CrossMark

ORIGINAL PAPER

Modeling heterogeneity: a praise for varying-coefficient models in causal analysis

Stefan Sperlich[1] · **Raoul Theler**[1]

Abstract This article considers the question of how to cope with heterogeneity when studying causal effects. The standard approach in empirical economics is still to use a linear model and interpret the coefficients as the average returns or effects. Nowadays, instrumental variables (IV) are quite popular to account for (unobserved) heterogeneity when estimating these parameters. First the inadequacy of these standard methods is illustrated. Then it is shown why varying-coefficient models have a strong natural potential to model heterogeneity in many interesting regression problems. Moreover, it is straight forward to develop alternative IV specifications in the varying-coefficient models framework. The corresponding modeling and implementation facilities that are nowadays available in R are studied.

Keywords Varying-coefficient models · Causal inference · Econometrics · Semiparametric modeling

1 Introduction

In economic theory, agents aim to achieve their highest level of welfare depending on the return to their characteristics and choices. This return is heterogeneous in the sense that it varies over agents even after having conditioned on all kind of observable covariates that reflect these choices and characteristics. Heterogeneous returns have always been an issue of modeling beyond the simple adding of additive independent errors to the mean regression (and apart from modeling these as heteroscedastic,

✉ Stefan Sperlich
stefan.sperlich@unige.ch

1   Geneva School for Economics and Management, Université de Genève, Bd du Pont d'Arve 40, 1211 Geneva, Switzerland

⚙ Springer

autocorrelated, spatial dependent, etc.). This is one of the reasons why the varying-coefficient models (Hastie and Tibshirani 1993, or Fan and Zhang 1999), the mixed- or random-effects models (see González Manteiga et al. 2013, for a recent review), and the Bayesian modeling (see for example Kneib and Fahrmeir 2006) have been attracting an increasing attention. This holds especially true for the area of non- and semiparametric statistics from which all these reference are taken. Like the well-known additive nonparametric model (Hastie and Tibshirani 1990; Mammen and Nielsen 1999) they have been proposed in order to bridge the gap between nonparametric models and the standard (generalized) linear model. They have been designed to help the non- and semiparametric methods to get from theory to practice. In fact, especially in empirical economics, non- and semiparametric methods have so far only been used for some prediction issues in finance. Nowadays, they are also proposed for matching in the treatment effect literature. But even there they do rather appear in the theory than in practice. However, there is an increasing awareness of problems that happen to arise with the identification, estimation and interpretation of average returns or effects. If heterogeneity in returns (to observable characteristics or choices) is present, then OLS estimation in linear models is inconsistent, and instrumental variable (IV) estimators lose their identification capacity; they identify only the average effects of non-identified subpopulations (the so-called LATE = local average treatment effect), if at all. Moreover, they need to satisfy quite unrealistic and non-testable conditions. Consequently, impact evaluation becomes much harder—no matter whether ex-post (i.e. inference) or ex-ante (i.e. prediction).

Micro foundations of economic theory are often concerned about mathematical modeling of agents under strong rational choice and utility optimization. When facing the problem of generalizing the individual results to the whole population, the existence of a mean or representative agent behavior is used to be assumed. While already this assumption has been a controversial issue in economic theory for half a century (Allais 1953), the related econometric problems have usually been underestimated. More precisely, as deviations from the mean behavior are typically endogenous, the simple traditional inference leads to biased and inconsistent estimates. A classical and well studied example is the resulting downward bias of the consumer price index, see for example Diewert (1998) or Crossley and Pendakur (2010).

Let us consider the linear regression model as the benchmark in econometric modeling,

$$Y_i = X_i^T \beta + \varepsilon_i \quad i = 1, \ldots, n,$$

where β is supposed to be the vector of coefficients quantifying the returns (or say, effect) to the covariates summarized in vector X_i. As a consequence, ε_i is an error term with mean 0, being mean-independent of X such that there is 'a priori' no endogeneity problem.[1] Heterogeneity comes from X and the error term, and the latter must capture all individual deviation from the mean return. Not only from a micro-econometric point of view this is a quite restricted heterogeneity which does not match with the modern

---

[1] A covariate X is said to be endogenous if the error terms is not mean independent of it. This typically leads to identification problems, biased estimators and invalids causal inference. Else it is called exogenous.

individual oriented modeling. For example, by considering the impact of the years of schooling on wage, it seems obvious that the return to schooling varies between individuals due their abilities. If these are not included in X, one might speak of an omitted variable problem. But if there is an interaction with education, it is also a modeling or functional form related problem. Consequently, a more realistic though too general model (if applied without further modeling) is

$$Y_i = X_i^T \beta_i + \varepsilon_i \quad i = 1, \ldots, n, \tag{1}$$

where the heterogeneity is also reflected in β_i. If one is only interested in the average return to (or effect of X), i.e. $E[\beta_i] = b$, the following model is actually the one of interest:

$$Y_i = X_i^T b + e_i, \quad e_i = X_i^T[\beta_i - b] + \varepsilon_i, \quad i = 1, \ldots, n. \tag{2}$$

This kind of heterogeneity does not invalidate the OLS estimation as long as the deviation from the mean is exogenous, i.e. $E[\beta_i - b | X_i] = 0$. However, this condition does typically not hold in economics as this would to some extent contradict the spirit of rational behavior of individuals. When agents choose their actions, they consider the personal abilities that impact the return of their choices. Students choose the number of years of schooling by taking into account their personal abilities which will affect their future wage in many ways. Therefore, the assumption of exogeneity of X_i is violated in this model in the sense that $E[e_i | X_i] \neq 0$ and the OLS estimation becomes inconsistent.

