33

Schriften aus der Fakultät Sozial- und Wirtschaftswissenschaften der Otto-Friedrich-Universität Bamberg

# KriMI: A Multiple Imputation Approach for Preserving Spatial Dependencies

## Imputation of Regional Price Indices using the Example of Bavaria

von Sara Bleninger

University of Bamberg Press

**33** Schriften aus der Fakultät Sozial- und Wirtschaftswissenschaften der Otto-Friedrich-Universität Bamberg

Schriften aus der Fakultät Sozial- und Wirtschaftswissenschaften der Otto-Friedrich-Universität Bamberg

Band 33

University of Bamberg Press
2017

# KriMI: A Multiple Imputation Approach for Preserving Spatial Dependencies

Imputation of Regional Price Indices
using the Example of Bavaria

von Sara Bleninger

University
of Bamberg
Press
**2017**

Bibliographische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Informationen sind im Internet über http://dnb.d-nb.de/ abrufbar.

Diese Arbeit hat der Fakultät Sozial- und Wirtschaftswissenschaften der Otto-Friedrich-Universität Bamberg als Dissertation vorgelegen.
1. Gutachter: Prof. Dr. Susanne Rässler
2. Gutachter: Prof. Dr. Trivellore Raghunathan
Tag der mündlichen Prüfung: 17.07.2017

Dieses Werk ist als freie Onlineversion über den Hochschulschriften-Server (OPUS; http://www.opus-bayern.de/uni-bamberg/) der Universitätsbibliothek Bamberg erreichbar. Kopien und Ausdrucke dürfen nur zum privaten und sonstigen eigenen Gebrauch angefertigt werden.

Herstellung und Druck: docupoint, Magdeburg
Umschlaggestaltung: University of Bamberg Press, Larissa Günther

© University of Bamberg Press, Bamberg, 2017
http://www.uni-bamberg.de/ubp/

ISSN: 1867-6197
ISBN: 978-3-86309-523-9 (Druckausgabe)
eISBN: 978-3-86309-524-6 (Online-Ausgabe)
URN: urn:nbn:de:bvb:473-opus4-502980
DOI: http://dx.doi.org/10.20378/irbo-50298

# Danksagung

Diese Dissertation ist nicht nur das Resultat meines Aufwands, sondern das Ergebnis der Geduld ganz vieler Menschen. Ich durfte die größtmögliche Unterstützung erfahren.

An erster Stelle muss ich mich bei meiner Doktormutter Susanne Rässler bedanken. Sie hat sich vor ungefähr fünf Jahren dafür mehr als verantwortlich gefühlt, dass ich meine Dissertation an der Universität Bamberg nicht nur einfach schreiben kann, sondern auch eine gute Betreuung und ein interessantes Thema habe. Sie hat ein Umfeld geschaffen, in dem ich mich frei entfalten konnte und mich sehr wohl fühle.
Zu diesem Umfeld gehören insbesondere meine nun ehemaligen Kollegen am Lehrstuhl. Besonders möchte ich mich dabei bei Alexandra Trojan bedanken, die mit mir zusammen in dem Projekt zu regionalen Preisen promoviert. Unseren kleinen privaten Workshop in unserem Esszimmer habe ich sehr genossen. Zudem möchte ich mich bei Thorsten Schnapp für seine andauernde R-Hilfe bedanken. Er hat eigentlich fast immer für meine R-Probleme Lösungen und Antworten gehabt. Das Gleiche gilt für Florian Meinfelder. Seine Hilfe war der Startpunkt für viele meiner R-Programmierungen. Sören Abbenseths geduldige Hilfe bei der Aufbereitung der Preisdaten darf nicht unerwähnt bleiben.

Rachel und Jenny schulde ich Dank für ihre unendliche Geduld, meine Arbeit auf sprachliche Mängel hin zu korrigieren. Sie haben das beide sehr ausdauernd gemacht und auch ertragen, was ich ihrer schönen Muttersprache in manchen Fällen angetan habe.

Aber die größte Hilfe bei allem war Philipp. Es ist von unschätzbarem Wert, uns zwei zu haben. Du hast mir hier nicht nur den Rücken frei gehalten, sondern mich durch die schwierigsten Situationen getragen und in allem auch ertragen. Auch unserem kleinen Jakob habe ich dafür zu danken, dass er so ein Sonnenschein ist, dass ich nebenbei meine Arbeit fertig schreiben konnte.

Zuletzt muss ich noch die Menschen im Hintergrund erwähnen, ohne deren Unterstützung ich das alles nicht geschafft hätte. Das sind auf jeden Fall meine beiden Eltern, die sich für mich immer ein Bein ausgerissen haben, die sich nie darüber beschwert haben, dass ich jahrelang studiert habe, die alles in jeglicher Hinsicht unterstützt haben, die immer da sind. Auch meiner Schwiegermama Susi gilt mein Dank, vor allem dafür, dass sie sich immer so toll um Jakob kümmert. Es ist aber auch schön, dass sie so stolz auf mich ist und bei Gott und der Welt mit ihrer Schwiegertochter angibt. Meinem Schwiegervater Manfred und Andrea muss ich vor allem für die tolle Unterstützung während meines Statistik-Studiums in München danken, aber auch für die vielen tollen Gespräche und das Interesse.

# Inhaltsverzeichnis

# Abbildungsverzeichnis

# Tabellenverzeichnis

# 1 Introduction

In the middle of every month the German Federal Statistical Office releases the new consumer price index reporting the current cost of living of a typical German household. Every month it is reported that the amount that a German household has to pay for a typical market basket changed at a certain degree given as a percentage number. The German media monthly report the current inflation rate for Germany based on this percentage number, which is called price index. Even though reported for years, every month there is still an important question left unanswered: Do price levels differ among regions? The public attention mostly focuses on the national inflation representing the change of the price level over time. A spatial point of view remains unobserved.

Exploring the regional disparities should be of great public interest. It goes beyond answering the question where we have to pay more or less. There are a lot of public transfer payments that are given according to needs. Of course the costs of these needs measured in money depends on the local price level. Recipients of the public transfer payments have to pay more for their livelihood in a high price region than persons living in a low price region. Also the salary of civil servants should be aligned to the local cost of living. For this reason, a Bavarian policeman took the state to court as he wanted his salary adjusted for the region he lived in. This lawsuit was one of the rare events when regional cost of living appeared in the public arena. The court refused the lawsuit.(BVerfG, 2007)

In science the question of regional price disparities plays a bigger role than in public discourse. There are a lot of publications about regional price levels. We are going to discuss some of them later. Economic and social scientists need regional cost of living for adjusting all kinds of income and payment variables in their studies. For example, if an economic scientist is analysing people's income, it is not the nominal amount which is of interest, but what we can buy for it. The latter is the real value, which is the nominal income adjusted for the regional price level.

Contrary to the need of regional price indices, there are not much data available on that high level of fragmentation. There are two relevant studies surveying regional data on prices in Germany, i.e. Rostin (1979) and Ströhl (1994), but these are rather old. The newer study of Ströhl (1994) only inspects prices in 50 cities, which is not enough for an investigation of regional price disparities. For these reasons there have been efforts to extend the data base by regression imputation, e.g. Roos (2006), or even with multiple imputation, e.g. Blien et al. (2009).
This thesis is written in the context of a project group with the objective of solving the various problems associated with computing regional price indices. There are several tasks to be solved: a regional price index needs to be developed, price data for all regions need to be surveyed or predicted, there are special problems with housing costs etc. For this reason, this thesis only plays a part in solving the various problems involved with acquiring a regional cost of living value. However, it is complete in providing statistical models to generate predictions of the prices in unsurveyed regions.

The single imputation by regression generates good results, i.e. means are correct in the long run. However, the interest is not to analyse regional prices itself, but to use the regional price level for subsequent studies. As regression imputation always underestimates variances of estimators, all succeeding significance levels are seriously biased. Donald B. Rubin developed a method to overcome the problem of underestimating variances by using Bayesian theory.(Rubin, 1987) The uncertainty brought in by the missing data is represented by imputing more times than once. Here, we define the regions where no price data were surveyed as missing values. The problem of unknown regional price levels becomes a problem of missing data which can be solved with missing data techniques, i.e. multiple imputation. In the second chapter we will introduce the argumentation and method of multiple imputation.

Actually, the problem of predicting the prices in unsurveyed regions is a matter of geostatistics. The models of geostatistics allow the modelling of spatial dependencies and inter-dependencies of neighbouring regions. Probably the prices in neighbouring regions are dependent on each other, making such a spatial modelling necessary. Geostatistics keeps a group of models at hand which are developed for predicting spatial dependent observations: kriging. In the third chapter, we will introduce kriging. The focus is on modelling spatial dependencies via correlation functions and covariances.

Both approaches solve one problem that we face when predicting unsurveyed regional prices. Multiple imputation allows proper inferences in succeeding statistical analysis. Kriging preserves spatial dependencies of the predicted values. If we want both, it seems a promising idea to combine both methods to get spatially dependent values for the unobserved prices that allow succeeding inference. In chapter 4 we derive two ways to implement kriging into the multiple imputation scheme. First, kriging is modelled as a mixed model where spatial dependencies are modelled in the variance-covariance matrix. The Bayesian formulation of the formed mixed model is then used to generate multiple imputations. Second, the familiarity of kriging to P-spline smoothing is exploited to derive a model that is closer to the given data. Again the Bayesian formulation is used to implement kriging into the multiple imputation model. We call these two models KriMI models as they are an amalgamation of kriging and multiple imputation.

In the seventh chapter we use all introduced statistical models to predict the unobserved prices in Bavarian regions. The Bavarian State Office for Statistics provided the data that are surveyed for the official consumer price index on a regional level. There are huge blank areas in the map of regional prices. For this reason the Bavarian data are a prime example to test our models. However, before we estimate and predict the missing price data, a short introduction to prices and price indices is given, as we need to know the observation object. We also mention briefly the special problems of regional price indices. For all statistical models we need to know relevant influencing factors. These are also derived in chapter 5. After this we create a baseline by imputing regional prices with a simple regression approach. Following sections identify the missing price data by multiple imputation and kriging to use the models in their pure form. In a final step we use the developed KriMI models for a multiple imputation of regional prices that preserves spatial dependencies. For all methods used, we report the regional price indices of the Bavarian regions.

All analyses and all graphs, including the maps, are made using the statistical software R Core Team (2015a), Belitz et al. (2015), and the geoinformation system GRASS Development Team (2012). We used several packages in R. All references for those R-packages can be found in the appendix.

# 2 Multiple Imputation

Multiple imputation is a method of handling missing data problems. (Zhang, 2003) As in single imputation, missing values are replaced by sensible values, but the replacement is done several times as „the inserted values should reflect variation within a model as well as variation due to the variety of reasonable models.“(Rubin, 1978). The aim is to still have a statistically valid inference in the case of missing data: In single imputation a correct variance imputation is no longer possible due to the replacements. Multiple imputation by contrast uses the differences between the several replacements to correct variance estimation to allow a valid statistical inference. Moreover, multiple imputation allows valid statistical inference in the case where the data analyst and the data manager are not the same person.(Rubin, 1996) These two arguments allow the use of multiple imputation in the case of missing regional price information.
Other techniques for incomplete data are based on the likelihood approach (e.g.the EM algorithm, the random effects model), but are inferior for the reason of greater complexity.(Zhang, 2003) As the problem of regionalised price indices can only be solved by providing completed data, imputation methods should be preferred to the likelihood based methods.

Zhang (2003) names three steps with which to conduct data analysis with multiple imputation:

1. Create $m > 1$ completed data sets by filling in each missing value $m$ times using a proper imputation model. Following the description of Schafer (1999) to get the multiple draws it becomes clear that it is not that easy:
   - First, specify a parametric model for the complete data.

- Second, according to Bayesian statistics, a prior has to be specified for the model parameters which are unknown, and
- $m$ independent draws from the conditional distributions are simulated.

2. Conduct a complete data analysis on each of the $m$ filled data sets.
3. Combine the statistics from step two by the combining rule.

Similarly, Rubin (1988) subsumes multiple imputation: „As its name suggests, multiple imputation replaces each missing value by a vector composed of $M \geq 2$ possible values. The $M$ values are ordered in the sense that the first components of the vectors for the missing values are used to create one completed data set, the second components of the vectors are used to create the second completed data set and so on; each completed data set is analysed using standard complete-data methods."(Rubin, 1988) A similar introduction can be found in Schafer (1999). Here, we emphasise step one while leaving the second step completely up to the researcher of regional prices. For step three, we give a short instruction. To do so we first describe the problem: In the next section we attend to the missingness mechanisms. Subsequently, a short insight is given into reasoning multiple imputation, followed by the combining rules. After having described the main body of multiple imputation, we spend some time on the details.

## 2.1 Describing the Missingness Mechanism

Before presenting a solution of missing values, we need to describe the problem itself. Ordinarily, there are data that are not observed when being collected. These data are missing. We distinguish three kinds of missingness described in this chapter.

Defining the variable R, with

$$r_{ij} = \begin{cases} 1 \text{ if } y_{ij} \text{ is observed} \\ 0 \text{ if } y_{ij} \text{ is missing} \end{cases}$$

with $i = 1, \dots, n$ denoting observations and $j = 1, \dots, p$ denoting variables. The probability distribution belonging to $R$ is $P\,(R|\varphi, Y)$ with the parameters $\varphi$ and the data matrix $Y$. The joint distribution of data and missingness indicator is (Zhang, 2003):

$$P\,(Y, R|\theta, \varphi) = P\,(Y|\theta)\,P\,(R|\varphi, Y)$$

where $\theta$ is a parameter or parameter vector for the data $Y$. We denote the missing data as $Y_{mis}$ meaning that these are the cases where $r_{ij} = 0$. Observed data are $Y_{obs}$ where $r_{ij} = 1$.
According to Rubin (1987), Zhang (2003), and several other authors, the following missingness mechanisms can be distinguished:

1. Missing Completely at Random (MCAR)(Rubin, 1987, Zhang, 2003)

   $$P\,(R|\varphi, Y) = P\,(R|\varphi)$$

   „The probability of missing data on one variable is not related to the value of that variable or other variables in the data set.“ (Patrician, 2002) By a comparison of the data of respondents to the data of non-respondents, the assumption of MCAR can be verified.(Patrician, 2002)

2. Missing at Random (MAR)(Rubin, 1987, Zhang, 2003)

   $$P\,(R|\varphi, Y) = P\,(R|\varphi, Y_{obs})$$

The probability of missingness of one value does not depend on the variable that is missing itself, but may depend on other variables included in the data set.(Patrician, 2002) According to Patrician (2002), the assumption MAR cannot be tested. MAR is relative to the variables included in the data set. All variables that are responsible for the missingness must be considered in the data set.(Schafer and Olsen, 1998)

3. Not Missing at Random (NMAR)(Rubin, 1987, Zhang, 2003)

$$P(R|\varphi, Y) \neq P(R|\varphi)$$

Following, we assume at least MAR, even though there are also possibilities to handle NMAR.(Rubin, 1987)

## 2.2 Introduction to Multiple Imputation

One way to identify appropriate values for missing data is multiple imputation (MI) introduced by Rubin (1987) and among others described in Little and Rubin (2002). The underlying idea is to impute $m$ credible values for each missing value to create $m$ filled data sets instead of imputing just one value to get one filled data set. Doing so, MI has all the advantages of single imputation thereby solving the problem of variances.

Multiple imputation uses the available information to find sensible values replacing the missing data. These are the observed data and information due to the data model (prior). A model connecting observed and unobserved values is needed to find proper imputations.(Rubin, 1987) This is done in the following subsubsection. In subsubsection 2.2.3 we describe how to find these proper values for missings, followed by a short insight into the special case of a monotone distinct missingness structure. We do not describe how to combine the multiple data sets until the following section.

### 2.2.1 Reasoning Multiple Imputation

Explicitly, MI means to repeatedly fill in values for the missing data $Y_{mis}$. In the end, we get: $Y^1_{mis},; Y^2_{mis},; \cdots, \; Y^m_{mis}$. The different sets of missing values are drawn from the posterior predictive.(Rubin, 1996) Rubin (1987) justifies this using the Bayesian Theory. One bunch of replacements for missings are got by drawing from the posterior distribution employing the Bayesian theory. As posteriors often cannot be computed, proper values for unobserved data are found by iterative simulation methods.(Little and Rubin, 2002) Due to drawing, Schafer (1997) states: „Like parameter simulation, multiple imputation is a Monte Carlo approach to the analysis of incomplete data. (...) solving an incomplete-data problem by repeatedly solving the complete data version.“(Schafer, 1997, S.104) Thereby, Bayesian theory is justification, theoretical background and prescription of MI.(Rubin, 1987)
In this section we first show the reason why inferential statistics can be done using imputed data instead of using the complete data. This is the justification and theoretical background of MI.

Little and Rubin (2002) and other authors of MI consider the complete data posterior:

$$P(\eta|Y_{obs}, Y_{mis}, X) \propto P(\eta)L(\eta|Y_{obs}, Y_{mis}, X). \tag{1}$$

In compliance with the three Bayes Postulates and especially the Bayes Corollary, inferences shall only be made on the basis of the complete data posteriori $P\,(\eta|Y,X) \;=\; P(\eta|Y_{obs}, Y_{mis}, X)$.(Rüger, 1999) Until now, we only formulated classic Bayesian inference. Returning to the missing data problem, it becomes clear that we cannot observe the complete-data posterior $P(\eta|Y_{obs}, Y_{mis}, X)$ due to the missing data. As we know the observed-data posterior, the relationship between observed-data-posterior and complete-data-posterior needs to be drawn (Little and Rubin, 2002, Zhang, 2003):

$$\begin{aligned} P(\eta|Y_{obs}, X) &= \int P(\eta, Y_{mis}|Y_{obs}, X)\, dY_{mis} \\ &= \int P(\eta|Y_{mis}, Y_{obs}, X)\, P(Y_{mis}|Y_{obs}, X) dY_{mis} \end{aligned} \tag{2}$$

First the rule of Total Probability is used to integrate $Y_{mis}$ out and, in a second step, again the Bayes Rule is employed.
Rubin (1987) summarises formula 2 in result 3.1: „Averaging the Complete-Data Posterior Distribution of (...) [$\eta$] over the Posterior Distribution of $Y_{mis}$ to obtain the Actual Posterior Distribution of (...) [$\eta$]“(Rubin, 1987, S.82). Rubin (1987) states that it „can be applied to simulate the actual posterior distribution of $Q$ (here: $\eta$) using repeated draws from the posterior distribution of the missing values.“(Rubin, 1987). As no additional information should be added to the estimation than the observed data and the prior knowledge expressed in the prior distribution, inferences should only be made on the basis of the observed data posterior. The derived link between observed-data posterior and complete-data posterior shows that this still holds if inference is made on the multiply imputed data.
Moreover, the integral 2 justifies Rubin's combining rules described later on in section 2.3, too.(Rubin, 1996) The integral is just approximated by the respective sum.

### 2.2.2 More than a Little Inaccuracy

Before turning to the sampling mechanism for the replacement of missing values, we need to confess a little inaccuracy. Correctly, Rubin (1987) adds in every formula mentioned above an indicator for inclusion to the survey $I$ and an indicator for response $R$. This is because of the statistical validity of the estimates.(Rubin, 1996) Rubin (1987) shows that it is not necessary to condition on the inclusion indicator under the assumption of an ignorable sampling mechanism. More importantly, he also shows the same for the indicator of non-response $R$ when the response mechanism is ignorable. The importance of this is due to the ignorance of most studies in respect to the possibility of non-response. Rubin (1987) states that most surveys just ignore non-response and, therefore, do not include conditioning on $R$. However, the assumption of ignorable non-response still remains even if $R$ is just ignored. To keep the necessary assumptions in mind, we note this inaccuracy.

However, the result implies more than just a simplification of notification. In ordinarily used statistical models the most employed assumption is the assumption of independent and identical distribution. For this very common and widely used assumption, Rubin (1987) states in his **Result 3.4**: „**The Equality of Completed-Data and Complete-Data Posterior Distributions When Using i.i.d. Models.**“(Rubin, 1987)

$$P(\eta|X, Y_{inc}, R_{inc}) \quad = \quad P(\eta|X, Y_{inc})$$

The subscipt $inc$ indicates whether the case was included in the study or not, $inc = (obs, mis)$. We show the proof of result 3.4, i.e. the equality of completed-data and complete-data posterior under the assumption of i.i.d. in the appendix.

Result 3.4 reasons that the same statistics may be computed on basis of the completed-data sets as would be done on the basis of the complete-data set. It is a further justification of the combining rules. Only if the same statistics can be computed, is it possible to estimate them by combining several completed data sets.

Later on, we assume ignorable non-response in the first instance. For all non-ignorable cases, this reformulation using Bayes' Rule is very useful:(Rubin, 1987)

$$\underbrace{P\left(\eta|X,Y_{inc},R_{inc}\right)}_{\text{completed-data posterior}} = \underbrace{P\left(\eta|X,Y_{inc}\right)}_{\text{complete-data posterior}} \underbrace{\frac{P\left(R_{inc}|X,Y_{inc},\eta\right)}{P\left(R_{inc}|X,Y_{inc}\right)}}_{\text{adjustment factor}}, \tag{3}$$

In the case of ignorable non-response, the adjustment factor becomes 1, which refers to the equality of both in that case (see Rubin's Result 3.3), otherwise the adjustment factor needs to be employed. According to Rubin (1978) „the trick is to tie the parameters for the different groups of units together so that the values we do see tell us something about the values we do not see.“(Rubin, 1978)

### 2.2.3 Getting Multiple Draws

However, we did not clarify yet how to actually get multiple replacements for the missing values. A short insight is given in this chapter. Let's start again with the underlying idea of drawing $m$-times from the posterior distribution $P\left(Y_{mis}|X,Y_{obs}\right)$ to get $m$ sets of imputations $Y^1_{mis},; Y^2_{mis}, \ldots, Y^m_{mis}$ and $m$ completed data sets $Y^1_{inc},\ Y^2_{inc}, \ldots, Y^m_{inc}$. (Rubin, 1987) Now we are engaged in creating one completed data set by finding sensible replacements for the missing values.
„[I]n order to insert sensible values for missing data we must rely on some model relating unobserved values to observed values“(Rubin, 1978) which is done when drawing from the posterior predictive which displays „the

sensitivity of inferences to reasonable choices of models“(Rubin, 1978). Therefore, the $m$ draws should be made from the posterior of $Y_{mis}$ which can be reformulated by the use of the rule of Bayes and the law of total probability (Rubin, 1987):

$$P\left(Y_{mis}|X,Y_{obs}\right)=\frac{P\left(X,Y\right)}{\int P\left(X,Y\right)dY_{mis}}$$

It is sufficient to consider $P\left(X,Y\right)$ to determine the model for the data.(Rubin, 1987, 1978) The aim is to formulate the prior knowledge of the data.(Rubin, 1978)
Then Rubin (1987) reformulates to get a connection to the parameter $\theta$ which is the parameter describing the distribution of $Y$ for using MCMC methods to simulate the complex distribution:

$$\begin{aligned}
P\left(X,Y\right) &= \int P\left(X,Y|\theta\right)P\left(\theta\right)d\theta \\
&= \int\left[\prod_{i=1}^{N} f\left(X_i,Y_i|\theta\right)\right]P\left(\theta\right)d\theta \\
&= \int\left[\prod_{i=1}^{N} f_{Y|X}\left(Y_i|X_i,\theta_{Y|X}\right)f_X\left(X_i|\theta_X\right)\right]P\left(\theta\right)d\theta
\end{aligned}$$

The last step using the rule of Bayes again does not seem necessary at the beginning, but as stated in Result 5.1 it greatly facilitates getting multiple imputations. It is not necessary to specify a model for the parameters. It is sufficient to model the conditional parameters and data models.(Rubin, 1987) For example, this can be done by simple regression.
After having specified the data model, we need to find a way to draw sensible values for the missing data. Starting again with the posterior predictive, which is the only distribution for drawing missing values, another facto-

risation is helpful. It can be found, among others, in Rubin (1987) and Zhang (2003):

$$\begin{aligned} P\left(Y_{mis}|Y_{obs},X\right) &= E_\theta[P(Y_{mis}|Y_{obs},X,\theta)] \\ &= \int P(Y_{mis}|Y_{obs},X,\theta)P(\theta|Y_{obs},X)d\theta \end{aligned} \tag{4}$$

The posterior predictive distribution now is factorised into:

1. conditional predictive distribution of $Y_{mis}$: $P(Y_{mis}|Y_{obs},X,\theta)$
2. observed data posterior distribution of $\theta$: $P(\theta|Y_{obs},X)$(Zhang, 2003)

To define the distribution of $Y$, we need to determine a prior $P(\theta)$ and a distribution for the data $P(Y_{obs}|\theta)$. Schafer (1997) also uses this expectation over the parameter $\theta$. He emphasises that a proper Bayesian multiple imputation uses the averaging over the observed data posterior to reflect the uncertainty of missing data given the parameters and the uncertainty of the model parameter themselves.(Schafer, 1997)
Using the Integral 4, Schafer (1999) describes the drawing scheme to get one set of imputations: first, a random draw of the parameter following the observed data posterior shall be done. Second, determined by the draw of the parameter, the missing values $Y_{mis}$ shall be drawn from the conditional predictive distribution.(Schafer, 1999)

According to Zhang (2003) and Schafer (1999) it is rarely possible to express the posterior predictive in closed form or to draw from it. However, he also states it is often easy to obtain the conditional predictive. This reasons the drawing mechanisms described later on. The drawing mechanisms are mostly MCMC-methods.(Schafer, 1999) In a lot of cases it is still quite difficult to simulate the draws from the observed data posterior. For this reason Rubin (1987) proposes a simplification in his Results 5.1 and 5.2. First, we should quote the

**Result 5.1 „The Imputation Task with Ignorable Nonresponse**“: „Given $\theta$, the $Y_{i,mis}$ are a posteriori independent with distribution depending on $\theta$ only through $\theta_{Y|X}$:“(Rubin, 1987)

$$P\left(Y_{mis}|X, Y_{obs}, \theta\right) = \prod_{ms} P\left(Y_{i,mis}|X_i, Y_{i,obs}, \theta_{Y|X}\right)$$

The proof is given in the appendix. In a lot of cases, it is much easier to compute $P\left(Y_{i,mis}|X_i, Y_{i,obs}, \theta_{Y|X}\right)$ than to compute $P\left(Y_{mis}|X, Y_{obs}, \theta\right)$. Just bear the possibility of linear regression in mind.
According to this result, it is possible to take $P\left(Y_{mis}|X, Y_{obs}, \theta_{Y|X}\right)$ into account instead of the unconditioned $P\left(Y_{mis}|X, Y_{obs}, \theta\right)$. Now, it all breaks down to specify $P\left(Y_{mis}|X, Y_{obs}, \theta_{Y|X}\right)$, which can often be simulated by MCMC methods. If the missings are ignorable, standard Bayes methods can be used.(Rubin, 1987) Zhang (2003) recommends simulating the observed data posterior distribution by a Gibbs-Sampling or to use the Predictive Model Method as alternative to the sampling.

In turn, Rubin (1987) exposes a simplification. He states this in **Result 5.2: „The Estimation Task with Ignorable Nonresponse When $\theta_{Y|X}$ and $\theta_X$ Are a Priori Independent**“:
„Suppose $\theta_{Y|X}$ and $\theta_X$ are a priori independent

$$Pr\left(\theta\right) = Pr\left(\theta_{Y|X}\right) Pr\left(\theta_X\right)$$

Then they are a posteriori independent; moreover, the posterior distribution of $\theta_{Y|X}$ involves only (a) the specifications $f_{Y|X}\left(\cdot|\cdot\right)$ and $Pr\left(\theta_{Y|X}\right)$ and (b) data from units with some $Y_{ij}$ observed.“(Rubin, 1987). For the reason of independence, it is possible to use only $P\left(\theta_{Y|X}|X, Y_{obs}\right)$, which simplifies computation:[1]

$$P\left(\theta_{Y|X}|X, Y_{obs}\right) \tag{5}$$

$$= \frac{\left[\prod_{i=1}^{N} \int f_{Y|X}\left(Y_i|X_i, \theta_{Y|X}\right) dY_{i,nob}\right] P\left(\theta_{Y|X}\right)}{\int \left[\prod_{i=1}^{N} \int f_{Y|X}\left(Y_i|X_i, \theta_{Y|X}\right) dY_{i,nob}\right] P\left(\theta_{Y|X}\right) d\theta_{Y|X}} \tag{6}$$

According to Rubin (1987), it is sufficient to multiply over the units with observed values. A further simplification is the case of univariate $Y_i$ as Rubin (1987) states in **Result 5.3:** „**The Estimation Task with Ignorable Nonresponse, $\theta_{Y|X}$ and $\theta_X$ a Priori Independent, and Univariate $Y_i$**
If $\theta_{Y|X}$ and $\theta_X$ are a priori independent and $Y_i$ is univariate so that the respondents have $Y_i$ and the non-respondents are missing $Y_i$, the posterior distribution of $\theta_{Y|X}$ involves only the respondents.“ For this reason, the conditional posterior becomes (Rubin, 1987):

$$P\left(\theta_{Y|X}|X, Y_{obs}\right) = \frac{\left[\prod_{obs} f_{Y|X}\left(Y_i|X_i, \theta_{Y|X}\right)\right] P\left(\theta_{Y|X}\right)}{\int \left[\prod_{obs} f_{Y|X}\left(Y_i|X_i, \theta_{Y|X}\right)\right] P\left(\theta_{Y|X}\right) d\theta_{Y|X}} \tag{7}$$

To subsume the last section, these steps were discussed:

1. $Y_{i,mis}$ are a posteriori independent.
2. The distribution of $Y_{i,mis}$ only depends on $\theta$ through $\theta_{Y|X}$.
3. $P\left(\theta_{Y|X}\right)$ and $P\left(\theta_X\right)$ are a priori independent.

In this case we only have to specify $f_{Y|X}$, $P\left(Y|X, \theta_{Y|X}\right)$ and data from observed units. In the case of univariate $Y$ we only need the data of the respondents.

[1] Jackman (2009) shows the same for regression models which is the main reason why Rubin's result makes computations easier.

### 2.2.4 The Gibbs-Sampler

The basic rule to get multiple draws for multiple imputation is stated in Conclusion 4.1 by Rubin (1987): „If imputations are drawn to approximate repetitions from a Bayesian posterior distribution of $Y_{mis}$ under the posited response mechanism and an appropriate model for the data, then in large samples the imputation model is proper.“(Rubin, 1987) Therefore, the drawing scheme should be based on a Bayesian idea to preserve correct variance estimation.(Rubin, 1987) Zhang (2003) describes the sampling idea to create multiple draws. First, draw $\widetilde{\theta}$ from its observed data posterior distribution given in equation 4. Then draw one set of $Y_{mis}$ from $P\left(Y_{mis}|Y_{obs}, X, \widetilde{\theta}\right)$. In order to create $m$ sets of missing data, repeat this $m$ times independently.(Zhang, 2003)
Under the assumption of normality the imputation problem can be solved by using a linear regression where the missing values are the predictions. The parameters of the regression are the distributional parameters being drawn to assess the model uncertainty.(Zhang, 2003)
Define $X$ as the data matrix with the variables observed and assume $Y \sim N\left(\mu, \sigma^2\right)$ and non-informative priors, then according to Zhang (2003) the observed data posterior distributions are:

- $\beta|Y_{obs}, \sigma^2 \sim N\left(\widehat{\beta}, \sigma^2\left(X'_{obs}X_{obs}\right)'\right)$
- $\sigma^2|Y_{obs} \sim \widehat{\epsilon}'\hat{\epsilon}\chi^2_{n-p}$

with the MLE estimator $\widehat{\beta} = \left(X'X\right)^{-1} X'Y_{obs}$ and the residual vector $\hat{\epsilon} = Y_{obs} - X\widehat{\beta}$. First, the $\sigma^2$ are drawn and, afterwards, $\beta$s are drawn from their observed data posterior distributions. Each set of randomly drawn parameters defines a regression leading to a set of predictions for the missing values.(Zhang, 2003)

If the case cannot be solved that easily, we need a solution based on Markov chains. In general, to conduct a Bayesian estimation, complex integrals have to be computed, but, for the reason of complexity a closed solution does not exist. Several techniques to find a solution have been introduced. An overview is given by Carlin and Louis (2008).
The most simple way to get such an approximate solution is the Gibbs sampling. It bases on the Monte Carlo integration. This is an iterative method where a Markov chain is generated. The idea underlying the Gibbs-Sampler is that we can simulate a Markov chain by drawing from the $p$ conditional posteriors instead of the joint posterior.(Albert, 2009) The output of the draws of the Markov chain corresponds to draws from the required density.(Carlin and Louis, 2008) Robert and Casella (2004) conclude: „The Gibbs Sampler is a technique for generating random variables from a (marginal) distribution indirectly, without having to calculate the density.“ Therefore, the Gibbs sampler is a Monte-Carlo-Markov-Chain-Method, where values of the parameter are sequentially drawn from a Markov Chain with a stationary distribution that corresponds to the density that cannot be computed.(Carlin and Louis, 2008) For generating draws, the conditional posteriors are used instead of the joint posterior.(Albert, 2009) The Gibbs-Sampler is an approximation of the joint posterior.(Hoff, 2009)
Schafer (1997, 1999) stresses the similarity of multiple imputation and data augmentation, because the most simple way to sample the $Y_{mis}$ from the factorisation given in equation 4 is by Gibbs Sampling. The stationary distribution of the sampling scheme is the target distribution $P\left(Y_{mis}|Y_{obs}\right)$. (Schafer, 1997, 1999)
We just want to draw directly from the posterior predictive, which is often not possible. Using Monte-Carlo-Integration we get:

$$P\left(y_{mis}|y_{obs}\right) = \int P\left(y_{mis}|\theta, y_{obs}\right) p\left(\theta|y_{obs}\right) d\theta \tag{8}$$

Now, we can approximate the posterior expectation by the mean of values drawn from the distributions inside the integral. As $\theta^1, \dots, \theta^S \sim P\left(\theta | y_{obs}\right)$ are i.i.d. draws, $P\left(\tilde{y} | y_{obs}\right)$ can be approximated by $\frac{1}{S}\sum_{s=1}^{S} P\left(y_{mis} | \theta^s\right)$. The draws of $P\left(y_{mis} | \theta^s\right)$ are generated by the following procedure:(Hoff, 2009)

$$
\begin{aligned}
\text{draw } \theta^{(1)} \sim P\left(\theta | y_{obs}\right) \quad &\rightarrow \quad \text{draw } \widetilde{y}^{(1)} \sim P\left(y_{mis} | \theta^{(1)}\right) \\
\text{draw } \theta^{(2)} \sim P\left(\theta | y_{obs}\right) \quad &\rightarrow \quad \text{draw } \widetilde{y}^{(2)} \sim P\left(y_{mis} | \theta^{(2)}\right) \\
&\vdots \\
\text{draw } \theta^{(S)} \sim P\left(\theta | y_{obs}\right) \quad &\rightarrow \quad \text{draw } \widetilde{y}^{(S)} \sim P\left(y_{mis} | \theta^{(S)}\right)
\end{aligned} \tag{9}
$$

The sequence $\left\{\left(\theta, y_{mis}\right)^{(1)}, \dots, \left(\theta, y_{mis}\right)^{(S)}\right\}$ consists of $S$ independent draws from the joint distribution $\left(\theta, y_{mis}\right)$ and the sequence $\left\{y_{mis}^{(1)}, \dots, y_{mis}^{(S)}\right\}$ consists of independent draws from the marginal posterior of $y_{mis}$. (Hoff, 2009)

To assure that draws are independent from each other, there are two possibilities: First, only every $k$th iteration should be taken, where $k$ is large enough to guarantee independence. Or , second, the other possibility is to run $m$ independent chains of length $k$ and take the last simulated value.(Schafer, 1997, 1999) The disadvantage of the latter approach is the computational burden to run that many chains. The disadvantage of the former procedure is that the chain can get stuck in a small subspace leading to realisations that are not from the whole parameter space.(Robert and Casella, 2004, Carlin and Louis, 2008)

### 2.2.5 Monotone Distinct Structure

A monotone distinct missing data structure helps to simplify multiple imputation. According to Rubin (1987) a monotone pattern of missing data is given if data can be sorted in the way that all cases of $Y_{[j]}$ are observed in $Y_{[i]}$ for all $i < j$. Rubin (1987) defines a monotone patter as: $Y_{[i]}$ is at least that observed as $Y_{[j]}$ is observed. Data having a monotone pattern of missingness look like a staircase.
A formal definition of a monotone missing data pattern is also given in Rubin (1987): Let $obs\,[j] = \left\{i|I_{ij}R_{ij} = 1\right\}$ be the set of units where the variable $Y_j$ is observed. Then a monotone pattern of missingness can be formally defined by:

$$obs\,[1] \supseteq obs\,[2] \supseteq \ldots obs\,[p] \tag{10}$$

The assumption of a distinct structure refers to the parameter space as the assumption of a monotone structure aims at the data. Rubin (1987) defines a distinct structure for the parameters $\theta_1, \theta_2, \ldots, \theta_p$ using the factorisation:

$$f\,(Y_i|X_i, \theta) = f_1\,(Y_{i1}|X_i, \theta_1)\, f_2\,(Y_{i2}|X_i, \theta_2) \ldots f_p\left(Y_{ip}|X_i, \theta_p\right) \tag{11}$$

then $\theta_1, \theta_2, \ldots, \theta_p$ are distinct, if they are a priori independent: $P\,(\theta) = \prod_{j=1}^{p} P\left(\theta_j\right)$. (Rubin, 1987) Reading Rubin (1974) clarifies this assumption: only if a distinct structure is assumed, a factorisation is possible that allows an estimation even on the ground of parts of the data. Following, we write $f_{ij} = f_j\left(Y_{ij}|X_i, \theta_j\right)$.

Rubin (1987) points out in two results that the estimation and the imputation tasks are much easier under the assumption of a monotone distinct pattern: „**Result 5.4 The Estimation Task with a Monotone-Distinct Structure**

Suppose the missingness-modelling structure is monotone-distinct. Then the estimation task is equivalent to a series of $p$ independent estimation tasks, each with univariate $Y_i$: the $j$th task estimates the conditional distribution of $Y_{[j]}$ given the more observed variables $\left(X, Y_{[1]}, \dots, Y_{[j-1]}\right)$ using the sets of units with $Y_{[j]}$ observed, $obs\,[j]$. Explicitly, the claim is that $\theta_1, \dots, \theta_p$ are a posteriori independent with

$$Pr\left(\theta_j | X, Y_{obs}\right) = \frac{\prod_{i \in obs[j]} f_{ij} Pr\left(\theta_j\right)}{\int \prod_{i \in obs[j]} f_{ij} Pr\left(\theta_j\right) d\theta_j} \tag{12}$$

“(Rubin, 1987)

If the pattern of missing data is monotone distinct, then we are allowed to formulate the posterior distributions independently.(Rubin, 1987) Moreover, if the pattern is monotone distinct it is also admissible to impute the missing data independently. Rubin (1987) states this in Result 5.5:
„**Result 5.5 The Imputation Task with a Monotone-Distinct Structure**
Suppose that the missingness-modelling structure is monotone-distinct. Then the imputation task is equivalent to a sequence of $p$ independent imputation tasks, each with univariate $Y_i$: the $j$th task independently imputes the missing values of $Y_{[j]}$ using their conditional distributions given $\theta$ and the observed values $X_i, Y_{i1}, \dots, Y_{ip-1}$ Explicitly, the claim is that the posterior distribution of $Y_{mis}$ given $\theta$ is

$$Pr\left(Y_{mis} | X, Y_{obs}, \theta\right) = \prod_{j=1}^{p} \prod_{i \in mis[j]} f_{ij} \tag{13}$$

where $mis\,[j] = \left\{i | I_{ij} = 1 \text{ and } R_{ij} = 0\right\}$ =the units missing variable $Y_{[j]}$ and (13) is the product of $p$ conditional distributions, each of which is formally equivalent to (...) the posterior distribution of $Y_{mis}$ given $\theta_{Y|X}$ with univariate $Y_i$.“(Rubin, 1987)

Since a monotone distinct pattern of missingness helps to simplify multiple imputation, Rubin (1987) proposes to discard data destroying monotonicity, if it is only a few data. Another way is to create blocks with monotone patterns of missingness.(Rubin, 1987) A hint of how to create these blocks can be found in Rubin (1974) even though in this text a possibility is described to estimate the parameters, and not a way of multiple imputation.

## 2.3 Combining Multiple Values

Each of the $m$ created data sets without missing values can be analysed with standard statistical methods. As methods are not determined, a lot of different methods can be used.(Rubin, 1987, Schafer, 1997, Rubin, 1996) How to combine and compute statistics on the basis of multiple imputed data is written in the following section.

We have $m$ completed data sets. For every completed data set it is possible to compute a complete data statistic. This delivers $m$ estimates $\hat{\eta}*_1, \hat{\eta}*_2, \cdots, \hat{\eta}*_m$ and $m$ estimated variances $\widehat{U}*_1, \widehat{U}*_2, \cdots, \widehat{U}*_m$(Rubin, 1987). We need to combine the estimates and the estimated variances to get one estimate and a corresponding estimated variance that accounts for the uncertainty brought in by missing values.

### 2.3.1 Multiple Imputation Estimation

Starting again with the integral described in formula 2: Rubin (1987) states here that the posterior distribution of the missing data can be simulated using the posterior distribution of the observed values. Accordingly, an estimator for the parameter of interest bases on the completed-data posterior. (Rubin, 1987) Interpreting the integral in 2 as an expectation over the missing values $Y_{mis}$, it becomes clear that a sensible approximation is the average. Doing so, Little and Rubin (2002) write: „Multiple imputation effectively approximates the integral (...) over the missing values as the average“(Little and Rubin, 2002):

$$P\left(\eta|Y_{obs},X\right)\approx\frac{1}{m}\sum_{l=1}^{m}P\left(\eta|Y_{mis}^{l},Y_{obs},X\right)$$

Ordinarily, Bayesian estimators are derived by taking the mode of the posterior distribution. Here this course of action is complicated. Because of the integral-relationship between posterior distribution of observed data and the completed data posterior distribution using expectations, the integrals can be interchanged. Using the mode which is the ordinary Bayesian estimator, this procedure is not possible. Rubin (1987) uses the expectation instead. Little and Rubin (2002) conclude that the mean and the variance are interpreted as „adequate summaries“(Little and Rubin, 2002) in this case.
The expectation is a Bayesian estimator on the background of decision theory. It can be shown (Rüger, 1999) that the expectation is the Bayes estimator using a quadratic loss function. In this case Rubin (1987) shows in **Result 3.2**:

$$\begin{aligned}E\left(\eta|Y_{obs},X\right) &= E\left(E\left(\eta|Y_{obs},Y_{mis},X\right)|Y_{obs},X\right)\\ &= E\left(\hat{\eta}|Y_{obs},X\right) \qquad (14)\end{aligned}$$

as the Bayesian estimator for the parameter of interest is: $\hat{\eta}=E\left(\eta|Y_{obs},Y_{mis},X\right)$ (Rubin, 1987). Equation 14 shows that the expectation of the interesting parameter equals the expectation of the estimator which is the same as the mean of the completed-data posterior. Schafer (1997) appoints this as one assumption necessary for inference with multiply imputed data. Zhang (2003) uses another argument: the mean as an estimator is derived by the moment summaries. The formula itself can be so interpreted

that in the long run an estimation via the completed-data posterior, e.g. the imputed data sets, is feasible. Similarly, Schafer (1999) argues that the mean over the $m$ completed data sets is the approximate expectation of the posterior.
Since the limiting value is (Rubin, 1987)

$$\bar{\eta}_{\infty} = lim_{m\to\infty} \sum_{l=1}^{m} \frac{\hat{\eta}_l}{m} = E\left(\hat{\eta}|Y_{obs}, X\right)$$

the first combining rule can be constituted. The $m$ different estimations of the parameter (for every completed data set one) can be combined according to the rules described in Rubin (1987), Rubin (1988), Schafer and Olsen (1998):

$$\bar{\eta}_m = \frac{1}{m} \sum_{l=1}^{m} \hat{\eta}_l \tag{15}$$

The integral is approximated by the average.(Little and Rubin, 2002) Rubin (1996) generalises that this is equivalent to equating the actual posterior distribution of $\eta$ and the average over the complete data posterior distribution of $\eta$.

### 2.3.2 Multiple Imputation Variance Estimation

Similar to the expectation, a combining rule for the variance is derived in **Result 3.2**: (Rubin, 1987, Zhang, 2003, Rubin, 1996)

$$\begin{aligned} Var\left(\eta|Y_{obs}, X\right) &= Var\left(E\left(\eta|Y_{obs}, Y_{mis}, X\right)|Y_{obs}, X\right) \\ &\quad + E\left(Var\left(\eta|Y_{obs}, Y_{mis}, X\right)|Y_{obs}, X\right) \\ &= Var\left(\hat{\eta}|Y_{obs}, X\right) + E\left(U|Y_{obs}, X\right) \end{aligned} \tag{16}$$

The first part of the sum is the variance between the expectations of the estimations based on the imputed data sets. The second part is the expectation of the variances. Using this reformulation and the following limiting values given in Rubin (1987), the combining rule for the variances is acquired. The limiting values are: (Rubin, 1987)

$$\overline{U}_{\infty} = lim_{m \to \infty} \sum_{l=1}^{m} \frac{U_l}{m} = E\left(U | Y_{obs}, X\right) \tag{17}$$

and

$$B_{\infty} = lim_{m \to \infty} \sum_{l=1}^{m} \frac{\left(\hat{\eta}_l - \overline{\eta}\right)' \left(\hat{\eta}_l - \overline{\eta}\right)}{m} = Var\left(\hat{\eta} | Y_{obs}\right) \tag{18}$$

Therefore, the second part of the sum above can be estimated by the mean of the posterior variance: (Rubin, 1987, Schafer and Olsen, 1998, Schafer, 1999)

$$\overline{U}_m = \frac{1}{m} \sum_{l=1}^{m} U_l \tag{19}$$

whereas $U_l$ is the variance of the variable in the $l$th completed data set (Rubin, 1987) and $U$ can be interpreted as the mean of the variances of the data sets. Again the integral is approximated by an average. The second part, the variance between the completed data sets, can be evaluated by: (Rubin, 1987)

$$B_m = \frac{1}{m-1} \sum_{l=1}^{m} \left(\hat{\eta}_l - \overline{\eta}_m\right)' \left(\hat{\eta}_l - \overline{\eta}_m\right) \tag{20}$$

in the multivariate case. Similarly, the between imputation variance in the univariate case is: (Rubin, 1988)

$$B = \sum_{l=1}^{m} \frac{\left(\eta_l - \overline{\eta}\right)^2}{(M-1)} \tag{21}$$

The formula for the within imputation variance is the same in the univariate and the multivariate case except that $U_l$ is a scalar in the former and a vector in the latter case. Combined together, it leads to the total variance of the statistic: (Rubin, 1987, Rubin, 1988, Little and Rubin, 2002)

$$T_\infty = \overline{U}_\infty + B_\infty \tag{22}$$

For small $m$, an improved approximation of the variance is (Little and Rubin, 2002):

$$V_m = \overline{U}_m + \left(1 + \frac{1}{m}\right) B_m \tag{23}$$

There is additional uncertainty brought in by the missing values and this is reflected by an increased variance. Schafer (1999) argues the same when justifying the inference for multiple imputed data. Schafer and Olsen (1998) interpret it as a correction factor for the simulation error.
According to Zhang (2003) $\left(1 + m^{-1}\right)$ accounts for the additional variance, because of the number of imputations being finite. The relative increase of variance that can be reasoned by missing data is:(Schafer, 1999, Schafer and Olsen, 1998)

$$r = \frac{\left(1 + m^{-1}\right) B}{\overline{U}}$$

An estimator for the amount of missing information can be derived from the first derivation of the log posterior of the parameter $\eta$: (Schafer, 1999, Schafer and Olsen, 1998)

$$\begin{aligned}\hat{\lambda} &= \frac{\frac{r+2}{\nu+3}}{r+1} \\ &= \frac{\left(\overline{U} - (\nu+1)(\nu+3)\,T^{-1}\right)}{\overline{U}}\end{aligned}$$

## 2.4 Requirements

Schafer and Olsen (1998) mention three assumptions made by multiple imputation: First, a probability model describing the complete data is necessary. Although multiple imputation is not very sensitive for data model misspecifications, Schafer and Olsen (1998) stress that the data model should be chosen carefully. We need to consider all variables that are highly correlated with the variables having missing values. In this case it is very easy to predict values for the missing data.(Rubin, 1978) Moreover, as subsequent analysis should not be restricted, as much variables as possible should be considered.(Schafer and Olsen, 1998) All important associations between the variables in the data set should be included. If they are not preserved, inference will be biased.(Patrician, 2002) For this reason, data imputers tend to include as many variables as possible, ruling out the case of an omitted variable bias.

Second, the prior distribution of the model parameter must be specified. This is reasoned by the underlying Bayesian theory. According to Schafer and Olsen (1998) a non-informative prior works well in most cases.

Third, the mechanism describing the missingness has to be ignorable. The easiest case is missing at random. Because of the MAR assumption it is possible to exploit the relationships between the observed data to estimate the missing values.(Schafer and Olsen, 1998, Patrician, 2002) According to Rubin (1978) it is possible to model the missingness with the help of variables highly correlated with the indicator of missingness $I$ to make it ignorable.(Rubin, 1978)

## 2.5 Advantages and Disadvantages

Let us start with the disadvantages and then go over to the manifold advantages of multiple imputation as a missing data technique.
Rubin (1987) only refers to the bigger efforts that are needed for conducting multiple imputation and for handling multiply imputed data. Of course the efforts to impute and to analyse data need to be increased as every step has to be repeated $m$-times.(Rubin, 1987) Today the argument that more memory space is needed is not convincing any more. However another problem arises through the repeated imputation and the ambiguity of multiple imputation. Through multiple imputation, some kind of random noise is brought into the analysis. Finally, it is difficult to understand and difficult to be accepted by unexperienced users.(Patrician, 2002)
More importantly, the modelling of non-response mechanisms and of the imputation scheme has to be done with great care: The validity of the analysis depends on correctly capturing the missingness mechanism.(Schafer, 1999) Moreover, the imputation model and analysis model must be compatible: If the imputation model is less restrictive then the analysis loses efficiency. If the imputation model is too restrictive and the model is not plausible, the inference is too conservative. Only in the case of correct assumptions, the analysis is efficient and unbiased. The major problem is that all variables which are not considered in the imputation model will not be correlated in every analysis.(Schafer, 1999)

However, multiple imputation does not suffer very serious disadvantages, indeed it has some very compelling advantages. First and most importantly for our application, suitable statistical methods are not determined. All complete data methods are possible. For this reason, any statistical software can be used for multiple imputed data sets.(Rubin, 1987, Rubin, 1988, Patrician, 2002, Schafer and Olsen, 1998) Moreover, data users may be very different with regard to their statistical knowledge and abilities. Each person can use those methods that he knows.(Rubin, 1996) Multiple imputation does not restrict the set of applicants, the application, the statistical methods or the statistical software. Moreover, as the missing dataproblem is handled once, consistency among different persons is assured.(Rubin, 1988) The knowledge of the data collector which is not applicable to the data analyst (e.g. for the reasons of confidentiality) can be used.(Rubin, 1996, Rubin, 1987, Patrician, 2002) This improves the quality of the imputation as a lot of information can be exploited.

Multiple imputation also has advantages in respect to statistical quality measures. First, it increases the efficiency of the estimation.(Rubin, 1987) Multiple imputation is highly efficient and shows in many cases excellent results with only a few imputed data sets.(Schafer and Olsen, 1998)

The problem of single imputation is that model uncertainty is not reflected.(Rubin, 1988) Through the multiple data sets, some kind of random variation is included in the imputation.(Patrician, 2002) Two modes of uncertainty can be considered: Sampling uncertainty and the reasons of missingness by choosing an appropriate model of non-response.(Rubin, 1987) The standard errors, p-values and so on, are valid.(Schafer and Olsen, 1998) The differences between the data sets allow valid inferential statistics.(Rubin, 1987)

# 3 Kriging

As noted above, multiple imputation needs to consider all relevant correlations. This also includes spatial relationships. For this reason, we need to look for a method that can be implemented in MI and accounts for spatial correlations. One such procedure is kriging.
Our application is the estimation of regional price indices. Obviously, the price level of one region partially depends on the price level of its neighbours, as there are local price alignments. For this reason, kriging seems to be a promising method for capturing the relevant information required to predict regional price levels.

Kriging itself is an optimal prediction using spatial correlations. Information of neighbours is considered in the estimation by a spatial correlation function generating an interpolation over a spatial random process, which is assumed to be stationary[2]. Doing so, the correlation depends on the distance between the region of interest and the other observation points.(Cressie, 1990) Therewith, the aim of kriging is not the estimation, but rather the prediction of values of unobserved regions. A regression type estimation extended by a spatial random process is done creating a linear prediction.(Cressie, 1990) The spatial correlation is represented in the regression by a parametric correlation function, which is a Gaussian random field.(Fahrmeir et al., 2007) As the model is just a regression it seems very straightforward to implement kriging into multiple imputation.

[2]Voltz and Webster (1990) note that the stationary assumption just needs to be local as only close neighbours have a significant influence.

## 3.1 Introduction into One-dimensional Kriging

Here, we describe the spatial modelling of one-dimensional kriging to get started. The one-dimensional case is more familiar and much easier than the two-dimensional case which we actually need for predicting regional price levels.

