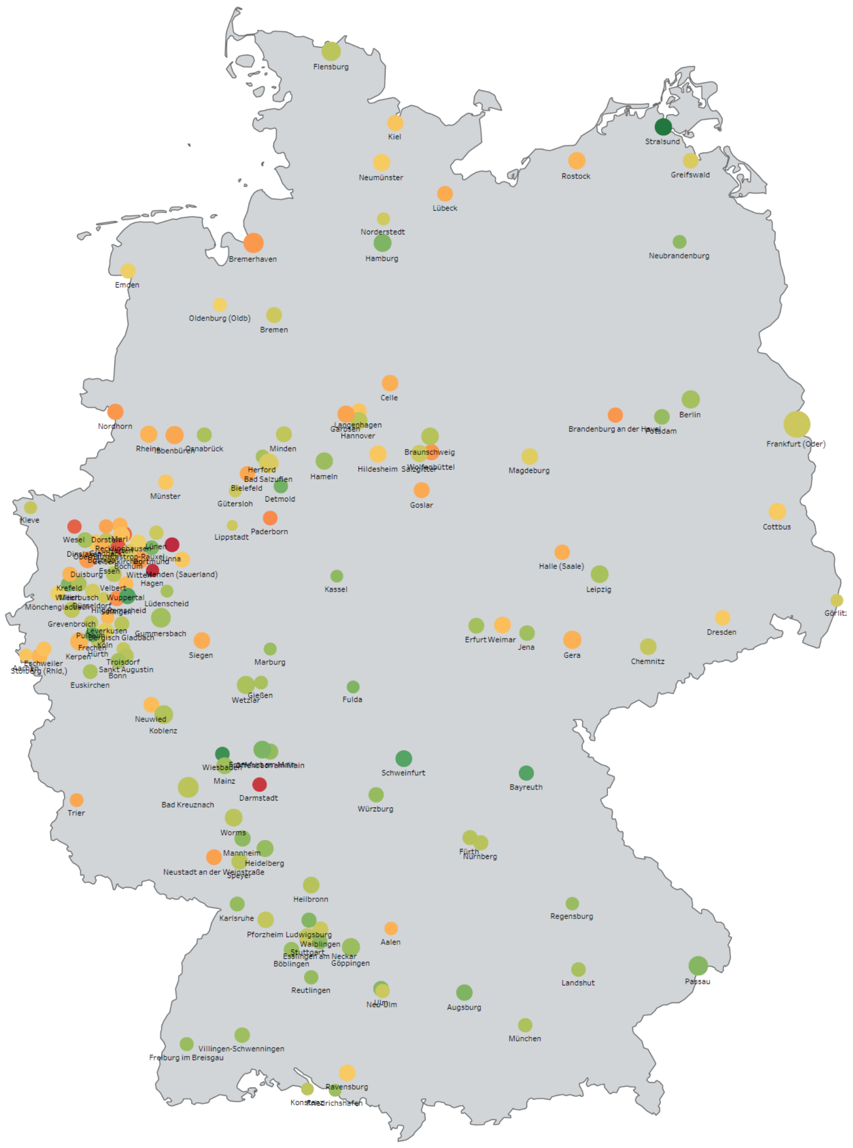
Table 3

City categorization in A-, B- and C-Cities

FINALRESULTS

RISK VS UNLEVERED RETURN SIMILAR FOR CITIES





Risk vs. Return for 2017

PLAUSIBLE RESULTS WITH ROOM FOR IMPROVEMENT

Caveats

- 21RE data is still limited in scope and only reflects offered prices, not the true transaction prices (these are only known by notaries and tax offices).
- The data quality of INKAR data is good, but only on yearly basis.
- Information on tile level had to be aggregated and only KPIs on city level are computed. In reality still big differences intracity possible (Apartment at Marienplatz vs. Giesing differs price-wise).
- Risk and return KPIs are our subjective choices and other KPIs might be better suited for investment decisions.