A popular way to deal with endogeneity in econometrics is to use a linear IV (or, equivalently, a two step least squares 2SLS) estimation. This means to be provided with an additional variable W_i such that $Cov(W_i, e_i) = 0$ but $Cov(W_i, X_i) \neq 0$ for $i = 1, \ldots, n$. Then predict X with W giving $\hat{X}_i(W_i)$ in a first step, and use these predictions instead of X_i for calculating an OLS estimator of b in Eq. (2). In practice, to find a set of observable variables W_i that strongly impact X_i but with $Cov(W_i, \beta_i) = 0$ is extremely unlikely. Already the existence of W_i that would fulfill $E[e_i | W_i] = 0$ is hard to justify unless W_i was completely random (like a lottery number). It must then still be related with X_i and have no other relation to β_i than that via X_i. Similar criticism applies if we interpret the heterogeneity in returns as an effect of omitted variables. As an example take Card (1995) who assumes that the proximity to a college has an effect on the years of schooling but no further effect on the wage. At a first glance, it might seem to be a valid instrument. However, the proximity to school is linked with the family's social background which in many ways will impact wage, also via the returns to schooling. Have in mind that most people working at the school live close to it, and at the same time care more about the education of their children than the average of parents would. So this is a typical example for an invalid IV.

To sum up, in presence of those 'endogenous' coefficients β_i, the traditional approaches of OLS and IV estimation fail in their attempt to estimate an average effect. Indeed, it is difficult to obtain an average effect without estimating directly the heterogeneity (nonparametrically). In a different setup, similar conclusions have been drawn by Deaton and Muellbauer (1980) in their studies on consumer behavior and consumer price index estimation.

An alternative way to deal with the dependence between β_i and X_i is to consider the use of a factor that controls for (at least most of) this dependency, say a third variable Z_i. If such a Z_i is available with $E[\beta_i|Z_i]$ being the best prediction of β_i given Z_i, then $\beta_i - E[\beta_i|Z_i] = u_i$ with $E[u_i|Z_i] = 0$ by definition. With $g(\cdot)$ being this best predictor, Eq. (2) becomes

$$Y_i = X_i^T E[\beta_i|Z_i] + X_i^T u_i + \varepsilon_i = X_i^T g(Z_i) + X_i^T u_i + \varepsilon_i, \quad i = 1, \ldots, n. \quad (3)$$

The variables comprised in Z are often called effect modifiers. This provides a natural extension of the above mentioned varying-coefficient model

$$Y_i = X_i^T g(Z_i) + \tilde{\varepsilon}_i, \quad E(\tilde{\varepsilon}_i|X_i, Z_i) = 0 \quad i = 1, \ldots, n. \quad (4)$$

For Z_i related to X_i one hopes to control by this means for the dependency between X_i and β_i. In other words, to estimate heterogeneity, the coefficient can be modeled as a nonparametric function (i.e. an unknown but deterministic smooth function) of an observed variable Z_i up to a random deviation u_i that is mean-independent from X_i. The effect modifier helps us to mitigate or solve the endogeneity that may occur due to heterogeneous returns. Note that variable Z must not be independent from X if it shall filter potential dependency between β_i and X_i. An extreme case is setting $Z = X$ giving basically the well-known nonparametric additive model.

Effective estimation of non- and semiparametric varying-coefficient models has been introduced in various articles, see Hastie and Tibshirani (1993) and Fan and Zhang (1999, 2000, 2008), including recursive estimation to improve efficiency (Cai et al. 2000), estimation under measurement errors (Chiang et al. 2001), models with generated regressors (Pendakur and Sperlich 2010), generalized varying-coefficient models (Mammen and Nielsen 2003; Roca-Pardinas and Sperlich 2010) or additive varying coefficients (Yang et al. 2006). The well-known time-varying-coefficient models for longitudinal data should be mentioned as well. For a recent review see Park et al. (2015). The extension of (4) to (3) is just an extension of the classical nonparametric varying-coefficient model toward a non- or semiparametric mixed effects model, see for example González Manteiga et al. (2013).

In empirical economics, however, these models are hardly known. The potential of applying this type of model has not been sufficiently emphasized. The aim of this article is to highlight the advantage of varying-coefficient models to deal with (endogenous, i.e. $X \not\perp Z$) heterogeneity, and to study the state-of-the-art of implementation and calculation. On this purpose, econometric justification as well as a discussion of computational means in the statistic software R (see R Development Core Team 2014) will be provided.

The rest of this article is structured as following: In the next section the problems that occur when using classical linear methods (including IV estimators) are illustrated. In Sect. 3 we show why and how varying-coefficient models can mitigate these problems. Section 4 briefly describes the implementation of these methods and the available R packages. These are then compared by simulation studies. Section 5 concludes.

2 Problems with linear OLS and IV estimation

Identifying an average effect of (or return to) a certain covariate is a multidimensional problem. When we interpret the coefficient in a linear model, we say Y would increase or decrease 'in average' by the value of the coefficient if one changed X by one unit. This is often misunderstood as being the average return (to X) in general. This is unfortunately incorrect as it refers only to the averaging-out of the individual deviations (i.e. the error term) iff this is mean-independent from X. When the heterogeneity of returns is stochastically related to X (causality is not required), this interpretation is definitively wrong. Let us therefore clarify this a bit.

Intuitively, to study an overall average effect, $E[\beta_i]$ makes sense if we can obtain the following equality from equation (2)

$$E[Y_i|X_i] = X_i^T E[\beta_i], \quad i = 1, \ldots, n.$$

This equation is satisfied if $E[e_i|X_i] = 0$, i.e. the mean independence due to the exogeneity assumption of X_i with respect to ε_i plus $E[\beta_i|X_i] = E[\beta_i] = b$. The latter is the mean-independence of β_i from returns X_i. Violation of this independence leads to inconsistency.