The model of one-dimensional kriging is:(Fahrmeir et al., 2007)

$$y_s = x_s'\beta + \epsilon_s;\ s \in \mathbb{R} \tag{24}$$

$\mu = x_t'\beta$ is the spatial trend. In the one-dimensional case, it may be a temporal trend, too. We assume that $\epsilon_t$ is multivariate normal with the expectation $E(\epsilon) = 0$ and the covariance matrix $Cov(\epsilon) = \sigma^2 I + \tau^2 R$, because of the underlying stochastic process which models the spatial or temporal dependencies.(Fahrmeir et al., 2007)
The underlying stochastic process is assumed to be Gaussian. It is the most usual assumption guaranteeing that the common distribution of observed and unobserved values is also normal. Moreover the expectation is linear in $y$ in this case.(Cressie, 1993) Model 24 borrows from time series analysis, modelling the correlation implicitly in the variance-covariance matrix. Here, we prefer a mixed model representation where the correlation is explicitly represented in the model. This can be done by adding a random process to the model, which we do later in this section.
Through the covariance we assume that errors are correlated in a typical way. (Fahrmeir et al., 2007, Cressie, 1990) The spatial correlation is expressed in $\tau^2 R$. $R$ itself is a correlation matrix consisting of the correlation function $\rho(\cdot,\cdot)$ which models how the distances between two observed points influence the correlation:(Fahrmeir et al., 2007)

$$R(s_i, s_j) = Corr\left(s_i, s_j\right) = \rho\left(s_i, s_j\right) \tag{25}$$

The covariance consists of two parts:

1. uncorrelated part: $\sigma^2 I$

2. correlated part modelling the spatial dependencies: $\tau^2 R$

The covariance is parameterised by three parameters: $\sigma^2$, $\tau^2$, and $\rho$. If all three are known, it is possible to estimate the parameter vector of fixed effects $\beta$ by GLS:(Fahrmeir et al., 2007)

$$\widehat{\beta} = \left(X'V^{-1}X\right)^{-1} X'V^{-1}y$$

where $X$ is a $n \times p$ design-matrix and $V$ defines the $n \times n$ covariance-matrix. $\beta$ and $y$ are as usually vectors. The parameter of the covariance can be estimated by appropriate methods as REML.(Fahrmeir et al., 2007)

## 3.2 The Method of Kriging

Kriging is a method that was developed for mining. In mining the aim is to find profit-yielding areas in a block. Before D. G. Krige, who worked at the Witwaterstand goldmines in South Africa (Schabenberger and Gotway, 2005), it was common practice to use the „sample mean of nearby core-sample assays to estimate the average grade in a prospective mining block. Those estimates were then used to mine selectively.“ (Cressie, 1990) Krige criticised this method defining a new approach that can now be named as a „two dimensional moving average (...) with predetermined radius“ (Cressie, 1990). Krige's contribution to the method kriging is to propose that covariances are used as weights in the BLUE estimation. (Cressie, 1990) A third component, defining the spatial BLUP characterising kriging was added by Matheron.(Cressie, 1990)[3] Matheron named the developed method kriging to honour D.G. Krige.(Schabenberger and Gotway, 2005) The covariances can be described by a spatial stochastic process. To sum up, Cressie (1990) enumerates three components characterising kriging:

1. covariances are used as weights,
2. the BLUE estimator $\hat{\mu}$ is used, and,
3. covariances are determined by a spatial stochastic process.(Cressie, 1990)

The second point is equivalent to assuming a normal distribution of the error term as this is the definition of the BLUE estimator.

### 3.2.1 Continuous Kriging

The model representation of kriging is a typical regression model emphasising the first three components of kriging:(Schabenberger and Gotway, 2005)

$$y(s) = x(s)'\beta + \gamma(s) + \epsilon(s) \tag{26}$$

where the covariables $x$ parameterise the spatial trend by $\mu = x'\beta$, also called the drift component (Kitanidis, 1997), $\gamma(s)$ is assumed in the easiest case to be a stationary Gaussian random field and the normal error $\epsilon(s)$ where $\epsilon(s) \sim N(0, \sigma^2)$ is independent from $\gamma(s)$.(Fahrmeir et al., 2007, Schabenberger and Gotway, 2005, Kitanidis, 1997) The model formulation differs from the model given above in making the underlying spatial process explicit.

[3] In meteorology the same method was developed by Gandin at the same time. Here it is called objective analysis.(Schabenberger and Gotway, 2005)

The spatial trend component can be estimated by a BLUE where the covariances determined by the assumed random field are the weights. The deterministic component $x(s)'\beta$ is also called „large scale component of spatial variability that can be represented with reasonable assurance as a determinist function.“ (Kitanidis, 1997) However in kriging the estimation of parameters is only the means to the end of predicting $y(s)$.(Schabenberger and Gotway, 2005) The prediction is described below.

The spatial dependent variation in $y(s)$ is characterised by a Gaussian random field.[4] (Stein, 1999) $\left\{\gamma(s), s \in \mathbb{R}^2\right\}$ for which we assume stationarity, meaning that:

- expectation function: $E(\gamma(s)) = \mu(s) = 0$
- variance function: $Var(\gamma(s)) = \tau^2(s) = \tau^2$
- correlation function: $Corr(\gamma(s), \gamma(t)) = \rho(s,t) = \rho(h)$, where $h$ is the distance between $s$ and $t$. (Fahrmeir et al., 2007)

The stationarity assumption defines the expectation of the random field and the variance of the random field as spatially constant quantities.(Fahrmeir et al., 2007)

In matrix notation we can write the model representation as (Fahrmeir et al., 2007)

$$y = X\beta + Z\gamma + \epsilon \tag{27}$$

[4]There are are lot of other possibilities to characterise the spatial random process describing the spatial dependencies.

where the vector $\gamma = (\gamma(s_{(1)}), \cdots, \gamma(s_{(d)}))'$ is comprised of the values of the stationary Gaussian random process at the $d$ spatial observation points $s_{(1)}, \dots, s_{(d)}$. $Z$ is the $(n \times T)$ matrix of incidence

$$Z[i,j] = \begin{cases} 1, & \text{if } y_i \text{ was observed at location } s_{(j)} \\ 0, & \text{else} \end{cases} .$$

(Fahrmeir et al., 2007)
The covariance matrix of $y_i$ can easily be derived:

$$\begin{aligned} Cov(y) &= Var\,(X\beta + Z\gamma + \epsilon) \\ &= Var(Z\gamma) + Var(\epsilon) = \tau^2 ZRZ' + \sigma^2 I \end{aligned}$$

where $R$ is the $T \times T$ correlation-matrix of the spatial effects, and not the covariance-matrix as Fahrmeir et al. (2007) define $R$.

$$\begin{aligned} R &= \left(Corr(\gamma(s_{(i)}), \gamma(s_{(j)}))\right) = \left(\rho(s_{(i)}, s_{(j)})\right) \\ &= \begin{pmatrix} 1 & \rho(s_1,s_2) & \dots & \rho(s_1,s_T) \\ \rho(s_1,s_2) & 1 & \rho(s_2,s_3) & \dots & \rho(s_2,s_T) \\ & & \ddots & & \\ \vdots & & & & \vdots \\ \rho(s_1,s_T) & & \dots & & 1 \end{pmatrix} \end{aligned}$$

transferring the spatial correlations from $\gamma$ to $y$.(Fahrmeir et al., 2007)

## 3.3 Prediction

The initial aim of kriging is an optimal prediction of unobserved values while using the information of observed points.(Fahrmeir et al., 2007, Stein, 1999, Schabenberger and Gotway, 2005) This aim even can be used as a definition of kriging, as is done by Stein (1999): „Best linear unbiased prediction is frequently used in spatial statistics where it is commonly called (...) kriging.“ (Stein, 1999) Almost the same definition can be found in Cressie (1993), where the author contrasts kriging from spatial prediction as the former is a mean squared error optimal spatial linear prediction and the latter is just a spatial prediction.(Cressie, 1993)
We can distinguish two cases: First, the case of the known mean $\mu$, which is called simple kriging, and, second, the case of the unknown $\mu$, called ordinary kriging and universal kriging.(Cressie, 1990, 1993) The definition of Stein (1999) makes the differences clearer: Simple kriging assumes a mean of 0, ordinary kriging assumes a constant, but unknown mean, and universal kriging assumes a „more general“ (Stein, 1999) mean. According to Cressie (1993), the difference between universal and ordinary kriging is: in ordinary kriging the mean of the spatially correlated variable is unknown, but a constant, and is implemented in the optimal prediction scheme via the assumption that predictions weights sum up to one. In universal kriging it is assumed that the mean can be estimated by a function of some other variables. The function of some other variables is the known linear predictor $X\beta$.(Cressie, 1993)
Here we only discuss predictions that can be made under the assumption of stationarity. The literature on kriging discusses two moment based predictors for the case of non-stationarity: the variogram and the covariogram predictors. A discussion of these two can be found among others in Schabenberger and Gotway (2005) and Cressie (1990).

To simplify notation we write for the vector of predictions

$$\widehat{y}_0' = \begin{pmatrix} \hat{y}(s_{01}) & \hat{y}(s_{02}) & \dots & \hat{y}(s_{0n_0}) \end{pmatrix},$$

where $s_{0i}$ denotes that the region is not observed. The observed values $y' = \begin{pmatrix} y(s_1) & y(s_2) & \cdots & y(s_n) \end{pmatrix}$ have the variance-covariance matrix $Var(y)$. The covariances of unobserved and observed values are denoted in the $(n \times n_0)$-matrix

$$Cov(y, y_0) = \begin{pmatrix} Cov(y, y(s_{01})) & Cov(y, y(s_{02})) & \dots & Cov(y, y(s_{0n_0})) \end{pmatrix} =$$

$$\begin{pmatrix} Cov(y(s_1), y(s_{01})) & Cov(y(s_1), y(s_{02})) & \dots & Cov(y(s_1), y(s_{0n_0})) \\ Cov(y(s_2), y(s_{01})) & Cov(y(s_2), y(s_{02})) & \dots & Cov(y(s_2), y(s_{0n_0})) \\ \vdots & \vdots & & \vdots \\ Cov(y(s_n), y(s_{01})) & Cov(y(s_n), y(s_{02})) & \dots & Cov(y(s_n), y(s_{0n_0})) \end{pmatrix}$$

If we write just $y(s_0)$, we refer to any unobserved region, which simplifies the subscription. Additionally, we define

$$\mu' = \begin{pmatrix} \mu(s_1) & \mu(s_2) & \dots & \mu(s_n) \end{pmatrix}$$

as the vector of expectations of the observed regions.

### 3.3.1 Simple Kriging Predictor

Let us start with the easiest case: the simplification of a fixed and known mean defining the simple kriging predictor. For the reason of simplifying readability, the derivations of this and the next sections can be found in the appendix.

As already mentioned, the aim is to predict the unobserved points using the information of the observed data points. The idea is to interpolate the observed data which means that the prediction is a linear mixture of the observed data points:

$$\hat{y}(s_0) = \lambda_0 + \lambda' y$$

As we are looking for the best prediction, $\widehat{y_0}$ should be as close as possible to the real data $y_0$. To measure how well data support the prediction, the mean squared prediction error (MSPE) is used. The MSPE is equivalent to the mean squared error of estimation when predicting. In the case of simple kriging the MSPE of one unobserved point is: (Stein, 1999, Cressie, 1993, Kitanidis, 1997)

$$\begin{aligned}
&E\left[\left(y(s_0) - \lambda_0 - \lambda' y\right)^2\right] \\
&= Var\,(y(s_0)) + E\,(y(s_0))^2 - 2\lambda_0 E\,(y(s_0)) - 2\lambda' Cov\,(y, y(s_0)) \\
&- 2\lambda' E\,(y)\,E\,(y(s_0)) \\
&+ \lambda_0^2 + 2\lambda_0\lambda' E\,(y) + \lambda' Var\,(y)\,\lambda + \lambda' E\,(y)^2\,\lambda \\
&= Var\,(y(s_0)) - 2\lambda' Cov\,(y, y(s_0)) + \lambda' Var\,(y)\,\lambda + \left(\mu(s_0) - \lambda_0 - \lambda'\mu\right)^2
\end{aligned}$$

where $\lambda_0 + \lambda' y$ is the linear predictor and $\lambda$ and $y$ are vectors of weights and observed values.

The MSPE has to be minimised as we want to get as close to the real data as possible. After partially derivating the MSPE to $\lambda$ and $\lambda_0$ we get the following estimators minimising the MSPE. :(Stein, 1999, Cressie, 1993, Schabenberger and Gotway, 2005)

$$\begin{aligned}
\widehat{\lambda_{0,sk}} &= \mu(s_0) - \hat{\lambda}'\mu \\
\widehat{\lambda_{sk}} &= Var(y)^{-1} Cov\,(y(s_0), y)
\end{aligned}$$

These estimators define the MSPE-optimal predictor: (Stein, 1999, Cressie, 1993, Schabenberger and Gotway, 2005)

$$\hat{y}(s_0) = \mu(s_0) + \left(Var\,(y)^{-1}\,Cov\,(y(s_0), y)\right)'\,(y - \mu) \tag{28}$$

The according MSPE is: (Cressie, 1990)

$$\begin{aligned} MSPE_{sk} &= E\left((y\,(s_0) - \hat{y}(s_0))^2\right) \\ &= Var(y(s_0)) - Cov\,(y(s_0), y)'\,Var\,(y)^{-1}\,Cov\,(y(s_0), y) \end{aligned}$$

The MSPE clarifies that the variance of the prediction, which is the latter term, has to be smaller than the variance of the data, because the MSPE is a positive value.(Schabenberger and Gotway, 2005)
From this point of view, is the simple kriging prediction a good prediction? To check this, we predict the observed values:(Schabenberger and Gotway, 2005)

$$\hat{y}(s) = \mu(s) + \Sigma\Sigma^{-1}\,[y(s) - \mu(s)] = y(s)$$

showing that the prediction is an interpolation of the observed values. Hence, the simple kriging predictor is an „exact interpolator“ or simple kriging predictors „honor the data“.(Schabenberger and Gotway, 2005)

### 3.3.2 Ordinary Kriging Predictor

In ordinary kriging we assume that the expectation $\mu_f$ is constant, but unknown with $E\left(y_0\right) = \mu_f i$, where $i$ is a vector of ones with the respective dimension and $\mu_f$ is a scalar denoting the constant mean. In the case of a fixed but unknown mean, which constrains to linear predictors, does not give a unique solution for minimising the MSPE.(Schabenberger and Gotway, 2005)

In the notation of Cressie (1990) the interpolation of ordinary kriging becomes evident: the uniform linear prediction that Cressie (1990) proposes starts with the idea that $y(s_0)$ is a weighted mixture of all other regions: $\sum_{s=1}^{n} \lambda_s y(s)$. The additional restriction is: $\sum_{s=1}^{n} \lambda_s = 1$, guaranteeing uniform unbiasedness.(Cressie, 1990) In matrix notation the interpolation is $\lambda' y$ and $\lambda' i$, where $i$ is of length $n$.

A justification of the unbiasedness-condition can be found in Schabenberger and Gotway (2005): an unbiased prediction means that

$$E\left(\lambda_0 + \lambda' y\right) = E\left(y(s_0)\right) = \mu(s_0)$$

which leads to the following condition:

$$\lambda_0 + \mu_f\left(\lambda' i - 1\right) = 0$$

The equality has to hold for all $\mu$, and therefore for $\mu_f = 1$, too, leading to

$$\lambda_0 = 0$$
$$\lambda' i = 1$$

which means that the interpolation weights should sum up to 1.(Schabenberger and Gotway, 2005) The mean is covered by $\mu_f$ making $\lambda_0$ equal to 0.

Again it boils down to a minimisation problem: the predictions should be as close as possible to the real data under the given restriction that the prediction is a linear mixture of the observed regions. Again the closeness is measured by the MSPE. The prediction is derived by minimising MSPE subject to $\lambda$:(Cressie, 1990, Schabenberger and Gotway, 2005)

$$E\left[\left(y(s_0) - \sum_{s=1}^{n} \lambda_s y(s)^2\right)^2\right] s.t. \sum_{i=1}^{n} \lambda = 1$$

As this is an optimisation problem under constraints, a Lagrange approach has to be applied. According to Schabenberger and Gotway (2005) the minimisation rule under constraint is:

$$\min_{\lambda} E\left[\left(\lambda' y - y(s_0)\right)^2\right] - 2m\left(\lambda' i - 1\right)$$

where $m$ is the Lagrange-multiplier. This is almost the same as in the foregoing equation, but in matrix notation. The optimisation leads to the following weights:(Cressie, 1990, Schabenberger and Gotway, 2005)

$$\widehat{\lambda'_{ok}} = \left[Cov\,(y, y(s_0)) + i\frac{1 - i'Var\,(y)^{-1}\,Cov\,(y, y(s_0))}{i'Var\,(y)^{-1}\,i}\right]' Var\,(y)^{-1}$$

Again all derivations are given in the appendix. The prediction of the unobserved value is then the interpolation of observed values using the given weights.

$$\widehat{y_0} = \widehat{\lambda'_{ok}} y$$
$$\left[ Cov\,(y, y(s_0)) + i \frac{1 - i' Var\,(y)^{-1}\, Cov\,(y, y(s_0))}{i' Var\,(y)^{-1}\, i} \right]' Var\,(y)^{-1}\, y$$

The mean $\mu_f$ is implicitly estimated. For this reason Schabenberger and Gotway (2005) prefer the explicit formulation by using the BLUP-estimation of the mean:(Cressie, 1990, Schabenberger and Gotway, 2005)

$$\begin{aligned} \widehat{y_{0,ok}} &= \lambda'_{ok} y \\ &= \widehat{\mu_f} + Cov\,(y, y(s_0))'\, Var\,(y)^{-1} \left( y - i\widehat{\mu_f} \right) \end{aligned}$$

In this formula the estimator of $\mu_f$ is explicit and not $\mu_f$ itself, which is the difference to the simple kriging estimator. The constant mean can be estimated by the GLS-estimator:(Schabenberger and Gotway, 2005)

$$\widehat{\mu_f} = \left( i' Var(y) i \right)^{-1} i' Var(y) y_s$$

To measure the uncertainty of the ordinary kriging predictor, the MSPE of the ordinary kriging predictor can be used. It is:

$$\begin{aligned} MSPE_{ok} =& Var(y(s_0)) - Cov(y, y(s_0))' Var(y)^{-1} Cov(y, y(s_0)) \\ &+ \frac{\left( 1 - i' Var(y)^{-1} Cov(y, y(s_0)) \right)^2}{i' Var(y)^{-1} i} \end{aligned}$$

### 3.3.3 Universal Kriging Predictor

Now we drop the assumption of a fixed mean. The mean varies spatially, i.e. it is a function in space: $\mu(s)$. The spatial trend $\mu(s)$ is assumed to be linear in covariables, i.e.

$$\mu(s) = X(s)\beta$$

where $X(s)$ is a matrix of spatially varying covariables and $\beta$ are the coefficients denoting fixed effects.(Schabenberger and Gotway, 2005) Kitanidis (1997) refers to this linear modelling as estimation with a drift. Later we discuss a non-parametric alternative to this parameterisation in the section about kriging as a smoothing spline. The model is now:

$$y(s) = X(s)\beta + e(s)$$

where the random variable $e(s)$ defines the spatial correlations.(Schabenberger and Gotway, 2005).

We prefer the formulation of Fahrmeir et al. (2007) where the spatial correlations are written down explicitly. Fahrmeir et al. (2007) extend the normal linear regression equation by a Gaussian random field parameterising the spatial dependencies that underly the kriging. Of course other random fields are possible, too, but for simplification we restrict modelling to the Gaussian case. Moreover, the Gaussian assumption is needed for the optimality of the kriging predictions.(Pollice and Bilancia, 2002) Later we use the conditional distribution to specify the random field as described by Besag (1974). This assumption leads to the correlation matrix defining the spatial dependencies.

The model of universal kriging is under the assumption of a Gaussian random field:(Fahrmeir et al., 2007, Pollice and Bilancia, 2002)

$$y = X\beta + Z\gamma + \epsilon \tag{29}$$

with the design-matrizes $X$ and $Z$, where $X$ includes all fixed effects that parameterise the spatial trend and the incidence-matrix $Z$:

$$Z[i,j] = \begin{cases} 1 \text{ if } y_i \text{ was observed in region } j \\ 0 \text{ else} \end{cases}$$

with the fixed effects $\beta$ and the random effects $\gamma \sim N(0, G)$ with the variance-covariance matrix

$$G = \tau^2 R = \tau^2 \begin{pmatrix} 1 & \rho(\cdot) & \cdots & \rho(\cdot) \\ \rho(\cdot) & 1 & \rho(\cdot) & \vdots \\ \vdots & & \ddots & \\ \rho(\cdot) & \cdots & \rho(\cdot) & 1 \end{pmatrix}$$

where $\rho(\cdot)$ is the correlation between two regions defined by a correlation function defined in chapter 3.4.2. $R$ is the correlation-matrix and $\tau^2$ is the nugget effect of the assumed random field. It is convenient to split the total variability in the systematic term $\gamma$, explaining the spatial variability, and the error term $\epsilon$.(Pollice and Bilancia, 2002)

Starting from a model based approach (Cicchitelli and Montanari, 2012), we need to specify the distribution of $y(s)$. Assume a common normal distribution for observed values and points for which we want to predict the unobserved values (Fahrmeir et al., 2007, Schabenberger and Gotway, 2005), then the common distribution of the observed values and the value that we want to predict is defined by

$$\begin{aligned} \begin{pmatrix} y \\ y(s_0) \end{pmatrix} \sim & \; N\left( \begin{bmatrix} \mu \\ \mu(s_0) \end{bmatrix}, \begin{bmatrix} Var(y) & Cov(y, y(s_0)) \\ Cov(y, y(s_0)) & Var(y(s_0)) \end{bmatrix} \right) \\ & = N\left( \begin{bmatrix} X\beta \\ x(s_0)\beta \end{bmatrix}, \begin{bmatrix} \tau^2 ZRZ + \sigma^2 I & \tau^2 r \\ \tau^2 r & \tau^2 + \sigma^2 \end{bmatrix} \right) \end{aligned}$$

where $r$ is the vector of $\rho(s_o, s_i)$. The parameters of the observed values $y$ are defined by the regression equation in the case of universal kriging (the mean is unknown), the parameters of the unobserved $y_0$ are the expectation and the variance $\tau^2 + \sigma^2$ as it is a prediction, and both have a covariance of $\tau^2 r$ as we assumed in the correlation function.(Fahrmeir et al., 2007, Schabenberger and Gotway, 2005)[5] If we consider several unobserved values, the given normal distribution has to be extended accordingly.
There arises a problem when using the given distribution for prediction. The predictors of universal kriging are defined such that they are exact " i.e. to the feature of predicting observed points by the points themselves. This is only reasonable when no random noise is assumed for the spatial process and predictions must coincide with observed measurements at sampled locations, but can lead to over-fitting of the predicted surface in the presence of measurement error" (Pollice and Bilancia, 2002). The solution is to use a spatial model with a covariance function assuming no error variance for unobserved regions.(Pollice and Bilancia, 2002) This procedure means to predict just $y(s_0) = x(s_0)'\beta + \gamma(s_0)$.(Fahrmeir et al., 2007) This leads to the following distribution:(Fahrmeir et al., 2007, Pollice and Bilancia, 2002)

$$\begin{pmatrix} y \\ y(s_0) \end{pmatrix} \sim N\left( \begin{bmatrix} \mu \\ \mu(s_0) \end{bmatrix}, \begin{bmatrix} Var(y) & Cov(y, y(s_0)) \\ Cov(y, y(s_0)) & Var(y(s_0)) \end{bmatrix} \right)$$
$$= N\left( \begin{bmatrix} X\beta \\ x(s_0)\beta \end{bmatrix}, \begin{bmatrix} \tau^2 ZRZ + \sigma^2 I & \tau^2 r \\ \tau^2 r & \tau^2 \end{bmatrix} \right)$$

In the case of simple kriging (the mean is known), the expectation $X\beta$ can be replaced by the known mean.(Cressie, 1990) Starting from this position: An easy device to get the optimal prediction of universal kriging is

[5] Note that there is a difference to the distribution referred in Fahrmeir et al. (2007): the authors only write down the parameters of $y(s_0) = x(s_0)\beta + \gamma(s_0)$ which is not the definition that we use. We use the prediction $y(s_0) = x(s_0)\beta + \gamma(s_0) + \epsilon$, defining another variance.

to replace the unknown $\mu$ by its linear predictor $X\beta$.(Stein, 1999, Cressie, 1993) As the unknown mean is substituted by the function of some other variables $X\beta$, it is called plug-in predictor.(Fahrmeir et al., 2007, Cressie, 1993, 1990)
The parameters of the mean function $\beta$ are estimated by GLS and inserted in the simple kriging predictor, defined in equation 28, leading to following estimator and predictor:(Cressie, 1993)

$$
\begin{aligned}
\widehat{\beta}_{gls} &= \left(X'Var(y)^{-1}X\right)^{-1} X'Var(y)y \\
\widehat{y}(s_0) &= \mu(s_0) + \left(Var\,(y)^{-1}\,Cov\,(y(s_0), y)\right)'\,(y - \widehat{X'\beta_{GLS}}) \\
\lambda' &= \Big\{Cov(y, y(s_0)) + X\left(X'Var(y)X\right)^{-1}\Big(x - X'Var(y)^{-1} \\
&\quad Cov(y, y(s_0)))'\Big\}\,Var(y)^{-1}
\end{aligned}
$$

This predictor is the best homogeneous linear unbiased predictor.(Cressie, 1993) Schabenberger and Gotway (2005) name the problems of this approach: first, for the GLS estimation of the trend parameters $\beta$ the variances have to be known. If they are unknown, the EGLS procedure is possible, but the estimation of the variances using the residuals questions the power of estimation. Moreover the uncertainty introduced by using an estimator $\widehat{\beta}$ instead of the known trend is not considered, resulting in an underestimation of the uncertainty of predicted values $\widehat{y}$.(Schabenberger and Gotway, 2005)

Another possibility is to derive the universal kriging predictor as we did before with the simple and the ordinary kriging predictor. We use an optimisation criterion for the prediction at point $y(s_0)$, given the observed points $y$. According to Cressie (1990), Schabenberger and Gotway (2005) and Stein (1999), the mean squared prediction error (MSPE) based on the quadratic loss function is the adequate optimisation criterion. Minimising

the MSPE provides a prediction that minimises the Bayes risk, that is unbiased [6] and that is linear in the case of a Gaussian random field. If the assumption of normality is dropped, a way to ensure the quality of estimation is to constrain the method to BLUPs.(Schabenberger and Gotway, 2005)
As we restrict ourselves to linear predictors, the prediction is again a mixture of the observed values: $a'y$. Again we minimise the MSPE:(Schabenberger and Gotway, 2005)

$$\begin{aligned} MSPE &= E\left[(a'y - y(s_0))^2\right] \\ &= a'Var(y)a + Var(y(s_0)) - 2a'Cov(y, y(s_0)) \\ &= a'\tau^2 ZRZ + \sigma^2 Ia + \tau^2 + \sigma^2 - 2a'\tau^2 r \end{aligned}$$

It is almost the same derivation as that for ordinary kriging. Therefore we need not write it down again. As we need an unbiased predictor which means that(Schabenberger and Gotway, 2005)

$$a'X\beta = x(s_0)'\beta$$

we can derive the unbiasdeness constraint(Schabenberger and Gotway, 2005)

$$a'X = x(s_0)'$$

To minimise the MSPE under the unbiasedness constraint we get the following Lagrange-equation:(Schabenberger and Gotway, 2005)

$$\min_a a'Var(y)a + Var(y(s_0)) - 2a'Cov(y, y(s_0)) + 2m'(X'a - x(s_0))$$

The optimisation, which is computed in the appendix, gives the following coefficients:(Schabenberger and Gotway, 2005)

[6]The unbiasedness can easily be shown by using the law of iterated expectations

- $m = X^{-1}Var(y)X'^{-1}\left(X'Var(y)^{-1}Cov(y, y(s_0)) - x(s_0)\right)$
- $a = \left[Var(y)^{-1} - Var(y)^{-1}X\left(X'Var(y)^{-1}X\right)^{-1}X'Var^{-1}\right]$

$$Cov(y, y(s_0)) + Var(y)^{-1}X\left(X'Var(y)^{-1}X\right)^{-1}x(s_0)$$

These coefficients lead to the universal kriging predictor:(Schabenberger and Gotway, 2005)

$$\widehat{y}_{uk}(s_0) = a'y = \widehat{x(s_0)'\beta}_{GLS} + Cov(y, y(s_0))Var(y)^{-1}\left(y - \widehat{X\beta}_{GLS}\right)$$

It is evident that in the case of universal kriging the prediction is the conditional mean at the unobserved data point $\widehat{x(s_0)'\beta}_{GLS}$ and a variance-weighted residual. To measure the uncertainty of the prediction, we need to look at the MSPE of the universal kriging predictor:(Schabenberger and Gotway, 2005)

$$\begin{aligned}
MSPE_{uk} &= Var(y(s_0)) - Cov(y, y(s_0))'Var(y(s_0))^{-1}Cov(y, y(s_0)) \\
&+ \left(x(s_0)' - Cov(y, y(s_0))'Var(y(s_0))^{-1}X\right)\left(X'Var(y(s_0))^{-1}X\right)^{-1} \\
&\left(x(s_0)' - Cov(y, y(s_0))'Var(y(s_0))^{-1}X\right)'
\end{aligned}$$

Schabenberger and Gotway (2005) give a generalisation for multiple unobserved regions: we define the first two moments as $E\begin{pmatrix} y \\ y_0 \end{pmatrix} = \begin{pmatrix} \mu \\ \mu_0 \end{pmatrix} = \begin{pmatrix} X\beta \\ X_0\beta \end{pmatrix}$ and $Var\begin{pmatrix} y \\ y_0 \end{pmatrix} = \begin{pmatrix} \Sigma_{yy} & \Sigma_{y0} \\ \Sigma_{0y} & \Sigma_{00} \end{pmatrix}$ which defines the multiple universal kriging predictor (Schabenberger and Gotway, 2005)

$$\widehat{y_{0,uk}} = \widehat{X_0\beta}_{GLS} + \Sigma_{0y}\Sigma_{yy}^{-1}\left(y - \widehat{X\beta}_{GLS}\right) \tag{30}$$

with the MSPE:

$$MSPE_{mult,uk} = \Sigma_{00} - \Sigma_{0y}\Sigma_{yy}^{-1}\Sigma_{y0} + \left(X_0 - \Sigma_{0y}\Sigma_{yy}^{-1}X\right) \left(X'\Sigma_{yy}^{-1}X\right)^{-1}\left(X_0 - \Sigma_{0y}\Sigma_{yy}^{-1}X\right)'$$

Another way of deriving the optimal predictor can be done via the distributional assumption.(Fahrmeir et al., 2007, Pollice and Bilancia, 2002) The optimal predictor is the conditional expectation when using the MSPE as loss-function.(Angrist and Pischke, 2009, Grimmett and Stirzaker, 2009). Fahrmeir et al. (2007) show that under normality the optimal predictor is the one given above. Setting the parameters of the normal distribution given above into formula 28, we get:(Fahrmeir et al., 2007)

$$\widehat{y_0} = E\left(y_0|y\right) = \mu_0 + \tau^2 r'(\tau^2 ZRZ' + \sigma^2 I)^{-1}(y - X\beta) \qquad (31)$$

The conditional variance of the estimators can be derived by the normal distribution and the law of iterated expectations as: (Fahrmeir et al., 2007)

$$Var\left(y_0|y\right) = \tau^2 + \sigma^2 - \tau^2 r'(\tau^2 ZRZ' + \sigma^2 I)^{-1} r\tau^2 \qquad (32)$$

Even without the assumption of normality, the prediction $\widehat{y_0}$ still fulfills some optimality properties: It is still the best linear unbiased prediction. (Fahrmeir et al., 2007, Pollice and Bilancia, 2002) [7]

The distributional assumption is also needed for the ML-estimation: $y(s) \sim N(X(s)\beta, Var(y;\theta))$.(Schabenberger and Gotway, 2005) Defining $Var\,(y;\theta) = \tau^2 ZRZ' + \sigma^2 I$ as the covariance matrix of the marginal mixed model with the parameterisation $\theta$, the log likelihood of the variance as

[7] It has to be noted that the difference of the normal distribution that we made compared to Fahrmeir et al. (2007) only is apparent in the variance of the predictor which can be reasoned with the error term having an expectation equal to 0.

well as the regression-parameters is:(Stein, 1999, Schabenberger and Gotway, 2005, Pollice and Bilancia, 2002)

$$
\begin{aligned}
l(\theta, \beta) &= -\frac{n}{2} - \frac{1}{2} \log detVar(y; \theta) && (33)\\
&\quad -\frac{1}{2}(y - X\beta)' Var(y; \theta)^{-1}(y - X\beta) && (34)
\end{aligned}
$$

Maximising equation 33 leads to the well known ML estimator(Stein, 1999, Schabenberger and Gotway, 2005)

$$
\widehat{\beta}(\theta) = \left(X' Var(y; \theta)^{-1} X\right)^{-1} X' Var(y; \theta)^{-1} y \qquad (35)
$$

which is the same as the often used GLS estimator.

### 3.3.4 Comparison by MSPEs

The three kriging predictors which now have been introduced can be compared by their MSPEs. The MSPE of the simple kriging predictor serves as the basis of comparison.
The ordinary kriging predictor differs by adding $\frac{\left(1 - i' Var(y)^{-1} Cov(y, y(s_0))\right)^2}{i' Var(y)^{-1} i}$.
For the reason of having quadratic terms in the nominator and the denominator, the ordinary kriging predictor has a higher MSPE than the simple kriging predictor. However keep in mind that we assumed that all assumptions are met for these MSPEs to be the true prediction error. The assumptions of the simple kriging predictor are much more restrictive.
The ordinary kriging predictor gets a smaller MSPE if the covariance structure of the unobserved points is more similar to the variance structure of the observed points. It means that the covariance structure is represented correctly by the observed points.

The universal kriging predictor adds something like a squared covariance adjusted residual to the MSPE of the simple kriging predictor representing the mixed effects model of universal kriging. Its MSPE is always as least as big as the MSPE of the simple kriging predictor. It equals the MSPE of the simple kriging prediction if the underlying regression has no residuals, i.e. all points are exactly on the regression line. The MSPE of the universal kriging and the ordinary kriging predictor cannot be compared without determining the vectors and the matrices.

### 3.3.5 Bayesian Prediction

Let us start with the underlying spatial Gaussian model (Pollice and Bilancia, 2002) defined in 29. Now we need to assume that the Gaussian random field is independent from $\beta$, for which we introduce a normal prior: $\beta \sim N\left(\mu_\beta, V_\beta\right)$.(Stein, 1999, Pollice and Bilancia, 2002) The variable $y$ has the marginal probability distribution: (Pollice and Bilancia, 2002)

$$f(y|\beta,\sigma^2,\tau^2,\phi,X) = \int f(y|\beta,\sigma^2,\gamma,X)P(\gamma|\sigma^2,\phi,X)d\gamma$$

The likelihood, together with an appropriate prior, defines the posterior distribution: (Pollice and Bilancia, 2002)

$$P(\beta,\sigma^2,\tau^2,\phi|y,X) = \frac{\int f(y|\beta,\sigma^2,\gamma,X)P(\gamma|\sigma^2,\phi,X)P(\beta)P(\sigma^2)P(\tau^2,\phi)d\gamma}{\int \dots \int f(y|\beta,\sigma^2,\gamma,X)P(\gamma|\sigma^2,\phi,X)P(\beta)P(\sigma^2)P(\tau^2,\phi)d\beta d\gamma d\sigma^2 d\tau^2 d\phi}$$

The joint distribution of the three random components of the model ($y, \beta, \gamma$) is: (Pollice and Bilancia, 2002)

$$\begin{pmatrix} y \\ \beta \\ \gamma \end{pmatrix} \sim N_{2n+p}\left(\begin{pmatrix} X\mu_\beta \\ \mu_\beta \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} XV_\beta X' + \Sigma_{yy} & XV_\beta & \tau^2 R \\ V_\beta X & V_\beta & \mathbf{0} \\ \tau^2 R & \mathbf{0} & \tau^2 R \end{pmatrix}\right)$$

where **0** is a vector or a matrix with the dimensions needed.
Treating the variance components as fixed, the posterior of $\beta$ is:(Stein, 1999, Pollice and Bilancia, 2002)

$$\beta|Y \sim N\left(\left(X'\Sigma_{yy}^{-1}X + V_\beta^{-1}\right)^{-1}\left(X'\Sigma_{yy}^{-1}y + V_\beta\mu_\beta\right), \left(X'\Sigma_{yy}^{-1}X + V_\beta^{-1}\right)^{-1}\right),$$

However, we are not that interested in estimating parameters using the given posterior, but in predicting unobserved values. The joint distribution of the parameters, observed values, and unobserved values can be denoted as:(Pollice and Bilancia, 2002)

$$\begin{pmatrix} y \\ y_0 \\ \beta \\ \gamma \\ \gamma_0 \end{pmatrix} \sim N_{2n+p}\left(\begin{pmatrix} X\mu_\beta \\ X_0\mu_\beta \\ \mu_\beta \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} XV_\beta X' + \Sigma_{yy} & XV_\beta X_0' + \Sigma_{y0} & XVar_\beta & \tau^2 R & \Sigma_{y0} \\ XV_\beta X_0' + \Sigma_{y0} & X_0 V_\beta X_0' + \Sigma_{00} & X_0 Var_\beta & \Sigma_{y0} & \tau^2 R \\ Var_\beta X & Var_\beta X_0 & Var_\beta & \mathbf{0} & \mathbf{0} \\ \tau^2 R & \Sigma_{y0} & \mathbf{0} & \tau^2 R & \Sigma_{y0} \\ \Sigma_{y0} & \tau^2 R & 0 & \Sigma_{y0} & \tau^2 R \end{pmatrix}\right)$$

We need the predictive distribution.(Pollice and Bilancia, 2002) The predictive distribution depends on the observed values and does not depend on unknown quantities. If something is unknown, there is the possibility to integrate it out. For example unknown parameters can be integrated out.(Hoff, 2009) This idea is the same as that which underlies multiple imputation and, therefore, it looks the same. The predictive distribution describes the distribution of the value $Y_f$ that has to be predicted under the knowledge of the observed values of that variable $Y$: (Hoff, 2009, Wagner, 03.02.2014, Pollice and Bilancia, 2002)

$$\begin{aligned} P\left(Y_f|Y\right) &= \int_{\Theta} P\left(Y_f, \theta|Y\right) d\theta \\ &= \int P\left(Y_f|\theta\right) P\left(\theta|Y\right) d\theta \end{aligned}$$

This is the posterior predictive as the data is already observed. If the prior and the likelihood are conjugated distributions, the posterior predictive is in closed form and can be computed.(Wagner, 03.02.2014, Hoff, 2009) If the posterior predictive is not in closed form, a sampling scheme can easily be created. First, parameters are sampled from $P\left(\theta|Y\right)$, and second the values $Y$ are sampled from $P\left(Y_f|\theta\right)$, where $\theta$ is specified through the draws in the first step.(Hoff, 2009) Ordinarily, variance parameters are unknown, too. Pollice and Bilancia (2002) propose to use an empirical Bayes estimator in the case of unknown variance.
In order to get a prediction, we need to derivate the posterior predictive. Stein (1999) proposes the use of a uninformative prior and the posterior predictive to be normal. According to Pollice and Bilancia (2002) the Bayesian predictor under normality is:

$$E\left(y_0|y\right) = X_0\mu_\beta + \left(XV_{\beta_0}X_0' + \Sigma_{y0}\right)\left(XV_{\beta_0}X_0 + \Sigma_{yy}\right)^{-1}\left(y - X\beta_0\right)$$

$$Cov\left(y_0|y\right) = X_0 V_\beta X_0' + \Sigma_{00} - \left(XV_{\beta_0}X_0' + \Sigma_{y0}\right)\left(XV_{\beta_0}X_0 + \Sigma_{yy}\right)^{-1}$$
$$\left(XV_\beta X_0 + \Sigma_{y0}\right)$$

Using a non-informative prior, the best prediction is again the conditional expectation with the conditional variance.(Stein, 1999)

A more general approach to find a Bayesian prediction is to use a loss function and to minimise the risk of the prediction. The aim is to minimise the risk which is the expected loss.(Wagner, 03.02.2014, Cressie, 1993)

$$E\left(R\left(\widehat{Y_f}, Y_f\right)\right) = \int R\left(\widehat{Y_f}, Y_f\right) P\left(\widehat{Y_f}, Y_f\right) dY_f$$

The MSPE can also be interpreted as a loss function. In the Bayesian sense, it corresponds to a quadratic loss function. Minimising the quadratic loss function leads to the expectation of the posterior predictive as optimal prediction.(Wagner, 03.02.2014)[8]

Cressie (1993) states that „simple kriging corresponds to a degenerate prior and universal kriging corresponds to a diffuse prior.“ (Cressie, 1993) The relationship is shown by Omre and Halvorsen (1989). The MSPE is used as the risk-function leading to nearly the same prediction-equation as in non-Bayesian kriging with the exception of not including the constraint of unbiasdeness. If the parameters of the spatial trend are estimated

[8] Cressie (1993) enumerates other possible loss functions that can be used:

- The absolute loss function $l\left(y, \hat{y}\right) = \left|y_0 - \hat{y}\right|$ leads to the conditional mean of the posterior predictive as optimal prediction.
- The 0-1 loss function $l\left(y, \hat{y}\right) = I\left(\left|y_0 - \hat{y}\right|\right)$ implies that the conditional mode is the optimal prediction.
- Asymmetric loss functions as the linex loss function can represent unequal weights of under-estimation and over-estimations.(Cressie, 1993) Metric loss functions as the linex loss function can represent unequal weights of under-estimations and over-estimations.(Cressie, 1993)

with the MSPE-risk-function, a distinction of cases can be made: If there is no uncertainty about the values of the coefficients, it results in a simple kriging approach. This congruency reflects that in simple kriging the mean structure is assumed known, hence there is no uncertainty about it. If there is no knowledge of the coefficients, the Bayesian approach of Omre and Halvorsen (1989) leads to universal kriging. In the latter case the uncertainty is reflected by a flat prior.(Omre and Halvorsen, 1989)

Cressie (1993) proposes an empirical Bayes predictor to overcome the problem of not knowing the parameters of the prior. Cressie (1993) also states that not only the mean parameter, but also the covariance parameter, can be determined by the empirical Bayes approach.

## 3.4 The Correlation Structure

The variance-covariance-matrix is a $n \times n$ matrix consisting of the variances on the diagonal and of the covariances of two regions. It specifies the correlations of the Gaussian random field and the variances of the data. Ordinarily the matrix is unknown, so it needs to be estimated. A structural identification needs to be done before the kriging predictors are estimated.(Zimmerman, 1989) Instead of estimating all $\frac{n(n-1)}{2}$ values of the matrix, it is convenient to assume a special kind of correlation to reduce the number of parameters that need to be estimated.(Schabenberger and Gotway, 2005) The specifications that are typical for the kriging case are described in the next section. Two factors determine the spatial correlation: the kind of assumed neighbourhood including the question of how to measure the grade of neighbourhood (inherent is the choice of a distance measure) and the assumptions determining the auto-regressive structure of the random field.(Schabenberger and Gotway, 2005) Later we describe how to estimate these parameters, before we show how they can be graphically represented by the variogram.

### 3.4.1 Neighbourships

Often the space is divided into discrete regions. In this case it is not possible to measure the distances between two regions continuously. A definition of neighbourhood becomes necessary.(Fahrmeir et al., 2007)

It is common sense that neighbours are defined as two regions that share a common border.(Fahrmeir et al., 2007). Using this idea we can order all regions by their degree of neighborhood. The first neighbours are regions sharing the same border. Second order neighbours are regions that are at least the first neighbours of the first neighbours. And so on. The following two graphs show the first neighbours of Munich. The next figure shows the second order neighbours of the city.

However, other concepts of measuring the degree of neighbourhood are possible. Some of them are described in Fahrmeir et al. (2007) The length of the shared border could be such a measure or the distance of the centroids.(Fahrmeir et al., 2007) The latter is the concept that was used here.

In a more common sense, the degree of neighbourhood can be interpreted as the degree of similarity of two regions, giving neighbourhood a more statistical than spatial connotation. Statistical because data and statistical methods such as data fusion could be used to find neighbours, or to define neighbours, respectively measure the degree of neighbourhood. Although it would be interesting to step into that issue more deeply, this is not the place to do so.

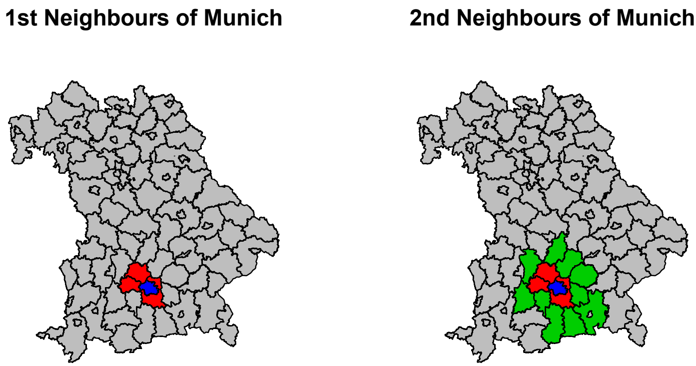


**Abbildung 1:** First and Second Neighbours of Munich

**Source:** Bayerische Vermessungsverwaltung (2015)

### 3.4.2 Correlation Functions

In geostatistics it is more common to specify the correlation function $R$ than the covariance $Var(y)$ of the stochastic process. Here we introduce some correlation functions useful for kriging.
The common assumptions for the correlation functions are:

1. stationarity: $\rho(t, s)$ only depends on the distances $h = |t - s|$. For this reason we can usually write $\rho(h)$.(Fahrmeir et al., 2007) This assumption is the weak stationarity, meaning that only the moments are stationary. Schabenberger and Gotway (2005) call it second order stationarity. The distribution need not be stationary, which would define the strict stationarity not assumed here.(Stein, 1999)

2. The correlation disappears for huge distances: $\rho(h) \to 0$ is for $h \to \infty$.

3. $\rho(h)$ is monotone decreasing. The further away two points are located, the less they are correlated.(Fahrmeir et al., 2007)

There are several correlation functions that fulfill the conditions. First, there is the **spherical correlation function**:

$$\rho\,(h,\phi) = \begin{cases} 1 - \frac{3}{2}\left|\frac{h}{\phi}\right| + \frac{1}{2}\left|\frac{h}{\phi}\right|^3 & \text{if } 0 \le h \le \phi \\ 0 \text{ if } h \ge \phi \end{cases}$$

where $\phi > 0$ is the range determining the distance where no correlation is assumed. The problem of this function is that it cannot be differentiated at $h = \phi$, which is a problem for all ML-methods.(Fahrmeir et al., 2007) Second, Fahrmeir et al. (2007) list the **potency exponential family**:

$$\rho\,(h,\phi,\kappa) = exp\left\{-\left|\frac{h}{\phi}\right|^{\kappa}\right\}$$

where $\phi > 0$ and $0 < \kappa < 2$. For $\kappa = 1$ the potency exponential family becomes the Gaussian correlation function. In the potency exponential family, all points are still correlated, even if they are located far away from each other. However, in reality the correlations are set to zero if they are too far away. This effective range is controlled by the scaling parameter $\phi$. The $AR(1)$-process can be interpreted as a discrete version of this family.(Fahrmeir et al., 2007)
The third family which is enumerated by Fahrmeir et al. (2007) is the **Matérn-family**. This is a quite general family basing on Bessel-functions. The covariance-function in this case is:(Schabenberger and Gotway, 2005)

$$C(h) = \sigma^2 \frac{1}{\Gamma(\kappa)}\left(\frac{\phi h}{2}\right)^{\kappa} 2K_{\kappa}(\phi h),\ \kappa > 0, \phi > 0$$

where $K_\kappa$ is the modified Bessel-function of the second kind of order $\kappa$ and $\phi$ determines the range defining spatial dependency.(Schabenberger and Gotway, 2005) $\kappa$ controls the respective range (Fahrmeir et al., 2007), which implies that it controls the degree of smoothness.(Stein, 1999) The smoothness increases with $\kappa$. If $\kappa \to \infty$ it becomes a Gaussian-Model:(Fahrmeir et al., 2007, Schabenberger and Gotway, 2005)

$$C(h) = \sigma^2 \exp\left\{-\phi h^2\right\}$$

Kitanidis (1997) describes the Gaussian case as a model family of its own. He reports the corresponding variograms, too. The scale parameter $\phi$ describes the point „at which the correlation is 0.05." (Kitanidis, 1997)
If $\kappa = \frac{1}{2}$ the family becomes the exponential model. The covariance function is:(Schabenberger and Gotway, 2005)

$$C(h) = \sigma^2 \exp\{-\phi h\}$$

and the respective correlation function:(Fahrmeir et al., 2007)

$$\rho\,(h; \phi, \kappa = 0.5) = exp\left(-\left|\frac{h}{\phi}\right|\right)$$

The exponential model is the well-known time-series model of a first order auto-regressive process.(Schabenberger and Gotway, 2005)
The Matérn family is the most flexible class of correlation functions, making them a standard.(Fahrmeir et al., 2007) Stein (1999) even mentions them as a canonical model. The spectral density can be found here, too. The respective auto-covariance function and the spectral density can be found in Stein (1999). For a further simplification, the parameter $\kappa$ is restricted to $\kappa = m + 0.5$ where $m$ is a non-negative integer.(Stein, 1999).

Dubrule (1983) proposes to use the following general covariance model, which is in fact the correlation function multiplied with the variance. Therefore he defines the correlation function this way:

$$Cov(h) = C_0\delta(h) - b_0|h| + b_s|h|^2 \log|h| + b_1|h|^3$$

where $C_0$ is the nugget effect. If $b_0$ and $b_1$ are assumed to equal zero, the covariance function defines a thin-plate leading to the equivalence of kriging and spline smoothing discussed later.
The thin-plate defines a radial basis-function in the spline approach. A radial basis-function uses the euclidean distance between observations $y = (y_1, y_2)$ and the knots $k = (k_1, k_2)$:(Fahrmeir et al., 2007)

$$B_k(y) = B(\|y - k\|) = B(r) \tag{36}$$

where $B(\cdot)$ is an appropriate function and $r = \|y - k\| = \left[(z_1 - k_1)^2 (z_2 - k_2)^2\right]^{0,5}$. The function defined by Dubrule (1983) is one such appropriate function. The functions define radial lines in a contour plot. All basis-functions have the same functional form and every single basis-function is located at one knot.(Fahrmeir et al., 2007)

### 3.4.3 Iteratively Re-weighted Generalised Least Squares

For a spatial prediction we not only have aim of estimating the $\beta$'s of the regression model, but also the parameter of the underlying random field, i.e. the parameters of the covariance-matrix defining the spatial correlations. Moreover it is necessary to estimate the parameters of the covariance matrix to be able to estimate the $\beta$'s by EGLS.(Schabenberger and Gotway, 2005) As the coefficients of the regression model parameterising the spatial trend are estimated using least squares, a method using the same idea is obvious.

The concept is to alternate between the estimation of the $\beta$'s by GLS and the estimation of $Var(y)$. For the GLS-estimation of the $\beta$'s the estimation of the variance-covariance-matrix is used and $Var(y)$ is estimated using the residuals that are computed using the $\beta$'s. This establishes the following iterated process.

1. The starting point is an OLS-estimation to get sensible estimations $\widehat{\beta}^0$.

2. The residuals are computed using $\widehat{\beta}^0$: $r = y(s) - X(s)\widehat{\beta}^0$.

3. The variogram is estimated using the residuals: The variogram is splitted into a deterministic part and a residual part: $\widehat{\omega}(h) = \omega(h;\theta) + e$. The residuals are used to compute the variance-covariance matrix of $e$. The parameters $\theta$ of the variogram are estimated by minimising the following least squares condition:

$$\widehat{\omega}(h) - \omega(h;\theta)' R(\theta) \left(\widehat{\omega}(h - \omega(h;\theta)\right)$$

   If the parameters of the variogram are known, $Var(y)^i$ can be determined.

4. A new GLS-estimation using $Var(y)^i$ gives $\widehat{\beta}^i$

After the first round, the process is iterated from the second to the last step until no relevant improvements compared to the former iteration are found.(Schabenberger and Gotway, 2005)

The problem of this procedure is that it is not known whether the process converges or not. After stopping we only know that it does not improve any more.(Schabenberger and Gotway, 2005) We are therefore not sure whether we have found the global optimum.

### 3.4.4 Maximum Likelihood (ML)

The estimator 35 described in the universal kriging chapter can be used to derive the profile log likelihood of $\theta$.(Schabenberger and Gotway, 2005) Inserting it into the log likelihood 33 and maximising this so-called profile log likelihood with respect to $\theta$ gives the ML estimator of the variance components $\theta$, which is biased.(Stein, 1999, Schabenberger and Gotway, 2005) For the reason of asymptotically unbiased estimators, the ML-estimator needs large sample sizes, as it can be severely biased in small samples.(Cressie and Lahiri, 1993, Pollice and Bilancia, 2002)
The minus twice the log likelihood profiled for $\beta$ is: (Schabenberger and Gotway, 2005)

$$ln\left(|\sigma^2 Var\,(y;\theta*)\,|\right) + nln(2\pi) + \sigma^2 r' Var\,(y;\theta*)^{-1}\, r$$

where $Var\,(y;\theta) = \sigma^2 Var\,(y;\theta*)$ and the GLS-residuals:(Schabenberger and Gotway, 2005)

$$r = y(s) - \left(X(s)' Var\,(y;\theta)^{-1}\, X(s)\right) X(s)' Var\,(y;\theta)^{-1}\, y(s)$$

Profiling for the variance $\sigma^2$ gives(Schabenberger and Gotway, 2005)

$$\widehat{\sigma^2_{ML}} = \frac{1}{n} r' Var\,(y;\theta*)^{-1}\, r$$

Profiling for both gives the optimisation problem for the parameters defining the correlation structure, which is minimised with respect to $\theta*$: (Schabenberger and Gotway, 2005)

$$ln\,(|Var\,(y;\theta*)\,|) + nln(\widehat{\sigma^2}) + nln(2\pi - 1)$$

Starting from the last optimisation problem, the ML-estimations can be computed following the profiling recursively.