Outlook

- Different KPIs for both risk and return
- More data and better data pipeline, data driven approaches dictate garbage in, garbage out
- Third, a analysis not only on city level but on tile or district level could be feasible. That approach would also need more data but would allow for single real estate objects to be evaluated more precise.
- SDE and risk model assumptions can be weakened and expanded, e.g. non-constant volatility for the SDE model or other additional factors.

Appendix

RISK MODEL – DATA PIPELINE

Input Variables

21st

INKAR

Indikatoren und Karten zur Raum- und Stadtentwicklung

Time series available

Constant over time
varying over cities

Logit Model:

$$y = \alpha + \beta * X + \varepsilon$$

- Interpolating / Filling Missing Data
- Calculating Growth Variables
- Quadrant idea

- Build y to capture risk

Location Factor Model:

$$y = \alpha + \beta * X + \varepsilon$$

- Aggregate

RISK MODEL – LOCATION FACTOR MODEL

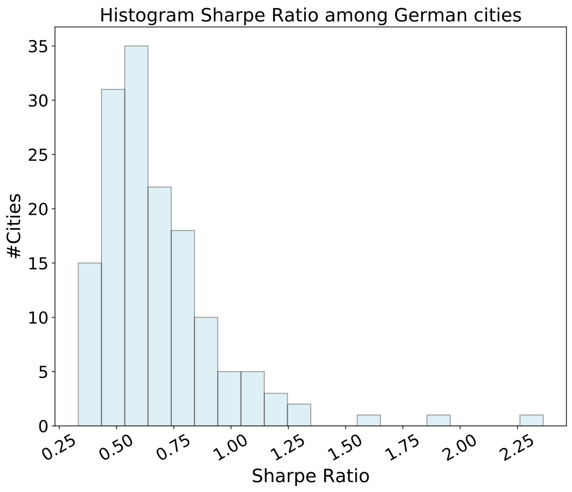
Risk as stability versus instability over time

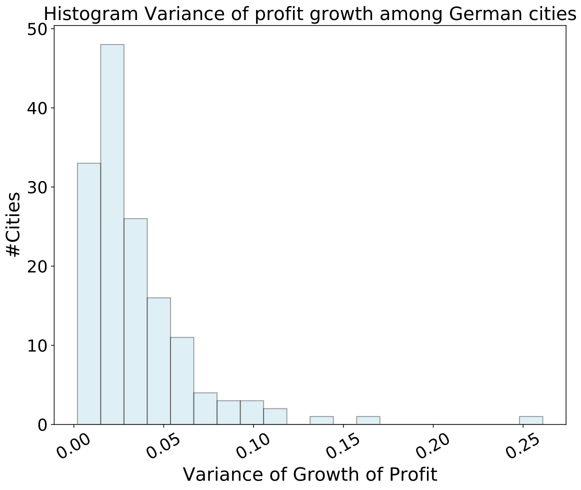
$$Sharpe\ Ratio_c = \frac{mean(profit_{c,qy})}{std(profit_{c,qy})}$$

$$Variance\ of\ profitgrowth_c = Var(\frac{profit_{c,qy} - profit_{c,qy-1}}{profit_{c,qy-1}})$$

Independent variables without timeseries X out of the data pipeline

$$y = \alpha + \beta * X + \varepsilon$$





Run regression pipeline for Sharpe Ratio and Variance of profitgrowth as y.

1. *Drop missing values*
2. *Remove correlated features*
3. *Include Interactions in X*
4. *Scale Features*
5. *Recursive Feature Elimination*

List of Figures

- Fig 1: https://www.berlin.de/tourismus/dampferfahrten/x/5233670-5433923-schiffstouren-im-regierungsviertel.html , Accessed 29.07.20, © JFL Photography - stock.adobe.com.
- Fig 2: https://www.airliners.de/analyse-verkehrszahekln/50268 , Accessed 29.07.20, © Fraport.
- Fig 3: "Autobahn Vector PNG" http://pluspng.com/png-99964.html , Accessed 29.07.20
- Fig 4: https://de.wikipedia.org/wiki/Datei:Siemens-logo.svg, Accessed 29.07.20, ©Siemens AG.