Proposition 1 *The OLS Estimator of the average effect $E[\beta_i]$ in model (2) is generally, i.e. without further assumptions, inconsistent.*

Proof Let $\{(Y_i, X_i)\}_{i=1}^n$ be a sample of independent random variables drawn model (2) with $E[\varepsilon_i|X_i] = 0$. Then,

$$\widehat{\beta}_{OLS} = (\Sigma X_i X_i^T)^{-1} \Sigma X_i Y_i \xrightarrow[n \to \infty]{p} E[\beta_i] + E^{-1}[X_i X_i^T] E[X_i e_i]$$

$$= E[\beta_i] + E^{-1}[X_i X_i^T] E\left[X_i X_i^T \{\beta_i - E[\beta_i]\} + X_i \varepsilon_i\right]$$

$$= E^{-1}[X_i X_i^T] E[X_i X_i^T \beta_i].$$

From the last line we will see that $\widehat{\beta}_{OLS} \xrightarrow{p} E[\beta_i]$ unless $E[\beta_i|X_i] = E[\beta_i]$, which is not true in general. □

Under the following simulation design, we get an idea of the consequences and the severity of this result:

$$Y_i = 0.6 + X_i \beta_i + \varepsilon_i$$
$$X_i = \gamma Z_i + u_i$$
$$\beta_i \sim Beta(1.659435 Z_i, 5)$$
$$\varepsilon_i \sim N(0, Var(X_i \beta_i))$$
$$Z_i \sim \log N(0, 1)$$
$$u_i \sim N(0, 1)$$
$$\gamma = \frac{Corr(X_i, Z_i)}{\sqrt{Var(Z_i) - Corr(X_i, Z_i)^2 Var(Z_i)}}.$$

Table 1 Simulated bias over the correlation between Z and X

$Corr(x_i, z_i)$	Intercept	Slope
−0.95	0.5169	0.4061
−0.75	0.0510	0.2426
−0.5	−0.0328	0.1140
−0.25	−0.0315	0.0315
0	−0.0024	0.0017
0.25	0.0288	0.0287
0.5	0.0359	0.1117
0.75	−0.0554	0.2413
0.95	−0.5198	0.4085
Parameter expectation	0.6	0.2857

We computed the average estimated bias of the OLS estimator $\frac{1}{m} \sum_{j=1}^{m} \widehat{\beta}_{OLS}^{j} - E[\beta_i]$ from $m = 1000$ simulation samples, each of size 100. Table 1 gives the biases for intercept (0.6) and the average slope (0.2857). As expected, the bias increases with the correlation between the coefficients and the regressors although there is not necessarily a (direct) causality relation from X on β. We see clearly that albeit a nontrivial dependency structure between β_i and X_i, the 'bias' of an OLS estimate compared to the average slope is quite severe.[2]

The result, that the OLS estimator is biased, is neither new nor surprising. The reason why this small simulation is presented here is to get aware of the role of g, Z_i and u_i when switching to model (3). This enables us to better understand source and size of the bias, to potentially model it, and to discuss ways of circumventing it—those that work and those that do not.

But is this simulated design of interest, does it reflect somehow any realistic situation, in particular in economics? We believe, it does. Take for example economies of scale—like they are generally accepted for the agricultural sector. The main interest is to estimate and understand the production function. 'Economies of scale' means that the size of an entity (enterprise, farm, …) changes the expected returns. Severance-Lossin and Sperlich (1999) studied the additive marginal effects and elasticities of inputs for farms in Wisconsin, USA. They applied a standard nonparametric additive model of the form

$$\log Y_i = \gamma_0 + \sum_{j=1}^{q} \gamma_j (\log X_{i,j}) + \varepsilon_i \qquad (5)$$

with all inputs X and output Y measured in US-Dollar, where all γ_j functions were nonparametric, corresponding therefore to our model with setting $\gamma_j(\log X_{i,j}) = g_j(\log X_{i,j}) \cdot \log X_{i,j}$ or $\gamma_j(\log X_{i,j}) = g_j(X_{i,j}) \cdot X_{i,j}$. They indeed found increasing returns $\sum_{j}^{q} \gamma_j' > 1$ for middle-sized and large farms; this means, *size* made the γ_j

---

[2] We put bias in quotation marks as the OLS estimate is actually not biased in the statistical sense, it does simply estimate something different than the average slope. Alternatively you may say that not the estimator is wrong but the interpretation that people often attach to it.

functions non-linear such that X_i would be endogenous in a standard log-linear model. *Size* was clearly driving both (at least partly), inputs X_i and β_i, and is therefore a natural effect modifier Z.

Similarly, in order to study the efficiency of labor offices in the Czech Republic during the transition years, Profit and Sperlich (2004) estimated a dynamic fixed effects panel model with time fixed effects η_t and individual fixed effects δ_i, i.e.

$$Y_{i,t} = \gamma_0 + \sum_{j=1}^{4} \gamma_j(\log X_{i,t-1,j}) + \sum_{1 \leq j < k \leq 4} \gamma_{j,k}(\log X_{i,t-1,j}, \log X_{i,t-1,k}) + \delta_i + \eta_t + \varepsilon_{i,t}$$
(6)

which is basically an extension of the additive interaction model of Sperlich et al. (2002) for longitudinal data. Equation (6) is a standard—apart from the fact that it is nonparametric—economic model to study job-matching with Y being log unemployment to job exits, and X summarizing the lagged Y, the log number of vacancies, the log number of inflows to unemployment, and the unemployment stock. Profit and Sperlich (2004) tested the returns to scale of the labor offices, studied time and space variation (i.e. the non-uniformity of Job-Matching), and afterward the sources of their variation. They found that the variation (i.e. the non-linearity of the γ_j functions) was due to borderland, and specific active labor market policies, say a vector of effect modifiers Z. At the same time, this vector Z was clearly correlated to X with different causality directions: e.g. being borderland had a positive impact on unemployment stock, but unemployment stock had an impact on the active labor market policy.