In most cases the profile likelihood of $\theta*$ cannot be optimised in closed form, necessitating the use of approximative methods.(Schabenberger and Gotway, 2005) For this reason, algorithms such as the well known Newton-Raphson have been established.
In the spatial setting there is the additional problem of dimensionality. To overcome this problem, Stein (1999) mentions two ways to facilitate computations: First, there is the method of Veccia. He proposes the partitioning of the likelihood according to the regions:

$$P\left(y_1, \dots, y_n\right) = P\left(y_1\right) \prod_{j=2}^{n} \left(y_j | y_1, \dots, y_{j-1}\right)$$

The latter is simplified by approximating the conditioning by the $m$ next neighbours. Of course, the fewer neighbours are taken into account, the less efficient is the analysis.(Stein, 1999) The second approach is more intuitive: The area of interest is divided into sub-regions. For every sub-region, an indiviudal likelihood is computed. The overall likelihood is approximated by the multiplication of the likelihoods of the sub-regions that have been computed.(Stein, 1999)

### 3.4.5 Restricted Maximum Likelihood

The REML estimation solves the problem of biasedness of the ML-estimator by considering the number of estimated mean parameters and leads to the well known estimator of the variance $\frac{1}{n-p}\left(y - X\beta\right)'\left(y - X\beta\right)$.(Fahrmeir et al., 2007, Stein, 1999, Schabenberger and Gotway, 2005)

The advantage of the REML-estimation is the possibility to estimate the auto-covariance function as well as to estimate the variance.(Stein, 1999) The disadvantages of the REML-estimation is that the functional form of the distribution of the depending variable needs to be specified for this extensive computations (Zimmerman, 1989), and this often leads to negative variance estimations(Kneib, Greven, 2011).

The data are transformed by $Ky(s)$ such that $E\,(Ky(s)) = 0$. $K$ is called the $(n-k) \times n$-matrix of error contrast.
In the case of simple and ordinary kriging, which is equivalent to assuming a constant mean, Schabenberger and Gotway (2005) propose the use of

$$\begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & -\frac{1}{n} & \cdots \\ & & \ddots & \end{pmatrix},$$

which means that $Ky(s)$ is the vector of differences.
$Ky(s) \sim N\left(\theta,\ KVar(y;\theta)K'\right)$ minus twice the log likelihood is:(Schabenberger and Gotway, 2005)

$$ln\left\{|KVar(y;\theta)K'|\right\} + (n-1)ln(2\pi) + y(s)'K'(KVar(y;\theta)K')^{-1}Ky(s)$$

which does not include the mean $\mu$. The REML-estimator for $\mu$ can be derived by plugging in the REML-covariance estimator into the GLS-estimator.(Schabenberger and Gotway, 2005)

A REML-estimator in the case of universal kriging means to optimise minus two-times the log likelihood of the transformed data:(Schabenberger and Gotway, 2005, Zimmerman, 1989, Cressie and Lahiri, 1993)

$$ln\left\{|KVar(y;\theta)K'|\right\} + (n-k)ln(2\pi) + y(s)'K'\left(KVar(y)K'\right)^{-1}Ky(s)$$

When using REML-estimations, the vector $\beta$ of the spatial trend seems to vanish. The optimisation of the likelihood function leads to the REML-estimators of the variance-components. This can be done by several well-known algorithms such as the Newton-Raphson or the Fisher-Scoring. The REML-estimator of $\beta$ means inserting the REML-estimations of the variance-components into the GLS-estimator of the regression coefficient.-(Schabenberger and Gotway, 2005) After some computations it seems as though the transformation $Ky(s)$ vanishes in the log likelihood due to the characteristics of the transformation. Minus twice the log likelihood is now:(Schabenberger and Gotway, 2005)

$$\begin{aligned} ln\left(|Var(y;\theta)|\right) \quad & +ln\left(|X(s)'Var(y;\theta)X(s)|\right) - ln\left(X(s)'X(s)\right) \\ & +rVar(y;\theta)^{-1}r + (n-k)ln\left(2\pi\right) \end{aligned}$$

where $r$ is the OLS-residual to differentiate the fixed effects away.(Searle et al., 2006) The REML-estimator becomes a ML-estimator based on the residuals. (Pollice and Bilancia, 2002)
Again profiling is possible where $\widehat{\sigma^2_{REML}} = \frac{1}{n-k} r'Var(y;\theta*)^{-1}r$ which gives minus twice the profiled log likelihood:(Schabenberger and Gotway, 2005)

$$\begin{aligned} ln\left(|Var(y;\theta*)|\right) \quad & +ln\left(|X(s)'Var(y;\theta*)X(s)|\right) + (n-k)ln(\widehat{\sigma^2_{REML}}) \\ & +(n-k)\left(ln\left(2\pi\right)\right) \end{aligned}$$

The estimated variance-parameters can be plugged into the estimators for the spatial trend. The MSPE and the variances of the estimators are slightly biased, but in „other words, the use of a plug-in estimator of the kriging variance is fine for most spatial problems with moderate to strong spatial auto-correlations.“ (Schabenberger and Gotway, 2005)

The REML-estimator can be compared to a Bayesian estimator with a non-informative prior. If prior independence of the variance parameters is assumed, the REML-estimator is the same as the mode of the posterior. The

REML-estimator, therefore, is a compromise between the ML-estimator and the Bayes-estimator under quadratic loss. The full likelihood is averaged over $\beta$ and afterwards maximised for the variance parameters.(Cressie and Lahiri, 1993)

### 3.4.6 Bayesian Estimation

Pollice and Bilancia (2002) note that under the most common case of a conjugate prior for $\beta$ and a flat prior for the variance parameters, the predictor is the same as discussed in the chapter about Bayesian prediction of kriging. Otherwise there are strong restrictions of the variance parameters needed to get a closed form predictor. A case is described in Pollice and Bilancia (2002). In general, numerical iterative methods are needed to get a Bayesian estimator of the variance parameters and the according estimators for $\beta$ and the prediction.(Pollice and Bilancia, 2002)

### 3.4.7 Variogram

The variogram, also called semi-variogram, measures the variance between two observation points. The semi-variance between the two units can be computed by:(Schabenberger and Gotway, 2005)

$$
\begin{aligned}
\omega(s_i, s_j) = \omega(h) &= \frac{1}{2} Var\left[y(s_i) - y(s_j)\right] \\
&= \frac{1}{2} Var\left[y(s_i)\right] + \frac{1}{2} Var\left[y(s_j)\right] - Cov\left[y(s_i), y(s_j)\right] \\
&= C(0) - C(s_i, s_j) = C(0) - C(h)
\end{aligned}
$$

where $C(\cdot)$ denotes the value of the covariance function.
If the underlying stochastic process is stationary, the variogram is a parameter of the process.(Schabenberger and Gotway, 2005) As we always

assume second order stationarity, the variogram is a parameter.
However, we do not use the variogram as a parameter that needs to be estimated. Instead we just use its graphical representation as a way to illustrate the variances and correlations. In Schabenberger and Gotway (2005) a detailed introduction to the variogram and its estimators can be found.
In figure 2 The intercept of the variogram is called the nugget effect, repre-

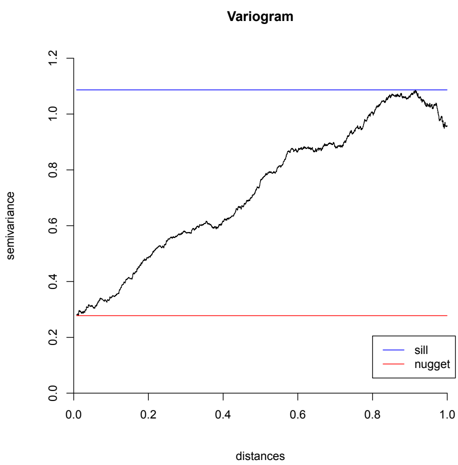


**Abbildung 2:** Example of a Variogram

**Source:** Own computation using a simulation example found in Ribeiro and Diggle (2015)

senting the chaotic part of the underlying random process.(Schabenberger and Gotway, 2005) It is the variance that is inherent to the spatial random field $\tau^2$. Kitanidis (1997) describes the nugget-effect as representing „microvariablilty in addition to measurement error.“
The asymptote of the variogram is the total variance $\tau^2 + \sigma^2$ of the stochastic model we look at. It is called sill. The lag at which the asymptote is approximately reached is called range.(Schabenberger and Gotway, 2005)
The covariance at lag $h$ is the distance between the variogram at lag $h$ to

the sill. It can be interpreted as the part of the variance at lag $h$ that is explained by the spatial modelling. The part under the variogram at lag $h$ is the unexplained part, i.e. the variability that two points with that distance have. Of course the bigger the lag, the smaller is the explained part, as we assume declining correlations.

There are several methods to estimate the variogram. Schabenberger and Gotway (2005) describe several of them. The easiest method is the estimator of Matheron (Schabenberger and Gotway, 2005):

$$\widehat{\omega}(h) = \frac{1}{2N(h)} \sum_{N(h)} \left(y(s_i) - y(s_j)\right)^2$$

where $N(h)$ is the number of observation pairs with the lag $h$. The problem with the Matheron's estimator is that it is very wiggly.
As we only want a graphical representation of the spatial correlation, this estimator seems appropriate, as it is simple to compute and unbiased. (Schabenberger and Gotway, 2005) To address the wigglyness, we smooth the function by using a LOESS estimator. The disadvantages of the proposed estimators are not relevant for our purpose, as we do not need the distributional features of the variogram.
We define the variogram in the same way as is done by Kitanidis (1997): „Plot the square difference" $\frac{1}{2}\left(y(s_i) - y(s_j)\right)^2$ „against the separation distance (...) for all measurement pairs (...). The experimental variogram is a smooth line through this scatter plot."(Kitanidis, 1997). Of course a sparsing is done by computing intervals for the lags as Kitanidis (1997) proposes.

# 4 KriMI: Multiple Imputation Using Kriging

To preserve both covariance and variance at once, MI can be combined with kriging. If correlation between observations are disregarded in a Bayesian estimation, the posterior inference becomes too precise. More seriously, predictions become inaccurate.(Gelman et al., 2004) The latter is most relevant for our case. If there are reasons to believe that data are spatially correlated, we need to reflect this in the modelling.
The buildup needs a Bayesian perspective on kriging: a Bayesian formulation of the kriging model is needed to implement the Bayesian estimation in the MI sampling scheme. There are two ways to formulate a Bayesian kriging scheme: first, kriging is interpreted as a mixed model that can be estimated using Bayesian methods. Second, the relationship of kriging and spline smoothing is exploited to find a Bayesian version of kriging.

## 4.1 Predicting Using Mixed Modelling

Using kriging as an imputation model by mixed modelling can also be found in Munoz et al. (2010), though in this article there are some serious mistakes in understanding multiple imputation and the underlying sampling mechanism. In Zhu et al. (2003) kriging is used as the imputation model, too. However the exact approach of multiple imputation using kriging is not described, the authors just note that they multiply impute using kriging to preserve spatial dependencies.(Zhu et al., 2003) This latter fact and the abscence of a description of MI using kriging make it necessary to describe the mixed model approach.

### 4.1.1 The Model

As shown in the kriging chapter, universal kriging can be reformulated as the following mixed model:(Fahrmeir et al., 2007)

$$y = X\beta + Z\gamma + \epsilon$$

with

- the design matrices $X$ and $Z$, where $X$ includes all fixed effects that parameterise the spatial trend. X has dimensions $n \times p$ and the incidence matrix $Z$ defining a $n \times S$ matrix ($n$ is the number of observations and $S$ is the number of observed regions) as

$$Z[i,j] = \begin{cases} 1 \text{ if } y_i \text{ was observed in region } j \\ 0 \text{ else} \end{cases}$$

- the effects $\beta \sim N(m, M)$ as a vector of fixed effects having length $p$ and the $S$ vector $\gamma \sim N(0, G)$ for the random effects defined as

$$G = \tau^2 R = \tau^2 \begin{pmatrix} 1 & \rho(\cdot) & \cdots & \rho(\cdot) \\ \rho(\cdot) & 1 & \rho(\cdot) & \vdots \\ \vdots & & \ddots & \\ \rho(\cdot) & \cdots & \rho(\cdot) & 1 \end{pmatrix}$$

  where $\rho(\cdot)$ is the correlation between two regions defined by a correlation function defined in the kriging chapter.

- the usually used error term is $\epsilon \sim N(0, \sigma^2)$.

The hyper-parameters of this mixed model are the variance parameters $\sigma^2$, $\tau^2$, and the parameters defining $\rho(\cdot)$, if they are seen as parameters, too. We propose to use the parameters defining $\rho(\cdot)$ as fixed parameters.(Fahr-

meir et al., 2007) Otherwise it seems to create too much difficults in finding a proper prior for these correlation parameters, in addition to the fact that in the data the observations are sparse that it is likely that we won't find a spatial correlation in the data.
As usual, we estimate the marginal model, and therefore the expectation of the response is $E\,(y) = X\beta$ and the variance-covariance is

$$Var\,(y) = Var\,(X\beta + Z\gamma + \epsilon) = \tau^2 ZRZ' + \sigma^2 I.$$

As $Z$ is not quite familiar it seems reasonable to show two short examples of how $Z$ and accordingly $Var\,(y)$. For a graphical representation look at table 1. where $y_{[i,j]}$ is observation $i$ in region $j$. The fields are four regi-

| | |
|---|---|
| $y_{[1,1]}$ | $y_{[2,2]}$ |
| $y_{[3,3]}$ | $y_{[4,4]}$ |

**Tabelle 1:** KriMI Example 1

**Source:** Own diagram

ons. This defines a matrix $Z = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. In this case the variance covariance matrix of $y$ is

$$Var(y) = \begin{pmatrix} \tau^2 + \sigma^2 & \tau^2\rho(\cdot) & \tau^2\rho(\cdot) & \tau^2\rho(\cdot) \\ \tau^2\rho(\cdot) & \tau^2 + \sigma^2 & \tau^2\rho(\cdot) & \tau^2\rho(\cdot) \\ \tau^2\rho(\cdot) & \tau^2\rho(\cdot) & \tau^2 + \sigma^2 & \tau^2\rho(\cdot) \\ \tau^2\rho(\cdot) & \tau^2\rho(\cdot) & \tau^2\rho(\cdot) & \tau^2 + \sigma^2 \end{pmatrix}$$

which is the quite familiar case of simple exponential correlation.
The second case shown in table 2 is a little more complicated as some regions are observed twice and one region is not observed at all. The incidence

| | |
|---|---|
| $y_{[1,1]}$ $y_{[2,1]}$ | $y_{[3,2]}$ |
| $y_{[4,3]}$ $y_{[5,3]}$ | |

**Tabelle 2:** KriMI Example 2

**Source:** Own diagram

matrix becomes: $Z = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ and it follows the variance covariance matrix of $y$:

$$Var(y) = \begin{pmatrix} \tau^2 + \sigma^2 & \tau^2 & \tau^2\rho(\cdot) & \tau^2\rho(\cdot) & \tau^2\rho(\cdot) \\ \tau^2 & \tau^2 + \sigma^2 & \tau^2\rho(\cdot) & \tau^2\rho(\cdot) & \tau^2\rho(\cdot) \\ \tau^2\rho(\cdot) & \tau^2\rho(\cdot) & \tau^2 + \sigma^2 & \tau^2\rho(\cdot) & \tau^2\rho(\cdot) \\ \tau^2\rho(\cdot) & \tau^2\rho(\cdot) & \tau^2\rho(\cdot) & \tau^2 + \sigma^2 & \tau^2 \\ \tau^2\rho(\cdot) & \tau^2\rho(\cdot) & \tau^2\rho(\cdot) & \tau^2 & \tau^2 + \sigma^2 \end{pmatrix}$$

The grouping according to regions is evident: observations in one region have a spatial correlation of 1, forming a covariance of $\tau^2$. On the diagonal we can find the variance of one observation, which consists of the variance of the spatial random process $\tau^2$ and the variance of the error term $\sigma^2$. The correlation of observations from different regions is described by the correlation function $\rho(\cdot)$ depending on the distance between two regions. In the kriging chapter, some correlation functions were introduced.

### 4.1.2 Bayesian Modelling

As variances are unknown, we need to develop a full Bayesian estimation to create a convenient Gibbs sampler to estimate the Bayesian kriging model. Gelman et al. (2004) propose to separate the problem in two steps: first, determine the regression coefficients with fixed variance parameters. Second, we define the variance parameters.(Gelman et al., 2004) The estimation of the Bayesian kriging model can be implemented in the multiple imputation scheme to simulate draws from the posterior predictive.[9] Here we use the mixed model representation of kriging similar to the sampling scheme of Liu et al. (2000) establishing a model for hierarchical data. Later we use the reformulation as P-Spline to get better predictions.

As it is prior knowledge, let us start with the prior on the parameters of the kriging model. For simplicity, we chose a conjugated Normal-Inverse-Gamma-model (Fahrmeir et al., 2007), which is a different choice from Liu et al. (2000) taking a flat prior for the fixed effects. In our case of regional prices it is necessary to exploit prior knowledge as the data base is very small. In our application the prior knowledge comes from all analysed products. We use the information of all products of a product-group defined by the COICOP-classification.

- The prior for $\beta$ is a normal distribution with the hyper-parameters $m$ and $M$: $\beta \sim N(m, M)$.
- For the random effect $\gamma$ we also assume a normal distribution:

  $$\gamma \sim N(0, \tau^2 R(\phi))$$

  The spatial dependencies are implemented in the correlation-matrix $R(\phi)$, which depends on the scaling parameter $\phi$, if we assume a Gaussian correlation function, which is the most simple case. If we assume another correlation function, additional hyper-parameters determining the correlation-function need to be specified.(Fahrmeir et al., 2007) It is necessary to find a parametric model for the spatial correlation to get a Bayesian estimation(Gelman et al., 2004) of the kriging model.

[9] In the Munoz et al. (2010) paper one serious mistake is made at this point: it seems that the authors do not distinguish between parameters and variables leading to confusion about the full conditionals. Munoz et al. (2010) write the Gibbs-sampler down as they draw from a full conditional of the random process after drawing the parameters of the random process. This procedure is not in the line of the imputation scheme legitimated by equation 4.

- For the error term we also assume a normal distribution which is homoscedastic and uncorrelated: $\epsilon \sim N(0, \sigma^2 I)$.
- The priors of the variance parameters are inverse gamma distributed: $\sigma^2 \sim IG(a_\sigma, b_\sigma)$ and $\tau^2 \sim IG(a_\tau, b_\tau)$.

The joint prior is:

$$\begin{aligned} P(\beta, \gamma, \sigma^2, \tau^2) &= P(\beta, \sigma^2) P(\gamma, \tau^2) \\ &= P(\beta | \sigma^2) P(\sigma^2) P(\gamma | \tau^2) P(\tau^2) \end{aligned}$$

as we assume that the parameters are a priori block-wise independent: the parameters of the fixed effects are independent from the parameters of the random effects.
Next, we specify the model for the data. Again, we assume a normal distribution:(Fahrmeir et al., 2007)

$$\begin{aligned} &P\left(y|\beta, \gamma, \tau^2, \sigma^2\right) = (2\pi)^{-\frac{p}{2}} \left|\sigma^2 I\right|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(y - X\beta - Z\gamma)' \left(\sigma^2 I\right)^{-1} \\ &(y - X\beta - Z\gamma)\} \\ &\propto \exp\left\{-\frac{1}{2}(y - X\beta - Z\gamma)' \left(\sigma^2 I\right)^{-1} (y - X\beta - Z\gamma)\right\} \end{aligned}$$

The result of multiplying the priors and the data likelihood is the posterior. The posterior of the mixed model representation of kriging is:

$$\begin{aligned} &P\left(\beta, \gamma, \tau^2, \sigma^2 | y_{obs}, X\right) \propto \left|\sigma^2 I\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y - X\beta - Z\gamma)' \left(\sigma^2 I\right)^{-1} \right. \\ &\left. (y - X\beta - Z\gamma)\right\} \left|\sigma^2 M\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\beta - m)' \left|\sigma^2 M\right|^{-1} (\beta - m)\right\} \\ &\left|\tau^2 R\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\gamma - 0)' \left|\tau^2 R\right|^{-1} (\gamma - 0)\right\} \\ &\left(\sigma^2\right)^{-(a_\sigma + 1)} \exp\left\{-\frac{b_\sigma}{\sigma^2}\right\} \left(\tau^2\right)^{-(a_\tau + 1)} \exp\left\{-\frac{b_\tau}{\tau^2}\right\} \end{aligned}$$

We need the full conditionals to create the appropriate Gibbs-sampler. The full conditional of $\beta$ is the derivation of the spatial case that can be found in the appendix.[10]

$$P\left(\beta|\gamma,\tau^2,\sigma^2,y_{obs},X\right) \propto \exp\{-\frac{1}{2}[\beta'(X'\left(\sigma^2 I\right)^{-1}X+ \left(\sigma^2 M\right))\beta] + \beta' X'\left(\sigma^2 I\right)^{-1}(y-Z\gamma) + \beta'\left(\sigma^2 M\right)^{-1} m\}$$

which is a multivariate normal with the variance $\Sigma_\beta = \left[(X'\left(\sigma^2 I\right)^{-1} X + \left(\sigma^2 M\right))\beta\right]^{-1}$ and mean vector $\mu_\beta = \Sigma X'\left(\sigma^2 I\right)^{-1}(y - Z\gamma) + (\sigma^2 M)^{-1} m$. The derivation in the appendix shows that the full conditional of $\beta$ is usually a mixture between data likelihood and the prior on $\beta$.(Fahrmeir et al., 2007)

The full conditional of the random effects $\gamma$ is a normal distribution, too:

$$P\left(\gamma|\beta,\tau^2,\sigma^2,y_{obs},X\right) \propto \exp\left\{-\frac{1}{2}\gamma'\left(Z'(\sigma^2 I)^{-1}Z + (\tau^2)^{-1}\right)\gamma + \gamma'\left[Z'(\sigma^2 I)^{-1}y - Z(\sigma^2 I)^{-1}X\beta\right]\right\}$$

Therefore, we can draw the random effects $\gamma$ from the normal distribution $N(\mu_\gamma,\Sigma_\gamma)$, where $\mu_\gamma = \Sigma_\gamma Z'(\sigma^2 I)^{-1}(y - X\beta)$ and $\Sigma_\gamma = \left(Z'(\sigma^2 I)^{-1}Z + (\tau^2 R)^{-1}\right)^{-1}$.

Fahrmeir et al. (2007) propose to use a slight informative inverse Gamma-distribution for the variance parameters $\tau^2$ and $\sigma^2$. Moreover we assume,

[10] A similar derivation of the full conditionals for a longitudinal case can be found in Fahrmeir et al. (2007).

that the variances are constant for all observations, i.e. we a priori exclude cases of heteroscedasticity. The full conditionals of the variance parameters are again inverse Gamma-distributions. For $\sigma^2$ it computes to:

$$P\left(\sigma^2|\beta,\gamma,\tau^2,y_{obs},X\right) = (\sigma^2)^{-(a_{\sigma^2}+n+1)}$$
$$\exp\left\{-\frac{1}{\sigma^2}\left[b_{\sigma^2}+\frac{1}{2}(y-X\beta-Z\gamma)'(y-X\beta-Z\gamma)+\right.\right.$$
$$\left.\left.\frac{1}{2}(\beta-m)'M^{-1}(\beta-m)\right]\right\}$$

which means that $\sigma^2 \sim IG(a_{\sigma^2}+n, b_{\sigma^2}+\frac{1}{2}(y-X\beta-Z\gamma)'(y-X\beta-Z\gamma)+\frac{1}{2}(\beta-m)'M^{-1}(\beta-m))$.
For $\tau^2$ we get:

$$P\left(\tau^2|\beta,\gamma,\sigma^2,y_{obs},X\right) = (\tau^2)^{-(a_{\sigma^2}+\frac{n}{2}+1)}\exp\left\{-\frac{1}{\tau^2}\left[b_{\tau^2}+\frac{1}{2}\gamma' R^{-1}\gamma\right]\right\}$$

That is an inverse Gamma-distribution, too:

$$\tau^2 \sim IG(a_\tau+\frac{n}{2}, b_\tau+\frac{1}{2}\gamma' R^{-1}\gamma)$$

The derivations of these full conditionals can be found in the appendix.

### 4.1.3 Criticism of Mixed Modelling of Kriging for Multiple Imputation

Munoz et al. (2010) conduct a simulation study and analyse the effect of MI under the kriging model on a salmon data set. For the simulation study the authors try different missing rates of an MAR missing mechanism. They compare the results to some other missing data methods. The authors report the more data are missing the worse all methods get, but compared to the other techniques the better the MI methods are. All MI results show relatively equal results with regards to coverage rates.(Munoz et al., 2010)

The idea to use the mixed model representation of kriging is also used by Munoz et al. (2010). In their point of view the Gaussian random field defines a „model for the variance-covariance matrix“, which does not hit the mark exactly. The linear model that underlies the kriging prediction is a mixed model. The fixed effects are the parametric trend and the random effects are the random field. In spite of this equivalence, the mixed model aims to estimate the fixed effects, which are the parameters of the parametric trend, whereas kriging is just interested in predicting values. Focusing on the mixed model approach, the estimation by a marginal model where the random effect representing the spatial correlation is located in the variance-covariance matrix is straight forward. However it does not match the idea of kriging to find the best prediction by interpolating from neighbours. A better representation of the idea of interpolation is given in the estimation by P-Splines in the next section.

## 4.2 Predicting Using P-Splines

Above we have chosen a known specification of $P(Y_{mis}|X, Y_{obs}, \theta)$ to impute unobserved values. According to Rubin (1987) the modelling of a regression can be used to parameterise the conditional posterior predictive:

$$y_i \sim N(X\beta, \Omega) \tag{37}$$

It can be generalised to $y_i \sim N(\mu, \Omega)$ and we parameterise the expectation $\mu = E(y|X) = h(\eta)$ with the response function $h(\cdot)$ and the linear predictor $\eta$.

### 4.2.1 P-Splines

If the functional form of the linear (in parameters) relationship between $X$ and $y$ is not known, it is only possible to specify the regression equation as (Lang and Brezger, 2001):

$$\eta = \gamma_0 + f_1(x_1) + \ldots + f_p(x_p) + v_i'\gamma \tag{38}$$

with $v_i'\gamma$ as the parametric part of the linear predictor.(Lang and Brezger, 2001) The model is a so-called additive model. The metric variable $y$ is explained by a smooth function $f_j(\cdot)$ which is unknown. Additionally, we assume a homoscedastic, normal error.(Lang and Brezger, 2004)

As the primary interest is not in the interpretation of the parameters, but rather to get a smooth function reflecting the data, the B-spline as well as the P-Spline approach are called nonparametric methods. We want to use the P-spline approach to implement kriging into the multiple imputation scheme. There are two reasons to do so: first, there is a link between the P-Spline approach and kriging as an exact interpolator. We will show this link in the following chapter. Second, splines can be used for an out-of-sample prediction as they are close to the data.[11]

[11] We have to distinguish this nonparametric modelling from the method described by Schafer (1999). He describes a nonparametric method to impute if no parametric assumption is made. This kind of nonparametric method is to be distinguished from the nonparametric regression approach that underlines our multiple imputation. Schafer (1999) does not mean the spline approach. He supposes to create a pool by drawing $a$ values from the respondents, if $a$ is the number of observed values in the data set of size $n$. A set of imputed values is created by drawing the $n - a$ missing values from the pool. Doing so, approximative draws from $P\ (Y_{mis}|Y_{obs})$ are made. This procedure simulates draws from a multinomial distribution. Schafer (1999) mentions two possibilities to use covariables: If the covariables are discrete, the data can be classified by them, and the proposed approximative bootstrap is conducted by every subgroup. If the covariables are continuous, Schafer (1999) proposes a logistic regression. The estimated probabilities are the grouping variables and the approximative Bayesian Bootstrap is done for the defined groups.(Schafer, 1999)

We need an identifying assumption for the functions $f_j(\cdot)$. Ordinarely, it is assumed that they are zero-mean function: $\frac{1}{range(x_j)} \int f_j(x_j)dx_j = 0$ (Lang and Brezger, 2004)[12]

According to the idea of splines, it is possible to approximate the unknown functions $f_j(\cdot)$ by a spline of order $l$:

$$f_j(x_j) = \sum_{k=1}^{d} \beta_{j,k} B_{j,k}(x_j) \tag{39}$$

with $B_{j,k}(x_j)$ as basis functions that are locally defined. Locally means that the basis functions are located between defined knots with $k = 1, \cdots, d$, and a specified number of neighbouring knots, usually $2 + l$ knots.(Fahrmeir et al., 2007, Lang and Brezger, 2004) $B_j(x_j)$ is equivalent to a $n \times k$ matrix of regressors including a number of explanatory variables. The $\beta_{j,k}$ can be interpretd as parameters of a linear predictor, hence, we can use ordinary ML-methods for estimation.(Lang and Brezger, 2004)
At every knot we try to approximate the function explaining the data by a weighted sum of basis-functions. Usually, the basis-functions are polynomial functions defined by their degree. A closer description of the basis-functions can be found in Fahrmeir et al. (2007). In the spatial case we need a two-dimensional equivalent: the tensor-product. These kind of bases lead to the so-called tensor-product or thin plate splines.(Hastie and Tibshirani, 1990, Wood, 2003) As already mentioned in chapter 3.4.2, Dubrule (1983) defines the correlation function of kriging by tensor product splines leading to a smooth estimation. This course of action leads to a smooth estimation of kriging, which is very similar to our parametrisation of kriging as a P-spline described in this chapter.

[12]This assumptions needs to be implemented in the MCMC-algorithm used later. Lang and Brezger (2004) propose to centralize the functions $f_j$ in every iteration step around 0. The posteriori needs to be unchanged. For that reason we need to add the means to the constant.(Lang and Brezger, 2004)

B-Splines tend to reproduce the data exactly making the identified function very hooked. To balance between the closeness to data and the smoothing of the estimated function, it is possible to implement penalising terms punishing for data closeness. The penalising terms account for similarity of the local parameters.(Fahrmeir et al., 2007)
The idea behind P-Splines is that the basis functions $B_k$ located at the neighbouring knots should be similar.(Fahrmeir et al., 2007) Otherwise the approximating function would connect all observation points making a wiggly function. However, the penalising term guarantees smoothness of the function. The function is smooth, if neighbouring parameters are similar.(Fahrmeir et al., 2007, Lang and Brezger, 2004)
To force the parameters to be similar it is necessary to add a penalisation term to the likelihood(Fahrmeir et al., 2007, Lang and Brezger, 2004)

$$L = l(y, \beta_1, \ldots, \beta p) - \lambda_1 \sum_{l=k+1}^{d} \left(\Delta^k \beta_{1l}\right)^2 - \ldots - \lambda_p \sum_{l=k+1}^{d} \left(\Delta^k \beta_{pl}\right)^2$$

where $\lambda_j$ is the so called smoothing parameter determining the roughness of the function. $\Delta^k$ define the $k$'th order differences of the parameters $\beta$.(Fahrmeir et al., 2007) The penalisation bases on the differences of neighbouring B-spline coefficients.(Lang and Brezger, 2004)
Equivalently it is possible to add a penalisation term to the least squares criterion:(Fahrmeir et al., 2007)

$$PLS = \sum_{i=1}^{n} \left( y_i - \sum_{j=1}^{d} \gamma_j B_j(z_i) \right)^2 + \lambda \sum_{J=k+1}^{d} \left(\Delta^k \gamma_j\right)^2$$

The likelihood has to be maximised whereas the least squares have to to minimised. For that reason the penalisation term is substracted in the former and is added in the latter formula.(Fahrmeir et al., 2007) To find the optimum of the penalised likelihood, Lang and Brezger (2004) enumerate two methods: first, the backfitting where the zero mean constraint of smooth functions is assured by centralising around zero described above. Second, the penalised likelihood can be maximised directly. Moreover, Lang and Brezger (2004) propose to choose the smoothing parameter $\lambda$ by crossvalidation. It is also possible to choose the smoothing parameter by Bayesian methods. For a description see Fahrmeir et al. (2007).

### 4.2.2 Bayesian P-Splines

To use the P-Splines in the multiple imputation scheme, we need a Bayesian perspective. To implement the penalisation we need to model it as a stochastic process.
To illuminate the idea, it is easier to start with a one-dimensional case, for example a temporal one, than to start with the two-dimensional spatial one. As we need a smooth curve, parameters of direct following periods should not differ a lot. Otherwise the estimated curve would be very wiggly. Instead the parameters should be similar, i.e. a parameter should depend on the foregoing one:(Fahrmeir et al., 2007, Lang and Brezger, 2004)[13]

$$\gamma_k = \gamma_{k-1} + u_k,\ u_k \sim N(0, \vartheta^2) \tag{40}$$

which is clearly the definition of a first order random walk. After a slight reformulation of the random walk, it becomes evident that the first order random walk defined here is analogous to the penalisation term using first differences. The parameter depends on the parameters of the foregoing period penalising for big differences. The analogy to the first differences

[13] It is also possible to define a second order random walk: $\gamma_k = 2\gamma_{k-1} - 2\gamma_{k-2} + u_k\ u_k \sim N(0, \tau^2)$ (Fahrmeir et al., 2007, Lang and Brezger, 2004)

can be seen more easy in the following formulation:(Fahrmeir et al., 2007, Lang and Brezger, 2004)

$$\gamma_k - \gamma_{k-1} = u_k,\ u_k \sim N(0, \vartheta^2) \tag{41}$$

The conditional distribution can be easily derived by computing its parameter: $\mu_{\gamma_k|\gamma_{k-1}} = E(\gamma_k|\gamma_{k-1}) = E(\gamma_{k-1} + u_k|\gamma_{k-1}) = \gamma_k$ as $E(u_k) = 0$ and $\gamma_{k-1}$ is a constant due to the conditioning on it.

$$\sigma^2_{\gamma_k|\gamma_{k-1}} = Var(\gamma_k|\gamma_{k-1}) = Var(\gamma_{k-1} + u_k|\gamma_{k-1}) = Var(u_k) = \vartheta^2$$

In the temporal case the conditional distributions can be derived from the random walk. In the two-dimensional, i.e. the spatial case this is not the case anymore, because of the mutual dependencies defined by the neighbourhood and the correlation function.
Equation 40 defines following conditional distribution:(Fahrmeir et al., 2007)

$$\gamma_k|\gamma_{k-1}, \dots \gamma_0 \sim N(\gamma_{k-1}, \vartheta^2) \tag{42}$$

The dependencies of the parameters $\gamma_k$ have the Markov-property: The conditional distribution of $\gamma_k$ only depends on the foregoing period $\gamma_{k-1}$ making the conditional expectation a constant term: $E\left(\gamma_k|\gamma_{k-1}, \dots, \gamma_1\right) = \gamma_{k-1}$ (Fahrmeir et al., 2007)
As the parameters have a distribution with location parameter depending on the parameters of the neighbours, the parameters are relatively similar, creating a smooth curve. The smoothing is done by choosing a smoothing prior.(Kimeldorf and Whaba, 1970) It is necessary to define a prior for the parameter of the first period. Fahrmeir et al. (2007) proposes to choose a non-informative one.
The joint density of the smooth effects can be derived from the conditional

distribution using the penalty matrix $K$:(Fahrmeir et al., 2007, Lang and Brezger, 2004)

$$
\begin{aligned}
P_{\gamma|\vartheta^2} &= \prod_{k=1}^{d} P(\gamma_k|\gamma_{k-1}, \dots, \gamma_1) \\
&= P(\gamma_1) \prod_{k=2}^{d} P(\gamma_k|\gamma_{k-1}) \\
&\propto \prod_{k=2}^{d} \frac{1}{\sqrt{2\pi\vartheta^2}} \exp\left\{\frac{1}{2\vartheta^2} \sum_{k=2}^{d} (\gamma_k - \gamma_{k-1})^2\right\} \\
&= \frac{1}{\left(2\pi\vartheta^2\right)^{\frac{d-1}{2}}} \exp\left\{\frac{1}{2\vartheta^2} \boldsymbol{\gamma}' K_1 \boldsymbol{\gamma}\right\}
\end{aligned}
$$

The common prior is a multivariate normal with mean 0 and the precision matrix $\frac{1}{\vartheta^2 K_1}$, where $K_1 = D_1 D_1$ where $D_1$ is a matrix of first order differences:(Fahrmeir et al., 2007, Lang and Brezger, 2004)
It is easy to adapt this to the two-dimensional case we need here. If we interpret the foregoing period as direct neighbours, we can use the conditional mean at a knot as the local linear fit.(Fahrmeir et al., 2007) Note that we still use the discrete case for simplification.
The Markov-property has to be generalised for the spatial case by Markov random fields (Lang and Brezger, 2004). The set of all regions is $D = \{1, \dots, s, \dots d\}$. $\gamma = \{\gamma_s, s \in D\}$ is a Markov random field, if the conditional distribution of $\gamma_s$ given all other effects $\gamma_r, r \neq s$ only depends on the direct neighbours. The according conditional density is $P(\gamma_s|\gamma_r, r \neq s) = P\left(\gamma_s|\gamma_r, r \in N(s)\right)$, where $N(s)$ is the set of neighbours.(Fahrmeir et al., 2007) The equivalence of the spatial to the temporal case described above is evident.
Whereas in the temporal case the conditional distribution can be derived immediately from the underlying random walk, it is not possible to do

so in the spatial case. It fails when trying to define the underlying spatial random walk due to the dimensionality.(Besag, 1974) Therefore it is conventional to define the random field by the common or by the conditional distribution of the random parameters. The conditions for a probability distribution to be a Markov random field are defined in the Hammersley-Clifford thereom.(Besag, 1974) Here we choose to define the random field by the conditional distribution and assume that the conditions are fulfilled.[14] The presentation by conditional distributions has the advantage of the direct link to Markov chains (Besag, 1974), which are important for the simulation of distributions not available in closed form.
In the simple case of normality we take a Gaussian random field:(Fahrmeir et al., 2007, Lang and Brezger, 2004)

$$\gamma_s|\gamma_r, r \in N(s) \sim N\left(\sum_{r:r\in N(s)} \frac{1}{|N(s)|}\gamma_r, \frac{\tau^2}{|N(s)|}\right). \tag{43}$$

where $|N(s)|$ denotes the number of neighbours. The conditional expectation is the arithmetic mean of the spatial effects in the neighbouring regions. The variance $\tau^2$ regulates, how much the spatial effect of one region $\gamma_s$ differs from the mean. Again, this is the stochastic form of the penalisation.(Fahrmeir et al., 2007)
Corresponding to the one-dimensional case, the joint distribution of all smooth parameters is:(Fahrmeir et al., 2007, Lang and Brezger, 2004)

$$p(\gamma|\vartheta^2) \propto \left(\frac{1}{2\pi\vartheta^2}\right)^{\frac{d-1}{2}} \exp\left\{-\frac{1}{2\vartheta^2}\gamma' K\gamma\right\} \tag{44}$$

[14] Due to Besag (1974) the most important condition is the positivity condition, assuming that the conditional distribution can be computed. The second condition is that the number of values that can be realised as one site is finite. Third the 0 can be realised at every site.

where the elements of $K$ are $k_{sr} = \begin{cases} N(s) \text{ if } s = r \\ -1 \text{ if } r \in N(s) \text{ or } s \in N(r) \\ 0 \text{ else} \end{cases}$ .(Lang and Brezger, 2004)

The matrix of neighbourships can be generalised by weighting neighbours according to their distance, to the length of their common frontier and so on. The conditional distribution is in the weighted case:(Fahrmeir et al., 2007)

$$\gamma_s | \gamma_r, r \in N \in N(s) \sim N\left( \sum_{r:r\sim s} \frac{w_{sr}}{w_{st}} \gamma_r, \frac{\vartheta^2}{w_{st}} \right). \tag{45}$$

with the symmetric weights $w_{sr} = w_{rs}$ and $w_{st} = \sum_{r:r\sim s} w_{sr}$.(Fahrmeir et al., 2007)

This leads to the global smoothness priors that define the Bayesian P-spline approach(Lang and Brezger, 2004):

$$\gamma | \vartheta^2 \propto \exp\left\{ -\frac{1}{2\vartheta^2} \gamma' K \gamma \right\}$$

with the according penalisation-matrix $K$.(Lang and Brezger, 2004)

$\vartheta^2$ is the variance parameter defining the variation of the conditional expectation. If it is small, no variation on the $\gamma_k$'s is possible. The estimation is constant in this case making the curve very smooth. The estimation gets more wiggly the bigger the variances $\vartheta^2$ are. The variance parameter can be interpretd as a inverse smoothing parameter. (Fahrmeir et al., 2007) For a full Bayes inference we need a hyper-prior for $\vartheta^2$. As it is a variance parameter the natural choice is an inverse gamma distribution.(Lang and Brezger, 2004)

$$\vartheta^2 \sim IG(a_j, b_j)$$

Lang and Brezger (2004) propose to take $a_j = 1$ and a small value for $b_j$ defining nearly a diffuse prior.(Lang and Brezger, 2004)

### 4.2.3 Kriging as a P-Spline

Cressie (1993) has already mentioned the possibility of estimating kriging via nonparametric methods. He stresses that kriging and splines are connected methods, but are used and interpreted differently. Nevertheless, he only names kernel or nearest-neighbour techniques.(Cressie, 1993)
Laslett (1994), Watson (1984) and Voltz and Webster (1990) discuss the differences between kriging and smoothing splines. According to the analysis of Laslett (1994), the former over-performs the latter under special circumstances. Voltz and Webster (1990) show the same, but refer to very small differences of the results of the two methods. However, Watson (1984) shows the similarities between the two approaches: smoothing splines and kriging are the same, if the underlying functions in smoothing are the covariance function of the model, but not all covariance functions can be used in spline smoothing. Laslett (1994) and Watson (1984) use smoothing splines, whereas we use P-splines. Following Fahrmeir et al. (2007) we show that kriging can be formulated as a P-Spline which is a different method from smoothing spline.

Here, we parameterise kriging as a P-Spline. We start with the mixed model described above:

$$y = X\beta + Z\gamma + \epsilon$$

with $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_d)$ as vector of $d$ different spatial random effects. If in all observed regions only one measurement is done, the incidence matrix becomes the identity matrix: $Z = I_d$.(Fahrmeir et al., 2007)

Now we re-parameterise the mixed model to get the connection to P-Splines (Fahrmeir et al., 2007):

$$\begin{aligned} y &= X\beta + ZRR^{-1}\gamma + \epsilon \\ &= X\beta + \widetilde{Z}\tilde{\gamma} + \epsilon \end{aligned}$$

where $\widetilde{Z} = ZR$ and $\tilde{\gamma} = R^{-1}\gamma$. As $\gamma$ still has the same distribution, the model did not change. In contrast, the interpretation of the design matrix $\widetilde{Z}$ changed:(Fahrmeir et al., 2007)

$$\widetilde{Z}\,[i,j] = \rho\left(s_i, s_j\right)$$

$\widetilde{Z}$ consists of the correlation function $\rho(\cdot)$. We interpret $\rho(\cdot)$ as basis function of a B-Spline-approach, i.e. $B_j(s_i) = \rho(s_i, s_j)$. In the case of an istotropic correlation function, especially when radial basis functions are used: $B_j(s_i) = \rho(\|s_i, s_j\|)$. The knots are the observational units. The underlying basis functions are:(Fahrmeir et al., 2007)

$$f(s_i) = \sum_{j=1}^{d} \widetilde{\gamma}_j \rho(s_i, s_j) \tag{46}$$

The parameters of all data points are considered according to their weights that are defined in the correlation matrix at every single data point. $\tilde{\gamma}$ are spatially correlated effects. The familiarities to the Bayesian P-Spline-approach are evident.(Fahrmeir et al., 2007)

Again it seems a good advice to look closely at the matrices to get a better insight into the method. As a simplification, we use a three region case as an example. In every region one data point has been observed. The correlation matrix is $R = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix} = B$, which is exactly the same as the matrix of basis-function as $Z$ is an identity matrix. In the case defined above if $Z = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ the matrix of basis-functions is $B = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}$.

The vector $\gamma' = \begin{pmatrix} \gamma_1 & \gamma_2 & \dots & \gamma_d \end{pmatrix}$ of random effects has the distribution $\gamma \sim N(0, G)$ with the covariance-matrix $G = \tau^2 R$ parameterising the spatial dependencies. After re-parameterisation $\tilde{\gamma}$ has another covariance:

$$
\begin{aligned}
Var(\tilde{\gamma}) &= Var(R^{-1}\gamma) \\
&= R^{-1} Var(\gamma) R^{-1} = R^{-1}\tau^2 R R^{-1} = \tau^2 R^{-1} \\
&= \frac{\tau^2}{3(1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23})} \\
&\quad \begin{pmatrix} 1 - \rho_{23}^2 & -\rho_{12} + \rho_{13}\rho_{23} & -\rho_{13} + \rho_{12}\rho_{23} \\ -\rho_{12} + \rho_{13}\rho_{23} & 1 - \rho_{13}^2 & -\rho_{23} + \rho_{12}\rho_{13} \\ -\rho_{13} + \rho_{12}\rho_{23} & -\rho_{23} + \rho_{12}\rho_{13} & 1 - \rho_{12}^2 \end{pmatrix}
\end{aligned}
$$

The last line is the simple case of three regions observed once. In general, $\tilde{\gamma}$ has the distribution $\tilde{\gamma} \sim N(0, \tau^2 R^{-1})$.(Fahrmeir et al., 2007)

The criterion to estimate the P-Spline is the following penalised least-squares rule: $PLS(\lambda) = (y - X\beta - \widetilde{Z}\tilde{\gamma})'(y - X\beta - \widetilde{Z}\tilde{\gamma}) + \lambda\tilde{\gamma}'R\tilde{\gamma}$ with the penalising matrix $R\left[s_i, s_j\right] = \rho\left(\left|s_i - s_j\right|\right)$ and the smoothing parameter $\lambda = \frac{\sigma^2}{\tau^2}$.(Fahrmeir et al., 2007)

Starting from the kriging side, it is possible to define the correlation function making kriging itself a spline smoothing technique. Dubrule (1983) defines the covariance function such that kriging becomes a thin-plate regression spline. We have already noted this approach when describing the covariance functions, which is in this case:

$$Cov(h) = |h|^2 log(h)$$

The similarity to the re-parameterisation of Fahrmeir et al. (2007) is evident. Whereas Fahrmeir et al. (2007) implement the correlation function explicitly in the formula by the re-parameterisation leading to correlation functions as basis functions, Dubrule (1983) defines the correlation function directly as basis function. The result is nearly the same, but with different basis functions. Dubrule (1983) notes that the difference between kriging and splines are that kriging yields an exact prediction and spline smoothing aims to find the best representation of the regression line. However, Fahrmeir et al. (2007) mention the possibility of defining a Matérn correlation function to get a Matérn-spline.
Using the thin-plate-regression-spline in the two-dimensional case defines a radial basis-function (Fahrmeir et al., 2007) as the Euclidean distance on which the thin-plates are defined defines a circular basis. Kitanidis (1997) notes that thin-plate splines are more elegant and flatter than cubic splines.
Another approach to implement smoothed spatial effects is described in Cicchitelli and Montanari (2012). The authors directly use the spatial coordinates as fixed effects and add a pseudo-variable consisting of transfor-

mation of the coordinates to construct a thin-plate regression spline as a random effect. Here the penalisation aims at the variability of the coefficients of the thin-plate-transformed random variables. If not all observed cases are used in the thin-plate-transformation, Cicchitelli and Montanari (2012) define a low-rank smoother, making estimation and prediction faster depending on the grade of simplification. Kriging itself becomes a full-rank smoother in this perspective.(Cicchitelli and Montanari, 2012) Note that this is Kriging itself, its re-parameterisation as a P-spline is something different.
Dubrule (1983) also uses thin-plate regression-splines to show the equivalence of kriging and splines.

### 4.2.4 Predictions

It is also possible to predict values of unobserved regions by the P-spline approach:(Fahrmeir et al., 2007)

$$y(s) = x(s)'\beta + f_{geo}(s) + \epsilon \tag{47}$$

where $f_{geo}(s) = \sum_{j=1}^{d} \widetilde{\gamma_j} B_j(s)$ represents the spatial effect. The basis-functions consist of the correlation function: $B_j(s) = \rho(s, s_{(j)})$. In our case, we assume an isotropic correlation function and radial basis functions making $B_j(s) = \rho(\|s - s_{(j)}\|)$. The penalising term $\tilde{\gamma}$ results from the correlation matrix $R$ and equals $\lambda\widetilde{\gamma}' K\tilde{\gamma} = \frac{\sigma^2}{\tau^2}\widetilde{\gamma}' R\tilde{\gamma}$.(Fahrmeir et al., 2007)

### 4.2.5 Choosing the Knots

When using the correspondence of kriging and splines, the knots are automatically the observed data points. In this case, the number of knots equals the sample size. Doing so, the problem is that a lot of coefficients have to be estimated. The number of knots needs to be reduced:(Fahrmeir et al., 2007)

$$D = \{\kappa_1, \dots, \kappa_m\} \subset C = s_1, \dots, s_n$$

Fahrmeir et al. (2007) propose to use covering criteria to test the quality of the subset. In the case of reduction the distribution of the vector of coefficients becomes $\tilde{\gamma} \sim N(0, \tau^2)R^{-1}$ with $R_{[i,j]} = \rho(\kappa_i, \kappa_j)$.(Fahrmeir et al., 2007)

#### 4.2.6 Bayesian Modelling and the Gibbs-Sampler

Here we derive the Bayesian perspective of the two-dimensional case. The one-dimensional case is described earlier.

The posterior we can derive for our case is a simple version of the posterior defined by Lang and Brezger (2004) as we do not have other smoothed effects than the spatial one and we do not assume an additional unstructured spatial effect:

$$P(\beta, \tilde{\gamma}, \sigma^2, \tau^2 | y, X) = L(\beta, \tilde{\gamma}, \sigma^2, \tau^2; y, X) P(\beta) P(\sigma^2) P(\tilde{\gamma}|\tau^2) P(\tau^2)$$

A simplification is that Lang and Brezger (2001) assume that the priors are independent.
For the fixed effects we choose a diffuse prior $\beta_j \sim const.$ as proposed by Lang and Brezger (2004, 2001). The prior $P(\sigma^2)$ is the inverse gamma $\sigma^2 \sim IG(a_\sigma, b_\sigma)$. As we assume a diffuse prior for the fixed effects $\beta$, conditioning on the variance parameter is not necessary.(Lang and Brezger, 2004, 2001)

In the two-dimensional, i.e. the spatial case, we also choose the smoothing prior to be normal as in the one-dimensional case of the B-spline approach:(Fahrmeir et al., 2007)

$$\tilde{\gamma} \sim N(0; \tau^2 R^{-1})$$

where $R$ is a $k \times k$ matrix and $k$ is the number of knots and the correlation matrix $R[i,j] = \rho(k_i, k_j)$. Conditioning on the variance hyper-parameters gives:(Fahrmeir et al., 2007)

$$P(\tilde{\gamma}|\tau) \propto \exp\left\{\frac{1}{2\tau^2}\tilde{\gamma}' R \tilde{\gamma}\right\}$$

This distribution equals the smoothing prior in a P-spline approach. The smoothing prior corresponds to the penalising term making the approach a P-spline. (Fahrmeir et al., 2007)
Fahrmeir et al. (2007) propose choosing a conjugate prior for the hyper-parameter $\tau^2$, which is a inverse Gamma-distribution $IG(a_{\tau^2}, b_{\tau^2})$. To get a diffuse-like hyper-prior, Lang and Brezger (2004) suggest to take $a_{\tau^2} = 1$ and $b_{\tau^2}$ small.
The full conditionals can be divided into independent blocks of parameters.(Lang and Brezger, 2004) According to Lang and Brezger (2001, 2004) the posteriors for the fixed effects parameters are normal distributions $\beta \sim N(m_\beta, Pre_\beta^{-1})$ with the precision matrix $Pre_\beta = \frac{1}{\sigma^2} X'X$ and the mean vector $m_\beta = \left(X'X\right)^{-1} X' \left(y - \tilde{\eta}\right)$ and $\tilde{\eta}$ is the part of the model depending on the other effects.(Lang and Brezger, 2001, 2004)

The re-parametrised spatial effects $\tilde{\gamma}$ are also normal with the precision-matrix that can be derived from Lang and Brezger (2004) by inserting the kriging correlation matrix as basis functions and as the penalty matrix:

$$\begin{aligned} Pre_{\tilde{\gamma}} &= \frac{1}{\sigma^2}B'B + \frac{1}{\tau^2}K \\ &= \frac{1}{\sigma^2}R'R + \frac{1}{\tau^2}R \end{aligned}$$

Ordinarily, this precision matrix is a sparse matrix and needs to be brought into the band structure needed by a reverse Cuthill-McKee algorithm. The mean vector is $m_{\tilde{\gamma}} = Pre_{\tilde{\gamma}}^{-1}\frac{1}{\sigma^2}R(y-\overline{\eta})$.(Lang and Brezger, 2004, 2001) The full conditionals of the variance parameters are inverse gamma distributions with the parameters $\overline{a}_{\tau^2} = a_{\tau^2} + \frac{rank(R)}{2}$ and $\overline{b}_{\tau^2} = b_{\tau^2} + \frac{1}{2}\tilde{\gamma}'R\tilde{\gamma}$ for $\tau^2$ and $\overline{a}_{\sigma^2} = a_{\sigma^2} + \frac{n}{2}$ and $\overline{b}_{\sigma^2} = b_{\sigma_2} + \frac{1}{2}e'e$ with $e$ as residuals for the variance component $\sigma^2$.(Lang and Brezger, 2004, 2001)

### 4.2.7 Other Approaches for Including Nonparametric Methods

In this section we describe two other methods to implement nonparametric regression methods into the multiple imputation scheme to complete the picture. First, we briefly subsume the smoothing spline implementation of He et al. (2010) being quite similar to our approach, but just using smoothing splines instead of P-splines. Second, we describe the implementation of the Nadaraya-Watson estimator by Aerts et al. (2002).

The approach of He et al. (2010) produced the idea to implement kriging via a spline approach in the multiple imputation. He et al. (2010) incorporate a so called functional mixed model into the sampling scheme of multiple imputation:

$$y_{ij} = X_{ij}\beta\left(t_{ij}\right) + Z_{ij}\alpha_i\left(t_{ij}\right) + \epsilon_{ij}$$

where:

- $X_{ij}$ and $Z_{ij}$ are the design matrices of the fixed and the random effects,
- $\beta\left(t_{ij}\right)$ and $\alpha_i\left(t_{ij}\right)$ are the fixed and random effects, and
- $\epsilon_{ij} \sim N\left(0, \sigma^2\right)$ is the idiosyncratic error term.