For both examples one can argue that a varying-coefficient model had been a reasonable alternative, allowing for the direct inclusion of Z in the model. Subsequently, a further step would be to consider $\beta_i - g(Z_i)$ as an unobservable disturbance that then may lead to a similar (though now mitigated) endogeneity problem. One could argue that Z might be hard to find or too weak to control for the dependency between X and the returns. As said, in econometrics a popular way to deal with all kind of endogeneity is the IV estimation based on a set of additional variables, so-called instruments W_i, that on the one hand is supposed to be exogenous in the sense that $E[e_i|W_i] = 0$, but on the other hand is correlated with X_i. As this is quite unlikely to exist, in empirical economics it is common to ignore the heterogeneity of β_i (i.e. to assume it to be fixed) and ask for $E[\varepsilon_i|W_i] = 0$, but

Proposition 2 *For a random sample* $\{W_i\}_{i=1}^{n}$ *of 'instruments' with* $E[\varepsilon_i|W_i] = 0$, $Cov(X_i, W_i) \neq 0$, *the linear IV estimator of* $b = E[\beta_i]$ *in model* (2) *is generally inconsistent except if* $E[e_i|W_i] = 0$. *Moreover,* $E[\beta_i - b|W_i] = 0$ *is not sufficient.*

Proof For simplification but without loss of generality let $dim(X) = dim(W)$. Then

$$\widehat{\beta}_{IV} = (\Sigma W_i X_i^T)^{-1} \Sigma W_i Y_i = (\Sigma W_i X_i^T)^{-1} \Sigma W_i \{X_i^T E[\beta_i] + e_i\}$$
$$\xrightarrow[n \to \infty]{p} E[\beta_i] + E^{-1}[W_i X_i^T] E\left[W_i X_i^T \{\beta_i - E[\beta_i]\}\right] + E^{-1}[W_i X_i^T] E[W_i \varepsilon_i]$$
$$= E^{-1}[W_i X_i^T] E[W_i X_i^T \beta_i],$$

where the last equality cannot be simplified without further assumptions like for example $\beta_i \perp (X_i, W_i)$ for all $i = 1, \ldots, n$.   □

We have added the comment that $E[\beta_i - b|W_i] = 0$ is not sufficient because it is a commmon mistake to believe it would be sufficient. In fact, assuming $\beta_i \perp W_i$ is not enough because this does typically not give $E[W_i X_i^T (\beta_i - b)] = 0$, what is needed for the consistency. Unfortunately, having a β_i that is independent of both, X_i and W_i, is very unlikely and even contradicts economic theory. For this latter case Imbens and Angrist (1994) defined the local average treatment effect (LATE). More specifically, under a set of additional non-testable assumptions, the linear IV estimator identifies the average return (or effect) for 'compliers'. These form the subpopulation of individuals that increased their value of X due to an increase in W. It is indispensable that all individuals move in the same direction (monotonicity assumption); therefore we can not relax the term 'increase' to simply 'change'. Its derivation relies on the identification of the expected potential outcome for a given value of the instrument w with the conditional expectation: $E[X_i^T (w)\beta_i] = E[Y_i|W_i = w]$ which works if for all compliers i one has $\beta_i \perp W_i$—being thus the crucial independence assumption for LATE. But as the compliers form only a specific subpopulation defined by the choice of instruments W, the LATE is not necessarily a relevant parameter. Finally one should mention that the concept was introduced for binary X and W. Extensions are tedious but straight for discrete X and W with finite support. The LATE becomes then a weighted average of the LATEs of compliers for a stepwise increase of W.

To sum up, an estimator based on a LATE argument would inherit the same defects as the classic IV estimator although it seems to allow to relax the independence assumption. Its interpretation is only clear for binary X and W, but even then may refer to an irrelevant parameter. In any case it will give only an average which in turn is little informative if heterogeneity in returns is important. The inadequacy of the tradition inference motivates the use of a new class of estimators. The proposal in the next section is based on a non-parametric modeling of the coefficient by considering the use of a set of observed variables Z_i, the so-called 'effect modifiers'.

3 The varying-coefficient model for heterogeneous effects

Let us now consider the varying-coefficient model (3). Introducing effects modifiers Z_i helps to mitigate the harmfulness of the association between β_i and X_i, and therefore β_i and W_i. One may even re-establish the validity of W_i as an instrument 'conditionally on $g(Z_i)$'. At this point, a crucial assumption is required for the present section: $u_i \perp X_i$ or $u_i \perp W_i$ for all $i = 1, \ldots, n$. In other words, β_i depends on X_i or W_i only through Z_i.

For brevity we discuss simply the two situations where (1) u_i is mean-independent from X_i, and (2) when (1) fails but either $\beta_i(z)$ or u_i are mean-independent from an instrument W_i which further fulfills $E[\varepsilon|W_i] = 0$ and $E[X_i|W_i]$ not being constant. Other cases like independence between X_i and Z_i are less interesting and not considered here.

3.1 When u is mean-independent from X

Recall that Eq. (3) was just a different way of writing $Y_i = X_i^T \beta_i + \varepsilon_i$ splitting β_i into a part conditioned on Z_i and a mean-zero remainder term u_i such that

$$E[Y_i|X_i] = X_i^T E[g(Z_i)|X_i] \neq X_i^T E[g(Z_i)]$$

unless X_i and Z_i are mean-independent. In this case it is hardly possible to compute an average effect without conditioning on Z, too. In contrast, doing so you get

$$E[Y_i|X_i, Z_i] = X_i^T g(Z_i),$$

or simply

$$Y_i = X_i^T g(Z_i) + e_i,$$

with $E[e_i|X_i, Z_i] = 0$. This is the simple varying-coefficient model (4) which can be estimated by different semiparametric methods.

It should be reminded that standard IV methods will fail here. This example illustrates the fact that if we want to obtain an overall parameter estimation, it is hardly avoidable to estimate the observable dependent heterogeneity in a first step. This is particularly true for cases where heterogeneity may be of first order. This is were varying-coefficient models are made for. Let us give two examples from the literature.

Making the Engle curves more flexible has been a central issue when analyzing consumer behavior in microeconomics. To our knowledge, in this context semiparametric varying-coefficient models have been applied first by Pendakur and Sperlich (2010). They studied consumer behavior, and modeled the expenditure shares by varying-coefficients models. More specifically, Y_i is now a vector of expenditure shares for M goods,

$$Y_i = exp.shares_i = \gamma(E_i^R) + \Gamma(E_i^R)\, P_i, \tag{7}$$

where vector $\gamma(\cdot) \in I\!R^M$ and matrix $\Gamma(\cdot) \in I\!R^{M \times M}$ are nonparametric functions of the *real* total expenditures E_i^R of household i (which are predicted from the model.)[3] P_i is the vector of the log prices of the M goods faced by household i. This allows one to e.g. check for Slutsky symmetry[4] or to determine the effective inflation index for any income group.[5] Asymmetry and different effective inflation indices imply that returns to prices must be heterogeneous, and this is what the authors found.

---

[3] The preposition *real* indicates that the numbers E^R assigned to the different households—in this example from all-over Canada—refer to the same purchase power.

[4] Simplified, this means that if a household switches from a product to a substitute because of a special discount, he will switch back if this discount is no longer conceded.

[5] The published consumer price index refers to a representative mean basket of goods and services. This basket clearly varies a lot over the different income groups, resulting in different effectively realized inflation rates for each group.

Similarly, Pendakur et al. (2010) considered the indirect utility in consumer demand systems. They modeled the utility for household i as

$$U_i = E_i - P_i^T \gamma(E_i) - \frac{1}{2} P_i^T \Gamma P_i, \tag{8}$$

where E_i it reported total expenditure of household i, P_i again the vector of log prices of all M items, Γ an unknown $M \times M$ matrix of unknown coefficients, and γ again a vector of M nonparametric functions. Roy's identity (Roy 1947) tells us that the vector of expenditure shares is given by $\frac{\partial U}{\partial P} / \frac{\partial U}{\partial E}$, i.e. in our case

$$exp.shares_i = \frac{\gamma(E_i) + \Gamma P_i}{1 - P_i^T \gamma'(E_i)}. \tag{9}$$

Equation (9) can be estimated from observed data, although not directly with standard methods. Again, the authors found heterogeneous returns to prices in the utility (8) of households. No direct method could have directly estimated the average returns to prices.

3.2 Dependence between u and X but existence of an instrument W

As indicated above, we consider the situation where W_i is mean-independent from u_i given Z_i. For this case a LATE type parameter can be identified. This is based on the joint conditioning of Z_i and W_i combined with the traditional monotonicity assumption related to this kind of inference. On this purpose, and for the sake of simplicity, X_i is considered to be a binary random variable. Let also W_i be a binary random variable. The extension to discrete variables with finite support is straight but rather tedious in notation and interpretation. One can then establish

Definition 1 Let $\{(Y_i, X_i, Z_i)\}_{i=1}^n$ be a random sample generated by model (3), and $\{W_i\}_{i=1}^n$ with $E[\varepsilon_i|W_i] = E[u_i|W_i] = 0$ such that X_i and W_i are dependent binary random variables with $X_i(W_i = 1) - X_i(W_i = 0) \geq 0$ almost sure for all i. The conditional LATE evaluated at $Z_i = z$ is

$$b_{LATE}(z) = \frac{E[Y_i|Z_i = z, W_i = 1] - E[Y_i|Z_i = z, W_i = 0]}{E[X_i|Z_i = z, W_i = 1] - E[X_i|Z_i = z, W_i = 0]}.$$

This is the expected average effect of $\beta_i(z) = g(z) + u_i$ in the population of compliers for the given value z of the effect modifier Z_i. If there are no compliers for $Z_i = z$, i.e. $E[X_i|Z_i = z, W_i = 1] - E[X_i|Z_i = z, W_i = 0] = 0$, then the $b_{LATE}(z)$ is not defined. We can estimate this conditional LATE now by combining the classical Wald estimator for endogenous dummies with a nonparametric estimator for varying-coefficient models.

Note, however, that also this LATE type estimator identifies only an average effect of the compliers. These do not necessarily constitute a clearly identified (sub)population or they might not be of interest. So how does this improve compared to the classic

LATE? It does in the following two aspects. First, depending on the choice and strength of the effect modifier Z, it is much easier to argue that W_i was independent from u_i than arguing it was independent from β_i. Second, if Z is a strong modifier, then we can control and thus study for most of the heterogeneity; the different parameters identified by the different instruments are expected to have a small variation, i.e. the choice of instruments should mainly effect the variance but hardly the mean of the estimator. A generalization to X_i, Z_i, W_i being continuous is therefore appealing. This can be done by the so-called control function approach, see Telser (1964). Newey et al. (1999) introduced a nonparametric extension and made it popular in econometric theory.

Let us consider the following simultaneous equations system:

$$Y_i = X_i^T g(Z_i) + X_i^T u_i + \varepsilon_i \tag{10}$$
$$X_i = h(W_i) + v_i, \quad i = 1, \ldots, n, \tag{11}$$

with the same assumptions as above but now X and W being (potentially) continuous and $E[v_i|W_i] = 0$. This is a natural extension of model (3). Note that the additivity in Eq. (11) guarantees the monotonicity of the effect of W on X (assumed in the binary case). Clearly, h can be nonparametrically estimated in order to predict v_i by $\widehat{v_i} := X_i - \hat{h}(Z_i)$. Then,

$$E[Y_i|X_i, Z_i, v_i] = X_i^T g(Z_i) + X_i^T E[u_i|X_i, Z_i, v_i].$$

As all dependence between X_i and u_i is reflected in v_i, and Z_i was independent from u_i by definition, we get for $E[u_i|X_i, Z_i, v_i] = E[u_i|v_i] =: \lambda(v_i)$

$$\begin{aligned} E[Y_i|X_i, Z_i, v_i] &= X_i^T g(Z_i) + X_i^T \lambda(v_i) \\ &= X_i^T g(Z_i) + X_i^T \lambda(\widehat{v_i}) + X_i^T [\lambda(v_i) - \lambda(\widehat{v_i})], \end{aligned}$$

where from Sperlich (2009) we know that the last term is asymptotically negligible. Moreover, it is known from the literature on additive varying coefficients that g_i and λ_i are identifiable up to additive and multiplicative constant terms.