Moreover, the fixed and random effects are modeled as smooth functions themselves. As a smoothing spline estimation also can be formulated as a mixed model, the smooth functions of the given model can be grouped into a fixed and a random effects part leading to a re-parameterisation by which a new sorting in fixed and random effects is generated.(He et al., 2010) The linear mixed model representation of the smoothed functions can be estimated by a restricted maximum likelihood approach.(He et al., 2010)
He et al. (2010) want to conduct a Gibbs sampling to get multiple draws for the missing values. By imposing the usual conjugate priors of normality for the slope parameters, inverse Gamma for the variance parameters in the univariate case, and inverse Wishart in the multivariate case, the Gibbs sampler is formulated as the well known iterative procedure by drawing successively from the full conditionals.(He et al., 2010) He et al. (2010) address the problem of slow convergence of a simple Gibbs sampler by a blocking strategy proposed by Chib and Carlin (1999). For the blocking strategy, a group of $X$ variables and a group of $Z$ variables is built. The modified Gibbs sampler has three blocks:

1. for the $X$ variables,

2. for the $Z$ variables, and

3. for the variance components.(He et al., 2010)

In a final step, the missing values can be drawn from the posterior predictive.
He et al. (2010) show in a simulation study as well as in an example that their functional multiple imputation approach works well. In the simulation study it becomes clear that their approach is reliably among the best in a lot of different scenarios. As the functional multiple imputation approach and no other method is stably reliable, one can conclude that it is the best. Moreover, the functional multiple imputation approach is better across the two missingness patterns: a monotone and a non-monotone one. However, He et al. (2010) also name the boundaries of their method: the investigated methods were all linear ones, only ignorable missing mechanisms were taken into account, and only cubic spline smoothing has been recognised.

Aerts et al. (2002) present an approach to include nonparametric smoothing methods in multiple imputation without Bayesian theory. They solve the problem of missing data via a bootstrap while still multiply imputing the data with a local re-sampling technique. The local semi-parametric resampling method is local for the reason of assuming local conditional distributions, but, allowing nonlinear mean structures. The nonparametric approach is implemented in the distributions for the bootstrapp sample.
The data setting of Aerts et al. (2002) is quite easy: They assume a complete observed variable $X$ and a variable $Y$ with some observations missing. Both variables are connectable via a regression. Aerts et al. (2002) are interested in a consistent estimation of the mean $\mu_Y$. Aerts et al. (2002) reason that with the nonparametric approach they avoid a parametric assumption and allow a more flexible regression design.

The starting point of Aerts et al. (2002) is to sample the missing $Y$ from their local distribution $L(X) = \sum_{j=1}^{n} w_j(x) I\{Y_j \leq y\}$. This hides the included nonparametric regression a little, but a regression technique is the way to determine the distribution of $Y$ conditional on $X$. Aerts et al. (2002) describe following sampling scheme:

1. Resampling step: The data are re-sampled in a nonparametric way. If a case is missing (i.e. the missing indicator $\delta_i = 1$) a replacement $Y_i * (l)$ is created by drawing from $L(X)$.
   This step is needed to create the additional variance in the data that is needed to catch the additional uncertainty by the missing data.

2. Imputation step:
   Given the data set that was filled in step 1, imputations are produced by local re-sampling or local semi-parametric re-sampling techniques. A second distribution $L * (X)$ is constructed on the basis of the data set of step 1 for a local re-sampling, or $(\mu*, \sigma*)$ are estimated for local semi-parametric re-sampling. Next, replacements of the missing data $Y^+$ are drawn from the distribution $L * (X)$ in the case of local re-sampling or drawn from $N(\mu*, \sigma*)$ in the case of local semi-parametric re-sampling.
   Repeating both steps $m$-times, creates $m$ imputations.(Aerts et al., 2002)

The technique of Aerts et al. (2002) is a nonparametric one by the way the re-sampling scheme for $L$ and $L*$ are defined. This is done by the definition of the weights $w_j$. Aerts et al. (2002) suggest two ways to define the weights:

- mean imputation: $w_j = \frac{\delta_j}{\sum_j \delta_j}$

- kernel density estimation: $w_j = \frac{\delta_j K_h(x-X_j)}{\sum_{k=1}^{n} \delta_k K_h(x-X_k)}$ where $K$ are symmetric uni-modal density functions $K_h(u) = \frac{\frac{u}{h}}{h}$ and $h$ is the parameter defining the bandwidth.

It is possible to use different weights and different bandwidths in both steps of imputation. If closer observations get a higher weight, a Nadaraya-Watson estimator is defined in this way.(Aerts et al., 2002) Additionally to the Nadaraya-Watson estimator, Aerts et al. (2002) also briefly discuss biased bootstrap weights as an alternative.
Aerts et al. (2002) state that the estimator $L$ is consistent, asymptotic normal and, in the case of missingness, the estimator is still consistent for the conditional distribution function. In the latter case, the mean of the estimator is the same as the Nadaraya-Watson estimator at $X = x$. The variance of $L$ is a nonparametric variance estimator of $L$.
In a simulation study, Aerts et al. (2002) show that their nonparametric multiple imputation technique is robust against model mispecifications. There are negligible differences between the specifications. There are negligible differences between the local re-sampling and the local semi-parametric re-sampling, but if model specification is correct the semi-parametric methods do not rule out the parametric multiple imputation.(Aerts et al., 2002)

# 5 Prices, Price Indices and the Official Price Statistics in Germany

## 5.1 Prices

A price is the the relation of exchange of two goods: The value of the amount of good A is measured in quantity units of good B. In this case good B is the currency and the quantity units of good B are the price.(Neubauer, 1996)
For the price-statistic, the definition of goods is broad: goods are physical products, services, housing rents and shop rents. In contrast interest rates and wages are not included in the definition of goods.(Neubauer, 1996) Heske (1992) defines products, fees, rents, and rates as relevant for the price-statistic. Very important for observing prices is that prices are not a characteristic of an object, i.e. a good, but of an act of purchase.(Neubauer, 1996) Therefore the observed units should not be the good, but the purchasing act. In fact, such an observation is not possible. The trick is to define the price observation as the price that was typically relevant for a certain period in time.

## 5.2 Price Indices

Price indices measure changes of the price-level over time, which is the mean change of prices of a representative shopping basket.(Neubauer, 1996) Price indices measure the price change of an aggregate of goods.(Heske, 1992) Contrary, price measure numbers mirror the variability of prices of one product. The former is always some kind of mean (arithmetic, geometric) of the latter. The price measure number is:(Vogel, 2000)

$$M_{t,s} = \frac{p_s}{p_t}$$

The definition of the price index depends on the chosen type of averaging of the price measure numbers to create the price of the typical market basket. Here we discuss the most common index types. If we use an arithmetic mean and the quantities of a fixed basis period is used, the index is a Laspeyres type index:(Heske, 1992, Vogel, 2000)

$$P_{t,s} = \frac{\sum p_s q_t}{\sum p_t q_t}$$

The change in value of the shopping basket is measured only by the price change as quantities are fixed over a specific basis period. As the arithmetic mean is chosen, the Laspeyres index can be aggregated. The disadvantage of the Laspeyers index is that it becomes obsolete.(Neubauer, 1996, Heske, 1992)

If the quantities of products that are relevant in the reporting period and the harmonic mean is used, the index becomes the Paasche price index: (Heske, 1992, Vogel, 2000)

$$P_{t,s} = \frac{\sum p_s q_s}{\sum p_t q_s}$$

The advantage of the Paasche index is that it does not become obsolete, as quantities of the reporting period are used. Moreover it also can be aggregated due to the harmonic averaging. However the Paasche index is more complicated to observe as every period a new shopping basket is needed. It is possible to compute a Paasche index from a Laspeyeres index by inversion.(Neubauer, 1996, Heske, 1992)

Using these indices, only the price changes are taken into account as the quantities of the shopping basket are fixed both times.(Heske, 1992) The Lowe index uses mean quantities (Vogel, 2000) Another well-known index is the Fisher Index, which is a superlative index we will introduce later.

The price index should be summable: the price level changes of the aggregate should be the sum of the single price level changes of the elements of the aggregate.(Heske, 1992) The change rate denotes the relative change and the difference the absolute change. Thereby the goods and their amounts have to be temporal constant.(Neubauer, 1996) This uniformity is one of the most crucial problems when translating the temporal price-indices to spatial price-indices. A discussion of an adequate regional shopping basket can be found below.

In general there are two types of price indices: the indices with a dynamic basis period and the ones with a fixed basis. The latter are relevant for long time observations. It is possible to compute a variable basis period from a fixed one by multiplication (chain-linking) simplifying for price measure numbers:(Heske, 1992)

$$P_{0,3} = P_{1,0}P_{2,1}P_{3,2} = \frac{p_1}{p_0}\frac{p_2}{p_1}\frac{p_3}{p_2} = \frac{p_3}{p_0}$$

One problem of the temporal price comparison is that the real price changes should be measured and only the nominal change can be reflected. The quality as well as the utility value of a product can vary over time. The observed price index only reflects the nominal change.(Heske, 1992) For example a record player had a higher utility value in the past, nowadays it is only kept for nostalgia reasons. Both groups pay high prices for record players even though it has no utility value any more. This fact is also relevant for the spatial price comparison. There are products which have regional importance but no value for other regions. For example a Spätzle maker has only some value in southern Germany. To face this problem the regional shopping basket should only include products that are regionally comparable.(Heske, 1992)[15]

## 5.3 Computing with Price Measure Numbers and Price Indices

It is possible to calculate an price index from two other indices. If the indices have the same basis period $i$, and if we want to use another basis period $l$, it is possible to rebase by the following formula:(Vogel, 2000)

$$P_{l,k} = \frac{P_{i,k}}{P_{i,l}}$$

The reporting period is $k$.(Vogel, 2000)
If the Indices have variable, but directly following periods, you can use chain-linking to compute another price index:(Vogel, 2000)

$$P_{l,k} = P_{l,i} P_{i,k}$$

The mean growth rate of prices $g$ can be computed as a geometric mean of the price indices:(Vogel, 2000)

$$G = \sqrt[n-1]{P_{1,2} P_{2,3} \dots P_{n-1,n}} = 1 + g$$

[15] In the temporal case methods to measure the change of the utility value can be used (i.e. hedonic methods.).(Heske, 1992)

## 5.4 Price Statistics in Germany

The aim of the price-statistic is to value goods. There are two possible ways to evaluate: to measure the price (market value) or to measure the utility (utility value). The price statistic is interested in the former.(Neubauer, 1996, Heske, 1992) The German offices of statistics cover and describe the data of price changes of produced and consumed goods.(Heske, 1992)

### 5.4.1 German Price Survey

The German offices of statistics survey the price data to compute the relevant price indices.(Neubauer, 1996) Their aim is to represent the price changes of different consumption patterns. We are only looking at the consumer price index (CPI) which covers the changes in the living costs of private households. A Laspeyres type index is computed and the basis is adjusted every fifth year. The market basket used to compute the CPI is supposed to represent the consumption pattern of a typical German household.(Heske, 1992, Neubauer, 1996)

The price data are observed several ways. It has to be noted that the sample is not a random sample at all.
The first possibility is that the price observer visits a shop which she has chosen herself in the region chosen by the statistical offices. She has a list of products that are described in detail. In the shop the delegate chooses the products that she considers to be typical for the products on the list. The price of the product is the observation. This procedure is repeated every month with the exact same goods in the shopping basket of that region.(Neubauer, 1996) As the products are exactly the same in one region across time and the price index is a quotient, differences between observa-

tion points due to quality differences cancel each other out. In the spatial setting the differences in quality become of crucial importance, because differences do not cancel out anymore. The price statistic is representative only of the price trend in a period and not for the level itself.(Neubauer, 1996) The second possibility is that the price observer interviews people who know the prices needed.(Neubauer, 1996) For example the regional manager of a chain store. Finally, there is the possibility to get price information from the internet.

The sample is chosen arbitrarily. The idea is to have a sample of typical cases (Neubauer, 1996), rather than a random one. The expert knowledge is decisive for constructing of the sample.(Neubauer, 1996) As Neubauer (1996) enumerates, the disadvantages of a random sample for the official price statistics are: only a true random sample can be more objective. Neubauer (1996) holds the view that the required random sampling scheme would be very complicated and, therefore is indanger of manipulation. Even more relevant, the knowledge for designing a sampling scheme is missing. For practical reasons the necessary sampling size cannot be achieved. Probably the random sample will have a lot of missing values or the chosen shop is not representative. The special products of which the prices are observed cannot be at all chosen by random. The local importance also has to be considered. The disadvantages of the arbitrary sample become less significant with the number of products and with the number of observations over time.(Neubauer, 1996)

### 5.4.2 The Consumer Price Index (CPI)

In this analysis we consider the consumer price index (CPI). The statistical offices also observe other price indices, such as the producer price index.

It is also possible to use the methods developed to fill in the gaps for those prices indices.
The CPI measures the average prices that private households pay for the products needed for their lifestyle. The consumption pattern should reflect the needs of a typical household.(Heske, 1992)

The communities in the sample of the CPI are chosen by the German Federal Statistical Office and the Statistical Offices of the States (Bundesländer). The underlying idea is to cover the cities with respect to their spatial location and their citizens.
Only cities with more than 5 000 citizens are in the sample. The number of observation points in the cities depends on the number of citizens: at least eight observation points in cities with more than 100 000 inhabitants and at least four in municipalities with less than 20 000 residents.(Neubauer, 1996) There are 118 cities in the survey.(Heske, 1992)
The shops where the price data are collected are chosen by the municipalities to consider local habits. The prices of supra-local providers are collected centrally by the statistical offices themselves.(Neubauer, 1996, Heske, 1992)
For the CPI, prices are observed monthly (mid-month) and housing rents every quarter.(Neubauer, 1996) In every community the price observer chooses arbitrarily relevant shops where she surveys the prices. To observe the price, the delegate looks at the price tag or asks about the price.(Neubauer, 1996)

The goods that are in the shopping basket of the price index are described in detail by the German Statistical Office.(Neubauer, 1996) The products in the shopping basket should represent the consumption behaviour of a typical consumer. The goods are not chosen accordingly to their importance in the national accounting system. However, as they should reflect

consumption patterns, the shopping basket is designed according to the consumption pattern surveyed in the national accounts.(Heske, 1992)
It is important that the descriptions of the products in the market basket are not too exact. The office does not instruct to observe a special product. The observer can choose the regionally typical product according to the given detailed descriptions, but she observes the same products every time,(Neubauer, 1996) so that differences between the observation points cancel out in the temporal view of price indices.
There are about 750 products and services for which the prices are covered. In the end there are about 300 000 prices in the data set.(Neubauer, 1996)

The statistic offices compute a Laspeyres type index with the shopping basket adjusted every fifth year.(Heske, 1992) To get the index, means of the surveyed prices categorized by COICOP-10 numbers are computed for cities, then for the states a mean of the community prices is also computed and in the end as a weighted mean of that state price indices the nation wide prices index is computed. Weights are the number of citizens of the states. Only in the last step indices were computed, before price measure numbers[16] for the single product groups are given. At the state level a correction for quality changes is made (hedonic price index), if necessary.(Neubauer, 1996, Heske, 1992, Statistisches Bundesamt, 2013)
The weighting scheme should reflect the consumption pattern of the typical private household. The average over one year is used to identify the weighting. As the basis changes every five years, the weighting is also adjusted every five years.(Heske, 1992)

[16] Actually it is not a price measure number, but some kind of weighted mean. The weights are determined by the shop category, number of shops in the community, and a multiplicator for the region. The exact description can be found in Bayerisches Landesamt fuer Statistik und Datenverarbeitung, Sachgebiet 35, Statistik der Verbraucherpreise (04.09.2013). As we do not use the dimension time the most adjustments described there are not relevant and not possible, e.g. the seasonal adjustment.

# 6 Regional Price Levels

Only price indices of relatively large regional units can be found.(Neubauer, 1996) The reason is that the observation of the price level is designed for temporal comparisons. The temporal price-level comparison outweights the spatial one.
At the highest spatial level, i.e. the state level, we call the price level the inflation rate. The comparison of inflation rates (mostly among countries) is the most widely used spatial comparison of price levels.(Neubauer, 1996)
If we want to construct a regional price level measure, we have to be aware of several problems. These come clear when reading the criticism of von der Lippe and Breuer (2009) on the Bavarian price level comparison study of 2003.
First, using the administrative price data, we have to be aware that this data is surveyed for a temporal comparison, which is different from the required data for a spatial comparison. To overcome this difficulty, the regional market basket and the weighting scheme are chosen with care.
Secondly, the regional shopping basket is a part of the normal shopping basket, excluding all products having identical prices from all over the country. For this reason, the regional price level comparison tends to overestimate the regional price differences. This fact has to be considered, when interpreting the regional price level.
The regional classification should not influence the chosen regional shopping basket. Here we solve this difficulty by using a broad classification of the products according to the COICOP-numbers and by implementing a statistical criterion for choosing the shopping basket. The criterion has been developed by Götz et al. (2009). The weighting scheme has to conform to the shopping basket. Von der Lippe and Breuer (2009) claim that the weighted mean of the regional price indices should be the same as the officially reported price index.(von der Lippe and Breuer, 2009)

## 6.1 Regional Price Indices

There are two reasons for computing regional price indices: On the one hand one is interested to compare the price level differences of two regions. On the other hand, researcher can aim to deflate nominal quantities for inter-regional differences.(Neubauer, 1996) For the former, it is sufficient to have an price index allowing bilateral comparisons. For the latter an index with multilateral comparisons is required. Von Auer (2012) defines a multilateral price index as a function that maps the prices of all products of a shopping basket of all regions to bilateral characteristic numbers.
A bilateral price index is $P_{bg}$, where $b$ is the reporting region and $g$ is the basis region. A multilateral index consists of all bilateral price indices:

$$\begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1R} \\ P_{21} & P_{22} & \cdots & P_{2R} \\ \vdots & & & \vdots \\ P_{R1} & & \cdots & P_{RR} \end{pmatrix}$$

Several price indices are discussed in the following section. First we introduce the relevant bilateral price indices. Afterwards we describe desirable properties of regional price indices. As a third step we give an overview of multilateral price indices and how the desirable properties are fulfilled.

### 6.1.1 Definition of Bilateral Regional Price Indices

Compared to the inter-temporal price level comparisons, the inter-regional price-indices have some special features. The variability of the price index numbers can be very high because of the different consumption patterns in the regions. For this reason a lot of goods should be taken into account.(Neubauer, 1996) The products in the shopping basket need to be comparable in all regions. According to Neubauer (1996), there are three solutions to this problem. Firstly, the shopping basket can be limited to the products that are available in all regions. In that case it is questionable if the selection of goods is representative. Secondly, a shopping basket that is representative for only one region can be chosen. This reduces the comparability of the other regions. Thirdly, for every region a representative index is computed by computing the mean of the Paasche and of the Laspeyres Index. This comes at a price of loosing interpretability.(Neubauer, 1996)

Inter-regional price measure numbers can be defined by:(Neubauer, 1996)

$$M_{bg}^{j} = \frac{p_{jb}}{p_{jg}}$$

which is the price measure number of product $j$ comparing the prices of regions $b, g$, with $g$ as basis region and $b$ the report region, and the prices $p_{jb}, p_{jg}$ of $m$ different products.(Neubauer, 1996) The regional price measure number refers to a unit. To combine the single unit numbers to a regional price measure number Götz et al. (2009) propose to compute the arithmetic mean of the unit numbers in the regional unit of interest.
The $m$ price measure numbers can be combined by the arithmetic mean which defines the **inter-regional Laspeyres price index**:(Neubauer, 1996, Götz et al., 2009, Kawka, 2009)

$$L_{bg} = \frac{\sum_{j=1}^{m} p_{jb} q_{jg}}{\sum_{j=1}^{m} p_{jg} q_{jg}}$$

The basis is a national constant shopping basket.
If the reference region is the national mean, the interpretation is: the average citizen in region $b$ pays ...% more/less than the nationwide average.(Götz et al., 2009) The latter defines the Lowe-type regional price index

we are going to use. In the two basic German studies on regional prices - the Rostin (1979)- and Ströhl (1994)-study - this modified regional Laspeyres price index is computed:

$$L_{bg,mof} = \frac{\sum_{j=1}^{m} \frac{p_{jb}}{p_{jg}} p_{jg} q_{jd}}{\sum_{j=1}^{m} p_{jg} q_{jd}} = \frac{\sum_{j=1}^{m} p_{jb} q_{jd}}{\sum_{j=1}^{m} p_{jg} q_{jd}}$$

where $q_{jd}$ is the national average weighting. The modified index has slightly different characteristics.(Kosfeld et al., 2007b, Rostin, 1979, Ströhl, 1994) If we want to weight the products differently in the regions, the Paasche index has to be computed. The Paasche type index can also be constructed as the harmonic mean of the price measure numbers.(Neubauer, 1996) The different preferences of consumers can be reflected by the different weightings.(Götz et al., 2009) The **inter-regional Paasche price index** is: (Neubauer, 1996, Götz et al., 2009)

$$P_{bg} = \frac{\sum_{j=1}^{m} p_{jb} q_{jb}}{\sum_{j=1}^{m} p_{jg} q_{jb}} \tag{48}$$

According to Neubauer (1996), the Paasche-type cannot face the problem of comparability of different regions due to the varying weights, as there is no order in the regions as it is in the temporal comparison.

Kosfeld et al. (2007b) notes that the Laspeyres price index is not a cost-of-living index (COLI), but a cost-of-goods index (COGI). The COLI is better for inter-regional comparisons. A COLI „compares the minimum expenditures (...) necessary to attain the same utility level in region r as in a base area b“(Kosfeld et al., 2007a). The comparison of the economic concept, which is the COLI, and the statistical concept, which is the COGI, in the regional case is discussed by Mehnert (1997). For example, the Paasche-type index is not a COLI, too.(Koo et al., 2000)

To construct a COLI, the utility of a market basket is made constant and the relative influence of price changes is investigated. The „true COL index is the comparison of the cost of purchasing the goods and services that provide the same utility in both a comparison area and a base area" (Koo et al., 2000). As an idea of utility is the basis, the result depends on the used utility function. Mehnert (1997) describe the case of a Cobb-Douglas and the more general case of a CES utility function.

However the COLI-approach is not operational in an empirical study. The results of Mehnert (1997) using the well known Ströhl-data show that is is not necessary to use the complicated economic indices as the results of the two index concepts do not differ a lot.

The solution to the problem to construct a COLI are superlative indices as the Fisher-, Törnquist- or Welsh-Index.(Kosfeld et al., 2007b, Koo et al., 2000, Kosfeld et al., 2007a) The OECD describes these kind of indices as: „Superlative" index numbers, which were developed as part of the economic approach to index numbers.[17]

The inter-regional Paasche and Laspeyres indices can be interpreted as approximations of the COLI, but including a substitution bias as both use fixed shopping baskets excluding substitution.(Koo et al., 2000) Here we only mention the concept of superlative indices as it would go beyond this thesis.

[17] Under this approach, the microeconomic theory of producers or consumers serves as a rationale for choosing between index numbers. Diewert (1976) introduced the notion of "flexible aggregators". These are functional forms that provide a second-order approximation to an arbitrary, twice differentiable linear homogenous function. Flexible aggregators can be interpreted as functional forms that cover a wide range of utility, production, distance, cost or revenue functions. Furthermore, Diewert calls index numbers "exact" when they can be directly derived from a particular flexible aggregator. For example, the Törnqvist index is exact for the translog flexible functional form – a widely used specification in empirical economics. Thus, if one accepts a translog form as an approximation to a production function, and uses standard assumptions about producer behaviour, the Törnqvist quantity index provides an exact formulation for inputs and outputs. An index that is exact for a flexible functional form is called „superlative".(OECD, 2001)

A compromise between Laspeyres and Paasche is the **inter-regional Fisher index**, which is the geometric mean of the both:(Götz et al., 2009)

$$F_{bg} = \sqrt{P_{bg} L_{bg}}$$

According to (Götz et al., 2009), the Fisher index accounts for both comparability and regional preferences. Moreover the Fisher-Index is an exact COLI under the assumption of a homogenous quadratic utility function. An exact COLI is also named a superlative index.(Koo et al., 2000) This approach mentions Neubauer (1996), as well, highlightning the compromise between comparability assured by the Laspeyeres index and the reflection of regional consumer habits assured by the Paasche Index.(Neubauer, 1996)

If a superlative-index for the assumption of a the translog utility function is needed, we get the **Törnqvist inter-area price index**, which is the geometric mean of the price ratios:(Koo et al., 2000)

$$T_{bg} = \prod_j \left( \frac{p_{jb}}{p_{jg}} \right)^s$$

where $s = \frac{1}{2} \frac{p_{jg} q_{jg}}{\left(\sum_j p_{jg}\right)\left(\sum_j g_{jg}\right)} + \frac{1}{2} \frac{p_{jb} q_{jb}}{\left(\sum_j p_{jb}\right)\left(\sum_j g_{jb}\right)}$ as the average expenditure for a good or service $j$.(Koo et al., 2000)

The **inter-regional Lowe index** is:(von der Lippe and Breuer, 2009)

$$LO_{bg}^{R} = \sum_j \frac{p_{jb}}{p_{jg}} \frac{p_{rj} q_{rj}}{\sum_j p_{jr} q_{jr}}$$

where the index $r$ denotes the reference region. The former quotient is a price measure number of the two regions of interest. The second quotient is the weighting computed using a region of reference. This solves the problem of choosing the weights according to the regions under consideration.(von der Lippe and Breuer, 2009) If the reference region is the average of all regions, the Lowe index becomes the modified Laspeyres index.(Mehnert, 1997)

### 6.1.2 Axioms for Regional Price Indices

For an inter-regional comparison the **transitivity** is crucial.(von Auer, 2012) It „ensures internal consistency.“ (Kosfeld et al., 2007a) Transitivity means $P_{AC} = P_{AB}P_{BC}$, making chain-linking possible that is to compute a price index on the basis of two other price indices including the price index.[18] The transitivity assures that the relative position of a region is independent from the support of observed regions.(von der Lippe and Breuer, 2009, von Auer, 2012) It is necessary to make the index consistent for an interregional comparison by defining a regional structure.(Kosfeld et al., 2007b, von Auer, 2012) Moreover the transitivity simplifies analysis, because not all possible comparisons need to be computed as it is possible to conclude from one index to the other. The index is multilaterally comparable.(Neubauer, 1996, Kosfeld et al., 2007b)

The Laspeyeres index does not fulfill this requirement, whereas the Lowe Index is transitive as the used weights are independent from the considered regions.(von der Lippe and Breuer, 2009, von Auer, 2012) The easiest way to construct a transitive index is to survey prices for identical goods in every region of interest. The index is transitive in the case of a common shopping basket, because the nominator and the denominator cancel. As

[18] We defined chain-linking in chapter 5.3.

the shopping basket is not representative in this case, this is not a good solution.(Neubauer, 1996)

Another requirement, which is linked to transitivity, is the location or regional **reversibility**(Kosfeld et al., 2007b): $P_{AB} = \frac{1}{P_{BA}}$. It means that „it is invariant to the choice of s or r (here A and B) as the base region in a bilateral comparison.“(Kosfeld et al., 2007a) Regional reversibility is not reached by Laspeyres.(von der Lippe and Breuer, 2009) The modified Laspeyres-index of Ströhl (1994) and Rostin (1979) is regional reversible. (Kosfeld et al., 2007b,a) According to von Auer (2012) the regionally reversibility is a test for transitivity.

The basic requirement also linked with transitivity (von Auer, 2012) is **identity**, meaning that a price index measures correctly. The identity is: $P_{AA} = 1$. Without the identity no sensible index is possible.
If identity and transitivity hold, the location reversibility is fulfilled. As the Lowe index is transitive and has the identity property, it is also location reversible.(von der Lippe and Breuer, 2009)

It is also common to claim that bilateral comparisons should be representative for two regions. The result should not depend on other regions. This feature is called **charactericity**. It is not independent from transitivity: if an index is transitive, all intermediate indices have an influence on the bilateral comparison, hence the charactericity is violated.(Kosfeld et al., 2007b, Koo et al., 2000, von Auer, 2012, Kosfeld et al., 2007a) Therefore the decision between transitivity and charactericity is a trade-off between the two.

Von Auer (2012) also mentions that a price index should be independent of the basis chosen. All indices using a hypothetical basis fulfill the **basis-**

**invariance** property. For example the Lowe index using the average of all regions satisfies it.(von Auer, 2012)

### 6.1.3 Multilateral Regional Price Index

A multilateral price index is more than the matrix of all bilateral price indices comparing the prices of every pair. As the natural ordering of temporal price indices cannot hold in the two-dimensional spatial case, the given axioms especially the transitivity cannot be easily fulfilled. Von Auer (2012) gives a good insight into the theory and methods of multilateral price indices without the discussion of superlative price indices. The methods try to harmonise the transitivity with the charactericity. Von Auer (2012) classifies the methods in three types:

1. **Correcting** means to start with fully characterised indices for every pair of regions. In a second step, changes to the given fully characterised indices are made to reach transitivity. These changes should be as small as possible.(von Auer, 2012) A familiar method of this approach is the EKS(Koo et al., 2000) or GEKS as von Auer (2012) names it, as he adds Gini to the list of authors.
   In the GEKS method, the first step consists of computing a Fisher-index for every region. Afterwards, the Fisher indices which fully satisfy charactericity are corrected to fulfill transitivity, too. The minimisation rule is based on squared residuals of the corrected indices to the original indices.(von Auer, 2012) Koo et al. (2000) interpret this procedure as minimising the distance to the Fisher-Index:

$$EKS_{ij} = \left( \prod \sum_{K} F_{iK} \prod_{K} F_{jK} \right)$$

where $K$ denotes the basis region and $i, j$ are the regions which we want to compare.[19]

2. The second class of multilateral price indices is the class of **chain-linking**. (von Auer, 2012) Here one region is chosen to be the basis for all bilateral price comparisons. All price-indices computed are fully characteristic. To get multilateral price comparisons, all other price indices are computed as quotients or products of the characteristic bilateral price indices.(Koo et al., 2000) To reduce the computational burden, we look for the minimal number of price indices needed to conclude to all other price-indices. After all needed price indices are computed in a way that they fulfill charactericity, the rest of the price indices are computed by division or multiplication to reach transitivity.(von Auer, 2012)

3. Von Auer (2012) names the third class **harmonisation** approach. Using this approach, prices and volumes are standardised to define artificial standard units. In a second step the prices and transformed volumes are used to compute a price index for every region. The most familiar index of this group is the Geary-Khamis-index. This class leads to generalised unit values.(von Auer, 2012)

## 6.2 A Shopping Basket for Regional Price Indices

The fundamental decision for a shopping basket for regional price indices is to decide between a regionally typical or an identical market basket. Neubauer (1996) names the most important criterion for this decision: the homogeneity between the regions is decisive for the degree of regionalisation of the market basket. As we only use data from Bavaria it is not

[19] Koo et al. (2000) develop their own new price index which is tailor-made for the data they use. It is a Fisher-type multilateral index as described above.(Koo et al., 2000) As it will not fit to our data problem here, we do not discuss the new COL of Koo et al. (2000).

necessary to regionalise the market basket, because regions are not that different.
According to von der Lippe and Breuer (2009), there are three problems: firstly products have to be comparable (their quality should be equal), secondly they must be available everywhere, and thirdly they must have different prices. Reading Götz et al. (2009) a fourth criterion gets evident: The regional shopping basket should reflect the overall shopping basket as good as possible. In the following sections, we describe how we are going to solve the four problems.

### 6.2.1 Products Should Be Comparable

We have to assure that the products in the shopping basket are comparable in all regions. This aim can be achieved by choosing just the products that are the same in every region. This course of action is taken by Rostin (1979) and Ströhl (1994), reasoned by the aim of identity. Rostin (1979) argues that all products need to be perfectly identical due to the fact that price differences could be caused by quality differences. However this leads to a shopping basket that is not representative any more, if the administrative data are used.
The opposite idea is to choose a shopping basket for every region reflecting the regional consumption behaviour, but then no comparisons can be drawn at all.(Neubauer, 1996) As we use data of the official statistic, it is not possible to only take prices of identical goods and services into account. Because the same products are surveyed every year, the identity is fulfilled in the temporal point of view.(Bayerisches Staatsministerium für Wirtschaft, Verkehr und Technologie, 2003) In the spatial point of view the products are not identical, but representative, because they are chosen to be typical for the region. Moreover in a study from the Bayerisches Staatsministerium für Wirtschaft, Verkehr und Technologie (2003) it is argued

that the goods need not to be perfectly identical, because most of them can be substituted. Secondly, as the most shops are chain stores of supra-regional brands, it is likely that the products are very similar among regions. These two facts should assure that the same products are available in every German region, and therefore the regional shopping baskets should be similar.(Bayerisches Staatsministerium für Wirtschaft, Verkehr und -Technologie, 2003) We have to contradict von der Lippe and Breuer (2009) claiming for identical products and identical shops for every observation point in the survey. Firstly, some of the differences can be controlled for in the estimation model, e.g. by including a variable for the shop quality. Secondly, the regional shopping basket should reflect the regional characteristics, allowing for some differences and some substitutes. Moreover we follow the arguments of Kawka (2009) that differences of prices due to different qualities can be interpreted as white noise to some degree.

The most important factor in making products comparable is ensuring that the goods have the same quality. The differences in the prices between regions should not reflect quality gaps. It needs to be distinguished between price and quality differences. To do so, Götz et al. (2009) propose to compute an intra-regional variance and an inter-regional variance for the single products. The former compared to the latter reflects possible quality differences. If the intra-regional variance is large compared to the inter-regional variance, the prices differences are generated by the quality.(Götz et al., 2009) These products should be excluded from the analysis. However, we have to keep in mind that the shopping basket should reflect the behaviour of a typical consumer in the region justifying some degree of quality differences.(Götz et al., 2009) Not all quality differences should cancel out. The authors von der Lippe and Breuer (2009) state that the different consumers habits are respected by the construction of price representatives. There is a trade-off between the aim of representativeness and the aim of identity. For the price observation there are officially only broad descriptions of the product for which the price should be surveyed. The local characteristics can be considered by choosing a product that is typical for the region following representativeness.(von der Lippe and Breuer, 2009)

As we use data of the price statistics of the Bavarian State Office for Statistics, we only can exclude cases that violate the comparability condition. It is not possible to survey data in a way to reach identity or comparability. Following the ideas of Götz et al. (2009), we construct a criterion to exclude products with large differences in quality. If the intra-regional variance is large compared to the inter-regional variance a great amount of variability is due to quality differences as shops differ intra-regional and the structure of chosen shops is the same across regions. The variance alone of one product is not meaningful, because the nature of the product itself could cause the high variance. The easiest way is to look at the quotient of the intra- and the inter-regional variance:

$$Q = \frac{Var_{intra}}{Var_{inter}}$$

The advantage of the quotient is its simple interpretation: If $Q = 1$ the inter- and the intra-regional variance are the same, there is no hint for quality differences. If $Q < 1$, the intra-regional variance is smaller than the inter-regional variance, which suggests that the product is a good indicator for regional price disparities. We want to exclude the cases where $Q$ is much bigger than 1, as it is a hint for quality differences. Here we choose $Q > 5$ as the relevant threshold.

### 6.2.2 Products Must Be Available in Every Region

The more precise the description are of the products in the shopping basket, the easier it is to exclude quality gaps, but the more difficult it is to find

the product in every region. In contrast, von der Lippe and Breuer (2009) states that at least one representative for every shopping basket post must be observed at the regions being subject to the estimation. This problem can be solved by using broad categories of the products.
By default the international COICOP numbers are used to create the classes. The COICOP-numbers define a hierarchical classification of the shopping basket: There are 12 main groups of products, composed of 40 product groups that can be split into 106 classes of products.(Götz et al., 2009) The COICOP-classification is part of the System of National Accounts and, therefore, it is a classification of consumer expenditures. The COICOP-classes are services, non-durables, semi-durables, and durables. The classification is not only made for expenditures of households, but also for non-profit organizations serving households, and general government.(Department of Economic and Social Affairs, 2000) For the CPI only households are taken into account, and therefore we just use the COICOP-numbers 1 to 12 representing expenditures of private households.

Another way to assure that enough goods are available in every region is to broaden the spatial classification. The broader the classification the more realistic it is to get the prices of all products in the regions needed.(Kawka, 2009) We choose districts (Kreise) as spatial classification. Kawka (2009) values this as a compromise between precise regional accountability and the aim to have enough measurements per region.
This problem is our initial theme. We want to fill-in the missing regional prices to obtain complete map. The definition of the shopping basket by the offices of statistics guarantees that all products are theoretically available everywhere in Germany. For these two reasons we do not need to consider the requirement that the products need to be available in every region. However we need to assure that the regional classification is broad enough to estimate prediction models, and that products are comparable

across regions. The former is done by using districts, the latter is done by the methods described in the previous chapter.

### 6.2.3 Products Must Have Different Prices

If products do not have different prices, a statistical analysis is not meaningful. In this case it is not possible to estimate or predict. Moreover, it is not necessary to compute regional prices, because they are always the same. There are no regional price differences. If the inter-regional variance is small, the price is not varying regionally and therefore it should be excluded from the regional shopping basket.(Götz et al., 2009) When computing the price index, however, the number of products whose prices do not vary regionally need to be taken into account by weighting, otherwise the regional price differences are overestimated.

We can use some information given by results of studies described above. The city comparison study of Ströhl (1994) gives some hints: the type of product seems to be a deciding factor. For durables Ströhl (1994) cannot find considerable regional differences. The differences are more prominent for commodities, and the largest differences are seen for services. (Ströhl, 1994)
A look at the Bayerisches Staatsministerium für Wirtschaft, Verkehr und - Technologie (2003) study is helpful, because it uses similar data as we are going to use. The authors exclude products with identical prices by excluding the „Z-Positions“. These are fixed prices such as for books, medicine, or and mail order products. According to von der Lippe and Breuer (2009) the exclusion of products with constant prices leads to an overestimation of regional differences. Of course, only including those products with unequal prices overvalues the differences, but for a prediction of the regional price levels it is better to have a larger variability. It is possible to combine

the predicted regional different prices with the spatially constant prices to a shopping basket that is relevant in a second step.
To guarantee that enough variance is in the data for an estimation of predicting models, we excluded all variables with a too small variance. We computed the variance coefficient to get a measure that is independent from the scale of the products. A list of the products can be found in the appendix. The changes in the shopping basket need to be considered in the weighting scheme.

#### 6.2.4 Regional Market Basket Must Be Representative

As the choice of a shopping basket is essential for calculating of regional price indices it is worth mentioning the criteria to construct it. According to Götz et al. (2009) there are two alternatives for constructing an optimal regional shopping basket: first, the number of products in the shopping basket is determined and then the most significant products are chosen. The second criterion uses the amount of explained variation as measure of the information included in the regional shopping basket.(Götz et al., 2009)[20]

[20] Here, we do not mean the following notion of representativity: Götz et al. (2009) ask „which goods are necessary“ for a good representation of the original shopping basket. The price index of the reduced shopping basket should be a good estimator of the price index based on the whole shopping basket. The products with the best content to the overall shopping basket should be in the reduced shopping basket.(Götz et al., 2009) This kind of representivity refers to the similarity between the original shopping basket and the shopping basket of the regional comparison, and does not refer to the representivity of regional consumer behaviour.

Götz et al. (2009) define the following equation, which is chosen ad-hoc:

$$min_{G_j \subset G} 1000 \sum_{i=1}^{n} a_i \left(p_{i,j} - p_i\right)^2 + \sum_{i \subset k}^{n} Var\left(p_{ij} - p_{kj}\right)$$

$$\text{under the restriction } \#G_j = c$$

where $c$ is the chosen number of products in the shopping basket and $a_i$ is the population share of a region. The first part of the equation refers to the deviation of the single price from the regional price, whereas the second part measures the variability of the batch of products. The latter is necessary as constant batches need fewer observations. The weight of 1000 for the first part of the sum is arbitrarily chosen. It is necessary because of the different scales of the two summands.(Götz et al., 2009, Bleninger and Trojan, 15.05.2014)
For the second approach, the quality of the different regional shopping baskets are measured via the correlation coefficient $R^2$ and the rank correlation coefficient. The price index of the overall shopping basket is regressed on the price index of the reduced product bunches. Then the $R^2$ are compared.(Götz et al., 2009)

## 6.3 The Price Data

We use data provided by the Bavarian State Office for Statistics. The data at hand are price data collected for the CPI of May 2011. There are 749 different products in the shopping basket. Some products are collected centrally. Prices of these products do not differ regionally. For this reason only the prices of 607 products are in the data set. In fact we face a number of 27,418 prices.(Bayerisches Landesamt fuer Statistik und Datenverarbeitung, Sachgebiet 35, Statistik der Verbraucherpreise, 04.09.2013)

The data set comprises of the regional price information of the single product, some regional codes and characteristics of the observed product such as the amount and the measurements. Additionally there is some information about the shops where the products are bought.(Bayerisches Landesamt für Statistik und Datenverarbeitung, Sachgebiet 35, Statistik der Verbraucherpreise, 2013) The Bavarian State Office for Statistics provided as an additional information the weighting scheme. The regional level of the data refer to the 96 Bavarian counties (administrative).

In a first step, we adjusted the data: as the product description for the price observers is broad, it was not possible to compare the prices of some product groups. For example the category services for funerals included a funeral as well as burials. These product groups including things that are not comparable at all were excluded from the analysis. Moreover products were observed using different measures (e.g. litres and grams) and in different amounts. We refined the data for these differences.[21] After cleaning there were 553 different products and 25 401 single prices left in the data set.

In a second step, we have chosen those products relevant for a regional price index according to the ideas described in chapter 6.2. Firstly, we excluded all products with fewer than 20 observations, due to the fact that a statistical analysis needs some observation. A prediction based on less than 20 observed prices cannot be useful. Secondly, all products with a quotient of intra-regional to inter-regional variance greater than 5 were removed. We defined this criterion above. Its aim is to assure that price differences are not caused by quality differences. After the necessary exclusion of data using this criterion, 465 products with 21,385 price observations were left. In a last step, we eliminated all products with a variance too small for a

[21] I have to thank Sören Abbenseth for his help at this point.

spatial analysis. We decided to remove all products with a variation coefficient smaller than 0.3. In the end the data set consists of 405 products with overall 18,922 price observations.

When computing the regional price index below, we needed to adjust the weighting scheme according to the changes made in the shopping basket. According to von der Lippe and Breuer (2009), the weighting scheme of regional prices should be external. As the weights of the official CPI are external, von der Lippe and Breuer (2009) and Kosfeld et al. (2007b) propose to use them. Kawka (2009) also uses these weightings.
The weights of the regional shopping baskets can be aligned according to the expenses of consumers living in that region as it is done for the overall CPI. This idea leads to individual shopping baskets for every region reflecting the consumption habits in every region.(Götz et al., 2009, Kawka, 2009) The alternative is, as Götz et al. (2009) mention, to use only similar goods leading to a homogeneous weighting scheme. The reason for this is that we are interested in differences of real price rather than of the cost of living. In the latter case a regional shopping basket is needed, in the former case a shopping basket is needed that is identical for all regions.(Kawka, 2009) In the Bayerisches Staatsministerium für Wirtschaft, Verkehr und - Technologie (2003) study a weighting scheme is described, but according to von der Lippe and Breuer (2009) it is very problematic.(von der Lippe and Breuer, 2009)[22]
Due to the fact that the national accounts do not report consumption behavior on a level smaller than the states level (Kawka, 2009), we choose an identical shopping basket for all regions. This supports the decision to use a statistical price index instead of a COLI. To use an external scheme, we

[22] Von der Lippe and Breuer (2009) give the following reasons: It is another weighting scheme as the one of the CPI of the German Federal Statistical Office. There are different weights used in the study and it is not reasoned. There are some absurd issues: rents become more important the more communities are included in the comparison. Other weights are adapted to the survey size. There is no relationship between the number of price observations and the used weights.

decided to use the weighting scheme of the official price statistic which is based on the consumer habits in Germany. We adjusted for the changes in the shopping basket made by re-weighting for the new sum (afterwards the sum of all weights is 1) and preserving the shares compared to each other.

## 6.4 Influencing Factors of the Price Level: A Regression Model for Regional Prices

In the following section a model determining regional price levels is developed. We will define the relevant influencing factors and the best functional form. The necessary literature view and theoretical background is partly based on an article which is a joint work with Alexandra Trojan and still work in progress. The empirical modelling and data analysis are not yet completed for the the joint article. As it is required here, the empirical evaluation is conducted by the author alone, but it is planned to be implemented in the article, too. To make the relevant passages accountable, every clause that bases on the joint work ends with a reference to the paper.

### 6.4.1 Discussion of Studies on Regional Price Levels

The most important studies on regional price levels in Germany are the two linked studies by Rostin (1979) and her successor by Ströhl (1994). Both studies collect data on regional prices in German cities with the aim of a inter-regional price comparison.
The older study of Rostin (1979) uses data from 31 west German cities to collect the data of the European purchasing power comparison. Rostin and his colleagues observed 60 000 prices of 400 products that were relevant for a regional price comparison.(Rostin, 1979) Housing costs were excluded

from the analysis as it was to difficult to find comparable flats. Of course this decreases the usefulness of the study. It can be assumed that price differences are underestimated, if housing is excluded.(Kawka, 2009) The selection of cities was done with respect to their spatial distribution, population size, population density, and closeness to the border. The reporting units were chosen to be equal with respect to quality in all cities. To control for seasonally varying qualities, seasonal products were collected at a very small time interval. Rostin (1979) computes a Laspeyres price index with Bonn as reference point.(Rostin, 1979, Bleninger and Trojan, 15.05.2014) The price differentials of the analysed product groups do not have the same value. Food products, clothing, shoes, and services are more expensive in bigger cities. Durable goods as electrical products, furniture, and drugstore articles are cheaper in larger cities than in smaller ones.

The study of Angermann was published in 1989. It uses data on regional prices that were surveyed to adjust salaries of employees of international organizations such as the European Community. Rents are included in the study, but only prices of high-end flats were explored. Moreover the analysis is restricted to a small number of cities.(Kawka, 2009, Angermann, 1989) As in the Rostin-study a Laspeyers-type index with weights representing the average German consumption are taken. Angermann (1989) reports large differences in prices in the four German cities which were observed.(Angermann, 1989)

The study of Ströhl (1994) is a replication of the Rostin (1979) study.(Neubauer, 1996) The successor study was undertaken 15 year later in 1993 and the results were published in Ströhl (1994). As a lot of other studies use the Ströhl-data for their analysis (Kawka, 2009), we describe it in more detail. The collaboration of the Federal Statistical Office, the Statistical Offices of the states and of the surveyed communities sampled price data in 50 cities

in Germany. All 31 cities of the Rostin-study and additional Eastgerman cities took part at the overall German price level comparison.
In this study 367 products were in the shopping basket. Again the federal weighting scheme and the Laspeyres-type index with the basis Bonn were used.(Ströhl, 1994, Bleninger and Trojan, 15.05.2014) In fact, a Lowe-type index is constructed as the weights are federal averages and the prices are from the basis and the reporting region.(Vogel, 2000, Mehnert, 1997) In turn, housing costs are not included. The cities itself were chosen by the German federal government to represent Germany spatially. Core areas as well as economically underdeveloped areas should be taken into account. Big cities are overrepresented. The cities were grouped according to their size to make homogenous classes. These classes are mostly important for the weighting. Although the CPI should reflect the prices a consumer in a region pays and therefore the acts of purchase of inhabitants should be surveyed in neighboring regions, only prices in the cities of interest can be observed.(Ströhl, 1994) The former definition of a regional CPI cannot be surveyed by a price observation in the shops. Instead a survey following consumers is necessary. However, the official price data reflect the supply prices in one region and not the demand prices.(Ströhl, 1994)
The shops themself were chosen by the communities, which were observed by the public offices. The retail structure was described and the observed shops were classified according to it. The official statistical offices prescribed some characteristics of the shops to prevent that they were too different.(Ströhl, 1994)
There are 367 products in the regional shopping basket. The goods should be the most popular product. To guarantee comparability, the description of the product was very precise. The very important housing is again not represented in the city comparison study as it was too difficult to find comparable flats.(Ströhl, 1994)
Ströhl (1994) uses two different weighting schemes: the first weighting scheme is for computing the means of the country, states and east-west

comparisons. Here the size classification is used. The cities of one class have the same weight according to portion of people living in that class of cities.(Ströhl, 1994)
For the weighting of the shopping basket the expenditure structure of the national accounts was used. As they are not available at a regional level, the weighting scheme is a national one. For that reason only a Laspeyres and not a Paasche index could be computed.(Ströhl, 1994)

The two studies - Rostin (1979) and Ströhl (1994) - surveyed very good data for a regional price index. Nowadays the data are worthless as they only represent that time. A temporal extrapolation as done by Haupt and Heinze (2012) seems not sensible on basis of that data.
To repeat the studies is also not possible: firstly, it has to be noted that a price survey for a regional price comparison is very cost and time intensive. Secondly, it has to be realised that even the two studies themselves have problems due to their inadequate regional basis. Moreover no rural areas are taken into account. Therefore, we need to find other ways to get a regional price level.
Of course there are similar efforts in other countries. Koo et al. (2000) describe two of those surveys. The American Chamber of Commerce Research Association (ACCRA) publishes a comparison of more than 250 cities including a lot of products and also rents. The aim is to represent the cost of living of a mid-level manager by 59 items in the shopping basket. The price comparison of one item is computed as the quotient of the average price in one region and the average price of all regions. The weights of the items which base on the Bureau of Labor Statistics (BLS) Consumer Expenditure Survey are constant for all regions defining a kind of Laspeyres-index. The market basket and the sample size vary over time making a time series impossible. However Koo et al. (2000) show that the ACCRA index has three problems: a sampling error, sampling bias, and an aggregation bi-

as.(Koo et al., 2000) Another US-survey on regional prices is conducted by the BLS itself. They tried to compute an index on the basis of the American CPI. Based on this data, Kokoski, Cardiff and Moulton computed an index which Koo et al. (2000) name a true Cost-of-living index. Their high-quality data minimizes the sampling error from ACCRA-index. A regional index for 44 areas is published. The computation of the index, which is a generalized EKS index, is quite complicated, but it has a smaller aggregation bias as the ACCRA-index. It is described and discussed by Koo et al. (2000).

Walden (1998) also uses data of the ACCRA. In this very early study, he imputes regional prices and also price indices by a regression approach, using the national price level, the population size, the geographical size of region, the population growth rate, the rate of persons having a managerial or professional occupation, an educational indicator, and the tax rate as explaining factors.(Walden, 1998) More interesting than the used co-variables is that the authors use seemingly unrelated regressions to face the problem that the prices of the single goods and services are correlated.(Walden, 1998) Of course, this issue is also problematic in our analysis. As we are going to estimate the prices of more than 400 products, it is not possible to establish a SUR-system of equations.

Kawka (2009) starts his study with a literature view of a hundred years of research on regional prices in Germany. In the 19th century some surveys on regional prices of single products have been conducted. Kawka (2009) notes that already in the studies of the Statistische Reichsamt in 1930 and 1941/42, which are large-scale surveys, agglomeration areas have a higher price level than rural areas and small communities have higher prices, if they have a special reputation (e.g. touristic spots). Later analysis are restricted to special topics or regions described above.(Kawka, 2009) At the end of the section Kawka (2009) subsumes a lot of international studies on regional prices. Instead of repeating it, we subsume the result by a list of possible influencing factors identified. The list can be found in the table at the end of this section.

Later in the Kawka (2009) study the influencing factors are analysed by regression analysis. The income per capital, the wages per employee, the population density, the number of hotel over nighstays per inhabitant, the population growth rate and a east-west dummy are identified as having a significant influence on the regional price level.(Kawka, 2009) As wages are part of the income a problem of multicollinearity may arise. However income measures the demand and wages measures the supply as it is a factor cost.(Kawka, 2009)
Kawka (2009) collects his price data from several sources over a time-span of 2 years (2006-2008). Overall there are 205 products which can be classified in 57 product groups. These data represent 73.2% of the shopping basket which is used by German offices of statistics.(Kawka, 2009) The several data sources are of different qualities and the long time-span of data collections are the reasons for not using that data here. The main result of Kawka (2009) is a classifications of products according to geographical pattern. Kawka (2009) defines following groups of products:

1. Products with homogenous prices:
   Food products (except those products that can be bought at a counter or at the market, e.g. bakery or butcher), products with fixed prices (e.g. books), cars and motorbikes

2. Products with a centre-periphery difference:
   Rents and procuration fees, hotel costs, local public traffic, sport clubs, recreational facilities (e.g. swimming baths), tax on land and buildings

3. Products with a east-west difference:
   Driving school costs, special-care home costs, craftsmen costs, electricity costs, tax accountant costs, taxi journeys

4. Products with a north-south difference:
   Newspaper, education costs (e.g. costs for adult evening classes), maps, betting fees, household insurance, heating oil, charges for car insurances

5. Products without any identifiable pattern:
   Petrol/diesel, drinking water, car insurance, bank account charges, commercial advert, cinema(Kawka, 2009)

The most important result of the Kawka (2009) study is that the lower prices in East Germany result from the structural disadvantages there.
Kawka (2009) also discusses the Bayerisches Staatsministerium für Wirtschaft, Verkehr und Technologie (2003) study that is a follow up of a study conducted by the GfK in 1982. Both studies analyze the price levels in 21 cities in Bavaria using the data of official statistics. Von der Lippe (2009) criticize vehemently the methods used.(Kawka, 2009, Bleninger and Trojan, 15.05.2014) Due to the fact that the number of different products that are surveyed in the regions differ according to their size and importance, Bayerisches Staatsministerium für Wirtschaft, Verkehr und Technologie (2003) defines three different observational groups with their own shopping basket. This leads to some odd results: the ranking between two regions can change, if other regions are included or excluded from analysis.[23] This and the criticism by the von der Lippe and Breuer (2009) study are discussed later in this chapter. The study Bayerisches Staatsministerium für Wirtschaft, Verkehr und Technologie (2003) use 17 sub-indices to compute an overall-index which refers to differences from the Bavarian average. The analysis uses three classes of products and groups of surveyed cities

[23] In rational choice literature this feature is named violation of independence of irrelevant alternatives.

Bayerisches Staatsministerium für Wirtschaft, Verkehr und Technologie (2003), which creates the most problems.(von der Lippe and Breuer, 2009) The study in 2003 is important for the reason of being cited in a court decision about about the adjustment of civil service pay to regional price levels.(Kawka, 2009) The Bundesverfassungsgericht has rejected the lawsuit of the Bavarian officers.(BVerfG, 2007)

We do not face these problems as we fill-up the shopping basket with all prices by using missing data techniques before computing the price index. For this reason we do not need different groups of surveyed cities and products.

The Grimmer and Schulz-Borck (1982) study evaluates the price levels of smaller communities and in rural areas. Kawka (2009) subsumes the results: Smaller municipalities have lower prices as long as transportation costs are not included in the analysis.(Kawka, 2009) Using empirical information on prices in rural areas Grimmer and Schulz-Borck (1982) make a model calculation of transportation costs and prices on the aim of utitilty optimisation. The underlying assumption is that the location constraints of small and rural communities lead to unrealised consumer needs. The data result from an own survey of prices in rural areas including 362 randomly chosen rural municipalities and 116 goods and services of the CPI shopping basket. Grimmer and Schulz-Borck (1982) identify the community size, the centrality, the population growth and density, the spending capacity and the north-south location as driving forces of the prices.(Kawka, 2009)
Even though quite old, the analysis of Grimmer and Schulz-Borck (1982) gives an interesting and unique insight into regional prices in peripherical areas. The results show that a positive effect of population size, population density, spending capacity, and closeness to agglomeration areas. Pri-

ces in southern regions are also assumed to be higher, but Grimmer and Schulz-Borck (1982) do not find empirical evidences for this. The effect of the growth of a community cannot be supported by the data.(Grimmer and Schulz-Borck, 1982) The used variables are classes of community size, Bundesland, rural, suburban and urban zones, development of population, per capita gross domestic product, communities per capita fiscal capacity and a measurement of spending capacity. In contrast, no effect of economic structure and the size of the primary sector can be found.(Grimmer and Schulz-Borck, 1982)
Another intersting aspect that Grimmer and Schulz-Borck (1982) highlight is that in 1982 nearly all products and services are available in the rural neighbourhood. We do not share the conclusion of a gap in the possibilies of fulfillment of demand as the most sizes of the shopping basket are available. However, the data are obsolete and nowadays the coverage of stores is better than in the eighties.

There are also many studies using the data of the studies described above. A short description of those studies can be found in Kawka (2009). Here we only describe those articles that are relevant for modelling.

The Roos (2006) study predicts three kinds of price index of all counties in Germany using the data of Rostin (1979) and Ströhl (1994). The price indices are the price index of services, the price index of goods, and the price index of durable goods. Roos (2006) imputes the missing price-level information by a regression for which he reports a quite good model fit.(Bleninger and Trojan, 15.05.2014)
Roos (2006) chooses his covariables according the following argumentation: if we assume a spatially segmented market structure without strategic market behaviour, immobile market participants and no transportation costs, the regional prices are determined by local demand and supply.

Then looking at the influencing factors of the demand side: it is determined by the local income and local preferences. Across Bavaria and presumably across whole Germany it is likely that preferences do not differ that much. The local income can be measured by the regional per capita income. The regional prices of the supply are determined by the local factor costs, which are retailing costs for trading goods, regional wages, rents for retail outlets, and the degree of local competition. Roos (2006) measures the demand side by the regional average wage level, the population size, the population density and a dummy for tourism. Tabuchi (2001) also reasons that the demand side is influenced how attractive a place is for tourists. He argues that the higher number of tourists and businessmen in cities, the higher the reservation price, which rises the price-level, of course.(Tabuchi, 2001) There are no variables representing local amenities and the competition structure. For the latter Roos (2006) tries to find empirical evidence, but he notes that its influence cannot be proven.(Roos, 2006, Bleninger and Trojan, 15.05.2014) In contrast, in a study about the influence of the competition structure on food prices in Sweden Asplund and Friberg (2002) find a significant effect. The authors measure the competition structure by the number of shops, size of shop and the branch affiliation structure. However, later in the analysis Asplund and Friberg (2002) state that the results depend on the rigid identity of the products that are compared.(Asplund and Friberg, 2002)[24]

The most important immobile good is housing which is very problematic, because comparable flats need to be found. As housing has a large weight in the shopping basket of German official statistic, the differences have a great influence on the price-level differences. We cannot prove whether the high differences of housing prices are caused by quality differences

[24] Asplund and Friberg (2002) use very rich data on food prices: there are about 1000 stores across Sweden with about 30 items. The analyzed products are precisely defined and exact the same products are surveyed in the shops. Second, a less exact data set is used. Asplund and Friberg (2002) also take 732 price observations in 40 towns, which are data of the Swedish government, into account.

(then we have a measurement problem) or by the impossibility of arbitrage.(Roos, 2006) This is important to bear in mind for the following analysis.
The two studies, Roos (2006) and Tabuchi (2001) complement each other. Tabuchi (2001) gives an equilibrium theory, which follows the same argument as Roos (2006), who explores the question empirically. In contrast, Götz (2012) tries to derive conditions of a market equilibrium. His analysis is quite theoretical and uses very strong assumptions. He underlines the influence of the degree of product differentiation and the intensity of the competition in the regional market. The analysis of Roos (2006) gives reason to doubt the influence of the competition structure.(Bleninger and Trojan, 15.05.2014) Here it is not possible to test the effect of the competition structure, because we do not have precise data. According to (Asplund and Friberg, 2002) we need exact data to analyse the effect of competition structure on price data.
Instead of a very theoretical model, Roos (2006) chooses those variables that are correlated with the price index and assumes a simple linear regression model:

$$\begin{aligned} CPI = & \alpha_0 + \alpha_1 \text{popsize} + \alpha_2 \text{GDP} + \alpha_3 \text{income} \\ & + \alpha_4 \text{retailrent} + \alpha_5 \text{density} + \alpha_6 \text{tourism} \\ & + \alpha_7 \text{east} + u \end{aligned}$$

where $u$ is a normal zero mean error term. Even though the model is quite simple, Roos (2006) reports a good model fit and good predictions measured by crossvalidation.(Roos, 2006, Bleninger and Trojan, 15.05.2014)

The already mentioned study of Asplund and Friberg (2002) on Swedish food prices stresses the importance of shop characteristics. The store type and size as well as the affiliation is controlled in the regression model.

Beside regional characteristics as regional income, wages, and population density (defined as a proxy for housing costs) a random effect on the store level is included in the anaylsis. However the emphasis of the study is the influence of the market structure on food prices making the study only comparable to a small part of our results as we are going to look at all prices of the market basket.(Asplund and Friberg, 2002)

Following economic theory, Kosfeld et al. (2007b) derive two regression models, one for housing rents and one for prices of consumer durables. The authors enrich utility and production functions under a maximization calculus with demographic characteristics. A more precise description can be found in the discussion paper Kosfeld et al. (2007a). On the demand side the ordinary demand-functions arising from the utility maximisation under the budget constraint are enriched by demographic characteristic using translation. If a Stone-Geary-type utility function is assumed, the established demand system becomes linear. From the supply side, a neoclassical production function is used where prices depend on regional characteristics, too.(Kosfeld et al., 2007a)
Even though the derivation of the models is strictly economic and quite complicted(Kosfeld et al., 2007a), it leads to a simple linear regression. Looking for an equilibrium gives the two econometric models: one for housing and one for consumption costs(Kosfeld et al., 2007b, Bleninger and Trojan, 15.05.2014) Here we only need the prices of durables:

$$CPI = \alpha_0 + \alpha_1 \text{income} + \alpha_2 \text{overnight} + \alpha_3 \text{wage} + \alpha_4 \text{density} + u$$

for the first three parameters Kosfeld et al. (2007b) assume a positive effect, whereas the direction of $\alpha_4$ cannot be assumed. The purchasing power is measured by disposable income and increases withthe demand of travellers measured by the number of hotel overnight stays. The concentration of demand is reflected by the population density.(Kosfeld et al., 2007a) As

wages, income and prices are simultaneously determined on the market, wages and income cannot be seen as strictly exogenous. To solve the endogeinity problem, Kosfeld et al. (2007b) propose the classic solution of an instrumental variable approach with the possible instruments: overnigth stays, population density, population size and human capital. The estimation proposed is a 2SLS.(Kosfeld et al., 2007b, Bleninger and Trojan, 15.05.2014, Kosfeld et al., 2007a) In our analysis later, it is not possible to consider the endogeneity problem, because there are no solutions to include it into the multiple imputation.
In a second step, Kosfeld et al. (2007a) update the data into future using inflation rates of the Bundesländer, to get regional prices beyond the year 1993. From our point of view, such an approach has too much uncertanties. For example it needs to be assumed that the regional price level changes parallelly to the state inflation.
Again the Ströhl (1994)-data were used as price data. For the explanatory variables time series data of the regional statistic division of the federal bureau of statistics and the federal bureau of labour are used. The available income is measured by the purchasing power as all other available data do not provide a complete time span. The wage level is represented by the average wage. The purchasing power of tourists is measured by the number of hotel overnights stays. The population density is included as demographic characteristic. The authors develop a panel dataset by predicting using regression analysis. The results are adjusted by the inflation rates of the Bundesländer. Afterwards the two estimated price indices are combined by weighted means. All parameters are significant and have the assumed direction.(Kosfeld et al., 2007b, Bleninger and Trojan, 15.05.2014)

Suedekum (2004) also derives conditions for a regional equilibrium from economic theory. He enriches the model of the new economic theory by housing to explain price differences between agglomerated areas and pe-

ripherical regions. The analysis only includes two regions to derive the conditions of a symmetric and of a full agglomerated distribution between the two regions. The formula and equations cannot be used to infer an estimation equation, but there are some hints towards it. First Suedekum (2004) names some variables that could be relevant: housing costs, size of the local population, number of jobs, nominal wage per employee, and per capita income. Moreover Suedekum (2004) takes a closer look at the supply side. To measure its specification, the number of producers, the number of farmers, and the supply of housing should be evaluated. Moreover he mentions that there is a relationship between the center-periphery differential and the housing costs.(Suedekum, 2004) We are going to consider this by including an interaction term of the two. The next problem that Suedekum (2004) mentions is that all variables measuring income and wage are not strictly exogeneous.(Suedekum, 2004, Bleninger and Trojan, 15.05.2014)
Blien et al. (2009) enhance the imputation idea of Roos (2006) by using multiple imputation to acount for the unceratinty of missing data. The authors analyze the agglomeration wage gap in West Germany. To distingiush between nominal and real wage Blien et al. (2009) need the regional price level. Blien et al. (2009) do not discuss their operationalization in detail. The model for multiple imputation is:(Blien et al., 2009, Bleninger and Trojan, 15.05.2014)

$$\begin{aligned} CPI = &\alpha_0 + \alpha_1 \log \text{size} + \alpha_2 \log \text{popsize} + \alpha_3 \log \text{landprice} \\ &+ \alpha_4 \log \text{landprice}^2 + \alpha_5 \text{toursim} + \alpha_6 \log \text{unemp} \\ &+ \alpha_7 \text{dummy}_{\text{district}} + \alpha_8 \text{dummy}_{\text{regtypes}} + \alpha_9 \text{wage} + u_i \end{aligned}$$

where the variables are:(Blien et al., 2009, Bleninger and Trojan, 15.05.2014)

- log size is the logarithmic territorial size of the regions
- log popsize is the logarithmic population size

- log landprice is the logarithmed price of building land
- tourism is an indicator of how attracttive a place is to tourists. It is computed as the number of overnight stay divided by the number of arrivals of visitors multiplied with the population density
- log unemp is the unemployment rate
- $\text{dummy}_{\text{district}}$ is an indictator variable for cities that are a region of their own
- regtypes are dummies to include the region type of the Federal Office for Building and Regional Planning (BBR). These characteristics allow us to distinguish between city centre and periphery, agglomerated and sparesly populated areas.[25]
- wage is the composition corrected wage (the correction bases on a Mincer-type wage regression: only the residuals of a wage regression based on human-capital theory are included in the imputation model. The explained part of the Mincer-type equation refers to other characteristics.)(Blien et al., 2009, Bleninger and Trojan, 15.05. 2014)

The only reason given to use the logarithm is that several models have been computed and this one is the best.(Blien et al., 2009)
Using the Ströhl-data for 32 West-German cities, $m = 5$ imputations are done. The analysis afterwards underlines the importance of having regional prices. Without regional prices it is not possible to correct sizes of interest, e.g Blien et al. (2009) study real wages. Only the comparison of real and nominal wages allows to answer the question of the agglomeration wage differential.(Blien et al., 2009, Bleninger and Trojan, 15.05.2014)

[25]The region types can be found on this web page: *http:// www.bbsr.bund.de/ BBSR/ DE/ Raumbeobachtung/ Raumabgrenzungen/ SiedlungsstrukturelleGebietstypen/ Regionstypen/ regionstypen.html*

The anaylsis from Mehnert (1997) tries to answer a similar question with the emphasis on a comparison of the economic and the statistical concepts of price indices. The results show that we do not need to use some kind of economic price index, which is very difficult to measure empirically, because results do not differ that much.(Mehnert, 1997)

The most recent study is the research report of Haupt and Heinze (2012). They also use the Ströhl-data to predict values of all 439 counties in Germany and extrapolate for the following twelve years using information about inflation rates of the Bundesländer and size of region, gross domestic product and unemploment rate on the regional level. It is disputable to predict 439 values along 12 years on the basis of the 51 price-level observations of the Ströhl-data.
To develop an econometric model Haupt and Heinze (2012) start with the CPI formulae with a fixed basis period, use the logarithm and derive a linear model. Due to the fact that the description in the reasearch report is very short and the authors often refer to a PhD-thesis not published yet, it is not possible to figure out the modelling.[26] More serious problems arise with the inherent assumption made: all changes are assumed to be stationary, all regions have the same covariance structure in time, changes in the Bundesländer over time are the same as in the counties, and, most remarkable the approximation error which is constructed via the econometric modelling is assumed to be zero.(Haupt and Heinze, 2012) It means that Haupt and Heinze (2012) assume that their estimation is always correct, which is a remarkable assumption.

As there are doubts in the quality of the prediction, we do not interprete the results. What can be learned from the study is, that we have to keep

[26]e.g. if there are some $\log$s missing or the authors estimate a random effect without including the respective fixed effect.

the contraints of our predictions and our assumption in mind when using the lean database.

However in Haupt and Heinze (2012) there is a good hint for the variable selection. All covariables that we use in the model to explain the prices need to be invariant to the population size of the region. Only if this is the case, is the analysis invariant to pooling of two regions. Using the total numbers, parameters would change, if we pool two regions, as prices (the explained variable) do not change and region characteristics changes (the explaining sizes).(Haupt and Heinze, 2012)
Table 3 summarizes the variables used in the studies previously discusssed.

| **Roos (2006)** | **(Haupt and Heinze, 2012)** | **Kawka (2009)** |
|---|---|---|
| average annual wage<br>population size<br>population density<br>dummy for tourism<br>GDP per capita<br>rental rate of retail outlets<br>east-west dummy | inflation rates of Bundesländer<br>regional size<br>per capita gross domestic product<br>unemployment rate<br>east-west dummy | income per capita<br>wages per employee<br>population density<br>number of hotel over night<br>stays per inhabitant<br>population growth rate<br>east-west dummy<br>regional classification<br>centre-periphery differential<br>north-south differential |
| **Grimmer and Schulz-Borck (1982)** | **Tabuchi (2001)** | **Suedekum (2004)** |
| community size<br>centrality/zone<br>population growth<br>population density<br>spending capacity<br>north-south location<br>Bundesland<br>per capita gross doemstic product | number of tourists<br>number of businessmen | housing costs<br>size of local population<br>number of jobs<br>nominal wage per employee<br>per capita personal income<br>number of producers<br>number of farmers<br>supply of housing |

| **Walden (1998)** | **others** cited by Kawka (2009) | **Blien et al. (2009)** |
|---|---|---|
| national price level<br>population size<br>regional size<br>population growth rate<br>number of managers<br>education<br>tax rate | population size<br>population densitiy<br>population growth rate<br>human capital<br>tourism<br>degree of urbanization<br>central or peripheral location<br>spending capacity<br>north-south location<br>wages per employee<br>income per capita<br>rents<br>regional amenities | population size<br>geographical size<br>land prices<br>tourism<br>unemployment rate<br>regional type (centre-periphery)<br>district as one region<br>composition corrected wage |
| **Kosfeld et al. (2007a)**<br>purchasing power<br>average wage<br>number of hotel overnightstays<br>population density | **Götz (2012)**<br>degree of product differentiation<br>competition structure | |

**Tabelle 3:** Summary of Influencing Factors Mentioned in Literature

### 6.4.2 The Data for the Influencing Factors

We use data from two data sets to get information about possible influencing factors of the regional prices: the Inkar-data provided by the Federal Institute for Research on Building , Urban Affairs and Spatial Development (Bundesinstitut für Bau-, Stadt- und Raumforschung, BBSR) and the spending capacity data as well as some base numbers of the GfK-Geomarketing.

Inkar is an abbreviation for „Indikatoren und Karten zur Stadt- und Raumentwicklung“, which means indicators and maps about the urban and spatial development. It includes about 600 variables on different regional levels of social, geographic, and economic development in Germany. The data originates from several public data sources.

GfK is one of the largest marketing and market research enterprises in the world. Its Geomarketing division provides data on the economic development on a regional level, too. Their data results from public as well as commercial data sources. Especially the data on regional spending capacity are of interest, as it includes more than just the local wages.

### 6.4.3 Identifying the Influencing Factors on Regional Prices

If we subsume table 3, we do not need that many factors as it may first appear.

**Income and Spending Capacity:**
We need to identify the wage per employee and the income per capita. The wage per employee is already in the Inkar-data set. Nevertheless it is dif-

ficult to distinguish the wage variable from the income variable. However it becomes clear when realizing the the wage reflects the factor costs making it a part of the supply side, whereas the income per capita measures the possible demand. The wage per employee is published in the Inkar data set. The income per capita should measure the ability of inhabitants to fulfill their needs, which identifies it as spending capacity. The spending capacity is part of the regional data of the GfK data. The spending capacity is defined as the prognosis of the regional consumption potential at the residence of the consumer which is the sum of all net income revenues. It is the income of all freelance and wage earnings, income from capital, and public transfer payments. Data from the wage and income statistics, statistics for public transfers, and a prognosis of economic research institutes are used.(GfK Geomarketing, 2013) We decided to use the data per capita and not per household.

**Housing:**

Housing costs are measured in a twofold manner. Housing rents are part of the CPI and for that reason cannot be used as explaining factor. We need to measure the market structure of housing. This can be done by including prices for building land, which is the same as to take some costs of investments into account. To measure the supply of housing an index was computed:

$$\text{housing index} = \frac{\text{number of vacant flats}}{\text{number of all flats}}$$

The idea is to measure the competition on the housing market: the higher the competition the higher the prices of housing, reducing the spending capacity for buying other products. On the other side, it could be possible that prices are higher due to the higher cost pressure. The data on the number of vacant flats and on the total number of flats are surveyed on a different regional level: some towns are reported independently from their counties, others are reported together with their counties. As the data are not missing, but only available on a larger scale for 17 counties, we decided to use the information for both, the county and the city.[27]

**Community Size:**

There are two ways to measure the community size: first, there is the geographical size. The area of the region is computed using the the R-package UScensus2010. It facilitates computation of the area of regions defined by a shapefile that bases on vectors. The shapefile can be downloaded from the Bayrische Vermessungsverwaltung.(Bayerische Vermessungsverwaltung, 2015, Almquist, 2014) Secondly, the population size is an interesting measure of the size of the regions. Of course we do not use the total amount, but the population density, due to the idea of Haupt and Heinze (2012). There are two highly correlated variables of the population density (r=0.9998). We decided to use the population density measured by the Census in 2011.

Another characteristic depending on the population size is the population growth rate. We measure the growth rate with the net internal migration, which is also given per capita.

**Tourism:**

The touristic attraction of the regions is reflected in the number of hotel overnight stays, which is a frequent operationalization. We find those in the Inkar-data set, where we also found information about the regional unemployment rate.

[27]There is one date missing: there is no information on the prices of building land of Bamberg. As it is just one date point, it does not seem necessary to use elaborated methods to fill in the gap. Instead we follow the idea, that the relationship between the rate of the price of building in the city of Bamberg and the price of building land of the county of Bamberg is nearly the same as the rate between a similar city and county. As Würzburg is a similar city, we use its information to single impute the price of buidling land in Bamberg.

**Centre-Periphery:**

The centre-periphery location is measured by the classification type. It has three classes: rural areas, areas with urbanised structures and urbanised areas. Moreover the rurality is measured by the following indicators: a small population density, missing centres, peripherical location, and a high number of people working in the primary sector.(Grimmer and Schulz-Borck, 1982) We also include the number of commuters, who also may have influence the prices according to Tabuchi (2001). Transportation costs are included by the distance to next regional metropolis.

**Amenities:**

It is very difficult to reflect the regional amenities that influence the regional price level. The regional human capital can be measured by education. To do so we included the percentage of highly-qualified persons in the equations. Another characteristic that we want to reflect is the occupational composition. We do that by using the classic grouping in the three occupational sectors agriculture (primary sector), industry (secondary sector), and services (third sector) that can be found in the Inkar- data set. We include the percentages of people working in the sectors. We only inlcude the secondary and third sector in the analysis, as from this it is possible to work out the percentage working in the primary sector.
We want to use the age structure as a third characteristic. To reflect the age composition we computed the mean, variance and skew of the age distribution using the Inkar data.

**Store Quality:**

In the von der Lippe and Breuer (2009) study the authors propose to estimate a location dummy-variable model to take the different qualities due to the particular shop into account (e.g. delicatessen store or discount supermarket). As we want to estimate the prices of the individual products

and not of groups of products there are only a few observations for each product making a location dummy-variable impossible. The same is true for a random intercept model and in particular for the econometric fixed effect model. It is also not possible to include a random or a fixed effect for the individual stores, which is proposed by Asplund and Friberg (2002), due to the small sample size per product. To control for the qualities of the shops, we used the information about the shops provided by the Bavarian State Office for statistics. There is a variable measuring the shop quality. Of course this information is only available for the observed data, and therefore cannot be used directly for making predictions. For prediction of regional prices, we use the store quality that is the mode of the observed data. There is no data available on shop size and affiliation as Asplund and Friberg (2002) include it in the regression model.
It is not necessary to estimate a multilevel model as we only have data from Bavaria. If data for Germany are of interest, a random effect is necessary to account for the Bundesländer.

Kawka (2009) notes that the influencing factors interfere with each other. For that reason we need to include interaction terms. The following interaction terms may have an influence: centrality may interact with housing costs (the idea to do so comes from Suedekum (2004)), wage, prices for building land. The wage variable and the income variable are also included with an interaction effect. Prices for building land and housing costs are included in the same way. The number of hotel overnight stays may interact with the indicator of being a county or a city.

Suedekum (2004) notes that there is an endogeneity problem for all variables measureing income and wage. For the latter problem he proposes an instrumental variable apporach, which Kosfeld et al. (2007b) solves with a 2SLS. It is not possible to solve the endogeinity problem in our analysis as a solution that works with multiple imputation needs to be developed, which goes far beyond of our work.

There are no data available to measure the competition structure of the markets as it was done by Asplund and Friberg (2002). However, their results show that using broader products categories makes the effect of competition structure vanish. To balance comparability and representativity, we defined broader product categories. Moreover, it is not assured that the effect really exists.(Roos, 2006) We measure the competition structure indirectly by representing the demand side.

The table on the next page subsumes the used variables and the influencing factor that is measured by them:

| **variable** | **variable name** |
|---|---|
| county size | area<br>popdens<br>mig |
| number of employees | emp |
| supply/factor costs | wage |
| demand | kk_ew (spending capacity) |
| real estate market | building<br>housing_index |
| number of hotel overnight stays | hoteln |
| unemployment rate | unemp |
| regional amenities | hi_quali<br>sec_sector<br>tert_sector<br>age_mean<br>age_var<br>age_scew |
| number of businessmen | commut |
| distance to regional metropolis | centre |
| shop quality | GKat (1=department store,<br>2=convenience store,<br>3=super market,<br>4=specialist discounter,<br>5=specialist store,<br>6=remaining retail (fuel stations, kiosk...),<br>7=service companies,<br>8=internet,mail order selling) |
| regional classification | Aggregat (urban municipality, county)<br>typ (1=urban region,<br>2=regions tending to urbanization,<br>3=rural region) |
| interaction terms | building*typ<br>wage*typ<br>wage*kk_ew<br>Aggregat*hoteln |

**Tabelle 4:** Influencing Factors Taken into Account

**Source:** Own research

# 7 Prediction of Regional Prices

In a first step, we predict the prices of all products in the shopping basket in all regions to get a full data set. We predict with several methods: first, we use single conditional mean imputation to get a baseline. Afterwards we do the same with multiple imputation. To include spatial dependencies, we predict missing price data with universal kriging. Last we want to preserve spatial dependencies as well as variance estimations. Therefore we predict with a multiple imputation scheme with implemented kriging (KriMI).

In a second step we compute the regional price index for the imputed data sets. As we only have the weighting scheme of the official CPI and no information of regional consumption patterns, it is only possible to compute the inter-regional Lowe index with an artificial reference region. This is the same as the modified Laspeyres index of Ströhl (1994). We identified some products that are not relevant for the regional price comparison as discussed above. The weights have been adjusted for this change of the shopping basket in such a way, that they again sum up to one by preserving the portions.

## 7.1 Single Imputation

The single imputation is done by conditional mean imputation. The conditional means are computed by using regressions that we derive in this chapter.

### 7.1.1 Regression Models on Regional Prices

As it is reported that effects on different prices may differ, we split analysis among groups of products. Due to the fact that we still have about 400 different products after selection of relevant products, it is not possible to find a model for each product. Instead we use the COICOP-classification with two digits, defining 12 broad categories, which have the following sample sizes after cleaning the price data:

- Food and non-alcoholic beverages, $n = 6143$
- Alcoholic beverages, tobacco and narcotics, $n = 310$
- Clothing and footware, $n = 1103$
- Housing, water, electricity, gas and other fuels, $n = 2513$
- Furnishings, household equipment and routine household maintenance, $n = 1997$
- Health, $n = 434$
- Transport, $n = 1991$
- Communication, $n = 41$
- Recreation and culture, $n = 2019$
- Education, $n = 24$
- Restaurants and hotels, $n = 746$
- Miscellaneous goods and services, $n = 1601$

To find the underlying functional form, we decide to take a look at the data: a generalised additive model on the the standardised prices was computed for every defined subgroup.[28] It was necessary to use the standardized prices, as within the groups the prices of different products have very different levels and ranges. The models identified in this manner will be used in the

[28] An exact description of that method can be found in Hastie and Tibshirani (1990) and Wood (2004)

prediction model for every single product. We regress on the prices not on the standardized prices when predicting.
As there were still too many parameters for an estimation and prediction of the single products compared to the sample size, we needed to identify the most relevent effects, even though multiple imputation needs to take as many effects into account as possible. To find the most relevant effects for every COICOP-class, we started with the full model including all polynoms and all possible interaction effects filtered in the literature analysis given above. Then we tried to find the best predictors by backward deletion. In the end we decided on the models which had only significant parameters.
We do not report the exact parameters as they are not used for predictions. We predict only on the basis of the estimations based on single products.

Doing so we identified following model for food and non-alcoholic beverages:

$$\begin{aligned}\text{price} =& \beta_0 + \beta_1\text{age\_skew} + \beta_2\text{area} + \beta_3\text{wage} \\ &+ \beta_4\text{popdens} + \beta_5\text{mig} \\ &+ \beta_6\text{commut} + \beta_7\text{commut}^2 + \beta_6\text{GKat} + \beta_7\text{typ}\end{aligned}$$

The table 11 reported in the appendix shows the results of the backward-deletion strategy for decreasing the numbers of parameters in the food category. The effect of the skew has a negative sign, indicating when more older people live in the region, the lower the prices are. Both measures of size, area and population density have a positive effect on price. The former effect does not confirm the analysis of Grimmer and Schulz-Borck (1982), whereas the latter does. The first can be explained by less competition in larger regions leading to smaller competition between shops which has the effect of higher prices. The second can be explained by higher demand raising prices too, because people cannot substitute buying food.

As wage is a factor cost, we expected a positive sign which was affirmed. The store quality is always cheaper than the reference category (warehousing) except for class number six (filling stations, kiosk) etc, which needs not to be discussed. Both regional types are more expensive than the city. This result is also found by Asplund and Friberg (2002). Perhaps this fact can be explained by higher transportation costs or smaller competition in less agglomerated regions. The only astonishing effect is the net internal migration. It is possible that when more people migrate to a region, the lower wages are and the lower the prices are as wages are a production factor. Contrary to our results, Asplund and Friberg (2002) report significant effects of population density, distance to central areas, and income. The data that these authors use are quite different and more important, their sampling size exceeds our sampling size enormously.

The equation for alcoholic beverages, tobacco and narcotics is:

$$\begin{aligned}\text{price} = &\beta_0 + \beta_1 \text{emp} + \beta_2 \text{wage} + \beta_3 \text{building} + \beta_4 \text{building}^2 \\ &+ \beta_5 \log \text{popdens} + \beta_6 \text{hoteln} + \beta_7 \text{commut} + \beta_6 \text{GKat}\end{aligned}$$

As for all the other outputs, the output of this class of products can be found in the appendix. There are not a lot things to interprete in the model for alcoholic beverages, tobacco and narcotics. The demand represented by the population density and the number of hotel overnight stays raises the prices. The negative effect of wage and the number of employees cannot be explained that easily. Looking at the graphs does not help in the case of number of employees as it seems to have a positive effect. The size of the effect is significant, but not very relevant as it is almost zero. The same can be said for the small effect of wage and the quadratic term of the prices of building land, too. The effects of the shop qualities are as expected.

The model of clothing and footware is:

$$\begin{aligned}\text{price} =&\beta_0 + \beta_1\text{hoteln} + \beta_2\text{age\_mean} + \beta_3\text{age\_mean}^2\\&+ \beta_4\text{commut} + \beta_5 \log\text{centre}\\&+ \beta_6\text{GKat}\end{aligned}$$

According to our estimation, tourism leads to price increases for clothing and footware. This can be explained by the higher demand. Grimmer and Schulz-Borck (1982) shows the same, but with different indicators of demand (population density and growth). The age has a twofold effect: firstly, youth leads to higher prices which can be explained by the enthusiasm for fashion among young people. Secondly, the latter increase can be reasoned by the better financial resources of the older age groups. The commuters seem to buy their clothes at home, diminishing the demand. Neither transportation costs, nor demand and supply can explain the negative effect of distance to the next regional metropolis.

We decided for this rather large model equation of housing, water, electricity, gas and other fuels:

$$\begin{aligned}\text{price} =&\beta_0 + \beta_1\text{area} + \beta_2\text{area}^2 + \beta_3\text{emp}\\&+ \beta_4\text{wage} + \beta_5\text{wage}^2\\&+ \beta_6\text{building} + \beta_7\text{popdens}^2 + \beta_8\text{hoteln}\\&+ \beta_9\text{sec\_sector} + \beta_{10}\text{tert\_sector}\\&+ \beta_{11}\text{age\_var} + \beta_{12}\text{age\_skew} + \beta_{13}\text{commut}\\&+ \beta_{14}\text{Aggreagt} + \beta_{15}\text{typ}\\&+ \beta_{16}\text{building} * \text{typ} + \beta_{17}\text{wage} * \text{typ} *\end{aligned}$$

As there are a lot of different types of products and services included in this COICOP-group, it is not possible to find explanations for most of the estimated parameters. As we just need the model for prediction and not for an interpretation, this does not create huge problems. Further information can be found in the appendix.

Furnishings, household equipment and routine household maintenance can be described with the following equation:

$$\begin{aligned} \text{price} = &\beta_0 + \beta_1 \text{area} + \beta_2 \text{emp} + \beta_3 \text{wage} \\ &+ \beta_4 \text{wage}^2 + \beta_5 \text{hi\_quali} + \beta_6 \text{GKat} \end{aligned}$$

The positive sign of the area parameter can be explained by the higher transportation costs. The U-shape of the variable wage cannot be reasoned that easily. Again the number of employees has a negative sign, near zero making its effect questionable. The positive sign of high qualified employees indicates that these kind of consumer are willing to pay more. The significant signs of the parameters for shop quality are those we expected.

The relevant model for products and services of health is

$$\text{price} = \beta_0 + \beta_1 \text{housing\_index} + \beta_2 \text{housing\_index}^2 + \beta_3 \text{typ}$$

There are only a few factors which influence health. The parabola describing the effect of the housing index opens downward. We cannot reason why only the scarcity of flats has an influence on the prices of health products. The variable typ indicates that the prices are significantly higher in rural than in urban areas, which can be justified by higher transportation costs in rural areas.

For transport we define the following model:

$$\begin{aligned}\text{price} =&\beta_0 + \beta_1\text{area} + \beta_2\text{emp} + \beta_3\text{wage}\\&+ \beta_4\text{housing\_index} + \beta_5\text{hoteln}\\&+ \beta_6\text{hi\_quali} + \beta_7\text{hi\_quali}^2\\&+ \beta_8\text{age\_mean} + \beta_9\text{age\_mean}^2 + \beta_{10}\text{age\_skew}\\&+ \beta_{11}\text{commut} + \beta_{12}\text{commut}^2 + \beta_{13}\text{GKat} + \beta_{14}\text{typ}\end{aligned}$$

It is important to note that the transportation COICOP class includes bikes, car spare ports, bus tickets car repairs and so on. It becomes evident that the population structure is a deciding factor for the transportation costs. Notably the age variables indicate that when a higher portion of younger people live in a region, the mobility is greater and transportation costs are higher. Also regions with a high amount of highly qualified people also have higher transportation prices. In the rural areas, prices of transportation services are higher due to less competition and higher transportation costs to other regions. The closer a regional metropolis is, the higher the transportation costs are. This can be reasoned with the pricing pressure in agglomerated areas.

The equation for communication is:

$$\begin{aligned}\text{price} =&\beta_0 + \beta_1\text{area} + \beta_2\text{emp} + \beta_3\text{wage}\\&+ \beta_4\text{kk\_ew} + \beta_5\text{kk\_ew}^2\\&+ \beta_6\text{popdens} + \beta_7\text{mig}^2\\&+ \beta_8\text{sec\_sector} + \beta_9\text{age\_mean}^2 + \beta_{10}\text{centre}\\&+ \beta_{11}\text{typ} + \beta_{12}\text{wage} * \text{kk\_ew}\end{aligned}$$

Only the price of a landline telephone is left in this COICOP-class. There are only a few things to be noted: the rural areas are more expensive. Perhaps this can be reasoned by more frequent and long distance calls. The higher the density and the net migration are (indicating a higher demand), the higher the prices are. The higher the regional mean of age is, the higher the prices are. It is plausible, because older people are less likely to use new media such as mobile phones or internet compared to younger people. This lessens the demand in regions with an older population.

We identified following model for recreation and culture:

$$\begin{aligned}\text{price} = &\beta_0 + \beta_1\text{emp} + \beta_2\text{wage} + \beta_3\text{popdens}\\ &+ \beta_4\text{popdens}^2 + \beta_5\text{unemp} + \beta_6\text{unemp}^2\\ &+ \beta_7\text{hi\_quali} + \beta_8\text{GKat}\end{aligned}$$

Whereas the U-shaped curve makes sense for describing the effect of the population density, there is no explanation for the parabola of the unemployment rate. The prices are high in very sparsely populated areas as there is little competition or no supply of recreation and culture opportunities. However, this effect overturns, if we look at agglomerated areas. Here the recreation and culture opportunities become more exlusive, and therefore more expensive. For example the swimming bath becomes a spa or adventure swimming bath with slides. As the wage is a factor cost and products of recreation and culture are produced and consumed on-site, the prices rise if factor costs rises.

The model for education is:

$$\begin{aligned}\text{price} =&\beta_0 + \beta_1\text{area} + \beta_2\text{area}^2 + \beta_3\text{area}^3\\ &+ \beta_4\text{kk\_ew} + \beta_5\text{building} + \beta_6\text{building}^2\\ &+ \beta_7\text{popdens} + \beta_8\text{hoteln} + \beta_9\text{hoteln}^2\\ &+ \beta_{10}\text{unemp} + \beta_{11}\text{unemp}^2 + \beta_{12}\text{age\_skew}\\ &+ \beta_{13}\text{age\_skew}^2\end{aligned}$$

As there are only 24 cases left, it is not necessary to interprete the model. Due to the fact that it was difficult to find comparable products, a lot of products where excluded from the analysis. Only classes in adult education centres are left.

The equation for restaurants and hotels is:

$$\begin{aligned}\text{price} =&\beta_0 + \beta_1\text{area} + \beta_2\text{area}^2 + \beta_3\text{area}^3\\ &+ \beta_4\text{wage} + \beta_5\text{wage}^2 + \beta_6\text{wage}^3\\ &+ \beta_7\text{kk\_ew} + \beta_8\text{housing\_index} + \beta_9\text{sec\_sector}\\ &+ \beta_{10}\text{tert\_sector} + \beta_{11}\text{age\_mean} + \beta_{12}\text{age\_mean}^2\\ &+ \beta_{13}\text{typ} + \beta_{14}\text{wage} * \text{kk\_ew}\end{aligned}$$

As products in this category are produced and consumed on-site, the factor costs are most relevant for regional prices. This is reflected in the significant parameters of wage. The spending capacity has a postive effect on prices. Apart from these variables, the population structure is most relevant for prices of restaurants and hotels. Again, the age is a downside-opened parabola, indicating the the very young and very old regions have lower prices than the middle-age regions, which can be explained by asking the question: Who consumes the most in restaurants and hotels? The parameters of the number of emoployees in the sectors point to the same fact as the parameters of regional type: in rural areas the prices of restaurants and hotels are lower than in urban areas.

The last model identified is for miscellaneous goods and service:

$$\begin{aligned} \text{price} = &\beta_0 + \beta_1 \text{area} + \beta_2 \text{building} + \beta_3 \text{sec\_sector} \\ &+ \beta_4 \text{tert\_sector} + \beta_5 \text{GKat} \end{aligned}$$

For this class, we cannot find a useful argument to explain the model.

Our results do not support the results of Grimmer and Schulz-Borck (1982). They identify the spending capacity and other similar variables as the most important factors. Here these effects can only be found in a few classes of products. We also show, that the positive effect of the community size on the regional price level is an effect of agglomeration and not of the community size itself, which is found by Mehnert (1997). If we control for the type of region and the agglomeration level as is done here, the effect of community size is not a significant factor in our models.

### 7.1.2 Single Imputation by Conditional Means

As in forgoing studies, e.g. Roos (2006), we simply impute by a single imputation approach. The best predictions for the unobserved prices are the conditional means computed by the regressions described above. As in the previous studies we have a promising result, that can be summarised and demonstrated by showing the Lowe price index for the 96 Bavarian regions. We have chosen the Lowe index for the reasons given above.

The map in figure 3 is coloured according to the regional price level in that region. The exact values are shown in a table in the appendix. The higher the price, the less intensive is the heat colour. There are two cities with very

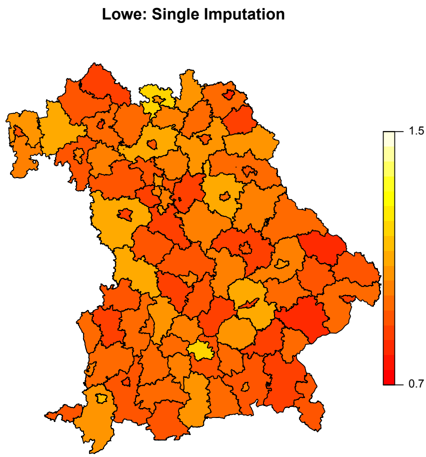


**Abbildung 3:** Regional Lowe Price Index by Single Regression Imputation

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

high prices: Munich, which is not surprising, and Coburg, which is quite astonishing. This supports the findings of Kawka (2009) that Munich is expensive and that the regions near the alps are high level regions as there are several southern regions under the high level regions. Moreover it seems that the counties around larger cities have relatively high price levels. This effect can be explained by the fact, that people earn more in city centres, where they work, and spend more in the regions around the centers, where they live. But this results contradict Kawka (2009) who finds that the prices are high in the center regions. The city centres itself have

an surprisingly low price level. Apart from these interpretations, we cannot find any other trends from the map. It just reports the predicted regional price level for Bavarian regions as the described influencing factors make us suggest.
Here is the list of top six most expensive and top six cheapest regions in Bavaria. The full ranked list can be found in the appendix. Periphery regi-

| Low level regions | |
|---|---|
| Hof, Stadt | 0.8201013 |
| Regen | 0.8359081 |
| Rottal-Inn | 0.8398376 |
| Schweinfurt, Stadt | 0.8603824 |
| Neuburg-Schrobenhausen | 0.8641514 |
| Wunsiedel i.Fichtelgebirge | 0.8682586 |
| **High level regions** | |
| Bamberg | 1.1372344 |
| Main-Spessart | 1.1376343 |
| Donau-Ries | 1.1488031 |
| Kempten (Allgäu), Stadt | 1.1513531 |
| München, Stadt | 1.2222956 |
| Coburg | 1.2497697 |

**Tabelle 5:** Highprice and Lowprice Regions by Single Regression Imputation

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

ons have a low price level, too, as shown by Kawka (2009).
The list shows that there are no spatial dependencies and no other trends to describe, if the regional price level is estimated with a single conditional mean imputation. The covariables used do not determine credible spatial dependencies. The same result is shown in the estimated variogram: The grey lines link the exact variogram values that are estimated. The red line is a loess smoother of the exact variogram values to make the function indicating spatial dependencies observable. As the grey lines do vary a lot

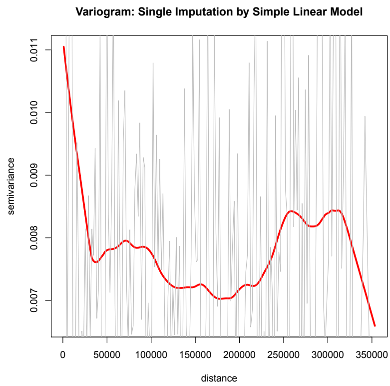


**Abbildung 4:** Variogram of the Regional Lowe Price Index by Single Regression Imputation

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

around the loess smoother, the smooth function of the variogram is not a good representation of the spatial dependecies.

Regional dependencies are a very feasible assumption in the case of regional prices. The price level in a regions is influenced by the neighbouring regions as people do not only shop in the region they live in, therefore neighbouring areas influence each other by the consumption behaviour and their structure. There are shopping trips in other cities and commuters, mixing the consumers of nearby regions. The variogram shows that there is no covariance explained by the spatial structure of the data. There is a relatively high nugget effect indicating a high variance, but, there is no hint for spatial dependencies as the variogram decreases very fast after the origin and then oscillating in an unordered way. The confusing spatial structure here is not credible at all for a regional price structure.

## 7.2 Universal Kriging

The price level of a region is influenced by the price level of neighbouring regions. The interdependence of the regional price level is caused by interregional shopping trips, commuters, similar transportation costs, mutual adaptation of the price level on the supply side and so on. Our aim has to be to preserve the regional correlation structure.
If the correlation structure is not considered in the model underlying the imputation of unsurveyed price data points, the spatial dependence structure is lost. A method for modelling spatial correlation is universal kriging as we decribed it at the beginning of this text. Kriging aims to predict values of unobserved locations, which meets our aim to impute prices of regions that are not in the sample of the official statistics.

### 7.2.1 The Universal Kriging Model

The universal kriging model consists of two parts: the first part $X\beta$ models the spatial trend which can be explained by regional sizes and $Z\gamma$ models the dependence of the neighbouring regions. The two parts define the following regression model:

$$y = X\beta + Z\gamma + \epsilon$$

where $\epsilon \sim N(0, \sigma^2)$. We use the same models for kriging as we used in the single imputation scheme. Thus the design matrix $X$ includes the variables that were selected for the product group defined by the COICOP-classification. The vector $\beta$ includes the respective parameters and needs to be estimated. The incidence matrix $Z$ is a $n$ matrix, where $n$ indicates the number of surveyed products and $r$ is the number of regions with observed prices. The vector of random effects is $\gamma \sim N(0, G)$. The spatial correlations are modelled in $G = \tau^2 R$, where $R$ is the correlation matrix and the correlation function $\rho(h; \cdot)$ defines the entries in $R$.

### 7.2.2 Estimating the Variance Parameters

In a first step, we need to determine the variance parameters of the model by the given data. As we grouped data according to the product, the sample sizes per product class are quite small, making the bad small sample properties of ML estimations of variance parameters relevant. For that reason, we have chosen to estimate the variance parameters by REML.

First, we assume a Gaussian correlation function, which is an exponential correlation function with the parameter $\kappa = 1$. The entries in $R$ are then (Fahrmeir et al., 2007)

$$\rho(h; \phi) = \exp\left\{-\left(\frac{h}{\phi}\right)\right\},$$

where $h$ is the distance of the centroids of the two regions of interest. $\phi$ is a scaling parameter and is the third variance parameter beneath $\tau$ and $\sigma$. According to Schabenberger and Gotway (2005) minus twice the loglikelihood in the REML case is:

$$\begin{aligned} &ln\left(|Var(y;\theta)|\right) + ln\left(|X(s)'Var(y;\theta)X(s)|\right) - ln\left(X(s)'X(s)\right) \\ &+rVar(y;\theta)^{-1}r + (n-k)ln\left(2\pi\right) \end{aligned}$$

with the OLS-residuals $r$.
Optimising the formula above leads to the strange result of negative variances, which is a well-known problem of REML-estimators. To solve this problem, a constrained optimisation or an optimisation with transformed parameters is proposed.(Kneib, Greven, 2011)

We have chosen to optimise the given REML-function in a constrained parameter-space, allowing only positive variances. The OLS residuals of the single imputation models have been used. There were some problems when optimising the likelihood. Not only do the results depend on the starting values, but also the existence of an optimum depends on the starting value. We tried several values and decided for the most stable ones across the groups. For three cases, different starting values were required for computational reasons. As several negative variances were defined as the result, we needed to restrict the parameter space to non-negative values. There where three cases were we needed to restrict the parameter space to positive values.
Later the reason for these three exceptions becomes clear. The products concerned are electricity prices. Due to the special market structure of electricity prices (de facto monoploy position of local suppliers, but regulation by the state), there is a high intraregional correlation creating a huge nugget effect which is the reason for the described computational problems.

### 7.2.3 Estimating the Parameters

The underlying assumptions of the kriging model are the same necessary for the mixed effect model, as Pollice and Bilancia (2002) state. Moreover, the kriging model can be interpreted as a mixed effect model, where the Gaussian random field is interpreted as a mixed effect defining a special covariance function. For this reason methods to estimate mixed effects models can be used.
The easiest way to account for the estimated covariances when estimating the parameters of the regression model is a Aitken-estimator.(Greene, 2012)

Under normality, the best way of estimation the kriging model is the GLS-estimator:(Schabenberger and Gotway, 2005)

$$\hat{\beta} = \left(X'Var(y;\theta)^{-1}X\right)^{-1} X'Var(y;\theta)^{-1}y$$

If the variance-covariance matrix $Var(y;\theta)$ is estimated itself, it becomes the EGLS-estimator by just inserting the estimators described in the chapter above. (Greene, 2012)
The Aitken-estimator is a simple way to program an EGLS-estimator. A decomposition of the $Var(y;\theta)^{-1} = P'P$ is needed and is inserted:

$$\hat{\beta} = \left(X'P'PX\right)^{-1} XP'Py$$

Obviously, the OLS estimation of the transformed model with $Py$ and $PX$ leads to the same result as the GLS-estimation that we need.(Greene, 2012)

### 7.2.4 Single Imputation by Kriging

We used the formula 30 to get imputation for prices in the unobserved regions. Of course, there have been a few negative imputations as we assume normal distributed prices. As the properties of the regions with negative Kriging predictions are seen to be low-price regions, we replace negative imputations by the minimum of the observed prices. It makes no sense to assume a price of 0.
The map of the kriging model can be seen in figure 5. Clearly the map is more varicoloured than the map of the single regression imputation representing the bigger range of the estimated Lowe price index. Munich is not in the top six of the most expensive regions anymore. There are regions at the edge of the study area with a higher price index. This can result from the used technique, which is in fact an interpolation technique, so there are often extreme values at the edges due to extrapolation.

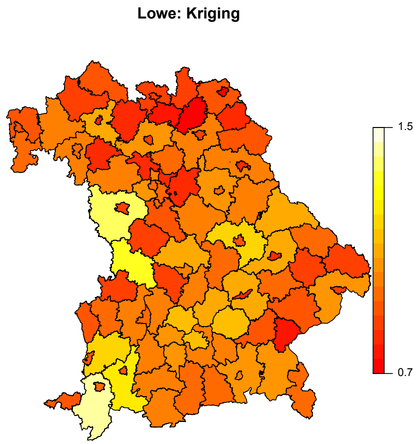


**Abbildung 5:** Regional Lowe Price Index by Kriging

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

The preservation of the spatial dependencies is evident, because neighbouring regions do not lead to extreme change in colour. The map has become a structure. It has a continous formation of regional price levels.

The table of the most expensive and cheapest six regions also shows, that the range is bigger in the kriging case than for the simple regression imputation. Moreover, most regions with extreme values are regions near to the border of Bavaria. The high prices in the Allgäu can be explained by tourism.

The variogram of the kriging model shows that the information we are adding to the model is preserved in the data when imputing missing values. Here we assume spatial dependencies between the regions. We implement

| Low level regions | |
|---|---|
| Hof, Stadt | 0.7400224 |
| Kulmbach | 0.7470393 |
| Lichtenfels | 0.7810324 |
| Altötting | 0.7810395 |
| Wunsiedel i.Fichtelgebirge | 0.8068562 |
| Weiden i.d.OPf., Stadt | 0.8163738 |
| **High level regions** | |
| Regensburg | 1.2116196 |
| Unterallgäu | 1.23106313 |
| Ostallgäu | 1.2982049 |
| Donau-Ries | 1.3520368 |
| Ansbach | 1.4156334 |
| Oberallgäu | 1.4646028 |

**Tabelle 6:** Highprice and Lowprice Regions by Kriging

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

this in the single imputation model by predicting prices of regions without data with kriging. The variogram is just as we assumed it would be: there is a nugget effect and the slow increasing interregional variance shows that the spatial dependence decreases with increasing distance between the regions.
The unsmoothed, real covariances represented by the grey lines are less wiggly represented by the grey lines, indicating that the course of the variogramm line is not an artifact of the loess smoothing represented by the red line. For the single regression imputation approach the real values of the variogramm vary more as there is no real spatial dependence resulting in an ordering of the values. The kriging imputation preserves spatial dependencies of the regional price level.

A possible criticism of using kriging and then showing the spatial dependence of the predictions is that the spatial correlations are just introduced

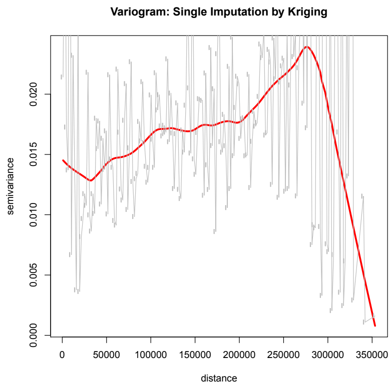


**Abbildung 6:** Variogram of the Regional Lowe Price Index by Kriging

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

through the kriging model. This is a true fact, but the assumption of spatial dependent regional price level is credible, and the data are very sparse resulting in a small degree of information. Due to the last fact, we need to find another source of information than the data, which is to impose assumptions.

## 7.3 Multiple Imputation

To get a reference from the multiple imputation side as well, we also estimated and predicted a simple multiple imputation model. We used the regression modelling of the simple imputation model to include all relevant variables in the shape to best explain regional prices.

### 7.3.1 Prior, Posterior, and Full Conditionals

Although in the normal linear regression case the posterior predictive is available in closed form, we sampled using a Gibbs-sampling scheme to be able to compare of the different methods discussed in this work. As the spatial model needs to be sampled using a Gibbs-sampler we also chose one for the simple regression multiple imputation under normality.

As we have some information about the distribution from the full data set using all products at once, noninformative priors are not a sensible choice. Instead we have chosen the conjugate model, where variance parameters follow an inverse Gamma distribution and the mean parameters are normal distributed. The hyperparameters of the normal distributions are the estimators of the whole data set. This course of action is not driven by wanting an empirical Bayes estimator, but by the idea that the whole data set has additional information on the regression parameters. We want to use this prior knowledge.

To establish a Gibbs-sampler we need to write down the posterior distribution first. Next we can define the full conditionals from which we sample. As already defined, the prior of the $\beta$ of the underlying regression model and the variance parameters. Due to the chosen normal-inverse-Gamma-model (NIG), the priors are:(Fahrmeir et al., 2007)

- joint prior: $P(\beta, \sigma^2) = P(\beta|\sigma^2)P(\sigma^2)$,
- conditional prior of $\beta$: $\beta|\sigma^2 \sim N(m, \sigma^2 M)$, with hyperparameters $m$ and $M$,
- and the marginal prior of $\sigma^2$: $\sigma^2 \sim IG(a, b)$, with the hyperparameters $a$ and $b$.

We decided against a noninformative prior as this requires a lot of data points and only a few parameters.(Gelman et al., 2004) In our case, the data points are few and quite spread out, but for multiple imputation of unobserved values it is necessary to include as many variables in the regression model as possible. Therefore, we use an informative prior. The data model is a normal with $X\beta$ and $\sigma^2 I$. Therefore the posterior is:(Fahrmeir et al., 2007)

$$\begin{aligned} P(\beta,\sigma^2|y,X) \quad \propto \quad & \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right\} \\ & \frac{1}{(\sigma^2)^{\frac{p}{2}}} \exp\left\{\frac{1}{2\sigma^2}(\beta-m)' M^{-1}(\beta-m)\right\} \\ & \frac{1}{(\sigma^2)^{a+1}} \exp\left\{-\frac{b}{\sigma^2}\right\} \end{aligned}$$

All relevant computations to derive the posterior and the following full conditionals can be found in the appendix. The full conditionals are:(Fahrmeir et al., 2007)

- $\beta|\cdot \sim N\left(\mu_\beta, \Sigma_\beta\right)$, where $\mu_\beta = \Sigma_\beta \left(\frac{1}{\sigma^2}X'y + \frac{1}{\sigma^2 M^{-1} m}\right)$ and $\Sigma_\beta = \left(\frac{1}{\sigma^2}X'X + \frac{1}{\sigma^2}M^{-1}\right)^{-1}$, and
- $\sigma^2|\cdot \sim IG\left(a', b'\right)$, where $a' = a + \frac{n}{2} + \frac{p}{2}$ and $b' = b + \frac{1}{2}(y - X\beta)'$ $(y-X\beta) + \frac{1}{2}(\beta - m)' M^{-1}(\beta - m)$

To get estimations of the parameters we conduct a Gibbs-Sampler.