4 Estimating varying-coefficient models in practice

We already presented above a list of references introducing different estimators for various versions of varying-coefficient models. In this section we provide a summary of existing implementations for varying-coefficient model estimators. Namely, we discuss some important R packages (R Development Core Team 2014) that contain those methods. Clearly, this review is not pretended to be complete or exhaustive. It rather serves as an illustration that practitioners are already today provided with the necessary tools for performing the modeling and estimation strategies discussed in the preceding section.

Among others, Fan and Zhang (1999, 2000, 2008) proposed a smoothing method for estimating g in a classical varying-coefficient model $Y_i = X_i^T g(Z_i) + \varepsilon_i$ with $E[\varepsilon_i|X_i, Z_i]$ for all $i = 1, \ldots, n$. Their method was based on kernel weighted local

polynomials. Intuitively it is quite appealing because it exhibits strong similarities with a weighted least squares M-estimator but only at a local level. In the local linear case one considers $g(Z) \approx \alpha_0 + \alpha_1^T (Z - z_0)$ in a neighborhood of z_0 with $\alpha_0 \in \mathbb{R}^{dim(X)}$ and $\alpha_1 \in \mathbb{R}^{dim(Z) \times dim(X)}$. More specifically, one tries to minimize the squared error

$$E \left[(Y - \{X^T \alpha_0 + X^T \alpha_1^T (Z - z_0)\})^2 | Z = z_0 \right].$$

Given a sample $\{(Y_i, X_i, Z_i)\}_{i=1}^n$ of independent observations from the same model, by convolution with a kernel function $K_h(\cdot)$ provided with bandwidth h around a point of interest z_0, one obtains the empirical counterpart:

$$(a_0^*, a_1^*) = \underset{(a_0, a_1)}{argmin} \frac{1}{n} \Sigma \left\{ Y_i - [X_i^T a + X_i^T b(Z_i - z_0)] \right\}^2 K_h(Z_i - z_0),$$

where a_0^* can be shown to be a consistent estimator for $g(z_0)$. Differently from this kernel smoothing approach, Hastie and Tibshirani (1993) proposed a smoothing spline method, also based on a penalized least squares minimization. There, each coefficient is parameterized on a cubic spline basis. Local maximum likelihood (e.g. Cai et al. 2000), other spline methods (e.g. Chiang et al. 2001), smooth backfitting (Mammen and Nielsen 2003; Roca-Pardinas and Sperlich 2010) and Bayesian structured additive models (Fahrmeir et al. 2004) have been introduced as well. For most of these methods asymptotic theory has been provided. Inference is typically based on resampling methods or on rough (parametric plug-in) approximations. But even more interesting for the empirical researcher is the question regarding implementation. Nowadays, several R packages can be used to estimate varying-coefficient models.

Hayfield and Racine (2008) have developed the so-called (np) package (nonparametric kernel smoothing methods for mixed data types). This package enables us to implement a kernel estimation for a wide range of variable types: continuous variables, ordered or unordered factors. The function npscoef provides estimators for a varying-coefficient model. An optimal bandwidth h can be obtained by a least-squares cross-validation method applying the function npscoefbw. For our purpose, however, a problem of this package is that it seems hard (if not impossible) to modify the set of effect modifiers from one varying coefficient to another, i.e. all varying coefficients depend on the same set Z. It is also not able to estimate additive varying coefficients. Note that the function predict only returns predictions for the expected Y. It does not predict the estimated coefficient functions at arbitrary values of the effect modifier. This makes a fair comparison with other packages a bit difficult.

Wood (2011, 2006, 2004, 2003, 2000) proposed a computationally fast algorithm for estimating generalized additive models based on penalized regression splines. This is implemented in his package mgcv. The available distribution families can be found with the function family.mgcv. The function considered in the simulation study below is gam.

Not exactly the same but quite related, at least in many situations, is the gamlss package. It was developed by Rigby and Stasinopoulos (2005), and originally devoted to the estimation of generalized additive models. The common exponential families are replaced by a more flexible distribution family allowing for extreme kurtotic,

skewed or over-dispersed continuous and discrete distributions. The available distribution families can be found with the function `gamlss.familly`. The special case of varying-coefficient models is handled by a maximum likelihood method based on penalized B-splines. The function studied in the following is `gamlss`.

In order to compare the three packages, several simulation designs are used. Due to the variety of methods used by these packages, the designs share common features to enable a comparison on the same basis: (a) the coefficient functions are the same whatever the design is; (b) a uniform distribution of the effect modifiers to prevent a bias at the tails when using a fixed bandwidth in kernel estimation, i.e. to make kernel and spline methods comparable; (c) bounded coefficient functions with trigonometric specification to stabilize the variance; (d) predictions of the coefficient functions at equidistant points on the interval $[-2, 2]$.