### 7.3.2 Gibbs-Sampler

To get multiple imputations under the assumption of the simple regression models defined by the single imputations, a Gibbs-sampler based on the full conditionals that we defined in the chapter above was used. We

sequentially draw from the full conditionals as described in Gelman et al. (2004), but in contrast we start with the regression parameters and draw the variance parameters afterwards. The explaining variables included in the data matrix $X$ are the same that we have chosen for the single imputation models. The same is the case for the functional form assumed.
However, parameter estimations are conducted in a Bayesian way. To get the estimations a Gibbs-sampler with 10.000 iterations is done. To reduce the computational burden, single iterations with enough distance were used to get $m = 5$ different parameter estimations for the multiple imputations. The different imputation parameters have a distance of 1.000 iteration steps, which guarantees their independence.
The sampler itself is based on the full conditionals defined in the previous chapter, which establishes a Gibbs-sampler by iterating between draws from the two full conditionals defined above to simulate draws from the joint posterior.

### 7.3.3 Multiple Imputation by Linear Regression

In the end, we are interested in draws from the posterior predictive to get imputation of prices in unobserved regions. We follow Gelman et al. (2004), to get draws of the $y_{mis}$. We first conduct the described Gibbs-sampler to determine the posterior predictive and then draw unobserved values by drawing out of the estimated posterior predictive.
The result of the prediction of multiple imputation is shown in map 7. Nearly all regions have the same colour (orange), indicating that there is much mass in the center of the distribution. As the multiple imputation price index is a mean over the five independent imputations of the Lowe index, the egalitarian result is expectable. On the other hand, there are some lighter coloured regions on the map, mostly around Munich. Munich and the surrounding regions are well known for its high prices, but the

differences is huge as Munich has a price level about 146% of the the Bavarian average. Even though a lot of the data points are in the middle of the distribution of price levels, the outlier values are extreme when using multiple imputation.
Another astonishing result is the high price level of Wunsiedel, for which we expected a very low price level as predicted with the foregoing methods. For the region of Ansbach it is the other way around. Kriging ranked it in the top six most expensive regions, however it is under the top six of the cheapest regions after multiple imputation.

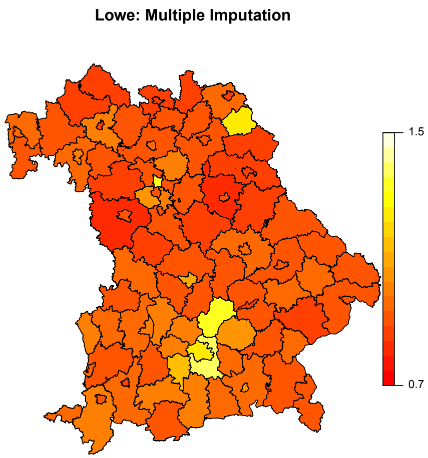


**Abbildung 7:** Regional Lowe Price Index by Multiple Imputation

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

The table of the top six most expensive and six cheapest regions confirms the north-south price differential that we also found using the single imputation approach: four of the expensive regions are in the south around Munich and all six low price regions are in the very north of Bavaria. However, the result of Wundsiedel and Ansbach are not the expected ones.

| Low level regions | |
|---|---|
| Ansbach | 0.8157848 |
| Amberg-Sulzbach | 0.8313815 |
| Hof, Stadt | 0.8512344 |
| Coburg, Stadt | 0.8590893 |
| Coburg | 0.8698144 |
| Weiden i.d.OPf., Stadt | 0.8744579 |
| **High level regions** | |
| Starnberg | 1.1605933 |
| München, Stadt | 1.23106313 |
| Wunsiedel i.Fichtelgebirge | 1.2681776 |
| Erlangen, Stadt | 1.3701361 |
| Freising | 1.3811412 |
| München | 1.4467345 |

**Tabelle 7:** Highprice and Lowprice Regions by Multiple Imputation

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

The variogram of the multiple imputation starts at a very low level compared to the foregoing one. There is nearly no nugget effect indicating only a small amount of variability of the regions itself. The variogram increases very fast. After reaching the maximum very early, it decreases fast, too. It seems that MI can only reproduce spatial dependencies over short distances.

Another important point is, that the real, unsmoothed values of the variogram fluctuate a lot around the smoothed variogram function. In reality, MI does not represent the spatial dependencies very well, as the smoothed variogram function just is an artifact of the smoothing.

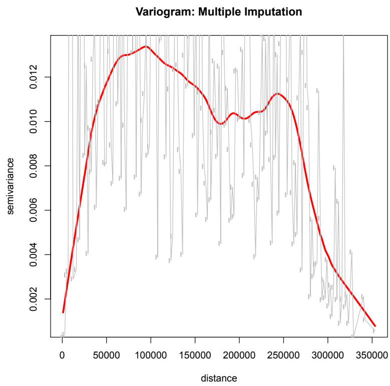


**Abbildung 8:** Variogram of Regional Lowe Price Index by Multiple Imputation

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

Multiple imputation promises to allow a feasible inference of statistics based on the completed data set. As the predicted regional prices shall be given to other researchers to conduct a statistical analysis, we want to preserve this feature. On the other hand, we want to implement the credible assumption of spatial dependencies that MI can only preserve in the short term.

## 7.4 KriMI: Multiple Imputation by Kriging

As already mentioned, kriging can be implemented using the multiple imputation scheme in two ways. Firstly, kriging can be interpreted as a mixed model approach, where the spatial dependencies are represented in a spatially correlated random effect. The mixed model approach can be estimated in a Bayesian way leading to a Gibbs-sampler described in the next paragraph. Secondly, the kriging approach can be interpreted as a Bayesian P-Spline leading to the second approach, which fits better to the idea of predicting values in unobserved regions.

### 7.4.1 KriMI by the Mixed Model Approach: Gibbs-sampler

All distributions needed are described in the chapter 4.1.2. For this reason we only describe the Gibbs-sampler that leads to parameters to compute multiple imputations of the prices in regions with no data. We just need to find a sensible order of the full conditionals already derived. We decided to draw the variance parameters from their full conditionals first. Afterwards we draw the random effects. In a last step draws are made from the full conditional of the fixed effects. In a nutshell the drawing scheme is:

1. Draw $\tau^2$ from $IG(a_\tau + \frac{n}{2}, b_\tau + \frac{1}{2}\gamma' R^{-1}\gamma)$
2. Draw $\sigma^2$ from $IG(a_{\sigma^2} + n, b_{\sigma^2} + \frac{1}{2}(y - X\beta - Z\gamma)'(y - X\beta - Z\gamma) + \frac{1}{2}(\beta - m)' M^{-1}(\beta - m))$
3. Draw $\gamma$ from $N(\mu_\gamma, \Sigma_\gamma)$ where $\mu_\gamma = \Sigma_\gamma Z'(\sigma^2 I)^{-1}(y - X\beta)$ and $\Sigma_\gamma = \left(Z'(\sigma^2 I)^{-1} Z + (\tau^2 R)^{-1}\right)^{-1}$
4. Draw $\beta$ from $N\left(\mu_\gamma, \Sigma_\gamma\right)$ where $\mu_\beta = \Sigma X'\left(\sigma^2 I\right)^{-1}(y - Z\gamma) + \left(\sigma^2 M\right)^{-1} m$ and $\Sigma_\beta = \left[\left(X'(\sigma^2 I)^{-1} X + (\sigma^2 M)\right)\beta\right]^{-1}$

The scaling parameter $\phi$ which determines the spatial correlation is explicitly not included in the sampling scheme. There are several reasons for not sampling $\phi$: firstly, the data at hand are so spatially rare that they do not allow an estimation of a parameter determining spatial dependencies. This fact underlines the advantage of the Bayesian approach chosen as the missing information in the data can be complemented by assumptions about the reality represented in prior knowledge. Determing $\phi$ means to make a guess about the spatial dependencies in the data. Secondly, $\phi$ is a hyperparameter. The assumption of a fixed $\phi$ does solve the problem of determing hyperparameters and the prior distributions needed for a spatial prediction. There is no knowledge about the distribution of the hyperparameter $\phi$.

### 7.4.2 KriMI by the Mixed Model Approach: Results

As there are no outliers as in the MI-approach, the map of regions coloured according to their regional price level is even more egalitarian than the foregoing map 8. This results from the kriging method in combination with the chosen mixed model technique leading to smooth predictions of unobserved regional prices. Kriging involves using price information from neighboring regions in a weighted average. Mixed modelling means to model the expectation which shifts the random component to the variance, leading to smooth mean predictions in the case of kriging as a mixed model. We already mentioned that the idea of mixed modelling does not fit well the idea of predicting missing values as it involves imputing some kind of mean with a specially modelled variance-covariance matrix.
However we will see that the mixed model approach has the great advantage of being very efficient due to the small computational burden. More importantly, it is less sensitive to inappropriate data cases. Moreover the alternative P-spline modelling discussed later has the disadvantage of a very high volatility of the predictions and of a small steadiness of estimation. The P-spline approach is fragile, whereas the mixed model approach is stable.
The result of the map is also reflected in the table of the top six high price and top six low price regions. It is well known that Munich and its exurbs are high price regions, whereas the regions near to the Bavarian border,

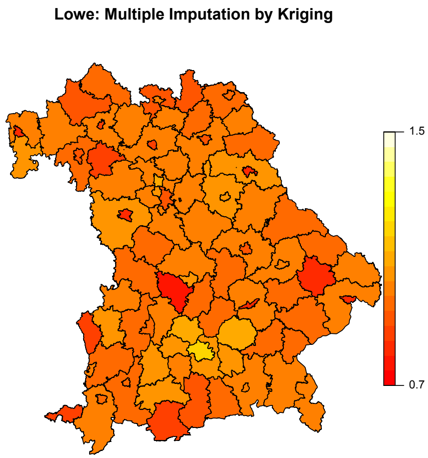


**Abbildung 9:** Regional Lowe Price Index by KriMI with Mixed Modelling

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

especially in the east, are low price regions. There are no outliers on both extremes as they are smoothed away in the twofold manner: the kriging and the mixed model technique.

The variogram is very similar to the variogram of the multiple imputation result. The semivariance increases up to a distance of 100 km and then decreases slowly. This trend is what we expect when assuming spatial dependencies of regional prices. Shopping trips beyond a threshold distance become too expensive, which explains the decreasing trend above the 100 km threshold.
The variogram of the KriMI approach using mixed modelling differs to the variogram of the MI approach by showing that there is a nugget effect.

| Low level regions | |
|---|---|
| Neuburg-Schrobenhausen | 0.7889400 |
| Weiden i.d.OPf., Stadt | 0.8117964 |
| Deggendorf | 0.8121880 |
| Passau, Stadt | 0.8140455 |
| Ansbach, Stadt | 0.8196061 |
| Landshut, Stadt | 0.8209351 |
| **High level regions** | |
| Amberg-Sulzbach | 1.0854961 |
| München | 1.0962311 |
| Erding | 1.1054661 |
| Dachau | 1.1282148 |
| Erlangen, Stadt | 1.1355460 |
| München, Stadt | 1.2205000 |

**Tabelle 8:** Highprice and Lowprice Regions by KriMI with Mixed Modelling

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

There is a variance of the spatial random process, making the KriMI prediction more plausible. That the prices in one region still differ is more plausible than to assume that they are the same.

### 7.4.3 KriMI by Baysian P-Splines: Priors, Posterior, Full Conditionals, and Gibbs-sampler

It is not easy to get random draws of the P-spline parameterization described above. Lukily the drawing scheme is described in Lang and Brezger (2001, 2004). In addition the sampling can be done by using BayesX and its implementation R2BayesX, which exactely does what we need here. (Umlauf et al., 2015, Belitz et al., 2015)

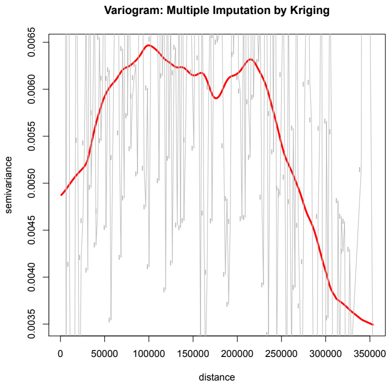


**Abbildung 10:** Variogram of the Regional Lowe Price Index by KriMI with Mixed Modelling

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

To get random draws, Lang and Brezger (2004) proposed the following drawing scheme defining a Gibbs-sampler, which is implemented in the software used. In step $t$

1. draw $\widetilde{\gamma}^{(t)}$ from the normal distribution with the precision matrix $Pre_{\widetilde{\gamma}}^{(t)}$ and mean vector $m_{\widetilde{\gamma}}1(t)$,

2. draw $\beta^{t}$ from the normal distribution $N(m_{\beta}^{(t)}, P_{\beta}^{-1(t)})$,

3. draw $\tau^2$ by a draw from the inverse gamma distribution $IG(\overline{a}_{\tau^2}^{(t)}, \overline{b}_{\tau^2}^{t})$,

4. draw $\sigma^2$ by a draw from the inverse gamma distribution $IG(\overline{a}_{\sigma^2}^{(t)}, \overline{b}_{\sigma^2}^{(t)})$,

5. update the distributions with the drawn parameters and start again. (Lang and Brezger, 2001)

To get the random draws of the coefficients we conduct random draws as proposed by Lang and Brezger (2001, 2004), Rue (2001):

1. Starting with the Cholesky decomposition: $Pre = LL'$,
2. Draw random numbers $Z \sim N(0,1)$ and solve $L'\beta = Z$ or $L'\tilde{\gamma} = Z$ depending on which parameters we are working, $Z$ are drawn for every parameter vector of their own and at every step.[29]
3. Solve $Pre_\beta m_\beta = \frac{1}{\sigma^2}X'(y - \overline{\eta})$ or $Pre_{\tilde{\gamma}} m_{\tilde{\gamma}} = \frac{1}{\tau^2}V'(y - \overline{\eta})$. To do so Lang and Brezger (2004), Rue (2001) propose to solve first $L\nu = \frac{1}{\tau^2}V'(y - \overline{\eta})$ by forward substituion to get $\nu$ and in a second step solve $L'm_{\tilde{\gamma}} = \nu$ by backward substitution to get the mean vector of interest. For $\beta$ use the equivalent formulae.(Rue, 2001)
4. Add the mean vectors to the values of second step.
5. As the precision matrix of $\tilde{\gamma}$ is a sparse matrix, it needs to be converted into a band matrix by the reverse Cuthill-McKee algorithm. (Lang and Brezger, 2004)

If we set in the matrices and vectors of the P-spline approach described above, we get an estimation and prediction tool of our problem. $\overline{\eta}$ means the predictor that we get using the rest of the linear model (not mentioned here for simplification). $V$ is the variance-covariance matrix of the universal kriging model.

### 7.4.4 KriMI by Bayesian P-Splines: Results

The geographic map information we used are free available at Bayerische Vermessungsverwaltung (2015) Unfortunatly the BayesX-software we need for P-spline approach has some problems with the shapefiles determining

[29] We just write $Z$ without an index for simplification.

the Bavarian regions. Shapefiles are vector-data defining edges of the polygon which represents a region in a map. Fortunately, these problems only affect a few regions. After cleaning the data for the problematic cases, it was no problem to use the geographic information of the Bayerische - Vermessungsverwaltung (2015).[30]
BayesX had some numerical problems to estimate and predict the prices of 15 products in our shopping basket. It was not possible to fix these problems. The products have an share of 2.6 percent on our reduced shopping basket. As this is a small share, we justify removing these products. The weightings were again adjusted such that they sum up to 100 percent. If BayesX had problems to estimate the variance parameters, we used the identity matrix for simplification. All choices of a variance parameter are arbitrary, so this choice is as good as any other possibility. Not many products were affected by this problem (about five), so it has not a huge influence on the result.

The map is getting more colourful again, indicating that the P-spline-method allows predictions close to the data which is a more wiggly function than is estimated by the mixed model method. The predicted price differences are bigger. The higher volatility does not disturb the spatial dependencies. There are clusters and chains of high price and low price regions. For example, there is a nest of low price level regions in the north midlands (in the south of Nuremberg) and a low price chain starting in the west and moving to the middle.
Again the price level of the cities is lower than the price level of the surrounding regions. The low price levels results probably from the higher competition and the lower transportation costs in urban areas.
Another consequence of the P-spline technique is that extreme values appear at the edge of the observational area. To predict at the edge of the defi-

[30]Thanks to Nikolaus Umlauf who helped me fixing this problem and making BayesX work for the data available.

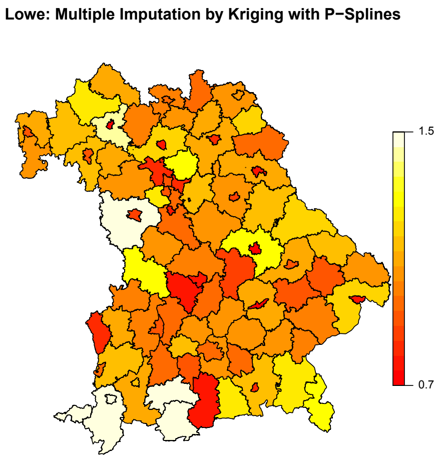


**Abbildung 11:** Regional Lowe Price Index by KriMI with P-Splines

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

nition area is a well known problem of smoothing techniques which arises from the missing information beyond that border. Predictions at the edge result mostly from an extrapolation of foregoing observations, making it more likely to become an outlier.
The same is represented in the table of the top six high level and low level regions. There are mostly peripherical regions at the Bavarian borders. There is a massive outlier: the region Ansbach has a price level about 190 percent of the Bavarian mean. This result is not credible at all, therefore it should not be interpreted and should be removed from analyis. The rest of the table shows the high variability of the P-spline approach as the range is a little higher than the range of the other kriging approaches implemented.

| Low level regions | |
|---|---|
| Regensburg, Stadt | 0.6843597 |
| Schweinfurt, Stadt | 0.7015048 |
| Landshut, Stadt | 0.7118566 |
| Ingolstadt, Stadt | 0.7533899 |
| Weiden i.d.OPf., Stadt | 0.7558110 |
| Bamberg, Stadt | 0.7638445 |
| **High level regions** | |
| Regensburg | 1.3051862 |
| Forchheim | 1.3093336 |
| Donau-Ries | 1.3247229 |
| Berchtesgadener Land | 1.3350977 |
| Schweinfurt | 1.5010405 |
| Ansbach | 1.9000075 |

**Tabelle 9:** Highprice and Lowprice Regions by KriMI with P-Splines

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

The variogram confirms the higher variability by showing a more prominent nugget effect than the one of the other price index predictions. Again it rises up to a 100 km distance and decreases afterwards. The spatial dependencies are evident up to this threshold, after reaching the sill, there is no spatial dependency anymore. The spatial dependency decreases with higher distances represented in the increasing variogram up to 100 km.

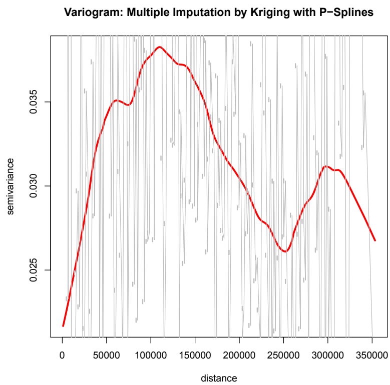


**Abbildung 12:** Variogram of the Regional Lowe Price Index by KriMI with P-Splines

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

# 8 Outlook

Firstly, let us recap the results of the analysis of the regional price levels in Bavaria. Secondly, we need to confess that the work is not finished yet. There are still many things to do and questions to be answered.
The single imputation by conditional means was the starting point for our evaluation of regional price levels as it is the method usually chosen. It works very well: it is easy to estimate and predict, the result shows sensible regional price levels. However there are two huge disadvantages already mentioned: it is not possible to use the results of single imputation to conduct any inferential statistical analysis afterwards, because the variances are seriously underestimated. Furthermore we wanted to preserve the assumed spatial dependencies, which is not possible with simple linear regression techniques.
For this reason we implemented the spatial dependencies by estabilishing a regression model that considers the influence of neighbouring regions. This is done using the kriging model prediction. The resulting regional price indices are still sensible. They vary a litte more. More important is that the spatial dependencies become apparent.
As the Kriging approach still does not account for the uncertainty brought about by the missing data, we wanted to establish a model that allows for both subsequent statistical inference and preservation of spatial dependencies. To do so, we started with multiple imputation by simple regression to get a baseline for the multiple impution part of our analysis. This first result makes it evident that the variability of the regional price level increases as we want it for the the reason of accounting for the addtional uncertainty. Even though the range of regional prices are huge in the MI case, the map shows, that the majority of the price indices lies in the middle of the distributions. Regional differentiation is impossible when using the multiple imputation approach.

In the end, we decided to implement kriging in the multiple imputation scheme, allowing for a regional differentiation, regional dependencies, and a reliable statistical inference of researchers using the price index. We established two ways to implement kriging in the MI-scheme. We call it KriMI. The mixed model approach, as well as the P-spline method create results with a higher variability as the MI result. Both account for the additional uncertainty allowing subsequent inferences. Their variograms are similar to the MI variogram, but their maps show, that results are more differentiated across regions. Both KriMI approaches differentiate better between regions than MI did.
If we compare the two KriMI approaches, there is a clear winner. The P-spline approach has two major disadvantages: firstly, it is very computationally demanding. One model needs about 30 minutes in our setting. This fact does not fit well for the MI analysis, as we need to estimate and predict every product several times, which multiplies computation time. Secondly, it is very accident sensitive. Models needed to be adjusted several times, before it was even impossible to make estimations and predictions for some products and for some regions. The mixed model approach is less sensitive and is much faster. Its stability outrivals the P-spline method, even though the latter seemed to fit the idea of predicting real values better.

Let us recap the threoretical part of the thesis, too. In the first part a short introduction into the MI-approach was given. Here the focus was on the justification of the multiple imputation idea. Afterwards, a short insight into the various kriging methods developed in geostatistics is given. We had two topics of focus in this chapter: firstly, we compared the quality of predictions by their MSPEs. Secondly, we described in detail the structures and parameters of the correlation- and covariance-matrices in the kriging case. The latter were needed in its following chapter, as the covariance-matrix is central for the Bayesian mixed model approach developed for the implementation of kriging into the multiple imputation scheme. However we go beyond the mixed model method in this last theoretical chapter, when modelling kriging as a Bayesian P-spline, which fits better to the multiple imputation idea than the expectation modelling idea of the mixed model method. Though matching theoretically better, the P-spline-approach is outruled by the mixed model method through its higher stability.

There are still a lot of unanswered questions and some unsolved problems. The integration of the regional price index computed as an interregional price index is not completed yet. The thesis at hand is embedded in a project on regional and interregional prices. As it is the first result of the project, there are some problems just named and described, but not solved at all. For this reason the price indices computed here stopped at the regional Lowe-index and are not advanced to interregionally transitive and characteristic price indices. This is the task of another part in the project and cannot be performed at once with the imputation of the regional prices. The same applies to the problem of housing costs. The rents, prices for building lands and so on have some very special features, that need to be considered and discussed. Again it is a thesis of its own and not solved here, but by some other researchers in the project team.
During analysis the theoretical framework of this thesis some other methodical problems arised, that are still unsolved. First of all, prices are not normally distributed at all. All methods used here rely on the assumption of normality. There are some approaches to solve the problem that the assumption of normality is not plausible in all data cases. The idea is to estimate the kriging model under the assumption of another distribution being part of the exponential family establishing a GLMM.(Cafarelli, 2002)
From econometric modelling, there arise two other problems to be analysed in the future. A correct econometric modelling of regional prices

requires a solution to the endogeneity problem of having variables measuring factor costs and income that depend on the price level itself. There are several approaches to solve this problem (e.g. the instrumental variable approach), but their impact on multiply imputed data are not examined yet. The same holds for the solution of the interdepence problems between prices. The prices of the products in the shopping basket are not independent from each other. They need to be estimated in a system of equations resulting in a seemingly unrelated regressions estimation (SUR). There is no technique developed to harmonise the SUR estimation when using the multiple imputation method.

Allthough there are still a lot things to do, we solved the problem of implementing regional dependencies in the multple imputation scheme. We call it KriMI. We showed that KriMI works well for the empirical example of predicting regional prices. Moreover we estimated quite reasonable regional price levels.

# Literatur

M. Aerts, G. Claesken, N. Hens, and G. Molenberghs. Local Multiple Imputation. *Biometrika*, 98(2):357–388, 2002.

J. Albert. *Bayesian Computation with R*. Use R! Springer, New York, 2nd edition, 2009.

O. Angermann. Vergleich des Verbraucherpreisniveaus in Bonn, Karlruhe, München und Berlin (West). *Wirtschaft und Statistik*, 4:258–261, 1989.

J. D. Angrist and J.-S. Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, 2009.

M. Asplund and R. Friberg. Food Prices and Market Structure in Sweden. *Scandinavian Journal of Economics*, 104(4):547–566, 2002.

Bayerische Vermessungsverwaltung. Verwaltungsgrenzen Bayern, 2015. URL `http://vermessung.bayern.de/opendata`.

Bayerisches Landesamt fuer Statistik und Datenverarbeitung, Sachgebiet 35, Statistik der Verbraucherpreise. Übersicht mit Kurzerläuterungen zu den Datenbeständen und Anlagen: Forschungsvorhaben Machbarkeitsstudie zur Schätzung und Erhebung regionaler Preisniveaus, 04.09.2013.

Bayerisches Landesamt für Statistik und Datenverarbeitung, Sachgebiet 35, Statistik der Verbraucherpreise. Datensatzbeschreibung, 2013.

Bayerisches Staatsministerium für Wirtschaft, Verkehr und Technologie. *Die reale Kaufkraft in Bayern 2002*. Bayerischen Staatsministerium für Wirtschaft, Verkehr und Technologie, München, 2003.

C. Belitz, A. Brezger, T. Kneib, S. Lang, and N. Umlauf. BayesX: Software for Bayesian Inference in Structured Additive Regression Models, 2015. URL `http://www.BayesX.org/`.

J. Besag. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2): 192–236, 1974.

R. Bivand. *classInt: Choose Univariate Class Intervals*, 2015. URL `http://CRAN.R-project.org/package=classInt`. R package version 0.1-22.

R. Bivand and N. Lewin-Koh. *maptools: Tools for Reading and Handling Spatial Objects*, 2015. URL `http://CRAN.R-project.org/package=maptools`. R package version 0.8-36.

R. Bivand and G. Piras. Comparing Implementations of Estimation Methods for Spatial Econometrics. *Journal of Statistical Software*, 63(18): 1–36, 2015. URL `URLhttp://www.jstatsoft.org/v63/i18/`.

R. Bivand, T. Keitt, and B. Rowlingson. *rgdal: Bindings for the Geospatial Data Abstraction Library*, 2015. URL `http://CRAN.R-project.org/package=rgdal`. R package version 0.9-2.

R. S. Bivand, J. Hauke, and T. Kossowski. Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geographical Analysis*, 45(2):150–179, 2013a. URL `http://www.jstatsoft.org/v63/i18/`.

R. S. Bivand, E. Pebesma, and V. Gomez-Rubio. *Applied Spatial Data Analysis with R*. Springer, New York, 2. edition, 2013b. URL `http://www.asdar-book.org/`.

S. Bleninger and A. Trojan. Analyzing Regional Prices. 15.05.2014.

U. Blien, H. Gartner, H. Stüber, and K. Wolf. Regional Price Levels and the Agglomeration Wage Differential in Western Germany. *Annals of Regional Science*, 43:71–88, 2009.

H. W. Borchers. *pracma: Practical Numerical Math Functions*, 2015. URL `http://CRAN.R-project.org/package=pracma`. R package version 1.8.3.

BVerfG. *2 BvR 556/04 vom 6.3.2007, Absatz-Nr. (1 - 76): http://www.bverfg.de/entscheidungen/rs20070306_2bvr055604.html.* 2007.

B. Cafarelli. Generalized Linear Mixed Models for the Interpolation of Non Gaussian Spatial Data: an Application to the Study of the Soil Fertility in Countryside of Capitanata, 2002. URL `http://old.sis-statistica.org/files/pdf/atti/RSMi0602p445-448.pdf`.

B. Carlin and T. Louis. *Bayesian Methods for Data Analysis*. Chapman & Hall/CRC, Boca Raton, 3. edition, 2008.

S. Chib and B. P. Carlin. On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing*, 9:17–26, 1999.

G. Cicchitelli and G. E. Montanari. Model-Assisted Estimation of a Spatial Population Mean. *International Statistical Review*, 2012.

N. Cressie. The Origins of Kriging. *Mathematical Geology*, 22(3):239–252, 1990.

N. Cressie and S. N. Lahiri. The Asymptotic Distribution of REML Estimators. *Journal of Multivariate Analysis*, 45:217–233, 1993.

N. A. Cressie. *Statistics for Spatial Data*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. Wiley, New York, 1993.

O. Dubrule. Two Methods with Different Objectives: Splines and Kriging. *Journal of the International Association for Mathematical Geology*, 15(2): 245–257, 1983.

L. Fahrmeir, T. Kneib, and S. Lang. *Regression: Modelle, Methoden und Anwendungen*. Springer, 2007.

Fortran code by H. Akima R port by Albrecht Gebhardt aspline function by Thomas Petzoldt interp2xyz and enhancements and corrections by Martin Maechler. *akima: Interpolation of irregularly spaced data*, 2013. URL `http://CRAN.R-project.org/package=akima`. R package version 0.5-11.

A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. Chapman & Hall/CRC Texts in Statistical Science. 3. edition, 2004.

A. Genz and F. Bretz. *Computation of Multivariate Normal and t Probabilities. Lecture Notes in Statistics*. Springer-Verlage, Heidelberg, 2009.

A. Genz, F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, and T. Hothorn. *mvtnorm: Multivariate Normal and t Distributions*, 2014. URL `http://CRAN.R-project.org/package=mvtnorm`. R package version 1.0-2.

GfK Geomarketing. Produktkatalog, 2013.

G. Götz. Chain-store Competition with Endogenous Entry: Localized vs. Uniform Pricing, 2012.

G. Götz, T. Krauskopf, and P. Winkler. Die Bestimmung regionaler Preisindizes - Das Beispiel Österreich, 2009.

GRASS Development Team. *Geographic Resources Anaylsis Support System (GRASS GIS) Software*. Open Source Geospatial Foundation, 2012. URL htttp://grass.osgeo.org.

W. H. Greene. *Econometric Analysis*. Prentice Hall, Boston, 7. edition, 2012.

S. Greven. Likelihood-Schätzung für lineare gemischte Modelle: Schätzung der Kovarianzparameter, 2011. URL www.statistik.lmu.de/institut/ag/fda///mixedmodels_sose11/.../kapitel3.pdf.

B. Grimmer and H. Schulz-Borck. Auswirkungen verschiedener Standorte auf die Lebenshaltung privater Haushalte. *Berichte über Landwirtschaft*, (60):543–563, 1982.

G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes*. Oxford Univ. Press, Oxford, 3. edition, 2009.

T. Hastie and R. Tibshirani. *Generalized Additive Models*, volume 43 of *Monographs on statistics and applied probability*. Chapman and Hall, London and New York, 1. edition, 1990.

H. Haupt and C. Heinze. Regionaler Preisindex für Deutschland: Abschlussbericht zur Beratung im Rahmen des Projekts Regionaler Preisindex für Deutschland, 2012.

Y. He, R. Yucel, and T. E. Raghunathan. A Functional Multiple Imputation Approach to Incomplete Longitudinal Data. *Statistics in Medicine*, 30 (10):1137–1156, 2010.

G. Heske. *Preisstatistik: Grundlagen und Praxis der volkswirtschaftlichen Preisstatistik*. Verl. Die Wirtschaft, Berlin, 1992. URL http://www.gbv.de/dms/hbz/toc/ht004340642.PDF.

P. D. Hoff. *A First Course in Bayesian Statistical Methods*. Springer texts in statistics. Springer, New York and London, 2009.

S. Jackman. *Bayesian Analysis for the Social Sciences*. Wiley Series in Probability and Statistics. Wiley, Chichester, U.K, 2009.

R. Kawka. Regionaler Preisindex. In Bundesinstitut für Bau-,Stadt- und Raumforschung (BBSR) im Bundesamt für Bauwesen und Raumordnung, editor, *Berichte*, volume 30. Bonn, 2009.

G. S. Kimeldorf and G. Whaba. A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines. *The Annals of Mathematical Statistics*, 41(2):495–502, 1970.

P. K. Kitanidis. *Introduction to Geostatistics: Applications to Hydrogeology*. Cambridge University Press, Cambridge and New York, 1997.

T. Kneib. *Bayes-Inferenz in generalisierten geoadditiven gemischten Modellen: Diplomarbeit (korrigierte Version)*. München.

J. Koo, K. R. Phillips, and F. D. Sigalla. Measuring Regional Cost of Living. *Journal of Business & Economic Statistics*, 18(1):128–136, 2000. URL `http://books.google.de/books?id=hY3EGwAACAAJ`.

R. Kosfeld, H.-F. Eckey, and J. Lauridsen. Disparities in Prices and Income across German NUTS 3 Regions, 2007a.

R. Kosfeld, H.-F. Eckey, and M. Schüßler. Ökonometrische Messung regionaler Preisniveaus auf der Basis örtlich beschränkter Erhebungen. In J. Möller, E. Hohmann, and D. Huschka, editors, *Der weiße Fleck - zur Konzeption und Machbarkeit regionaler Preisindizes*, volume 324 of *Gutachten*, pages 91–124. IAB-Bibliothek, Nürnberg, 2007b.

S. Lang and A. Brezger. Bayesian P-Splines, 2001.

S. Lang and A. Brezger. Bayesian P-Splines. *Journal of Computational and Graphical Statistics*, 13(1):183–212, 2004.

G. M. Laslett. Kriging and Splines: An Empirical Comparison of Their Predictive Performance in Some Applications. *Journal of the American Statistical Association*, 89(426):391, 1994.

R. J. A. Little and D. B. Rubin. *Statistical analysis with missing data*. Wiley series in probability and statistics. Wiley, Hoboken, N.J, 2. edition, 2002.

M. Liu, J. Taylor, and T. Belin. Multiple Imputation and Posterior Simulation for Multivariate Missung Data in Longitudinal Studies. *Biometrics*, 56(4):1157–1163, 2000.

A. D. Martin, K. M. Quinn, and J. H. Park. MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software*, 42(9):22, 2011. URL `http://www.jstatsoft.org/v42/i09/`.

A. Mehnert. Das reale Einkomen im interregionalen Vergleich - Der Einfluß des Preisniveaus auf die regionalen Einkommensdisparitäten unter Anwendung des statistischen und des ökonomischen Indexkonzeptes. *IWH-Diskussionspapiere*, (66):1–27, 1997.

B. Munoz, V. M. Lesser, and R. A. Smith. Applying Multiple Imputation with Geostatistical Models to Account for Item Nonresponse in Environmental Data. *Journal of Modern Applied Statistical Methods*, 9(1):274–286, 2010.

W. Neubauer. *Preisstatistik*. Vahlen, München, 1996.

OECD. Measuring Productivity: Measurement of Aggregate and Industry-Level Productivity Growth. *OECD-Manual*, 2001.

H. Omre and K. B. Halvorsen. The Bayesian Bridge between Simple and Universal Kriging. *Mathematical Geology*, 21(7):767–786, 1989.

Original S code by Richard A. Becker and Allan R. Wilks. R version by Ray Brownrigg. Enhancements by Thomas P Minka. *maps: Draw Geographical Maps*, 2014. URL `http://CRAN.R-project.org/package=maps`. R package version 2.3-9.

P. A. Patrician. Multiple Imputation for Missing Data. *Research in Nursing & Health*, 25(1):76–84, 2002.

R. B. Pebesma, E.J. Classes and Methods for Spatial Data in R. *R News*, 5 (2):22, 2005. URL `http://cran.r-project.org/doc/Rnews/`.

J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*, 2015. URL `http://CRAN.R-project.org/package=nlme`. R package version 3.1-120.

A. Pollice and M. Bilancia. Kriging with Mixed Effects Models. *Statistical Methods in Medical Research*, 62(3):405–429, 2002.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2015a. URL `http://www.R-project.org/`.

R Core Team. *foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ...*, 2015b. URL `http://CRAN.R-project.org/package=foreign`. R package version 0.8-63.

P. Ribeiro and P. Diggle. *geoR: Analysis of Geostatistical Data*, 2015. URL `http://CRAN.R-project.org/package=geoR`. R package version 1.7-5.1.

C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer texts in statistics. Springer, New York, 2. edition, 2004.

Roger Bivand. *spgrass6: Interface between GRASS 6 and R. R package version 0.8-6*, 2014.

M. W. Roos. Regional Price Levels in Germany. *Applied Economics*, 38, 2006. URL `http://books.google.de/books?id=lNCjPgAACAAJ`.

W. Rostin. Zwischenörtlicher Preisvergleich des Verbraucherpreisniveaus in 31 Städten. *Wirtschaft und Statistik*, 3:403–410, 1979.

D. B. Rubin. Multiple Imputations in Sample Surveys: A Phenomenological Bayesian Approach to Nonresponse. *The Proceedings of the Survey Research Methods section of the American Statistical Association*, pages 20–34, 1978.

D. B. Rubin. *Multiple Imputation for Nonresponse in Surveys*. Wiley, Hoboken, 1987.

D. B. Rubin. Characterizing the Estimation of Parameters in Incomplete-Data Problems. *Journal of the American Statistical Association*, 69(346): 467–474, 1974.

D. B. Rubin. An Overview of Multiple Imputation. *The Proceedings of the Survey Research Methods section of the American Statistical Association*, pages 79–84, 1988.

D. B. Rubin. Multiple Imputation After 18+ Years. *Journal of the American Statistical Association*, 91(434):473–489, 1996.

H. Rue. Fast Sampling of Gaussian Markov Random Fields. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 63(2):325–338, 2001.

B. Rüger. *Test- und Schätztheorie*. Oldenbourg, München and Wien, 1999.

O. Schabenberger and C. A. Gotway. *Statistical Methods for Spatial Data Analysis*. Taylor & Francis, 2005. URL `http://books.google.de/books?id=iVJuVLArmZcC`.

J. L. Schafer. *Analysis of Incomplete Multivariate Data*. Chapman and Hall/CRC, Boca Raton, 1997. URL `http://books.google.de/books?id=7GrWPQAACAAJ`.

J. L. Schafer. Multiple Imputation: a Primer. *Statistical Methods in Medical Research*, (8):3–15, 1999.

J. L. Schafer and M. K. Olsen. Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective. *Multivariate Behavioral Research*, 33(4):545–571, 1998.

S. R. Searle, G. Casella, and C. E. McCulloch. *Variance Components*. Wiley-Interscience paperback series. Wiley, Hoboken, NJ, 2006.

K. Soetaert. *shape: Functions for plotting graphical shapes, colors*, 2014. URL `http://CRAN.R-project.org/package=shape`. R package version 1.4.2.

Statistisches Bundesamt. Berechnungshandbuch: Verbundprogramm der Verbraucherpreisstatistik: Version 02.03.00, 2013.

M. L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, 1999.

G. Ströhl. Zwischenörtlicher Vergleich des Verbraucherpreises in 50 Städten. *Wirtschaft und Statistik*, page 415, 1994. URL 434.

J. Suedekum. Agglomeration and Regional Cost-of-Living, 2004. URL `gfr.ersa.org/sommerseminar_2006/agglom.pdf`.

T. Tabuchi. On Interregional Price Differentials. *The Japanese Economic Review*, 52(1):104–115, 2001.

N. Umlauf, D. Adler, T. Kneib, S. Lang, and A. Zeileis. Structured Additive Regression Models: An R Interface to BayesX. *Journal of Statistical Software*, 63(21):1–46, 2015. URL `http://www.jstatsoft.org/v63/i21/`.

W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, 4. edition, 2002. URL `http://www.stats.ox.ac.uk/pub/MASS4`.

F. Vogel. *Beschreibende und schließende Statistik*. Oldenbourg, München, 12 edition, 2000.

M. Voltz and R. Webster. A Comparison of Kriging, Cubic Splines and Classification for Predicting Soil Properties from Sample Information. *Journal of Soil Science*, 41(3):473–490, 1990.

L. von Auer. Räumliche Preisvergleiche: Aggregationskonzepte und Forschungsperspektiven. *Wirtschafts- und Sozialstatistisches Archiv*, 6(1): 27–56, 2012.

P. M. von der Lippe and C. C. Breuer. Regionale Kaufkraftvergleiche in Deutschland: Bedarf, Methoden und Machbarkeit. *AStA Wirtschafts- und Sozialstatistisches Archiv*, 3(1):25–40, 2009.

H. Wagner. Einführung in die Bayes-Statistik: Vorlesungsskript WS 2010/11. 03.02.2014.

M. J. Walden. Geographic Variation in Consumer Prices: Implications for Local Price Indices. *The Journal of Consumer Affairs*, 32(2):204–226, 1998.

G. S. Watson. Smoothing and Interpolation by Kriging and with Splines. *Journal of the International Association of Mathematical Geology*, 16 (6):601–615, 1984.

S. N. Wood. Thin Plate Regression Splines. *Journal of the American Statistical Association. Series B*, 65(1):95–114, 2003.

S. N. Wood. Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models. *Journal of the American Statistical Association*, 99(467):673–686, 2004.

P. Zhang. Multiple Imputation: Theory and Method. *International Statistical Review*, 71(3):581–592, 2003.

L. Zhu, B. P. Carlin, and A. E. Gelfand. Hierarchical Regression with Misaligned Spatial Data: Relating Ambient Ozone and Pediatric Asthma ER Visits in Atlanta. *Environmetrics*, 14(5):537–557, 2003.

D. L. Zimmerman. Computationally Efficient Restricted Maximum Likelihood Estimation of Generalized Covariance Functions. *Mathematical Geology*, 21(7):655–672, 1989.

# A Proofs and Computation

## A.1 R-packages used

- spgrass6:
  - Roger Bivand (2014)
- foreign:
  - R Core Team (2015b)
- classInt:
  - Bivand (2015)
- maps:
  - Original S code by Richard A. Becker and Allan R. Wilks. R version by Ray Brownrigg. Enhancements by Thomas P Minka (2014)
- maptools:
  - Bivand and Lewin-Koh (2015)
- nlme:
  - Pinheiro et al. (2015)
- spdep:
  - Bivand and Piras (2015)
  - Bivand et al. (2013a)
- shape:
  - Soetaert (2014)
- geoR:
  - Ribeiro and Diggle (2015)

- mvtnorm:
  - Genz et al. (2014)
  - Genz and Bretz (2009)
- MCMCpack:
  - Martin et al. (2011)
- MASS:
  - Venables and Ripley (2002)
- pracma:
  - Borchers (2015)
- sp:
  - Pebesma (2005)
  - Bivand et al. (2013b)
- akima:
  - Fortran code by H. Akima R port by Albrecht Gebhardt aspline function by Thomas Petzoldt interp2xyz and enhancements and corrections by Martin Maechler (2013)
- rgdal:
  - Bivand et al. (2015)

## A.2 Multiple Imputation

**proof of Rubin's Result 3.2**
As shown allready in the text, using the law of iterated expectations and variances:

$$\begin{aligned} E\left(\eta|Y_{obs,X}\right) &= E\left(E\left(\eta|X, Y_{obs}, Y_{mis}\right)|Y_{obs}, X\right) \\ &= E\left(\hat{\eta}|Y_{obs}, X\right) \end{aligned}$$

and

$$\begin{aligned} Var\left(\eta|Y_{obs}, X\right) &= Var\left(E\left(\eta|Y_{obs}, Y_{mis}, X\right)|Y_{obs}, X\right) && (49) \\ &\quad +E\left(Var\left(\eta|Y_{obs}, Y_{mis}, X\right)|Y_{obs}, X\right) \\ &= Var\left(\hat{\eta}|Y_{obs}, X\right) + E\left(U|Y_{obs}, X\right) && (50) \end{aligned}$$

where $\eta$ is the parameter of interest, $\hat{\eta}$ is its estimator, and $U$ is its variance.

**proof of Rubin's Result 3.4**
Before proving the result, Rubin (1987) shows by employing de Finettis rule that the joint distribution of $P(X, Y, R)$ can be reformulated in i.i.d. form such that the distribution of $Y$ is a product of realizations of a random variable with a distribution depending on the parameter $\theta$:

$$\begin{aligned} P(Y, X, R) &= \int \prod_{i=1}^{N} f(X_i, Y_i, R_i|\theta) P(\theta) d\theta \\ &= \int \prod_{i=1}^{N} f_{XY}(X_i, Y_i, R_i|\theta_{XY}) \\ &\quad f_{R|XY}(R_i|X_i, Y_i, \theta_{R|XY}) P(\theta) d\theta && (51) \end{aligned}$$

whereas $\theta_{XY}$ and $\theta_{R|XY}$ are distributional parameters and functions of the parameter $\theta$ itself. This result is essential for the following proof of Result 3.4.

Rubin (1987) just gives the procedure of the proof and does not execute it, but he lists following assumptions: „Suppose that:

1. the sampling mechanism is ignorable, and the joint distribution of $(X, Y, R)$ is modeled i.i.d. form (...), where
2. conditional on $\theta_{XY}$, the completed-data and complete-data posterior distributions of $Q$ (here:$\eta$) are equal (...) and
3. the completed-data and complete-data posterior distributions of $\theta_{XY}$ are equal (...).“(Rubin, 1987)

(2.) means that $P\left(\eta|X, Y_{inc}, R_{inc}, \theta_{XY}\right) = P\left(\eta|X, Y_{inc}, \theta_{XY}\right)$ and (3.) means that $P\left(\theta_{XY}|X, Y_{inc}, R_{inc}\right) = P\left(\theta|X, Y_{inc}\right)$. According to Rubin (1987), these two have to be multiplied and then intergrated over $\theta_{XY}$:

$$\int P\left(\eta|X, Y_{inc}, R_{inc}, \theta_{XY}\right) P\left(\theta_{XY}|X, Y_{inc}, R_{inc}\right) d\theta_{XY} =$$

$$= \int P\left(\eta|X, Y_{inc}, \theta_{XY}\right) P\left(\theta|X, Y_{inc}\right) d\theta_{XY}$$

$$\begin{aligned} E_{XY}\left[P\left(\eta|X, Y_{inc}, R_{inc}, \theta_{XY}\right)\right] &= E_{XY}\left[P\left(\eta|X, Y_{inc}, \theta_{XY}\right)\right] \\ P\left(\eta|X, Y_{inc}, R_{inc}\right) &= P\left(\eta|X, Y_{inc}\right) \end{aligned} \tag{52}$$

**q.e.d**

The proof given in Rubin (1987) follows directly from the formulae given in the definition.(Rubin, 1987) A more general proof is given in Rubin (1996):
**proof of result 4.1**(Rubin, 1996)
since:

- $E\left(\hat{\eta}|X, Y\right) \to \eta$
- $E\left(\overline{\eta}_{\infty}|X, Y, I\right) \to \hat{\eta}$

$$\begin{aligned} E\left(\overline{\eta}_{\infty}|X, Y\right) &= E\left(E\left(\overline{\eta}_{\infty}|X, Y, I\right)|X, Y\right) \\ &= E\left(\hat{\eta}|X, Y\right) = \eta \end{aligned}$$

and, second for the variances, since (Rubin, 1996)

- $E(U|X,Y) \rightarrow \ \mathrm{Var}(\hat{\eta}|X,Y)$
- $E\left(\overline{U}_\infty|X,Y,I\right) \rightarrow U$
- $E(B_\infty|X,Y,I) \rightarrow \ \mathrm{Var}\left(\overline{\eta}_\infty|X,Y,I\right)$

$$
\begin{aligned}
E(T_\infty|X,Y) &= E\left(\overline{U}_\infty|X,Y\right) + E(B_\infty|X,Y) \\
&= E\left(E\left(\overline{U}_\infty|X,Y,I\right)|X,Y\right) + \\
&\quad E(E(B_\infty|X,Y,I)|X,Y) \\
&= E(U|X,Y) + E\left(\mathrm{Var}\left(\overline{\eta}_\infty|X,Y,I\right)|X,Y\right) \\
&= \mathrm{Var}(\hat{\eta}|X,Y) + E\left(\mathrm{Var}\left(\overline{\eta}_\infty|X,Y,I\right)|X,Y\right) \\
&= \mathbf{var}\left(E\left(\overline{\eta}_\infty|X,Y,I\right)\right) + E\left(\mathrm{Var}\left(\overline{\eta}_\infty|X,Y\right)|X,Y,I\right) \\
&= \mathbf{var}(\overline{\eta}|X,Y)
\end{aligned}
$$

which equals approximately $\left(\eta - \overline{\eta}_\infty\right) \sim N(0, T_\infty)$.(Rubin, 1996)
**q.e.d.**

Result 5.1 can be proven in the following way.
**proof of Rubin's Result 5.1**

$$
\begin{aligned}
P(Y_{mis}|X,Y_{obs},\theta) &= \frac{P(X,Y_{mis},Y_{obs}|\theta)}{P(X,Y_{obs}|\theta)} \\
&= \frac{\int P(X,Y|\theta)\,dY_{exc}}{\int P(X,Y|\theta)\,dY_{nob}}
\end{aligned}
$$

as in the denominator all not observed cases and in the numerator the excluded are banned. Then using the i.i.d. assumption and the law of total probability leads to

$$
= \frac{\int\int\left[\prod_{i=1}^{N} f_{XY}(X_i,Y_i|\theta)\right]P(\theta)\,d\theta dY_{exc}}{\int\int\prod_{i=1}^{N} f_{XY}(X_i,Y_i|\theta)\,P(\theta)\,d\theta dY_{nob}}
$$

Afterwards, we exchange integrals and extract $P(\theta)$ from the inner integral:

$$= \frac{\int \int \left[\prod_{i=1}^{N} f_{XY}(X_i, Y_i|\theta)\right] dY_{exc} P(\theta)\, d\theta}{\int \int \prod_{i=1}^{N} f_{XY}(X_i, Y_i|\theta)\, dY_{nob} P(\theta)\, d\theta}$$

As the inner integrals in numerator and denominator are constant in $\theta$, it is possible to pull them out of the integral over $\theta$ and reduce afterwards:

$$= \frac{\int \int \left[\prod_{i=1}^{N} f_{XY}(X_i, Y_i|\theta)\right] dY_{exc}}{\int \int \prod_{i=1}^{N} f_{XY}(X_i, Y_i|\theta)\, dY_{nob}}$$

According to Rubin (1987) it is convenient to write:

$$f_{XY}(X_i, Y_i|\theta) = f_{Y|X}\left(Y_i|X_i, \theta_{Y|X}\right) f_X(X_i|\theta_X).$$

Using this relation, we get:

$$= \frac{\int \int \left[\prod_{i=1}^{N} f_{Y|X}\left(Y_i|X_i, \theta_{Y|X}\right) f_X(X_i|\theta_X)\right] dY_{exc}}{\int \int \prod_{i=1}^{N} f_{Y|X}\left(Y_i|X_i, \theta_{Y|X}\right) f_X(X_i|\theta_X)\, dY_{nob}}$$

Again, we pull constants out of the integral and reduce:

$$= \frac{\int \int \left[\prod_{i=1}^{N} f_{Y|X}\left(Y_i|X_i, \theta_{Y|X}\right)\right] dY_{exc}}{\int \int \prod_{i=1}^{N} f_{Y|X}\left(Y_i|X_i, \theta_{Y|X}\right) dY_{nob}}$$

According to the rule of Bayes, this is the same as

$$= P\left(Y_{mis}|X, Y_{obs}, \theta_{Y|X}\right)$$

**q.e.d.**(Rubin, 1987)

A short version of the following proof can be found in Rubin (1987). The following proof uses Fubinis Rule and Tonellis Rule.
**proof of Result 5.2**

$$
\begin{aligned}
P(\theta|X, Y_{obs}) &= \frac{P(Y_{obs}, X)\, P(\theta)}{P(\theta)} \\
&= \frac{\left[\prod_{i=1}^{N} f_{XY}\left(Y_{i,obs}, X_i|\theta\right)\right] P(\theta)}{\int \left[\prod_{i=1}^{N} f_{XY}\left(X_i, Y_{i,obs}|\theta\right)\right] P(\theta)\, d\theta} \\
&= \frac{\left[\prod_{i=1}^{N} f_{Y|X}\left(Y_{i,obs}, X_i|\theta_{Y|X}\right) f_X(X_i|\theta_X)\right]}{\int\int \left[\prod_{i=1}^{N} f_{Y|X}\left(Y_{i,obs}|X_i, \theta_{Y|X}\right) f_X(X_i|\theta_X)\right]} \\
&\quad \frac{P\left(\theta_{Y|X}\right) P(\theta_X)}{P\left(\theta_{Y|X}\right) P(\theta_X)\, d\theta_{Y|X} d\theta_X} \\
&= \frac{\left[\prod_{i=1}^{N} f_{Y|X}\left(Y_{i,obs}, X_i|\theta_{Y|X}\right)\right] P\left(\theta_{Y|X}\right)}{\int \left[\prod_{i=1}^{N} f_{Y|X}\left(Y_{i,obs}|X_i, \theta_{Y|X}\right)\right] P\left(\theta_{Y|X}\right) d\theta_{Y|X}} \\
&\quad \frac{\left[\prod_{i=1}^{N} f_X(X_i|\theta_X)\right] P(\theta_X)}{\int \left[\prod_{i=1}^{N} f_X(X_i|\theta_X)\right]) P(\theta_X)\, d\theta_X} \\
&= \frac{\left[\int \prod_{i=1}^{N} f_{Y|X}\left(Y_i, X_i|\theta_{Y|X}\right)\right] P\left(\theta_{Y|X}\right) dY_{nob}}{\int\int \left[\prod_{i=1}^{N} f_{Y|X}\left(Y_i|X_i, \theta_{Y|X}\right)\right] P\left(\theta_{Y|X}\right) dY_{nob} d\theta_{Y|X}} \\
&\quad \frac{\left[\prod_{i=1}^{N} f_X(X_i|\theta_X)\right] P(\theta_X)}{\int \left[\prod_{i=1}^{N} f_X(X_i|\theta_X)\right]) P(\theta_X)\, d\theta_X} \\
&= \frac{\prod_{i=1}^{N} f_X(X_i|\theta_X)\, P(\theta_X)}{\int \prod_{i=1}^{N} f_X(X_i|\theta_X)\, P(\theta_X)\, d\theta_X} \\
&\quad \frac{\int \prod_{i=1}^{N} f_{Y|X}\left(Y_i|X_i, \theta_{Y|X}\right) P\left(\theta_{Y|X}\right) dY_{nob}}{\int\int \prod_{i=1}^{N} f_{Y|X}\left(Y_i|X_i, \theta_{Y|X}\right) P\left(\theta_{Y|X}\right) dY_{nob} d\theta_{Y|X}} \\
&= \dots \frac{\int P\left(Y|X, \theta_{Y|X}\right) P\left(\theta_{Y|X}\right) dY_{nob}}{\int\int P\left(Y|X, \theta_{Y|X}\right) P\left(\theta_{Y|X}\right) dY_{nob} d\theta_{Y|X}}
\end{aligned}
$$

$$= \frac{\prod_{i=1}^{N} f_X\left(X_i|\theta_X\right) P\left(\theta_X\right)}{\int \prod_{i=1} N f_X\left(X_i|\theta_X\right) P\left(\theta_X\right) d\theta_X}$$
$$\frac{P\left(Y_{obs}|X, \theta_{Y|X}\right) P\left(\theta_{Y|X}\right)}{\int P\left(Y_{obs}|X, \theta_{Y|X}\right) P\left(\theta_{Y|X}\right) d\theta_{Y|X}}$$
$$= \frac{P\left(\theta, X\right)}{P\left(X\right)} \frac{P\left(Y_{obs}, \theta_{Y|X}|X\right)}{P\left(Y_{obs}\right)}$$
$$= P\left(\theta_X|X\right) P\left(\theta_{Y|X}|X, Y_{obs}\right)$$
$$= P\left(\theta_X|X, Y_{obs}\right) P\left(\theta_{Y|X}|X, Y_{obs}\right)$$

$\theta_X$ and $\theta_{Y|X}$ are a posteriori independent.
**q.e.d.**
The **proof of result 5.3** as univariate case follows direct from result 5.2.

Rubin (1987) proves the result 5.4 in the following manner:
**proof of result 5.4**
According to the rule of Bayes and the rule of total probability we can write:

$$P\left(\theta|X, Y_{obs}\right) = \frac{P\left(Y_{obs}|X, \theta\right) P\left(\theta\right)}{\int P\left(Y_{obs}|X, \theta\right) P\left(\theta\right) d\theta} \tag{53}$$

In $P\left(Y_{obs}|X, \theta\right) P\left(\theta\right)$, the notobserved are integrated out:

$$P\left(Y_{obs}|X, \theta\right) = \int P\left(Y|X, \theta\right) dY_{nob} \tag{54}$$

As the probability on the right side is a product according to the distinct structure,

$$P\left(Y|X, \theta\right) = \prod_{j=1}^{p} \prod_{i=1}^{N} f_{ij}$$

we can write

$$dY_{nob} = \prod_{j=1}^{p} \prod_{i \in nob[j]} dY_{ij}$$

Then 54 becomes

$$
\begin{aligned}
P\left(Y_{obs}|X,\theta\right) &= \int \dots \int \left[\prod_{j=1}^{p}\prod_{i=1}^{N} f_{ij}\right]\left[\prod_{j=1}^{p}\prod_{i\in nob[j]} dY_{ij}\right] \\
&= \int \dots \int \left[\prod_{j=1}^{p}\prod_{i\in obs[j]} f_{ij}\right]\left[\prod_{j=1}^{p}\prod_{i\in nob[j]} f_{ij} dY_{ij}\right] \\
&= \left[\prod_{j=1}^{p}\prod_{i\in obs[j]} f_{ij}\right] \underbrace{\int \dots \int \left[\prod_{j=1}^{p}\prod_{i\in nob[j]} f_{ij} dY_{ij}\right]}_{=1} = \prod_{j=1}^{p}\prod_{i\in obs[j]} f_{ij} \qquad (55)
\end{aligned}
$$

As the parameters $\theta$ are distinct and inserting 55 into 53 yields

$$
P\left(\theta|X,Y_{obs}\right) = \prod_{j=1}^{p} \frac{\prod_{i\in obs[j]} f_{ij} P\left(\theta_j\right)}{\int \prod_{i\in obs[j]} f_{ij} P\left(\theta_j\right) d\theta_j}
$$

**q.e.d.**

The following proof of result 5.5 is given in Rubin (1987).
**proof**

$$
\begin{aligned}
P\left(Y_{mis}|X,Y_{obs},\theta\right) &= \frac{P\left(Y_{mis},Y_{obs}|X,\theta\right)}{P\left(Y_{obs}|X,\theta\right)} \\
&= \frac{\prod_{j=1}^{p}\prod_{i\in inc} f_{ij}}{\prod_{j=1}^{p}\prod_{i\in obs[j]} f_{ij}} \\
&= \prod_{j=1}^{p} \frac{\prod_{i\in inc} f_{ij}}{\prod_{i\in obs[j]} f_{ij}} = \prod_{j=1}^{p}\prod_{i\in mis[j]} f_{ij}
\end{aligned}
$$

**q.e.d.**

## A.3 Kriging

All derivations and proofs looking at one region base on Cressie (1990), Schabenberger and Gotway (2005), Stein (1999). The other computations of this section are conducted by the author and inspired by the given literatur.

### A.3.1 Simple Kriging

If we look at one unobserved region, the MSPE is:

$$\begin{aligned} MSPE(\lambda_0, \lambda) = E\left[\left(y(s_0) - \lambda_0 - \lambda' y\right)^2\right] & \\ &= Var\left(y(s_0)\right) + E\left(y(s_0)\right)^2 - 2\lambda_0 E\left(y(s_0)\right) \\ &- 2\lambda' Cov\left(y, y(s_0)\right) - 2\lambda' E\left(y\right) \\ &E\left(y(s_0)\right) + \lambda_0^2 + 2\lambda_0 \lambda' E\left(y\right) + \lambda' Var\left(y\right) \lambda \\ &+ \lambda' E\left(y\right)^2 \lambda \\ &= Var\left(y(s_0)\right) - 2\lambda' Cov\left(y, y(s_0)\right) + \lambda' Var\left(y\right) \lambda \\ &+ \left(\mu(s_0) - \lambda_0 - \lambda' \mu\right)^2 \end{aligned}$$

Partially deviating to the parameter $\lambda_0$ and the parameter-vector $\lambda$ gives

- $\frac{\partial MSPE(\cdot)}{\partial \lambda} = -2Cov\left(y, y(s_0)\right) + 2Var(y)\lambda - 2\lambda'\left(\mu(s_0) - \lambda_0 - \lambda' \mu\right)$
- $\frac{\partial MSPE(\cdot)}{\partial \lambda_0} = -2\left(\mu(s_0) - \lambda_0 - \lambda' \mu\right)$

Setting the latter equal to zero we can compute the optimal $\lambda_0$

$$\widehat{\lambda_0} = \mu(s_0) - \lambda' \mu$$

which can be used in the equation for $\lambda$, which gives MSPE-optimal predictor-coefficients $\lambda$:

$$\hat{\lambda} = Var(y)^{-1} Cov(y, y(s_0))$$

If we insert these parameters in the equation or the predictor, we get the MSPE-optimal predictor:

$$\begin{aligned}\hat{y}(s_0) =& \widehat{\lambda_0} + \hat{\lambda}'y \\ &= \mu(s_0) - \hat{\lambda}'\mu + (Var(y)^{-1}Cov(y, y(s_0)))'y \\ &= \mu(s_0) - (Var(y)^{-1}Cov(y, y(s_0)))'\mu + (Var(y)^{-1} \\ &Cov(y, y(s_0)))'y \\ &= \mu(s_0) + Cov(y, y(s_0))'Var(y)^{-1}(y - \mu)\end{aligned}$$

The MSPE of the optimal predictor of one region becomes after inserting the optimal estimators $\lambda_0$ and $\lambda$

$$MSPE_{sk,sing} = Var(y(s_0)) - Cov(y, y(s_0))'Var^{-1}Cov(y, y(s_0))$$

If we look at all unobserved regions at once, the MSPE can be derivated by:

$$\begin{aligned}&E\left[\left(y_0 - i\lambda_0 - i\lambda'y\right)\left(y_0 - i\lambda_0 - i\lambda'y\right)'\right] = \\ &\quad E\left(y_0y_0' - y_0\lambda_0 i' - i\lambda' y_0 y' - i\lambda_0 y_0' \right.\\ &\quad \left. i\lambda_0\lambda_0' i' + i\lambda_0 y'\lambda i' - i\lambda' y y_0 + i\lambda' y\lambda_0' i' + i\lambda' yy'\lambda i'\right) \\ &\quad = E\left(y_0y_0'\right) - 2\lambda_0 iE\left(y_0'\right) - 2i\lambda' E\left(y_0y'\right) + i\lambda_0^2 i' + 2i\lambda_0 E\left(y'\right)\lambda i' + \\ &\quad i\lambda' E\left(yy'\right)\lambda i' \\ &\quad = Var\left(y_0\right) + \mu_0\mu_0' - 2\lambda_0 i\mu_0' - 2i\lambda'\left[Cov(y, y_0) + \mu_0\mu\right] + i\lambda_0^2 i' + \\ &\quad 2i\lambda_0\mu'\lambda i' + i\lambda'\left[Var(y) + E(y)E(y)'\right]\lambda i' \\ &\quad = Var\left(y_0\right) - 2i\lambda' Cov(y, y_0) + i\lambda' Var(y)\lambda i' + \\ &\quad \left[\mu_0 - i\left(\lambda_0 + \lambda'\mu\right)\left(\lambda_0 + \lambda'\mu\right)'\right]\end{aligned}$$

Again we minimize the MSPE and get the two partial derivatives:

- $\frac{\partial MSPE(\cdot)}{\partial \lambda} = -2Cov\left(y, y_0\right) + 2Var(y)\lambda i' i - 2\mu i'(\mu_0 - i\lambda_0 - i\lambda'\mu)$
- $\frac{\partial MSPE(\cdot)}{\partial \lambda_0} = -2i\left(\mu_0 - i\lambda_0 - i\lambda'\mu\right)'$

Setting the latter formula equal to 0, we get

$$i\lambda_{0,sk,mult} = \mu_0 - i\lambda'\mu$$

Using this result in the partial derivative of $\lambda$ and setting again to zero delivers:

$$-Cov(y,y_0)i' + Var(y)\lambda i'i = 0$$

From which follows that

$$\lambda_{sk,mult} = Var(y)^{-1}Cov(y,y_0)i(i'i)^{-1}$$

Inserting this result in the formula of the optimal predictor gives:

$$\begin{aligned}
\widehat{y_{sk,mult}} &= i\lambda_0 + i\lambda' y \\
&= \mu_0 - i\lambda'(y-\mu) \\
&= \mu_0 - Cov(y,y_0)'Var(y)^{-1}(y-\mu)
\end{aligned}$$

The correspondent MSPE is:

$$\begin{aligned}
MSPE_{sk,mult} &= Var(y_0) - 2i\left(Var(y)^{-1}Cov(y,y_0)i(i'i^{-1})\right)' \\
&\quad Cov(y,y_0) \\
&\quad + i\left(Var(y)^{-1}Cov(y,y_0)i(i'i)^{-1}\right)' Var(y)Var(y)^{-1} \\
&\quad Cov(y,y_0)i(i'i)^{-1}i' \\
&\quad + (\mu_0 - \mu_0 + i\lambda'\mu - i\lambda'\mu)(\mu_0 - \mu_0 + i\lambda'\mu - i\lambda'\mu)' \\
&= Var(y_0) - 2Cov(y,y_0)'Var(y)^{-1}Cov(y,y_0) \\
&\quad + Cov(y,y_0)'Var(y)^{-1}Cov(y,y_0) \\
&= Var(y_0) - Cov(y,y_0)'Var(y)^{-1}Cov(y,y_0)
\end{aligned}$$

### A.3.2 Ordinary Kriging

The following computations are based on the formula and ideas given in Schabenberger and Gotway (2005).
The optimization of the MSPE under the constraint of a linear mixture gives the Lagrange equation:

$$
\begin{aligned}
&E\left[\left(\lambda' y - y(s_0)\right)\left(\lambda' y - y(s_0)\right)'\right] - 2m\left(\lambda' i - 1\right) = \\
&\quad = E\left[\lambda' y y' \lambda - 2\lambda' y y(s_0) + y(s_0)^2\right] - 2m\left(\lambda' i - 1\right) \\
&\quad = \lambda' E\left(yy'\right)\lambda - 2\lambda' E\left(yy(s_0)\right) + E\left(y(s_0)^2\right) - 2m\left(\lambda' i - 1\right) \\
&\quad = \lambda' Var\left(y\right)\lambda + \lambda' E\left(y\right) E\left(y\right)\lambda - 2\lambda' Cov\left(yy(s_0)\right) - 2\lambda' E\left(y\right) \\
&\quad E\left(y(s_0)\right) + Var\left(y(s_0)\right) + E\left(y(s_0)\right)^2 - 2m\left(\lambda' i - 1\right) \\
&\quad = \lambda' Var\left(y\right)\lambda + \mu_y^2 \underbrace{\lambda' i i' \lambda}_{=1} - 2\lambda' Cov\left(yy(s_0)\right) - 2\mu_y^2 \underbrace{\lambda' i}_{=1} \\
&\quad + Var\left(y(s_0)\right) + \mu_y^2 - 2m\left(\lambda' i - 1\right) \\
&\quad = \lambda' Var\left(y\right)\lambda - 2\lambda' Cov\left(yy(s_0)\right) + Var\left(y(s_0)\right) - 2m\left(\lambda' i - 1\right)
\end{aligned}
$$

The partial derivations are:

- $\frac{\partial MSPE(\cdot)}{\partial \lambda} = 2Var(y)\lambda - 2Cov\left(y, y(s_0)\right) - 2mi$
- $\frac{\partial MSPE(\cdot)}{\partial m} = 2\left(\lambda' i - 1\right)$

Setting the first derivation to zero and solving to $\lambda$ gives

$$\lambda = Var(y)^{-1}\left(Cov\left(y, y(s_0)\right) + im\right)$$

Inserting $\lambda$ into the second partial derivation set to zero and solving for $m$ gives

$$\widehat{m}_{ok} = \left(1 - iVar(y)^{-1} Cov(y, y(s_0))\right)\left(i' Var(y)^{-1} i\right)^{-1}$$

After inserting the optimal $m$ into the foregoing formula we get the result for the mixture coefficients:

$$\widehat{\lambda}_{ok} = Var(y)^{-1}\left(Cov\,(y, y(s_0)) + i\left(1 - iVar(y)^{-1}Cov(y, y(s_0))\right)\right.$$
$$\left.\left(i'Var(y)^{-1}i\right)^{-1}\right)$$

This defines the ordinary kriging predictor:

$$\widehat{y}_{ok}(s_0) = \left(Cov\,(y, y(s_0)) + i\left(1 - iVar(y)^{-1}Cov(y, y(s_0))\right)\right.$$
$$\left.\left(i'Var(y)^{-1}i\right)^{-1}\right)' Var(y)^{-1}y$$

As the mean has to be estimated implicitely in this formula, Schabenberger and Gotway (2005) proposes to use the GLS-estimator of the mean:

$$\widehat{\mu} = \left(i'Var(y)^{-1}i\right)^{-1} i'Var(y){-}1y$$

and then reformulate the formula above in the following manner:

$$\widehat{y}_{ok}(s_0) = (\;Cov\,(y, y(s_0)) + i\left(1 - iVar(y)^{-1}Cov(y, y(s_0))\right)$$
$$\left(i'Var(y)^{-1}i\right)^{-1}\;)'\;Var(y)^{-1}y$$
$$= \left(Cov\,(y, y(s_0)) + i\left(i'Var(y)^{-1}i\right)^{-1}\right.$$
$$\left. -i\left(i'Var(y)^{-1}i\right)^{-1}\left(iVar(y)^{-1}Cov(y, y(s_0))\right)\right)'$$
$$Var(y)^{-1}y$$
$$= Cov\,(y, y(s_0))'\;Var(y)^{-1}y$$
$$+ \left(i'Var(y)^{-1}i\right)^{-1} i'Var(y)^{-1}y$$
$$- Cov(y, y(s_0))Var(y)^{-1}i\left(i'Var(y)^{-1}i\right)^{-1} i'Var(y)^{-1}y$$

$$= \underbrace{\left(i'Var(y)^{-1}i\right)^{-1} i'Var(y)^{-1}y}_{\widehat{\mu_y}}$$

$$+ Cov\,(y, y(s_0))'\,Var(y)^{-1}y$$

$$- Cov(y, y(s_0))Var(y)^{-1}i\,\underbrace{\left(i'Var(y)^{-1}i\right)^{-1} i'Var(y)^{-1}y}_{\widehat{\mu_y}}$$

$$= \widehat{\mu_y} + Cov\,(y, y(s_0))'\,Var(y)^{-1}y$$

$$- Cov(y, y(s_0))Var(y)^{-1}i\widehat{\mu_y}$$

The MSPE of the ordinary kriging predictor can be derived by inserting the predictor into the MSPE formula above:

$$\begin{aligned} MSPE_{ok} = Var(y(s_0)) + \Bigg(&Var(y)^{-1}\,(Cov(y, y(s_0)) \\ &+ i\frac{\left(1 - i'Var(y)^{-1}Cov(y, y(s_0))\right)}{i'Var(y)^{-1}i}\Bigg)\Bigg)' \\ &Var(y)\Bigg(Var(y)^{-1}\,(Cov(y, y(s_0)) \\ &+ i\frac{\left(1 - i'Var(y)^{-1}Cov(y, y(s_0))\right)}{i'Var(y)^{-1}i}\Bigg)\Bigg) \\ &- 2\Bigg(Var(y)^{-1}\,(Cov(y, y(s_0)) \\ &+ i\frac{\left(1 - i'Var(y)^{-1}Cov(y, y(s_0))\right)}{i'Var(y)^{-1}i}\Bigg)\Bigg)'\,Cov(y, y(s_0)) \\ =&Var(y(s_0)) + Cov(y, y(s_0))'Var(y)^{-1}Cov(y, y(s_0)) \\ &+ Cov(y, y(s_0))'Var(y)^{-1} \\ &\left(i\frac{\left(1 - i'Var(y)^{-1}Cov(y, y(s_0))\right)}{i'Var(y)^{-1}i}\right) + \end{aligned}$$

$$+\left(i\frac{\left(1-i'Var(y)^{-1}Cov(y,y(s_0))\right)}{i'Var(y)^{-1}i}\right)'$$

$$Var(y)^{-1}Cov(y,y(s_0))$$

$$+\left(i\frac{\left(1-i'Var(y)^{-1}Cov(y,y(s_0))\right)}{i'Var(y)^{-1}i}\right)'Var(y)^{-1}$$

$$\left(i\frac{\left(1-i'Var(y)^{-1}Cov(y,y(s_0))\right)}{i'Var(y)^{-1}i}\right)$$

$$-2Cov(y,y(s_0))'Var(y)^{-1}Cov(y,y(s_0))$$

$$-2\left(i\frac{\left(1-i'Var(y)^{-1}Cov(y,y(s_0))\right)}{i'Var(y)^{-1}i}\right)'$$

$$Var(y)^{-1}Cov(y,y(s_0))$$

$$= Var(y(s_0)) - Cov(y,y(s_0))'Var(y)^{-1}Cov(y,y(s_0))$$

$$+\frac{\left(1-i'Var(y)^{-1}Cov(y,y(s_0))\right)}{i'Var(y)^{-1}i}i'Var(y)^{-1}i$$

$$\frac{\left(1-i'Var(y)^{-1}Cov(y,y(s_0))\right)}{i'Var(y)^{-1}i}$$

$$=Var(y(s_0)) - Cov(y,y(s_0))'Var(y)^{-1}Cov(y,y(s_0))$$

$$+\frac{\left(1-i'Var(y)^{-1}Cov(y,y(s_0))\right)^2}{i'Var(y)^{-1}i}$$

### A.3.3 Universal Kriging

The starting points and the results of the following computations can be found in Schabenberger and Gotway (2005). The MSPE and the constraint yield the following Langrange function to be minimized:

$$
\begin{aligned}
MSPE = &\min_a a'Var(y)a + Var(y(s_0)) - 2a'Cov(y, y(s_0)) \\
&+ 2m'(X'a - x(s_0))
\end{aligned}
$$

The partial derivations are:

- $\frac{\partial MSPE(\cdot)}{\partial m} = 2(X'a - x(s_0))$
- $\frac{\partial MSPE(\cdot)}{\partial a} = 2Var(y)a - 2Cov(y, y(s_0)) + 2Xm$

Setting to zero and solving the first equation for $a$ gives:

$$a = X'^{-1}x(s_0)$$

Inserting this into the second yields:

$$Var(y)X'^{-1}x(s_0) + Xm = Cov(y, y(s_0))$$

from which follows that

$$
\begin{aligned}
m &= X'^{-1}Cov(y, y(s_0)) - X^{-1}Var(y)X'^{-1}x(s_0) \\
&= X^{-1}Var(y)X'^{-1}\left(X'Var(y)^{-1}Cov(y, y(s_0)) - x(s_0)\right)
\end{aligned}
$$

Inserting $m$ into the second equation solved for $a$ gives the optimal mixture coefficients:

$$\begin{aligned}
a &= Var(y)^{-1}\left(Cov(y, y(s_0)) - Xm\right)\\
&= Var(y)^{-1}\Big[Cov(y, y(s_0)) - X\left(X^{-1}Var(y)X'^{-1}\right.\\
&\left.\left(X'Var(y)^{-1}Cov(y, y(s_0)) - x(s_0)\right)\right)\Big]\\
&= \left[Var(y)^{-1} - Var(y)^{-1}X\left(X'Var(y)^{-1}X\right)^{-1}X'Var^{-1}\right]\\
&Cov(y, y(s_0)) + Var(y)^{-1}X\left(X'Var(y)^{-1}X\right)^{-1}x(s_0)
\end{aligned}$$

The optimal universal kriging predictor is then:

$$\begin{aligned}
\widehat{y}_{uk}(s_0) &= a'y\\
&= x(s_0)'\underbrace{\left(X'Var(y)^{-1}X\right)^{-1}X'Var(y)^{-1}y}_{\widehat{\beta}_{GLS}}\\
&+ Cov(y, y(s_0))Var(y)^{-1}\\
&\left(y - X\underbrace{\left(X'Var(y)^{-1}X\right)^{-1}X'Var(y)^{-1}y}_{\widehat{\beta}_{GLS}}\right)\\
&= x(s_0)'\widehat{\beta}_{GLS} + Cov(y, y(s_0))Var(y)^{-1}\left(y - X\widehat{\beta}_{GLS}\right)
\end{aligned}$$

and its MSPE can be computed in the following manner:

$$\begin{aligned}
MSPE_{uk} &= a'Var(y)a + Var(y(s_0)) - 2a'Cov(y, y(s_0))\\
&+ 2m'(X'a - x(s_0)) =
\end{aligned}$$

$$= Var(y(s_0)) + \Big[Cov(y, y(s_0))' Var(y(s_0))^{-1}$$
$$- Cov(y, y(s_0))' Var(y(s_0))^{-1} X(X' Var(y(s_0))^{-1} X)^{-1}$$
$$X' Var(y(s_0))^{-1}$$
$$+ x(s_0) \left(X' Var(y(s_0))^{-1} X\right)^{-1} X' Var(y(s_0))^{-1}\Big]$$
$$Var(y(s_0)) \Big[Var(y(s_0))^{-1} Cov(y, y(s_0))$$
$$- Var(y(s_0))^{-1} X(X' Var(y(s_0))^{-1} X)^{-1} X' Var(y(s_0))^{-1}$$
$$Cov(y, y(s_0))$$
$$+ Var(y(s_0))^{-1} X(X' Var(y(s_0)^{-1} X)^{-1} x(s_0)\Big]$$
$$- 2\Big[Cov(y, y(s_0))' Var(y(s_0))^{-1}$$
$$- Cov(y, y(s_0))' Var(y(s_0))^{-1} X(X' Var(y(s_0))^{-1} X)^{-1}$$
$$X Var(y(s_0))^{-1}$$
$$+ x(s_0) \left(X' Var(y(s_0))^{-1} X\right)^{-1}$$
$$X' Var(y(s_0))^{-1}\Big] Cov(y, y(s_0))$$
$$= Var(y(s_0)) - Cov(y, y(s_0))' Var(y(s_0))^{-1} Cov(y, y(s_0))$$
$$- Cov(y, y(s_0))' Var(y(s_0))^{-1} X(X' Var(y(s_0))^{-1} X)^{-1}$$
$$X' Var(y(s_0))^{-1} Cov(y, y(s_0))$$
$$+ Cov(y, y(s_0))' Var(y(s_0))^{-1} X(X' Var(y(s_0))^{-1} X)^{-1} x(s_0)$$
$$- Cov(y, y(s_0))' Var(y(s_0))^{-1} X(X' Var(y(s_0))^{-1} X)^{-1}$$
$$X' Var(y(s_0))^{-1} Cov(y, y(s_0))$$
$$+ Cov(y, y(s_0))' Var(y(s_0))^{-1} X(X' Var(y(s_0))^{-1} X)^{-1}$$
$$X' Var(y(s_0))^{-1} X$$
$$(X' Var(y(s_0))^{-1} X)^{-1} X' Var(y(s_0))^{-1} Cov(y, y(s_0))$$
$$- Cov(y, y(s_0))' Var(y(s_0))^{-1} X(X' Var(y(s_0))^{-1} X)^{-1}$$
$$X' Var(y(s_0))^{-1} X$$
$$(X' Var(y(s_0))^{-1} X)^{-1} x(s_0)$$
$$+ x(s_0)' (X' Var(y(s_0))^{-1} X)^{-1} X' Var(y(s_0))^{-1} Cov(y, y(s_0))$$

$$
\begin{aligned}
&- x(s_0)'(X'Var(y(s_0))^{-1}X)^{-1}X'Var(y(s_0))^{-1}X \\
&(X'Var(y(s_0))^{-1}X)^{-1}X'Var(y(s_0))^{-1}Cov(y, y(s_0)) \\
&+ x(s_0)'(X'Var(y(s_0))^{-1}X)^{-1}X'Var(y(s_0))^{-1} \\
&X(X'Var(y(s_0))^{-1}X)^{-1}x(s_0) \\
&- 2Cov(y, y(s_0))Var(y(s_0))^{-1}Cov(y, y(s_0)) \\
&+ 2Cov(y, y(s_0))Var(y(s_0))^{-1}X(X'Var(y(s_0))^{-1}X)^{-1}X' \\
&Var(y(s_0))^{-1}Cov(y, y(s_0)) \\
&- 2x(s_0)'(X'Var(y(s_0))^{-1}X)X'Var(y(s_0))^{-1}Cov(y, y(s_0)) \\
&= Var(y(s_0)) - Cov(y, y(s_0))'Var(y(s_0))^{-1}Cov(y, y(s_0)) \\
&+ x(s_0)'X(X'Var(y(s_0))^{-1}X)^{-1}x(s_0) \\
&+ Cov(y, y(s_0))'Var(y(s_0))^{-1}X(X'Var(y(s_0))^{-1}X)^{-1} \\
&X'Var(y(s_0))^{-1}Cov(y, y(s_0)) \\
&+ Cov(y, y(s_0))'Var(y(s_0))^{-1}X(X'Var(y(s_0))^{-1}X)^{-1}x(s_0) \\
&- Cov(y, y(s_0))'Var(y(s_0))^{-1}X(X'Var(y(s_0))^{-1}X)^{-1}x(s_0) \\
&+ x(s_0)X(X'Var(y(s_0))^{-1}X)^{-1}X'Var(y(s_0))^{-1}Cov(y, y(s_0)) \\
&- x(s_0)X(X'Var(y(s_0))^{-1}X)^{-1}X'Var(y(s_0))^{-1}Cov(y, y(s_0)) \\
&- 2x(s_0)'X(X'Var(y(s_0))^{-1}X)^{-1}X'Var(y(s_0))^{-1}Cov(y, y(s_0)) \\
&= Var(y(s_0)) - Cov(y, y(s_0))'Var(y(s_0))^{-1}Cov(y, y(s_0)) \\
&+ x(s_0)'(X'Var(y(s_0))^{-1}X)^{-1}x(s_0) \\
&- x(s_0)'(X'Var(y(s_0))^{-1}X)^{-1}Cov(y, y(s_0))'Var(y(s_0))^{-1}X \\
&- Cov(y, y(s_0))'Var(y(s_0))^{-1}X(X'Var(y(s_0))^{-1}X)^{-1}x(s_0) \\
&+ Cov(y, y(s_0))'Var(y(s_0))^{-1}X(X'Var(y(s_0))^{-1}X)^{-1} \\
&X'Var(y(s_0))^{-1}Cov(y, y(s_0)) \\
&= Var(y(s_0)) - Cov(y, y(s_0))'Var(y(s_0))^{-1}Cov(y, y(s_0)) \\
&+ \left(x(s_0)' - Cov(y, y(s_0))'Var(y(s_0))^{-1}X\right)\left(X'Var(y(s_0))^{-1}X\right)^{-1} \\
&\left(x(s_0)' - Cov(y, y(s_0))'Var(y(s_0))^{-1}X\right)'
\end{aligned}
$$

## A.4 Full Conditionals of the NIG Regression Model

The posterior in the NIG-case is:(Fahrmeir et al., 2007)

$$
\begin{aligned}
P(\beta,\sigma^2|y,X) \propto \; & \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\,(y-X\beta)'\,(y-X\beta)\right\} \\
& \frac{1}{(\sigma^2)^{\frac{p}{2}}} \exp\left\{\frac{1}{2\sigma^2}\,(\beta-m)'\,M^{-1}\,(\beta-m)\right\} \\
& \frac{1}{(\sigma^2)^{a+1}} \exp\left\{-\frac{b}{\sigma^2}\right\}
\end{aligned}
$$

from which we can compute the following full conditionals:(Fahrmeir et al., 2007)

$$
\begin{aligned}
P(\beta|\sigma^2,y,X) \propto \; & \exp\left\{-\frac{1}{2\sigma^2}\,(y-X\beta)'\,(y-X\beta)\right\} \\
& \exp\left\{\frac{1}{2\sigma^2}\,(\beta-m)'\,M^{-1}\,(\beta-m)\right\} \\
= & \exp\left\{-\frac{1}{2\sigma^2}y'y+\frac{1}{\sigma^2}\beta'X'y-\frac{1}{2\sigma^2}\beta'X'X\beta\right\} \\
& \exp\left\{-\frac{1}{2\sigma^2}\beta'M^{-1}\beta-\frac{1}{2\sigma^2}m'M^{-1}m+\frac{1}{\sigma^2}\beta'M^{-1}m\right\} \\
\propto & \exp\left\{\frac{1}{\sigma^2}\beta'X'y-\frac{1}{2\sigma^2}\beta'X'X\beta\right\} \\
& \exp\left\{-\frac{1}{2\sigma^2}\beta'M^{-1}\beta+\frac{1}{\sigma^2}\beta'M^{-1}m\right\} \\
= & \exp\Bigg\{\,\beta'\left(\frac{1}{\sigma^2}X'y+\frac{1}{\sigma^2}M^{-1}m\right) \\
& -\frac{1}{2}\beta'\left(\frac{1}{\sigma^2}X'X+\frac{1}{\sigma^2}M^{-1}\right)\beta\Bigg\}
\end{aligned}
$$

which is a multivariate normal with $\mu_\beta = \Sigma_\beta\left(\frac{1}{\sigma^2}X'y + \frac{1}{\sigma^2}M^{-1}m\right)$ and $\Sigma_\beta = \left(\frac{1}{\sigma^2}X'X + \frac{1}{\sigma^2}M^{-1}\right)^{-1}$.
The full conditional for the variance is:

$$
\begin{aligned}
P(\sigma^2|\beta, y, X) \quad & \propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right\} \\
& \frac{1}{(\sigma^2)^{\frac{p}{2}}} \exp\left\{\frac{1}{2\sigma^2}(\beta - m)' M^{-1}(\beta - m)\right\} \\
& \frac{1}{(\sigma^2)^{a-1}} \exp\left\{-\frac{b}{\sigma^2}\right\} \\
= \quad & \frac{1}{(\sigma^2)^{a+\frac{n}{2}+\frac{p}{2}-1}} \exp\left\{-\frac{1}{\sigma^2}\left[\frac{1}{2}(y - X\beta)'(y - X\beta)\right.\right. \\
& \left.\left.+\frac{1}{2}(\beta - m)' M^{-1}(\beta - m) + b\right]\right\}
\end{aligned}
$$

which is an Inverse Gamma distribution with $a' = \frac{n}{2} + \frac{p}{2} + a$ and $b' = \frac{1}{2}(y - X\beta)'(y - X\beta) + \frac{1}{2}(\beta - m)' M^{-1}(\beta - m) + b$

## A.5 Full Conditionals of the KriMI Model

The posterior of the KriMI-model is:

$$
\begin{aligned}
&P\left(\beta,\gamma,\sigma^2,\tau^2|y,X\right) = P\left(y|\beta,\gamma,\sigma^2,\tau^2,X\right)P\left(\beta,\gamma,\sigma^2,\tau^2\right)\\
&= P\left(y|\beta,\gamma,\sigma^2,\tau^2,X\right)P\left(\beta,\sigma^2\right)P\left(\gamma,\tau^2\right)\\
&= P\left(y|\beta,\gamma,\sigma^2,\tau^2,X\right)P\left(\beta|\sigma^2\right)P(\sigma^2)P\left(\gamma|\tau^2\right)P(\tau^2)\\
&\propto det(\sigma^2 I)^{-\frac{1}{2}}\exp\{-\frac{1}{2}(y-X\beta-Z\gamma)'\left(\sigma^2 I\right)^{-1}\\
&(y-X\beta-Z\gamma)\}\\
&det(\sigma^2 M)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\beta-m)'\left(\sigma^2 M\right)^{-1}(\beta-m)\right\}\\
&det(\tau^2 R)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\gamma-0)'\left(\tau^2 R\right)^{-1}(\gamma-0)\right\}\\
&\frac{1}{\sigma^{2(a_\sigma+1)}}\exp\left\{-\frac{b_\sigma}{\sigma^2}\right\}\frac{1}{\tau^{2(a_\tau+1)}}\exp\left\{-\frac{b_\tau}{\tau^2}\right\}
\end{aligned}
$$

Now, we can derive the full conditionals of the particular parameters by leaving out all fixed parameters and looking for a well known distribution. The full conditional of $\beta$ is:

$$
\begin{aligned}
&P\left(\beta|\gamma,\sigma^2,\tau^2,y,X\right) \propto \exp\{-\frac{1}{2}(y-X\beta-Z\gamma)'\left(\sigma^2 I\right)^{-1}\\
&\quad(y-X\beta-Z\gamma)-\frac{1}{2}(\beta-m)'\left(\sigma^2 M\right)^{-1}(\beta-m)\}\\
&=\exp\left\{-\frac{1}{2}\left[y'\left(\sigma^2 I\right)^{-1}y-2\beta'X'\left(\sigma^2 I\right)^{-1}y+\beta'X'\left(\sigma^2 I\right)^{-1}X\beta\right.\right.\\
&\quad+2\beta'X'\left(\sigma^2 I\right)^{-1}Z\gamma\\
&\quad+2y'\left(\sigma^2 I\right)^{-1}Z\gamma+\gamma'Z'\left(\sigma^2 I\right)^{-1}Z\gamma\\
&\quad\left.\left.+\ \beta'\left(\sigma^2 M\right)^{-1}\beta-2\beta'\left(\sigma^2 M\right)^{-1}m+m'\left(\sigma^2 M\right)^{-1}m\right]\right\}\\
&\propto\exp\left\{-\frac{1}{2}\left[-2\beta'X'\left(\sigma^2 I\right)^{-1}y+\beta'X'\left(\sigma^2 I\right)^{-1}X\beta+2\beta'X'\left(\sigma^2 I\right)^{-1}\right.\right.\\
&\quad\left.\left.Z\gamma+\ \beta'\left(\sigma^2 M\right)^{-1}\beta-2\beta'\left(\sigma^2 M\right)^{-1}m\right]\right\}
\end{aligned}
$$

$$= \exp\left\{-\frac{1}{2}\left[\beta' \underbrace{\left(X'\left(\sigma^2 I\right)^{-1} X + \left(\sigma^2 M\right)^{-1}\right)}_{\Sigma_\beta^{-1}} \beta\right]\right.$$

$$\left. + \beta' X' \left(\sigma^2 I\right)^{-1} y - \beta' X' \left(\sigma^2 I\right)^{-1} Z\gamma + \beta' \left(\sigma^2 M\right)^{-1} m\right\}$$

$$= \exp\left\{\frac{1}{2}\beta'\Sigma_\beta^{-1}\beta + \beta'\left[\Sigma_\beta^{-1}\Sigma_\beta X'\left(\sigma^2 I\right)^{-1}(y - Z\gamma) + \left(\sigma^2 M\right)^{-1} m\right]\right\}$$

which is a multivariate normal with:

- $\mu_\beta = \Sigma_\beta \left[X'\left(\sigma^2 I\right)^{-1}(y - Z\gamma) + \left(\sigma^2 M\right)^{-1} m\right]$
- $\Sigma_\beta = \left(X'\left(\sigma^2 I\right)^{-1} X + \left(\sigma^2 M\right)^{-1}\right)^{-1}$

Next, we derive the form of the full conditional of the spatial effects:

$$P\left(\gamma|\beta, \sigma^2, \tau^2 y, X\right) \propto \exp\left\{-\frac{1}{2}(y - X\beta - Z\gamma)'\left(\sigma^2 I\right)^{-1}\right.$$

$$\left.(y - X\beta - Z\gamma) - \frac{1}{2}\gamma'\left(\tau^2 R\right)^{-1}\gamma\right\} =$$

$$= \exp\left\{-\frac{1}{2}\left[y'\left(\sigma^2 I\right)^{-1} y - 2X'\beta'\left(\sigma^2 I\right)^{-1} y\right.\right.$$

$$- 2\gamma' Z'\left(\sigma^2 I\right)^{-1} y + \beta' X'\left(\sigma^2 I\right)^{-1} X\beta + 2\beta' X'\left(\sigma^2 I\right)^{-1} Z\gamma$$

$$\left.\left.\gamma' Z'\left(\sigma^2 I\right)^{-1} Z\gamma + \gamma'\left(\tau^2 R\right)^{-1}\gamma\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\gamma' \underbrace{\left[Z'\left(\sigma^2 I\right)^{-1} Z + \left(\tau^2 R\right)^{-1}\right]}_{\Sigma_\gamma^{-1}} \gamma\right.$$

$$\left. + +\gamma' Z'\left(\sigma^2 I\right)^{-1} y - \gamma' Z'\left(\sigma^2 I\right)^{-1} X\beta\right\}$$

$$= \exp\left\{-\frac{1}{2}\gamma'\Sigma_\gamma^{-1}\gamma + \gamma'\Sigma_\gamma^{-1}\Sigma_\gamma\left[Z'\left(\sigma^2 I\right)^{-1}(y - X\beta)\right]\right\}$$

defining a multivariate normal with:

- $\mu_\gamma = \Sigma_\gamma Z' \left(\sigma^2 I\right)^{-1} (y - X\beta)$
- $\Sigma_\gamma = \left[Z' \left(\sigma^2 I\right)^{-1} Z + \left(\tau^2 R\right)^{-1}\right]^{-1}$

The full conditionals of the variance parameters are inverse gamma distributions, which can be shown by:

$$
\begin{aligned}
&P(\sigma^2|\beta,\gamma,\tau^2,y) \propto det(\sigma^2 I)^{-\frac{1}{2}} \\
&\quad \exp\left\{-\frac{1}{2}(y - X\beta - Z\gamma)' \left(\sigma^2 I\right)^{-1} (y - X\beta - Z\gamma)\right\} \\
&\quad det(\sigma^2 M)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\beta - m)'(\sigma^2 M)^{-1}(\beta - m)\right\} \\
&\quad \left(\sigma^2\right)^{-(a_\sigma+1)} \exp\left\{-\frac{b_\sigma}{\sigma^2}\right\} \\
&= \left(\sigma^2\right)^{-(\underbrace{a_\sigma + n}_{a'_\sigma}+1)} \exp\{-\frac{1}{\sigma^2} \\
&\quad \underbrace{\left[b_\sigma + \frac{1}{2}(y - X\beta - Z\gamma)'(y - X\beta - Z\gamma) + \frac{1}{2}(\beta - m)' M^{-1}(\beta - m)\right]}_{b'_\sigma}\}
\end{aligned}
$$

and

$$P(\tau^2|\beta,\gamma,\sigma^2,y) \propto det(\tau^2 R)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\gamma' \left(\tau^2 R\right)^{-1} \gamma\right\}$$
$$\left(\tau^2\right)^{-(a_\tau+1)} \exp\left\{-\frac{b_\tau}{\tau^2}\right\}$$
$$= \left(\tau^2\right)^{-(\underbrace{a_\tau + \frac{n}{2}}_{a'_\tau}+1)} \exp\left\{-\frac{1}{\tau^2}\left[\underbrace{b_\tau + \frac{1}{2}\gamma' R^{-1}\gamma}_{b'_\tau}\right]\right\}$$

The parameters of the inverse gamma distributions are noted in the derivations.

# B Price Data

**Tabelle 10:** Data Used in the Analysis

| COICOP | Name of product | Observations | Observed Regions |
|---|---|---|---|
| 111110100 | Reis | 46 | 19 |
| 111211100 | Weissbrot | 45 | 19 |
| 111211200 | Toastbrot | 50 | 19 |
| 111212100 | Roggenbrot oder Mischbrot | 48 | 19 |
| 111213200 | Koernerbrot oder Vollkornbrot | 50 | 19 |
| 111215100 | Knaeckebrot | 50 | 19 |
| 111217100 | Frische Broetchen | 49 | 18 |
| 111217200 | Broetchen zum Fertigbacken | 43 | 19 |
| 111229100 | Salzstangen oder andere Dauerbackwaren | 46 | 19 |
| 111310100 | Nudeln | 48 | 19 |
| 111321100 | Nudelfertiggericht | 35 | 19 |
| 111410100 | Tortenboden | 53 | 19 |
| 111431100 | Kuchen oder Torte, tiefgefroren | 45 | 19 |
| 111433200 | Hefegebaeck | 52 | 18 |
| 111433300 | Kuchen oder Torte, frisch | 50 | 19 |
| 111611100 | Weizenmehl | 47 | 19 |
| 111615100 | Griess, Roggenmehl oder Aehnliches | 50 | 19 |
| 111621100 | Backmischung fuer Kuchen | 47 | 19 |
| 111625100 | Cornflakes, Muesli oder Aehnliches | 46 | 19 |
| 112110100 | Rindfleisch zum Kochen | 53 | 19 |
| 112120100 | Roulade oder Lende vom Rind | 48 | 19 |
| 112150100 | Rindfleisch zum Schmoren oder Braten | 51 | 19 |
| 112210500 | Kotelett oder Schnitzel vom Schwein | 52 | 19 |
| 112290100 | Kassler oder anderes Schweinefleisch | 50 | 19 |
| 112290300 | Schweinebraten | 54 | 19 |
| 112300100 | Lammfleisch | 43 | 15 |
| 112410100 | Frisches Gefluegelfleisch | 52 | 19 |
| 112450100 | Tiefgefrorenes Gefluegelfleisch | 49 | 19 |
| 112511200 | Salami, Cervelatwurst oder andere Dauerwurst | 53 | 19 |
| 112512200 | Bratwurst | 52 | 19 |
| 112513100 | Leberwurst oder Blutwurst | 50 | 19 |
| 112515100 | Gekochter Schinken oder anderes gegartes Fleisch | 51 | 19 |
| 112515200 | Roher Schinken, Schinkenspeck oder Bauchspeck | 50 | 19 |
| 112520100 | Leber oder andere Innereien | 48 | 19 |
| 112610100 | Feinkostsalat auf Fleischbasis | 36 | 16 |
| 112630100 | Tiefgefrorenes Fleischfertiggericht | 47 | 19 |
| 112650200 | Fleischfertiggericht in Konserven | 48 | 19 |
| 112690100 | Gyros oder andere fertige Fleischpfanne | 47 | 19 |
| 112710100 | Kaninchenfleisch oder anderes Wildfleisch | 39 | 17 |

| COICOP | Name of product | Observations | Observed Regions |
|---|---|---|---|
| 112720100 | Hackfleisch | 52 | 19 |
| 113111300 | Forelle | 84 | 17 |
| 113131100 | Frisches Fischfilet | 39 | 15 |
| 113300100 | Raeucherfisch | 46 | 18 |
| 113411100 | Fischkonserve | 37 | 18 |
| 113431100 | Fischstaebchen oder Aehnliches | 39 | 18 |
| 113431200 | Tiefgefrorene Fischzubreitung | 40 | 18 |
| 113439100 | Eingelegte Matjes, Fischsalat oder Aehnliches | 31 | 15 |
| 114150100 | H-Milch | 69 | 19 |
| 114400100 | Joghurt | 68 | 19 |
| 114510100 | Hartkaese | 52 | 17 |
| 114530100 | Weichkaese | 65 | 19 |
| 114550100 | Quark | 70 | 19 |
| 114630100 | Fertigdessert | 68 | 19 |
| 114700100 | Eier | 65 | 19 |
| 115100100 | Butter | 53 | 19 |
| 115210100 | Margarine | 54 | 19 |
| 115250100 | Pflanzenfett zum Braten und Backen | 51 | 19 |
| 115400100 | Sonnenblumenoel, Rapsoel oder Aehnliches | 51 | 19 |
| 116110100 | Orangen | 57 | 19 |
| 116150100 | Grapefruits | 32 | 13 |
| 116300100 | Aepfel | 56 | 19 |
| 116500100 | Kirschen | 61 | 19 |
| 116500200 | Avocados | 53 | 18 |
| 116650100 | Weintrauben | 53 | 19 |
| 116700200 | Kiwis, Ananas oder Mangos | 51 | 18 |
| 116700300 | Melonen | 36 | 10 |
| 116830100 | Erdnuesse, Studentenfutter oder Aehnliches | 51 | 19 |
| 116890100 | Suesse Mandeln, Kokosraspeln oder Aehnliches | 51 | 19 |
| 116912100 | Sauerkirschen oder andere Steinobstkonserve | 43 | 19 |
| 116917100 | Ananaskonserve, Erdbeerkonserve oder Aehnliches | 45 | 19 |
| 117110100 | Kopfsalat oder Eisbergsalat | 59 | 19 |
| 117190200 | Lauch oder Sellerie | 38 | 13 |
| 117190300 | Feldsalat oder Rucola | 95 | 15 |
| 117230100 | Weisskohl | 32 | 11 |
| 117290200 | Brokkoli | 103 | 17 |
| 117310100 | Tomaten | 58 | 19 |
| 117330100 | Paprika | 57 | 19 |
| 117350100 | Gurken | 58 | 19 |
| 117390200 | Auberginen | 42 | 18 |
| 117390300 | Kuerbisse oder Mais | 70 | 19 |
| 117410100 | Zwiebeln, Knoblauch oder Aehnliches | 58 | 19 |
| 117420100 | Champignons oder andere Pilze | 56 | 18 |
| 117610100 | Tiefgefrorener Spinat | 48 | 19 |
| 117711100 | Gurkenkonserve | 46 | 18 |
| 117713100 | Sauerkrautkonserve | 38 | 19 |

| COICOP | Name of product | Obser-vations | Observed Regions |
|---|---|---|---|
| 117714100 | Pilzkonserve | 36 | 18 |
| 117715100 | Erbsenkonserve | 46 | 19 |
| 117719400 | Stangenspargel oder andere Gemuesekonserve | 45 | 18 |
| 117800300 | Neue Ernte (Speisefruehkartoffeln) | 64 | 16 |
| 117911100 | Pommes frites | 51 | 19 |
| 117912100 | Kartoffelklossmehl oder Kartoffelpueree | 32 | 13 |
| 117913100 | Kartoffelchips oder Kartoffelsticks | 52 | 19 |
| 118210100 | Marmelade, Konfituere oder Gelee | 47 | 19 |
| 118230100 | Bienenhonig | 45 | 19 |
| 118310100 | Schokoladentafel | 55 | 19 |
| 118390100 | Riegel oder andere Erzeugnisse aus Schokolade | 52 | 19 |
| 118410100 | Pralinen | 47 | 17 |
| 118450100 | Bonbons | 63 | 19 |
| 118490100 | Kaugummi, Gummibaerchen oder Aehnliches | 50 | 19 |
| 118500100 | Speiseeis in Packungen | 42 | 18 |
| 118500200 | Eis am Stiel oder anderes portioniertes Speiseeis | 27 | 12 |
| 118610100 | Kakaohaltiger Brotaufstrich | 43 | 19 |
| 119100200 | Tomatenketchup oder Gewuerzketchup | 34 | 19 |
| 119100300 | Sossenpulver, Sossenbinder oder Aehnliches | 24 | 11 |
| 119100400 | Grillsosse, Sojasosse, Salatdressing oder Aehnliches | 27 | 19 |
| 119210100 | Salz | 34 | 19 |
| 119230100 | Gewuerze | 35 | 19 |
| 119313100 | Dosensuppe | 20 | 11 |
| 119340100 | Puddingpulver | 24 | 18 |
| 119353100 | Fertignahrung fuer Saeuglinge oder Kleinkinder | 39 | 18 |
| 119410100 | Essig | 36 | 19 |
| 119420100 | Mayonnaise | 29 | 18 |
| 119490100 | Senf | 35 | 19 |
| 119490200 | Vitamintabletten oder Aehnliches | 41 | 18 |
| 121111100 | Bohnenkaffee | 59 | 19 |
| 121115100 | Instantkaffee oder Aehnliches | 44 | 19 |
| 121210200 | Schwarzer Tee oder gruener Tee, in Aufgussbeuteln | 43 | 19 |
| 121300100 | Kakaopulver oder Aehnliches | 44 | 19 |
| 122100100 | Mineralwasser | 50 | 19 |
| 122210100 | Colagetraenk | 59 | 19 |
| 122290100 | Koffeinfreie Limonade | 53 | 19 |
| 122310100 | Apfelsaft oder aehnlicher Fruchtsaft | 56 | 19 |
| 122320100 | Orangensaft oder aehnlicher Fruchtsaft | 56 | 19 |
| 122340100 | Multivitaminsaft | 56 | 19 |
| 122380100 | Diaetfruchtsaft | 37 | 17 |
| 122400100 | Gemuesesaft | 57 | 19 |
| 211030100 | Korn oder Doppelkorn | 61 | 19 |
| 211050100 | Likoer | 63 | 19 |
| 212110100 | Weisswein | 62 | 19 |
| 212130100 | Rotwein oder Rosewein | 62 | 19 |

| COICOP | Name of product | Observations | Observed Regions |
|---|---|---|---|
| 212210100 | Sekt, Prosecco, Champagner | 62 | 19 |
| 312111100 | Herrenanzug | 21 | 6 |
| 312127100 | Herrenlederjacke | 28 | 6 |
| 312129300 | Herrensakko, F/S | 42 | 6 |
| 312132200 | Herrenjeans | 41 | 6 |
| 312149300 | Herrenmantel, F/S | 31 | 6 |
| 312151100 | Pullover oder Strickjacke fuer Herren | 24 | 6 |
| 312191100 | Herrenbusinesshemd | 26 | 6 |
| 312194100 | Herrenschlafanzug | 31 | 6 |
| 312211200 | Damenkostuem, Kleid oder Hosenanzug, F/S | 57 | 6 |
| 312223200 | Damenrock, F/S | 27 | 6 |
| 312227100 | Damenstoffhose | 21 | 6 |
| 312231200 | Damenbluse, F/S | 44 | 6 |
| 312313200 | Kinderhose, F/S | 29 | 6 |
| 312341200 | Sportbekleidung fuer Kinder | 36 | 6 |
| 312352200 | Kindershirt, F/S | 45 | 6 |
| 312359300 | Struempfe, Socken oder Strumpfhose fuer Kinder | 46 | 6 |
| 312361100 | Strampelanzug oder Zweiteiler fuer Saeuglinge | 29 | 6 |
| 313011200 | Muetze, Kappe oder Hut, F/S | 26 | 6 |
| 313019100 | Fahrradhelm | 36 | 17 |
| 313079100 | Reissverschluss | 26 | 7 |
| 314110100 | Aenderungsschneiderarbeit | 37 | 18 |
| 314210100 | Chemische Reinigung oder Faerben von Bekleidung | 45 | 18 |
| 314250200 | Waschen und Buegeln von Bekleidung | 35 | 17 |
| 321110100 | Klassische Herrenschuhe | 39 | 6 |
| 321110300 | Herrenfreizeitschuhe, F/S | 59 | 6 |
| 321140200 | Herrenhausschuhe | 20 | 6 |
| 321150100 | Herrensportschuhe | 21 | 6 |
| 321210300 | Damenfreizeitschuhe, F/S | 51 | 6 |
| 321250200 | Damensportschuhe | 21 | 6 |
| 321310300 | Kinderschuhe, F/S | 39 | 6 |
| 321310400 | Kleinkinderschuhe, F/S | 31 | 6 |
| 322000200 | Schuhreparatur | 39 | 18 |
| 411021200 | Miete Altbauwohnung bis 70 sqm, freifinanziert | 158 | 30 |
| 411021300 | Miete Altbauwohnung ab 70 sqm, freifinanziert | 94 | 25 |
| 411022200 | Miete Neubauwohnung bis 70 sqm, oeffentlich gefoerdert | 73 | 37 |
| 411022400 | Miete Neubauwohnung ab 70 sqm, freifinanziert | 648 | 50 |
| 411022500 | Miete Neubauwohnung bis 70 sqm, freifinanziert | 834 | 51 |
| 411022600 | Miete Einfamilienhaus ab 100 sqm | 52 | 13 |
| 431010100 | Tapete | 46 | 19 |
| 431030100 | Dispersionsfarbe | 41 | 19 |
| 431030200 | Acrylfarbe | 40 | 19 |
| 431070100 | Spachtelmasse oder Gips | 52 | 19 |
| 441000100 | Frischwasser, monatliche Abnahme 6 hm | 28 | 27 |
| 441000200 | Frischwasser, monatliche Abnahme 15 hm | 28 | 27 |