We simulated four different designs, see below. From each we draw 1500 independent random samples for sample sizes $n = 50, 100, 250$ (i.e. 500 simulation runs for each size) to obtain 500 replicates at fixed values of the effect modifiers Z for each design and sample size. From them we calculated 90 % simulation intervals (i.e. cut the
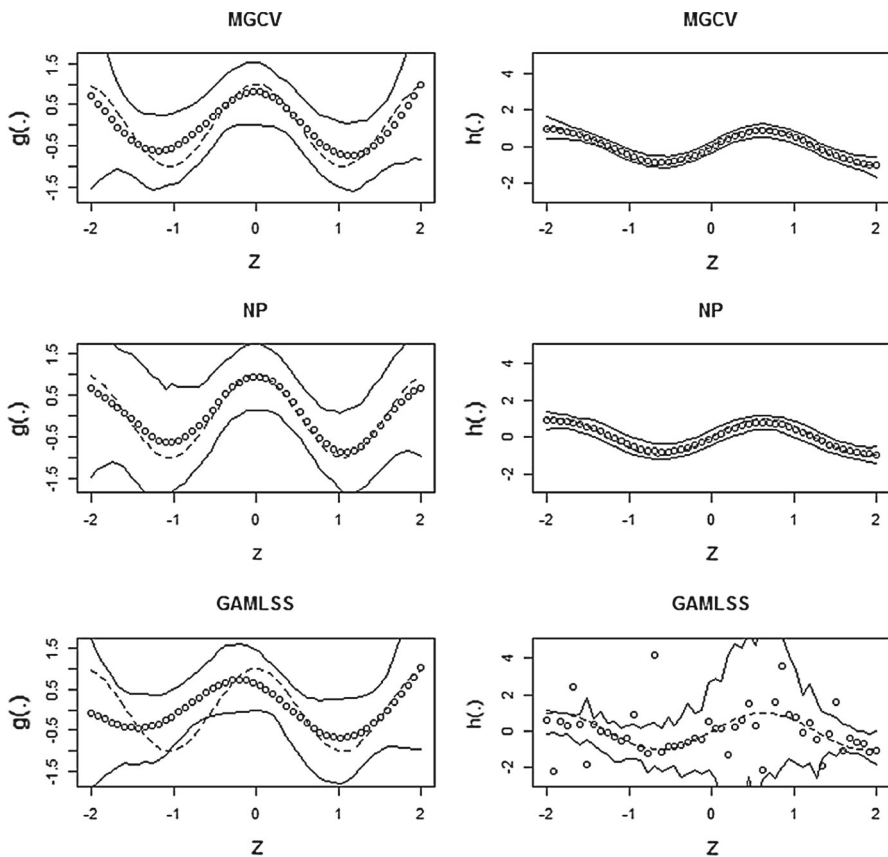

Fig. 1 Design 1, 90 % simulation intervals for sample size $n = 50$ for g on the *left*, h on the *right*
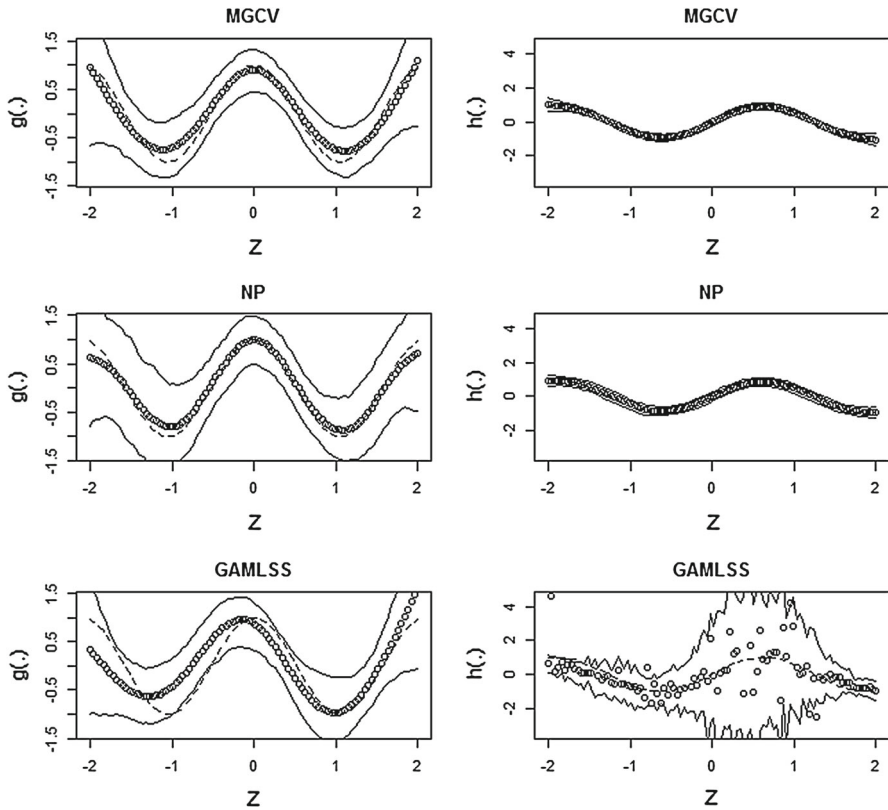
Fig. 2 Design 1, 90 % simulation intervals for sample size $n = 100$ for g on the *left*, h on the *right*

outer 5 % from above and from below), see Figs. 1 to 12, and the average bias (ABias) and average mean squared error (AMSE) for the varying coefficient functions. In all figures the dotted lines in the center are the 'true' data generating function, the circles are the averages of the 500 predictions, and the bands are constructed from the 5 and 95 % quantiles of the 500 simulations. Note that for notational convenience the two nonparametric coefficient functions in Design 1–3 are denoted by $g(.), h(.)$ instead of $g_1(.), g_2(.)$.