| COICOP | Name of product | Observations | Observed Regions |
|---|---|---|---|
| 442000100 | Muellabfuhr, kleine Muellmenge | 28 | 27 |
| 442000200 | Muellabfuhr, grosse Muellmenge | 28 | 27 |
| 444037100 | Grundsteuer B | 37 | 19 |
| 451015200 | Strom, monatliche Abnahme 200 kWh | 28 | 27 |
| 451015300 | Strom, monatliche Abnahme 325 kWh | 28 | 27 |
| 451015400 | Strom, monatliche Abnahme 1275 kWh | 28 | 27 |
| 452130100 | Gas, monatliche Abnahme 1000 kWh | 28 | 27 |
| 452130200 | Gas, monatliche Abnahme 1600 kWh | 28 | 27 |
| 452130300 | Gas, monatliche Abnahme 2300 kWh | 28 | 27 |
| 452200100 | Fluessiggas in Flaschen | 27 | 17 |
| 453010100 | Leichtes Heizoel | 77 | 18 |
| 454030200 | Kohlebriketts | 31 | 16 |
| 454070100 | Brennholz, Holzpellets oder Aehnliches | 23 | 12 |
| 511011100 | Stuhl oder Eckbank | 35 | 11 |
| 511015100 | Schrankelement fuer Einbaukueche | 39 | 11 |
| 511017100 | Kuechenzeile oder Einbaukueche | 54 | 11 |
| 511023100 | Lattenrost oder Sprungrahmen | 35 | 7 |
| 511024100 | Matratze mit Federkern | 49 | 7 |
| 511025100 | Matratze aus Latex oder Schaumstoff | 41 | 7 |
| 511032100 | Schlafsofa | 43 | 10 |
| 511042100 | Wohnzimmertisch oder Esszimmertisch | 43 | 10 |
| 511043100 | Wohnzimmerschrank | 69 | 10 |
| 511052100 | Badezimmermoebel | 44 | 10 |
| 512070200 | Laminat, Fertigparkett, Linoleum oder Aehnliches | 45 | 6 |
| 512090100 | Verlegen und Fixieren von Bodenbelaegen | 31 | 15 |
| 513050100 | Abschleifen und Versiegeln von Parkettfussboden | 52 | 17 |
| 520011200 | Wolldecke oder Fleecedecke | 40 | 16 |
| 520012100 | Bettdecke oder Kopfkissen | 47 | 19 |
| 520014100 | Bettbezuggarnitur | 44 | 18 |
| 520015100 | Bettlaken | 39 | 18 |
| 520031100 | Gardine oder Vorhang | 30 | 14 |
| 520033200 | Innenjalousie oder Rollo | 39 | 17 |
| 520040100 | Badezimmerteppich oder Badezimmergarnitur | 35 | 18 |
| 520061100 | Handtuch | 41 | 17 |
| 531210200 | Waschmaschine | 28 | 8 |
| 531320100 | Mikrowellenherd | 30 | 9 |
| 531510100 | Staubsauger | 45 | 10 |
| 532010100 | Toaster, Waffeleisen oder Aehnliches | 35 | 6 |
| 532020300 | Kaffeevollautomat, Padmaschine oder Aehnliches | 33 | 9 |
| 532030100 | Elektrisches Ruehrgeraet oder Mixgeraet | 23 | 6 |
| 540111100 | Trinkglaeser | 43 | 18 |
| 540121100 | Porzellanservice | 31 | 15 |
| 540121200 | Tafelgeschirr aus Porzellan | 41 | 18 |
| 540131100 | Auflaufform | 36 | 18 |
| 540230100 | Essbesteck, Tortenheber oder Aehnliches | 42 | 17 |
| 540373100 | Babyflasche oder Aehnliches | 38 | 17 |

| COICOP | Name of product | Observations | Observed Regions |
|---|---|---|---|
| 551050100 | Schlagbohrer | 31 | 7 |
| 551050200 | Akkuschrauber oder Akkubohrschrauber | 33 | 7 |
| 552010200 | Gartenschere, Spaten oder Astschere | 31 | 19 |
| 552020100 | Hammer, Schraubendreher oder Aehnliches | 32 | 19 |
| 552020500 | Malerpinsel oder Farbroller | 33 | 19 |
| 552031300 | Vorhaengeschloss oder Kofferschloss | 32 | 18 |
| 552032200 | Halogenlampe | 45 | 19 |
| 552033100 | Energiesparlampe | 37 | 18 |
| 552034100 | Steckdose, Stecker, Kabel oder Aehnliches | 44 | 17 |
| 552035100 | Batterien | 49 | 18 |
| 561120100 | Weichspueler, Staerke oder Aehnliches | 43 | 19 |
| 561150100 | Sanitaerreinger | 43 | 18 |
| 561190100 | Metallpflegemittel oder anderes Pflegemittel | 32 | 17 |
| 561190300 | Allzweckreiniger oder anderes Reinigungsmittel | 44 | 18 |
| 561219100 | Filterpapier, Pappbecher oder Aehnliches | 46 | 18 |
| 561231100 | Naegel, Schrauben oder Aehnliches | 25 | 15 |
| 561232100 | Besen oder Buerste | 42 | 17 |
| 561239200 | Klebstoff, Zuendhoelzer oder Aehnliches | 35 | 14 |
| 611090100 | Melissengeist, Gesundheitsbaeder oder Aehnliches | 41 | 18 |
| 612030100 | Wundpflaster oder andere Verbandsstoffe | 29 | 16 |
| 612050100 | Fieberthermometer, Waermflasche oder Aehnliches | 39 | 18 |
| 612090200 | Kondome oder Schwangerschaftstest | 39 | 18 |
| 613012200 | Blutdruckmessgeraet oder Aehnliches | 45 | 19 |
| 613031100 | Brillenglas, GKV | 38 | 18 |
| 613032200 | Brillenglas, PKV | 38 | 18 |
| 613032300 | Kontaktlinsen | 29 | 17 |
| 622012100 | Zahnersatz, Krone, PKV | 31 | 18 |
| 622012200 | Zahnersatz, Bruecke, PKV | 31 | 18 |
| 623220100 | Physiotherapie, PKV | 44 | 18 |
| 623320100 | Haeusliche Krankenpflege, PKV | 30 | 17 |
| 713000100 | Fahrrad | 59 | 18 |
| 721011100 | Pkw-Reifen | 74 | 19 |
| 721031100 | Autobatterie | 38 | 17 |
| 721031200 | Zuendkerzen | 38 | 18 |
| 721039200 | Zubehoer oder Ersatzteile fuer Kraftfahrzeuge | 72 | 19 |
| 721060100 | Autowachs, Lackpflegemittel oder Aehnliches | 41 | 17 |
| 721071100 | Reifen oder Schlauch fuer Fahrraeder | 34 | 18 |
| 721079100 | Zubehoer oder Ersatzteile fuer Fahrraeder | 36 | 18 |
| 722013100 | Superbenzin, 95 Oktan | 201 | 19 |
| 722013300 | Superbenzin, 98 und mehr Oktan | 108 | 19 |
| 722015100 | Diesel, unter 60 Cetan | 110 | 19 |
| 722015300 | Diesel, 60 und mehr Cetan | 44 | 18 |
| 722017100 | Autogas | 43 | 18 |
| 722051100 | Motorenoel | 51 | 17 |
| 723013100 | Lackieren eines Kotfluegels | 68 | 19 |

| COICOP | Name of product | Observations | Observed Regions |
|---|---|---|---|
| 723015100 | Pkw-Inspektion | 70 | 18 |
| 723017100 | Wechseln der Bremsfluessigkeit bei Pkw | 71 | 18 |
| 723017200 | Wechseln der Bremskloetze bei Pkw | 70 | 18 |
| 723017300 | Wechseln der Auspuffanlage bei Pkw | 71 | 18 |
| 723017400 | Wechseln von Stossdaempfern bei Pkw | 71 | 18 |
| 723017500 | Wechseln der Kupplung bei Pkw | 69 | 18 |
| 723017600 | Wechseln der Reifen bei Pkw | 36 | 18 |
| 723018100 | Pkw-Waesche | 36 | 18 |
| 723039100 | Fahrradreparatur | 33 | 18 |
| 724010100 | Fahrschulunterricht fuer Pkw, Praxis | 53 | 18 |
| 724010200 | Fahrschulunterricht fuer Pkw, Theorie | 37 | 18 |
| 724060100 | Miete von Garage, Stellplatz vom Wohnungsvermieter | 138 | 26 |
| 724060200 | Miete von Garage, Stellplatz | 46 | 19 |
| 724080100 | Parkuhrgebuehr oder Parkscheingebuehr | 30 | 19 |
| 724080200 | Parkhausgebuehr | 37 | 18 |
| 735011100 | Verbund-Einfache Fahrt/Erwachsener, Tarif 1 | 20 | 18 |
| 735013100 | Zeitkarte Verbundverkehr, Ausbildungstarif | 28 | 17 |
| 735015100 | Verbund-Monatskarte/Erwachsener, Tarif 1 | 22 | 18 |
| 736055100 | Moebeltransport | 36 | 18 |
| 820010300 | Festnetztelefon | 41 | 18 |
| 911111100 | Tragbarer Radiorecorder | 29 | 8 |
| 911121100 | Hi-Fi-Anlage | 31 | 9 |
| 911123100 | Uhrenradio | 24 | 6 |
| 911131100 | Autoradio oder stationaeres Navigationsgeraet | 24 | 7 |
| 911142200 | MP3-Player oder Aehnliches | 28 | 7 |
| 911192100 | Lautsprecherboxen fuer Rundfunk oder Hi-Fi | 26 | 7 |
| 911193100 | Kopfhoerer | 25 | 7 |
| 911210200 | Fernsehgeraet | 32 | 9 |
| 911220100 | DVD-Recorder | 30 | 8 |
| 911230100 | DVD-Player oder Blu-ray-Player | 28 | 8 |
| 911240100 | SAT-Anlage | 33 | 9 |
| 912113100 | Digitale Kamera | 47 | 9 |
| 912120100 | Digitaler Camcorder | 25 | 8 |
| 913015500 | Organizer oder mobiles Navigationsgeraet | 29 | 8 |
| 913016100 | Taschenrechner oder Tischrechner | 32 | 10 |
| 914011100 | DVD-Rohlinge, unbespielte Mini-Disc oder Aehnliches | 32 | 16 |
| 914013100 | USB-Stick oder Speicherkarte | 37 | 17 |
| 914013200 | CD-Rohlinge | 41 | 18 |
| 914021100 | Unterhaltungsmusik auf CD | 39 | 15 |
| 914021200 | Klassische Musik auf CD | 33 | 17 |
| 914023100 | Film auf DVD | 53 | 18 |
| 914024100 | PC-Spiel | 49 | 18 |
| 914030200 | Fotoalbum | 39 | 19 |
| 915010100 | Reparatur an Unterhaltungselektronik | 38 | 18 |
| 931011200 | Gesellschaftsspiel | 47 | 18 |

| COICOP | Name of product | Observations | Observed Regions |
|---|---|---|---|
| 931012100 | Spielekonsole | 41 | 17 |
| 931012200 | Spiel fuer Spielekonsole | 42 | 17 |
| 931013100 | Elektrische Modelleisenbahn oder Zubehoer | 24 | 10 |
| 931014100 | Kunststoffbaukasten | 35 | 17 |
| 931014200 | Experimentierkasten oder Modellbausatz | 37 | 16 |
| 931016100 | Dreirad, Roller oder anderes Kindersportfahrzeug | 38 | 18 |
| 931017100 | Puppe | 43 | 18 |
| 931018100 | Teddybaer oder anderes Plueschtier | 45 | 18 |
| 932018200 | Inline-Skates, Schlittschuhe oder Rollschuhe | 27 | 17 |
| 932018300 | Fussballschuhe oder andere Spezialsportschuhe | 40 | 17 |
| 932019100 | Planschbecken, Taucherbrille oder Aehnliches | 35 | 18 |
| 932024100 | Zelt | 34 | 17 |
| 932029300 | Luftmatratze oder anderer Campingartikel | 32 | 16 |
| 933012100 | Topfpflanze | 50 | 19 |
| 933040100 | Blumentopf oder Blumenkasten | 37 | 17 |
| 933051100 | Blumenduenger | 26 | 14 |
| 941021100 | Musikunterricht | 25 | 10 |
| 941030100 | Gebuehr fuer Fitnessstudio | 39 | 10 |
| 942120100 | Theaterbesuch | 48 | 18 |
| 942120200 | Abonnement fuer Theater | 39 | 15 |
| 942150100 | Kinobesuch | 61 | 18 |
| 942210100 | Museumsbesuch oder Zoobesuch | 53 | 18 |
| 942370100 | Leihgebuehr fuer eine DVD | 29 | 16 |
| 942430100 | Filmentwicklung oder Pauschale fuer Digitalisierung | 40 | 18 |
| 942430200 | Abzug eines Bildes | 40 | 18 |
| 954031200 | Briefumschlaege, Briefblock oder Briefpapier | 41 | 18 |
| 954035100 | Schulheft, Malblock oder Zeichenblock | 47 | 17 |
| 954039100 | Druckerpapier | 37 | 18 |
| 954055200 | Fuellhalterpatronen | 39 | 17 |
| 954057100 | Farbkasten | 44 | 16 |
| 1040000100 | Lehrgangsgebuehr VHS | 24 | 18 |
| 1111011220 | Fleischgericht, Hotel | 22 | 13 |
| 1111011230 | Fleischgericht, Restaurant oder Cafe | 24 | 10 |
| 1111012120 | Fischgericht, Hotel | 28 | 17 |
| 1111012130 | Fischgericht, Restaurant oder Cafe | 29 | 15 |
| 1111013120 | Teig- oder Eierspeise, Hotel | 28 | 16 |
| 1111013130 | Teig- oder Eierspeise, Restaurant oder Cafe | 27 | 12 |
| 1111014120 | Suppe oder Eintopf, Hotel | 32 | 18 |
| 1111014130 | Suppe oder Eintopf, Restaurant oder Cafe | 30 | 15 |
| 1111016120 | Eisbecher oder Dessert, Hotel | 28 | 18 |
| 1111017220 | Gericht anderer Art, Hotel | 27 | 17 |
| 1111017230 | Gericht anderer Art, Restaurant oder Cafe | 28 | 14 |
| 1111030310 | Pizza zum Mitnehmen, Restaurant oder Cafe | 24 | 11 |
| 1111030320 | Pizza zum Mitnehmen, Schnellrestaurant oder Aehnliches | 27 | 13 |
| 1111051120 | Kaffee, Tee, Kakao oder Aehnliches, Hotel | 29 | 18 |

| COICOP | Name of product | Observations | Observed Regions |
|---|---|---|---|
| 1111052120 | Fruchtsaft oder Gemuesesaft, Hotel | 29 | 18 |
| 1111052130 | Fruchtsaft oder Gemuesesaft, Restaurant oder Cafe | 28 | 14 |
| 1111053120 | Mineralwasser, Hotel | 29 | 18 |
| 1111053130 | Mineralwasser, Restaurant oder Cafe | 28 | 14 |
| 1111054120 | Erfrischungsgetraenk, Hotel | 29 | 18 |
| 1111055120 | Spirituose, Hotel | 32 | 18 |
| 1111055130 | Spirituose, Restaurant oder Cafe | 32 | 15 |
| 1111056120 | Bier, Hotel | 31 | 18 |
| 1111056130 | Bier, Restaurant oder Cafe | 31 | 15 |
| 1111057130 | Wein, Restaurant oder Cafe | 29 | 15 |
| 1112010100 | Verzehr von Speisen in Kantinen | 30 | 9 |
| 1112010200 | Mensaessen, gaengigste Kategorie | 35 | 17 |
| 1211011100 | Friseur fuer Herren | 63 | 18 |
| 1211011200 | Friseur fuer Kinder | 60 | 18 |
| 1211015100 | Friseur fuer Damen; Waschen, Schneiden, Foehnen | 63 | 18 |
| 1211015200 | Friseur fuer Damen; Dauerwelle | 62 | 18 |
| 1211015300 | Friseur fuer Damen; Faerben oder Toenen | 63 | 18 |
| 1211031100 | Gebuehr fuer Sonnenstudio | 23 | 14 |
| 1211032100 | Kosmetikbehandlung oder Aehnliches | 55 | 19 |
| 1212010100 | Haartrockner oder anderes Haarpflegegeraet | 32 | 6 |
| 1212050100 | Elektrischer Rasierapparat | 32 | 6 |
| 1212070100 | Elektrische Zahnbuerste | 28 | 7 |
| 1213012100 | Haarbuerste, Kamm oder Haarspange | 45 | 16 |
| 1213013100 | Personenwaage | 42 | 17 |
| 1213014100 | Zahnbuerste, nicht elektrisch | 57 | 18 |
| 1213017100 | Nassrasierer, Rasierklingen oder Aehnliches | 38 | 18 |
| 1213020100 | Eau de Toilette oder Parfuem | 46 | 18 |
| 1213032200 | Haarfarbe oder Haartoenung | 35 | 16 |
| 1213040100 | Handcreme | 44 | 18 |
| 1213040300 | Kindercreme | 43 | 18 |
| 1213051100 | Zahncreme | 57 | 18 |
| 1213070100 | Lippenstift oder Lippenpflegestift | 41 | 17 |
| 1213070200 | Nagellack | 43 | 18 |
| 1213083200 | Duschgel, Duschbad oder Badezusatz | 54 | 18 |
| 1213091100 | Toilettenpapier | 55 | 18 |
| 1213092100 | Papiertaschentuecher | 46 | 18 |
| 1213093100 | Windeln fuer Saeuglinge oder Kleinkinder | 50 | 18 |
| 1213099100 | Tampons, Kosmetiktuecher oder andere Hygieneartikel | 42 | 18 |
| 1231053100 | Damenarmbanduhr | 36 | 18 |
| 1231053200 | Herrenarmbanduhr oder Taschenuhr | 35 | 17 |
| 1231070100 | Wecker, Stoppuhr oder Aehnliches | 38 | 18 |
| 1232111100 | Damenhandtasche | 30 | 17 |
| 1232152100 | Aktenkoffer, Aktentasche oder Aktenmappe | 26 | 16 |
| 1232153100 | Schulranzen oder Rucksack | 27 | 16 |

| COICOP | Name of product | Observations | Observed Regions |
|---|---|---|---|
| 1232154300 | Koffer, Reisetasche oder Aehnliches | 29 | 16 |
| 1232221100 | Kinderwagen | 32 | 17 |
| 1232223100 | Autokindersitz | 39 | 15 |
| 1232261100 | Sonnenbrille | 57 | 18 |
| 1240030300 | Essen auf Raedern | 33 | 19 |

**Source**: Own computation using data of the Bayerisches Landesamt für Statistik und Datenverarbeitung

# C Outputs and Results

## C.1 Outputs of the Linear Regression Estimation by COICOP-Classification

**Tabelle 11:** Results for Food and Non-alcoholic Beverages

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.1526 | 0.3073 | 0.50 | 0.6194 |
| age_skew | -0.6555 | 0.2762 | -2.37 | 0.0177 |
| area | 0.0004 | 0.0001 | 2.97 | 0.0030 |
| wage | 0.0002 | 0.0001 | 2.96 | 0.0031 |
| popdens | 0.0001 | 0.0000 | 5.45 | 0.0000 |
| mig | -0.0176 | 0.0050 | -3.50 | 0.0005 |
| commut | 0.0059 | 0.0021 | 2.83 | 0.0046 |
| I(commut^2) | -0.0001 | 0.0000 | -4.24 | 0.0000 |
| GKat2 | -0.8162 | 0.1847 | -4.42 | 0.0000 |
| GKat3 | -0.5337 | 0.1849 | -2.89 | 0.0039 |
| GKat4 | -1.4857 | 0.1842 | -8.07 | 0.0000 |
| GKat5 | -0.1193 | 0.1866 | -0.64 | 0.5226 |
| GKat6 | 0.5306 | 0.1997 | 2.66 | 0.0079 |
| GKat7 | -0.4406 | 0.2701 | -1.63 | 0.1029 |
| typ2 | 0.2243 | 0.0598 | 3.75 | 0.0002 |
| typ3 | 0.2342 | 0.0634 | 3.69 | 0.0002 |

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

**Tabelle 12:** Results for Alcoholic Beverages, Tobacco, and Narcotics

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -0.2640 | 0.8511 | -0.31 | 0.7567 |
| emp | -0.0000 | 0.0000 | -2.06 | 0.0401 |
| wage | -0.0007 | 0.0003 | -2.71 | 0.0070 |
| building | -0.0031 | 0.0016 | -1.96 | 0.0509 |
| I(building^2) | 0.0000 | 0.0000 | 2.01 | 0.0454 |
| I(log(popdens)) | 0.5454 | 0.1954 | 2.79 | 0.0056 |
| hoteln | 0.0348 | 0.0162 | 2.15 | 0.0324 |
| commut | -0.0117 | 0.0052 | -2.25 | 0.0253 |
| GKat3 | -0.0384 | 0.1288 | -0.30 | 0.7655 |
| GKat4 | -0.9185 | 0.1020 | -9.00 | 0.0000 |
| GKat5 | 1.6051 | 0.2599 | 6.17 | 0.0000 |
| GKat6 | 1.5074 | 0.2284 | 6.60 | 0.0000 |

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

**Tabelle 13:** Results for Clothing and Footware

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 162.2165 | 45.2186 | 3.59 | 0.0003 |
| hoteln | 0.0557 | 0.0154 | 3.63 | 0.0003 |
| age_mean | -7.5687 | 2.0740 | -3.65 | 0.0003 |
| I(age_mean^2) | 0.0883 | 0.0238 | 3.71 | 0.0002 |
| commut | -0.0162 | 0.0035 | -4.67 | 0.0000 |
| centre | -0.0203 | 0.0080 | -2.53 | 0.0115 |
| GKat2 | -0.8858 | 0.1834 | -4.83 | 0.0000 |
| GKat4 | -0.7777 | 0.0988 | -7.87 | 0.0000 |
| GKat5 | 0.4144 | 0.0771 | 5.37 | 0.0000 |
| GKat7 | 0.0925 | 0.1040 | 0.89 | 0.3743 |

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

**Tabelle 14:** Results for Housing, Water, Electricity, Gas and other Fuels

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 19.0204 | 4.5633 | 4.17 | 0.0000 |
| area | 0.0031 | 0.0007 | 4.20 | 0.0000 |
| I(area^2) | -0.0000 | 0.0000 | -4.51 | 0.0000 |
| emp | 0.0000 | 0.0000 | 1.66 | 0.0977 |
| wage | -0.0173 | 0.0029 | -6.07 | 0.0000 |
| I(wage^2) | 0.0000 | 0.0000 | 5.49 | 0.0000 |
| building | -0.0011 | 0.0005 | -2.09 | 0.0369 |
| popdens | -0.0003 | 0.0001 | -2.34 | 0.0192 |
| hoteln | -0.0102 | 0.0046 | -2.21 | 0.0274 |
| sec_sector | 0.0734 | 0.0228 | 3.23 | 0.0013 |
| tert_sector | 0.0568 | 0.0216 | 2.63 | 0.0087 |
| age_var | 0.0072 | 0.0017 | 4.13 | 0.0000 |
| age_skew | 3.0075 | 0.6134 | 4.90 | 0.0000 |
| commut | -0.0035 | 0.0017 | -2.11 | 0.0348 |
| AggregatLandkreis | -1.3732 | 0.2721 | -5.05 | 0.0000 |
| typ2 | -7.9040 | 1.5299 | -5.17 | 0.0000 |
| typ3 | -7.0966 | 1.6156 | -4.39 | 0.0000 |
| building:typ2 | 0.0022 | 0.0007 | 3.17 | 0.0015 |
| building:typ3 | 0.0038 | 0.0007 | 5.61 | 0.0000 |
| wage:typ2 | 0.0024 | 0.0005 | 4.50 | 0.0000 |
| wage:typ3 | 0.0020 | 0.0006 | 3.59 | 0.0003 |

**Source:** Own computation using Bayerische Vermessungs-verwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

**Tabelle 15:** Results for Furnishings, Household Equipment and Routine Household Maintenance

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 10.8850 | 2.8301 | 3.85 | 0.0001 |
| area | 0.0005 | 0.0001 | 3.86 | 0.0001 |
| emp | -0.0000 | 0.0000 | -3.17 | 0.0016 |
| wage | -0.0076 | 0.0019 | -4.05 | 0.0001 |
| I(wage^2) | 0.0000 | 0.0000 | 4.06 | 0.0001 |
| hi_quali | 0.0366 | 0.0137 | 2.66 | 0.0078 |
| GKat2 | -0.0848 | 0.1141 | -0.74 | 0.4573 |
| GKat3 | -0.1154 | 0.1854 | -0.62 | 0.5335 |
| GKat4 | -0.2140 | 0.0884 | -2.42 | 0.0155 |
| GKat5 | 0.3694 | 0.0898 | 4.11 | 0.0000 |
| GKat7 | 0.0336 | 0.1597 | 0.21 | 0.8333 |

**Source:** Own computation using Bayerische Vermessungs-verwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

**Tabelle 16:** Results for Products and Services of Health

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 3.3046 | 0.9175 | 3.60 | 0.0004 |
| housing_index | -184.0443 | 53.1206 | -3.46 | 0.0006 |
| I(housing_index^2) | 2232.6906 | 685.8243 | 3.26 | 0.0012 |
| typ2 | 0.0899 | 0.1696 | 0.53 | 0.5964 |
| typ3 | 0.4330 | 0.1973 | 2.19 | 0.0287 |

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

**Tabelle 17:** Results for Transport

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -218.3077 | 74.3370 | -2.94 | 0.0034 |
| area | -0.0010 | 0.0002 | -4.90 | 0.0000 |
| emp | 0.0000 | 0.0000 | 5.75 | 0.0000 |
| wage | -0.0005 | 0.0001 | -3.51 | 0.0005 |
| housing_index | -19.1716 | 6.0011 | -3.19 | 0.0014 |
| hoteln | 0.0337 | 0.0092 | 3.67 | 0.0002 |
| hi_quali | 0.0958 | 0.0650 | 1.47 | 0.1404 |
| I(hi_quali^2) | -0.0104 | 0.0035 | -3.00 | 0.0027 |
| age_mean | 10.2015 | 3.4199 | 2.98 | 0.0029 |
| I(age_mean^2) | -0.1189 | 0.0393 | -3.03 | 0.0025 |
| age_skew | -3.1917 | 0.8741 | -3.65 | 0.0003 |
| commut | -0.0066 | 0.0037 | -1.76 | 0.0780 |
| I(commut^2) | 0.0001 | 0.0001 | 2.63 | 0.0086 |
| GKat4 | 0.1662 | 0.2082 | 0.80 | 0.4249 |
| GKat5 | 0.8007 | 0.1979 | 4.05 | 0.0001 |
| GKat6 | 0.8261 | 0.1944 | 4.25 | 0.0000 |
| GKat7 | 0.9445 | 0.1932 | 4.89 | 0.0000 |
| typ2 | 1.2020 | 0.1591 | 7.56 | 0.0000 |
| typ3 | 0.9804 | 0.1424 | 6.88 | 0.0000 |

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

**Tabelle 18:** Results for Communication

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 677.4657 | 147.9970 | 4.58 | 0.0001 |
| area | -0.0069 | 0.0030 | -2.31 | 0.0286 |
| emp | 0.0000 | 0.0000 | 1.80 | 0.0828 |
| wage | -0.2047 | 0.0484 | -4.23 | 0.0002 |
| kk_ew | -0.0456 | 0.0095 | -4.78 | 0.0001 |
| I(kk_ew^2) | 0.0000 | 0.0000 | 4.07 | 0.0004 |
| popdens | 0.0023 | 0.0006 | 3.72 | 0.0009 |
| mig | 0.1746 | 0.0507 | 3.44 | 0.0019 |
| sec_sector | -0.1232 | 0.0321 | -3.84 | 0.0007 |
| age_mean | 1.7353 | 0.4540 | 3.82 | 0.0007 |
| centre | 0.3145 | 0.0945 | 3.33 | 0.0025 |
| typ2 | 9.3154 | 2.5525 | 3.65 | 0.0011 |
| typ3 | 8.1966 | 2.4665 | 3.32 | 0.0026 |
| wage:kk_ew | 0.0000 | 0.0000 | 4.27 | 0.0002 |

**Source:** Own computation using Bayerische Vermessungs-verwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

**Tabelle 19:** Results for Recreation and Culture

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -0.3583 | 0.5353 | -0.67 | 0.5033 |
| emp | -0.0000 | 0.0000 | -3.19 | 0.0014 |
| wage | 0.0003 | 0.0001 | 2.36 | 0.0184 |
| popdens | -0.0004 | 0.0002 | -1.92 | 0.0551 |
| I(popdens^2) | 0.0000 | 0.0000 | 2.93 | 0.0034 |
| unemp | -0.3407 | 0.1335 | -2.55 | 0.0108 |
| I(unemp^2) | 0.0347 | 0.0122 | 2.84 | 0.0046 |
| hi_quali | 0.0526 | 0.0242 | 2.18 | 0.0297 |
| GKat2 | -0.3786 | 0.1204 | -3.14 | 0.0017 |
| GKat3 | -0.2892 | 0.2511 | -1.15 | 0.2496 |
| GKat4 | -0.0334 | 0.0928 | -0.36 | 0.7186 |
| GKat5 | 0.2163 | 0.0942 | 2.30 | 0.0218 |
| GKat7 | 0.0593 | 0.1014 | 0.59 | 0.5584 |

**Source:** Own computation using Bayerische Vermessungs-verwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

**Tabelle 20:** Results for Education

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 14.8944 | 6.1065 | 2.44 | 0.0349 |
| area | -0.0022 | 0.0052 | -0.43 | 0.6774 |
| I(area^2) | 0.0000 | 0.0000 | 0.96 | 0.3614 |
| I(area^3) | -0.0000 | 0.0000 | -1.85 | 0.0939 |
| kk_ew | -0.0007 | 0.0002 | -3.75 | 0.0038 |
| building | 0.0462 | 0.0063 | 7.29 | 0.0000 |
| I(building^2) | -0.0000 | 0.0000 | -6.26 | 0.0001 |
| popdens | -0.0052 | 0.0007 | -7.20 | 0.0000 |
| hoteln | 1.2157 | 0.2014 | 6.03 | 0.0001 |
| I(hoteln^2) | -0.0636 | 0.0106 | -6.03 | 0.0001 |
| unemp | -4.3863 | 1.4869 | -2.95 | 0.0145 |
| I(unemp^2) | 0.4312 | 0.1246 | 3.46 | 0.0061 |
| age_skew | 23.2343 | 12.0826 | 1.92 | 0.0834 |
| I(age_skew^2) | -56.9101 | 25.7641 | -2.21 | 0.0517 |

**Source:** Own computation using Bayerische Vermessungs-verwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

**Tabelle 21:** Results for Restaurants and Hotels

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -5778.2672 | 1891.0720 | -3.06 | 0.0023 |
| area | -0.0775 | 0.0325 | -2.39 | 0.0173 |
| I(area^2) | 0.0002 | 0.0001 | 2.55 | 0.0110 |
| I(area^3) | -0.0000 | 0.0000 | -2.61 | 0.0093 |
| wage | 1.0505 | 0.3679 | 2.86 | 0.0044 |
| I(wage^2) | -0.0004 | 0.0001 | -2.86 | 0.0044 |
| I(wage^3) | 0.0000 | 0.0000 | 2.86 | 0.0043 |
| kk_ew | 0.0098 | 0.0034 | 2.93 | 0.0036 |
| housing_index | 350.5532 | 141.3831 | 2.48 | 0.0134 |
| unemp | -1.7169 | 0.6716 | -2.56 | 0.0108 |
| sec_sector | 4.6847 | 1.5676 | 2.99 | 0.0029 |
| tert_sec | 4.7094 | 1.5889 | 2.96 | 0.0031 |
| age_mean | 190.9659 | 61.7049 | 3.09 | 0.0020 |
| I(age_mean^2) | -2.1588 | 0.6964 | -3.10 | 0.0020 |
| typ2 | -2.7170 | 1.5494 | -1.75 | 0.0799 |
| typ3 | -11.8705 | 4.8433 | -2.45 | 0.0145 |
| wage:kk_ew | -0.0000 | 0.0000 | -2.90 | 0.0038 |

**Source:** Own computation using Bayerische Vermessungs-verwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

**Tabelle 22:** Results for Miscellaneous Goods and Service

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 17.0438 | 4.3774 | 3.89 | 0.0001 |
| area | -0.0006 | 0.0002 | -3.10 | 0.0020 |
| building | 0.0003 | 0.0001 | 2.77 | 0.0056 |
| sec_sector | -0.1751 | 0.0447 | -3.92 | 0.0001 |
| tert_sec | -0.1724 | 0.0437 | -3.95 | 0.0001 |
| GKat2 | 0.2038 | 0.1410 | 1.45 | 0.1486 |
| GKat3 | 0.8835 | 0.2055 | 4.30 | 0.0000 |
| GKat4 | -0.0426 | 0.1145 | -0.37 | 0.7097 |
| GKat5 | 0.4176 | 0.1182 | 3.53 | 0.0004 |
| GKat6 | -0.4718 | 0.9701 | -0.49 | 0.6268 |
| GKat7 | 0.1591 | 0.1171 | 1.36 | 0.1744 |

**Source:** Own computation using Bayerische Vermessungsverwaltung (2015), Bayerisches Landesamt für Statistik und Datenverarbeitung, GfK, Inkar

## C.2 Results of the Price Index Imputation

### C.2.1 Single Imputation

1. Hof, Stadt 0.8201013
2. Regen 0.8359081
3. Rottal-Inn 0.8398376
4. Schweinfurt, Stadt 0.8603824
5. Neuburg-Schrobenhausen 0.8641514
6. Wunsiedel i.Fichtelgebirge 0.8682586
7. Rhön-Grabfeld 0.8697298
8. Landshut, Stadt 0.8700655
9. Regensburg 0.8717883
10. Passau, Stadt 0.8734108
11. Ansbach, Stadt 0.8752414
12. Günzburg 0.8795940
13. Roth 0.8800637
14. Freising 0.8847081
15. Mühldorf a.Inn 0.8886298
16. Nürnberger Land 0.8897167
17. Kaufbeuren, Stadt 0.8912579
18. Traunstein 0.8938806
19. Fürth 0.8982893
20. Deggendorf 0.9024807
21. Erlangen-Höchstadt 0.9030428
22. Eichstätt 0.9054934
23. Bad Kissingen 0.9065943
24. Garmisch-Partenkirchen 0.9113360
25. Bayreuth, Stadt 0.9159729
26. Regensburg, Stadt 0.9185315
27. Bamberg, Stadt 0.9188365
28. Würzburg 0.9194801
29. Lindau (Bodensee) 0.9196970
30. Dillingen a.d.Donau 0.9229229
31. Weissenburg-Gunzenhausen 0.9258624

32. Rosenheim 0.9322860
33. Ostallgäu 0.9357981
34. Neustadt a.d.Aisch-Bad Windsheim 0.9381133
35. München 0.9435648
36. Pfaffenhofen a.d.Ilm 0.9440977
37. Berchtesgadener Land 0.9448994
38. Freyung-Grafenau 0.9461147
39. Aschaffenburg, Stadt 0.9490933
40. Weiden i.d.OPf., Stadt 0.9499808
41. Miesbach 0.9542831
42. Hassberge 0.9559813
43. Schweinfurt 0.9564005
44. Landsberg am Lech 0.9568721
45. Passau 0.9577202
46. Hof 0.9594739
47. Amberg, Stadt 0.9642055
48. Weilheim-Schongau 0.9644596
49. Fürstenfeldbruck 0.9662892
50. Kulmbach 0.9686585
51. Neustadt a.d.Waldnaab 0.9699859
52. Dingolfing-Landau 0.9742159
53. Ebersberg 0.9744197
54. Unterallgäu 0.9759437
55. Neu-Ulm 0.9791494
56. Altötting 0.9826024
57. Augsburg 0.9910478
58. Cham 0.9950909
59. Augsburg, Stadt 0.9970627
60. Nürnberg, Stadt 1.0006459
61. Lichtenfels 1.0013193
62. Dachau 1.0030026
63. Würzburg, Stadt 1.0032934
64. Miltenberg 1.0135324
65. Starnberg 1.0148694
66. Forchheim 1.0150967

67. Ingolstadt, Stadt 1.0165845
68. Kitzingen 1.0210430
69. Straubing-Bogen 1.0249661
70. Fürth, Stadt 1.0289576
71. Schwandorf 1.0339173
72. Kelheim 1.0411541
73. Neumarkt i.d.OPf. 1.0411696
74. Rosenheim, Stadt 1.0456579
75. Memmingen, Stadt 1.0457657
76. Erlangen, Stadt 1.0500228
77. Coburg, Stadt 1.0538633
78. Bayreuth 1.0568918
79. Straubing, Stadt 1.0602405
80. Oberallgäu 1.0674316
81. Kronach 1.0701077
82. Aschaffenburg 1.0727959
83. Bad Tölz-Wolfratshausen 1.0784502
84. Erding 1.0815833
85. Tirschenreuth 1.0856733
86. Schwabach, Stadt 1.0900519
87. Aichach-Friedberg 1.0931598
88. Ansbach 1.1110141
89. Landshut 1.1186288
90. Amberg-Sulzbach 1.1337360
91. Bamberg 1.1372344
92. Main-Spessart 1.1376343
93. Donau-Ries 1.1488031
94. Kempten (Allgäu), Stadt 1.1513531
95. München, Stadt 1.2222956
96. Coburg 1.2497697

### C.2.2 Universal Kriging

1. Hof, Stadt 0.7400224
2. Kulmbach 0.7470393
3. Lichtenfels 0.7810324
4. Altötting 0.7810395
5. Wunsiedel i.Fichtelgebirge 0.8068562
6. Weiden i.d.OPf., Stadt 0.8163738
7. Straubing, Stadt 0.8197419
8. Kitzingen 0.8213983
9. Landshut, Stadt 0.8306510
10. Hassberge 0.8339324
11. Erlangen-Höchstadt 0.8353357
12. Schweinfurt, Stadt 0.8357465
13. Nürnberger Land 0.8399433
14. Kronach 0.8652720
15. Passau, Stadt 0.8686054
16. Neuburg-Schrobenhausen 0.8735058
17. Ansbach, Stadt 0.8750864
18. Bad Kissingen 0.8759788
19. Coburg 0.8823855
20. Bayreuth, Stadt 0.8859142
21. Coburg, Stadt 0.8861515
22. Deggendorf 0.8870476
23. Bamberg, Stadt 0.8876583
24. Freyung-Grafenau 0.8908867
25. Dillingen a.d.Donau 0.8979981
26. Weissenburg-Gunzenhausen 0.8995407
27. Hof 0.9131504
28. Aschaffenburg, Stadt 0.9145170
29. Mühldorf a.Inn 0.9167630
30. Regensburg, Stadt 0.9177454
31. Lindau (Bodensee) 0.9268650
32. Memmingen, Stadt 0.9287908
33. Aschaffenburg 0.9302854

34. Kaufbeuren, Stadt 0.9303793
35. Tirschenreuth 0.9341325
36. Neu-Ulm 0.9370757
37. Rhön-Grabfeld 0.9394187
38. Roth 0.9444295
39. Rottal-Inn 0.9496702
40. Schwabach, Stadt 0.9503267
41. Berchtesgadener Land 0.9559814
42. Würzburg, Stadt 0.9559832
43. Amberg, Stadt 0.9606257
44. Bad Tölz-Wolfratshausen 0.9607877
45. Miltenberg 0.9621867
46. Dingolfing-Landau 0.9633806
47. Günzburg 0.9638614
48. Nürnberg, Stadt 0.9652820
49. Kempten (Allgäu), Stadt 0.9671195
50. Forchheim 0.9688031
51. Augsburg, Stadt 0.9753414
52. Kelheim 0.9766654
53. Miesbach 0.9781232
54. Fürth, Stadt 0.9788373
55. Regen 0.9899673
56. Rosenheim, Stadt 1.0002519
57. Weilheim-Schongau 1.0035412
58. Neumarkt i.d.OPf. 1.0036293
59. München 1.0046428
60. Landsberg am Lech 1.0080139
61. Main-Spessart 1.0109298
62. Ingolstadt, Stadt 1.0128720
63. Erlangen, Stadt 1.0186250
64. Neustadt a.d.Waldnaab 1.0227139
65. Ebersberg 1.0227225
66. Pfaffenhofen a.d.Ilm 1.0259036
67. Neustadt a.d.Aisch-Bad Windsheim 1.0267400
68. Garmisch-Partenkirchen 1.0273244

69. Aichach-Friedberg 1.0365034
70. Augsburg 1.0368716
71. Schwandorf 1.0428309
72. Bayreuth 1.0436250
73. Würzburg 1.0471122
74. Passau 1.0693591
75. Rosenheim 1.0698395
76. Amberg-Sulzbach 1.0753136
77. Starnberg 1.0754271
78. Traunstein 1.0800953
79. Fürstenfeldbruck 1.0818399
80. Bamberg 1.0929002
81. Fürth 1.0958611
82. Straubing-Bogen 1.1082838
83. Eichstätt 1.1092413
84. Cham 1.1130193
85. Landshut 1.1197975
86. Freising 1.1268388
87. Schweinfurt 1.1488127
88. Dachau 1.1529093
89. Erding 1.1596339
90. München, Stadt 1.1996054
91. Regensburg 1.2116196
92. Unterallgäu 1.2310631
93. Ostallgäu 1.2982049
94. Donau-Ries 1.3520368
95. Ansbach 1.4156334
96. Oberallgäu 1.4646028

### C.2.3 Multiple Imputation

1. Ansbach 0.8157848
2. Amberg-Sulzbach 0.8313815
3. Hof, Stadt 0.8512344
4. Coburg, Stadt 0.8590893

5. Coburg 0.8698144
6. Weiden i.d.OPf., Stadt 0.8744579
7. Schwandorf 0.8764382
8. Neustadt a.d.Waldnaab 0.8783245
9. Kronach 0.8837102
10. Neustadt a.d.Aisch-Bad Windsheim 0.8842282
11. Landshut, Stadt 0.8884899
12. Rottal-Inn 0.8890747
13. Weissenburg-Gunzenhausen 0.8902455
14. Tirschenreuth 0.8910003
15. Kaufbeuren, Stadt 0.8939619
16. Schweinfurt, Stadt 0.8943557
17. Bad Kissingen 0.8947357
18. Rhön-Grabfeld 0.8952988
19. Neumarkt i.d.OPf. 0.8977737
20. Landshut 0.9029385
21. Straubing, Stadt 0.9046834
22. Fürth, Stadt 0.9060522
23. Kulmbach 0.9079136
24. Regen 0.9093858
25. Traunstein 0.9099704
26. Passau 0.9102050
27. Mühldorf a.Inn 0.9119506
28. Cham 0.9147976
29. Pfaffenhofen a.d.Ilm 0.9176307
30. Passau, Stadt 0.9194583
31. Hassberge 0.9198975
32. Bayreuth 0.9217868
33. Ansbach, Stadt 0.9243879
34. Dillingen a.d.Donau 0.9249514
35. Main-Spessart 0.9267658
36. Roth 0.9279831
37. Lichtenfels 0.9293229
38. Amberg, Stadt 0.9305949
39. Erlangen-Höchstadt 0.9353063

40. Kelheim 0.9354105
41. Schwabach, Stadt 0.9357450
42. Bamberg 0.9364269
43. Garmisch-Partenkirchen 0.9396885
44. Freyung-Grafenau 0.9400581
45. Bayreuth, Stadt 0.9409262
46. Kitzingen 0.9409813
47. Kempten (Allgäu), Stadt 0.9417210
48. Straubing-Bogen 0.9417645
49. Nürnberger Land 0.9418182
50. Berchtesgadener Land 0.9421836
51. Miltenberg 0.9429403
52. Neuburg-Schrobenhausen 0.9443866
53. Unterallgäu 0.9444988
54. Eichstätt 0.9481401
55. Landsberg am Lech 0.9485602
56. Altötting 0.9528472
57. Ostallgäu 0.9567148
58. Rosenheim 0.9585101
59. Hof 0.9638036
60. Deggendorf 0.9657974
61. Regensburg, Stadt 0.9664604
62. Bamberg, Stadt 0.9667237
63. Memmingen, Stadt 0.9673677
64. Miesbach 0.9677986
65. Günzburg 0.9688502
66. Lindau (Bodensee) 0.9708342
67. Aschaffenburg, Stadt 0.9738562
68. Aschaffenburg 0.9752837
69. Dingolfing-Landau 0.9761280
70. Donau-Ries 0.9811705
71. Würzburg 0.9850011
72. Regensburg 0.9920760
73. Augsburg 0.9936966
74. Ebersberg 0.9943708

75. Oberallgäu 1.0008725
76. Forchheim 1.0009756
77. Weilheim-Schongau 1.0012959
78. Schweinfurt 1.0106999
79. Bad Tölz-Wolfratshausen 1.0196792
80. Würzburg, Stadt 1.0212690
81. Neu-Ulm 1.0256271
82. Fürstenfeldbruck 1.0293434
83. Aichach-Friedberg 1.0298816
84. Augsburg, Stadt 1.0396664
85. Nürnberg, Stadt 1.0422722
86. Dachau 1.0467329
87. Fürth 1.0539098
88. Rosenheim, Stadt 1.0686367
89. Erding 1.0806009
90. Ingolstadt, Stadt 1.1102873
91. Starnberg 1.1605933
92. München, Stadt 1.2681776
93. Wunsiedel i.Fichtelgebirge 1.2787221
94. Erlangen, Stadt 1.3701361
95. Freising 1.3811412
96. München 1.4467345

### C.2.4 KriMI by Mixed Modelling

1. Neuburg-Schrobenhausen 0.7889400
2. Weiden i.d.OPf., Stadt 0.8117964
3. Deggendorf 0.8121880
4. Passau, Stadt 0.8140455
5. Ansbach, Stadt 0.8196061
6. Landshut, Stadt 0.8209351
7. Aschaffenburg, Stadt 0.8491591
8. Schweinfurt, Stadt 0.8568685
9. Lindau (Bodensee) 0.8772655
10. Kitzingen 0.8883496

11. Bamberg, Stadt 0.8888034
12. Garmisch-Partenkirchen 0.8936197
13. Neu-Ulm 0.8965289
14. Würzburg, Stadt 0.9012724
15. Bayreuth, Stadt 0.9042244
16. Bad Kissingen 0.9092169
17. Bad Tölz-Wolfratshausen 0.9145489
18. Regensburg, Stadt 0.9186289
19. Coburg 0.9329673
20. Kronach 0.9352008
21. Nürnberg, Stadt 0.9491606
22. Freising 0.9510322
23. Weissenburg-Gunzenhausen 0.9544836
24. Pfaffenhofen a.d.Ilm 0.9553494
25. Unterallgäu 0.9615638
26. Würzburg 0.9625840
27. Rottal-Inn 0.9633555
28. Rhön-Grabfeld 0.9667342
29. Miesbach 0.9685322
30. Mühldorf a.Inn 0.9702106
31. Altötting 0.9717845
32. Augsburg 0.9732860
33. Regen 0.9737754
34. Augsburg, Stadt 0.9751769
35. Ostallgäu 0.9779378
36. Hof, Stadt 0.9808575
37. Cham 0.9848543
38. Dingolfing-Landau 0.9853766
39. Donau-Ries 0.9854393
40. Kelheim 0.9855111
41. Tirschenreuth 0.9872493
42. Hof 0.9872831
43. Kulmbach 0.9889437
44. Kempten (Allgäu), Stadt 0.9949644
45. Kaufbeuren, Stadt 0.9993353

46. Schweinfurt 1.0006330
47. Aschaffenburg 1.0035408
48. Rosenheim, Stadt 1.0045714
49. Bayreuth 1.0046195
50. Berchtesgadener Land 1.0098494
51. Hassberge 1.0129669
52. Oberallgäu 1.0145326
53. Main-Spessart 1.0165864
54. Regensburg 1.0195066
55. Traunstein 1.0212460
56. Freyung-Grafenau 1.0218610
57. Rosenheim 1.0252858
58. Schwandorf 1.0271336
59. Wunsiedel i.Fichtelgebirge 1.0271448
60. Eichstätt 1.0287191
61. Lichtenfels 1.0300595
62. Aichach-Friedberg 1.0317225
63. Forchheim 1.0320385
64. Bamberg 1.0338027
65. Passau 1.0341159
66. Neustadt a.d.Aisch-Bad Windsheim 1.0344998
67. Nürnberger Land 1.0361172
68. Schwabach, Stadt 1.0367119
69. Starnberg 1.0375873
70. Straubing-Bogen 1.0378001
71. Memmingen, Stadt 1.0382724
72. Landshut 1.0384655
73. Dillingen a.d.Donau 1.0394563
74. Roth 1.0394928
75. Erlangen-Höchstadt 1.0396638
76. Straubing, Stadt 1.0403278
77. Neumarkt i.d.OPf. 1.0439256
78. Fürth 1.0442225
79. Coburg, Stadt 1.0458942
80. Fürth, Stadt 1.0524002

81. Miltenberg 1.0570129
82. Günzburg 1.0576249
83. Neustadt a.d.Waldnaab 1.0585496
84. Amberg, Stadt 1.0605862
85. Landsberg am Lech 1.0608176
86. Fürstenfeldbruck 1.0621905
87. Ansbach 1.0628830
88. Weilheim-Schongau 1.0793769
89. Ebersberg 1.0806050
90. Ingolstadt, Stadt 1.0850374
91. Amberg-Sulzbach 1.0854961
92. München 1.0962311
93. Erding 1.1054661
94. Dachau 1.1282148
95. Erlangen, Stadt 1.1355460
96. München, Stadt 1.2205000

### C.2.5 KriMI by P-Splines

1. Regensburg, Stadt 0.6843597
2. Schweinfurt, Stadt 0.7015048
3. Landshut, Stadt 0.7118566
4. Ingolstadt, Stadt 0.7533899
5. Weiden i.d.OPf., Stadt 0.7558110
6. Bamberg, Stadt 0.7638445
7. Passau, Stadt 0.7639893
8. Bad Tölz-Wolfratshausen 0.7792010
9. Schwabach, Stadt 0.7808078
10. Neuburg-Schrobenhausen 0.7889952
11. Erlangen-Höchstadt 0.8141803
12. Rosenheim, Stadt 0.8143169
13. Bayreuth, Stadt 0.8272663
14. Neu-Ulm 0.8493104
15. Ansbach, Stadt 0.8593051
16. Amberg, Stadt 0.8918926

17. Kelheim 0.8960941
18. Augsburg, Stadt 0.9008758
19. Aschaffenburg, Stadt 0.9140104
20. Würzburg, Stadt 0.9179867
21. Fürth, Stadt 0.9226042
22. Erlangen, Stadt 0.9235155
23. Straubing, Stadt 0.9346584
24. Starnberg 0.9405913
25. Deggendorf 0.9450302
26. Dingolfing-Landau 0.9464613
27. Coburg, Stadt 0.9523998
28. Memmingen, Stadt 0.9536040
29. Roth 0.9566290
30. Landsberg am Lech 0.9668490
31. Tirschenreuth 0.9679247
32. Pfaffenhofen a.d.Ilm 0.9705580
33. München, Stadt 0.9762547
34. Kronach 0.9785627
35. Nürnberg, Stadt 0.9835377
36. Aichach-Friedberg 0.9865170
37. Ebersberg 0.9885502
38. Coburg 1.0007806
39. Hassberge 1.0014120
40. Augsburg 1.0068642
41. Rottal-Inn 1.0079472
42. Dillingen a.d.Donau 1.0143884
43. Altötting 1.0180981
44. Eichstätt 1.0492635
45. Neustadt a.d.Aisch-Bad Windsheim 1.0545480
46. Mühldorf a.Inn 1.0571326
47. Freyung-Grafenau 1.0582837
48. Amberg-Sulzbach 1.0591144
49. Freising 1.0709741
50. Hof, Stadt 1.0718184
51. Erding 1.0738424

52. Weissenburg-Gunzenhausen 1.0758676
53. München 1.0781244
54. Neumarkt i.d.OPf. 1.0786010
55. Dachau 1.0837267
56. Miltenberg 1.0911579
57. Lichtenfels 1.0916932
58. Hof 1.0947637
59. Kulmbach 1.1007703
60. Landshut 1.1103514
61. Aschaffenburg 1.1122951
62. Neustadt a.d.Waldnaab 1.1131825
63. Ostallgäu 1.1182465
64. Straubing-Bogen 1.1212420
65. Günzburg 1.1272133
66. Bayreuth 1.1315846
67. Kitzingen 1.1354440
68. Rhön-Grabfeld 1.1466130
69. Nürnberger Land 1.1521608
70. Main-Spessart 1.1526720
71. Würzburg 1.1670117
72. Unterallgäu 1.1705233
73. Kaufbeuren, Stadt 1.1714508
74. Regen 1.1739556
75. Schwandorf 1.1789801
76. Fürstenfeldbruck 1.1808380
77. Rosenheim 1.1956889
78. Wunsiedel i.Fichtelgebirge 1.2022298
79. Bamberg 1.2107640
80. Cham 1.2158785
81. Passau 1.2335421
82. Fürth 1.2533359
83. Bad Kissingen 1.2834810
84. Miesbach 1.2907140
85. Traunstein 1.2953992
86. Regensburg 1.3051862
87. Forchheim 1.3093336
88. Donau-Ries 1.3247229
89. Berchtesgadener Land 1.3350977
90. Schweinfurt 1.5010405
91. Ansbach 1.9000075

University
of Bamberg
Press

Multiple imputation is a method to handle the problem of missing values in a dataset. As it accounts for the uncertainty brought in by the missing data, it is possible to conduct reliable statistical tests after this method has been implemented. Kriging uses neighbourhood effects to predict values of unobserved regions. It can be seen as an imputation technique. The unobserved regions are missing data points, and the kriging predictions are the imputations. Due to the fact of being a single imputation technique, no proper statistical inferences are possible after filling the dataset. If spatially dependent data face the problem of missing data and a proper statistical inference is needed, a modelling of the spatial correlation in the multiple imputation model is needed. Here this is prevailed by implementing kriging in the model used for multiple imputation. We call the resulting method KriMI. The exact problem can be found when looking at regional price levels in Bavaria. The Bavarian State Office for Statistics surveys the prices which are needed to compute the price index only in a few regions. The prices of the unobserved regions are treated as missing data.

ISBN 978-3-86309-524-6

www.uni-bamberg.de/ubp/