4.1 Design 1

As previously mentioned, prediction of the coefficient isn't implemented in the np package. Further, this package does not seem to be designed to consider different sets of effect modifiers for each coefficient function. However, a preliminary simulation study allows to appreciate the performance of this package in case these problems are irrelevant for the practitioner. This is the aim of the first design where the effect modifier is considered as being a fixed grid on the interval $[-2, 2]$. Else, we simulate the model
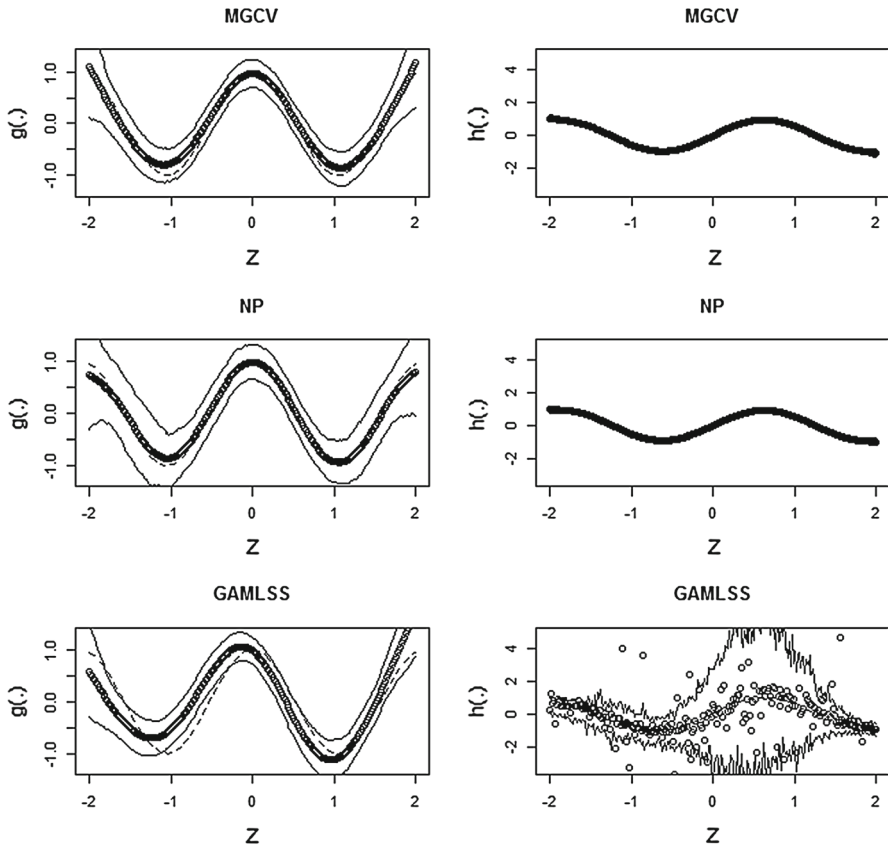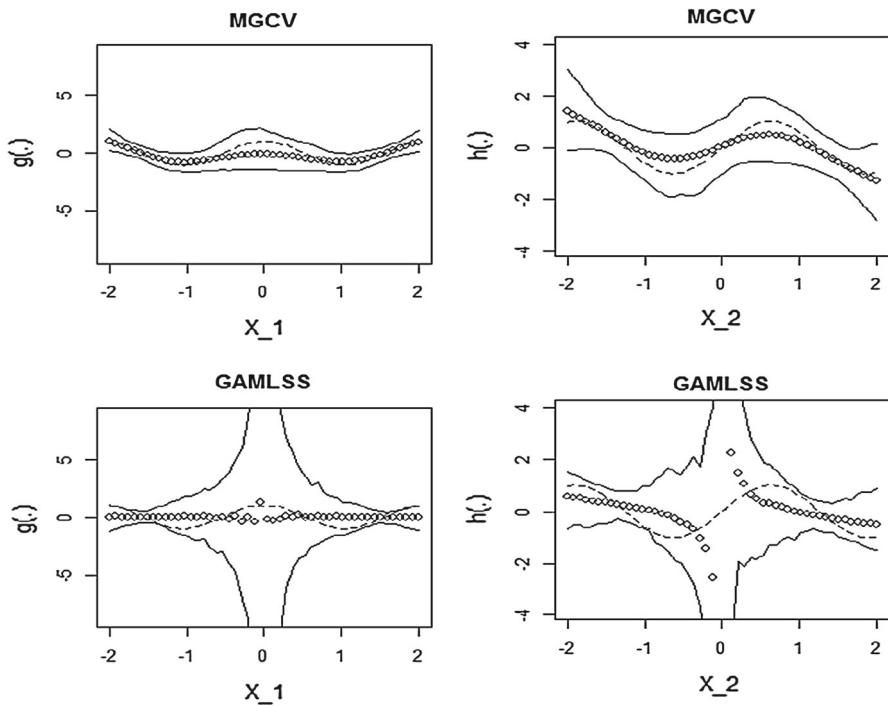
Fig. 3 Design 1, 90 % simulation intervals for sample size $n = 250$ for g on the *left*, h on the *right*

$$Y = g(Z) + h(Z)X + \varepsilon$$
$$g(Z) = \cos(3Z), \ h(Z) = \sin(2.5Z)$$
$$X = 2W_1 - 0.5W_2^2 + 2Z, \ Z = \left\{-2 + i\frac{4}{n} | i = 1, \dots, n\right\}$$
$$W_1 \sim N(0, 1), W_2 \sim N(0, 1), \varepsilon \sim N(0, 1).$$

The results are given in Figs. 1, 2 and 3, and Table 2. The figures lead us to the conclusion that the np and mgcv perform pretty well. The mgcv package seems to dominate other packages in terms of ABias and AMSE when the sample size is small ($n = 50$). For larger sample sizes, the np package performs much better. Indeed, when the curvature is important ($g(.)$), there is a slight bias that is probably due to the smoothing effect related to any nonparametric estimation. The np package tends to reduce this as the sample size increases. The gamlss package estimation has clearly a problem for $Z \leq 0$ when estimating the 'intercept' function $g(.)$, and is rather imprecise at values $Z \in [0, 1]$ when estimating $h(.)$. We could not find out what

Table 2 Results calculated from 500 samples drawn from Design 1

Sample size	Package	$ABias[\hat{g}(.)]$	$ABias[\hat{h}(.)]$	$AMSE[\hat{g}(.)]$	$AMSE[\hat{h}(.)]$
50	mgcv	0.0262	0.0035	0.0641	0.0088
	np	0.0518	0.0073	0.0370	0.0179
	gamlss	0.0262	0.2088	0.2348	4.0957
100	mgcv	0.0364	−0.0021	0.0267	0.0032
	np	0.0251	0.0003	0.0168	0.0077
	gamlss	0.0364	−2.2790	0.1460	517.4362
250	mgcv	0.0425	−0.0014	0.0129	0.0014
	np	0.0150	0.0005	0.0073	0.0034
	gamlss	0.0425	−0.6461	0.1335	1348.8040

Fig. 4 Design 2, 90 % simulation intervals for sample size $n = 50$ for g on the *left*, h on the *right*

exactly is the source of the problem. This imprecision tends to even increase when the sample size increases.

4.2 Design 2

The second design is to study the situation where the effect modifiers are simply the regressors of the model, i.e. $X = Z$. Again, the functions g, h are univariate but refer
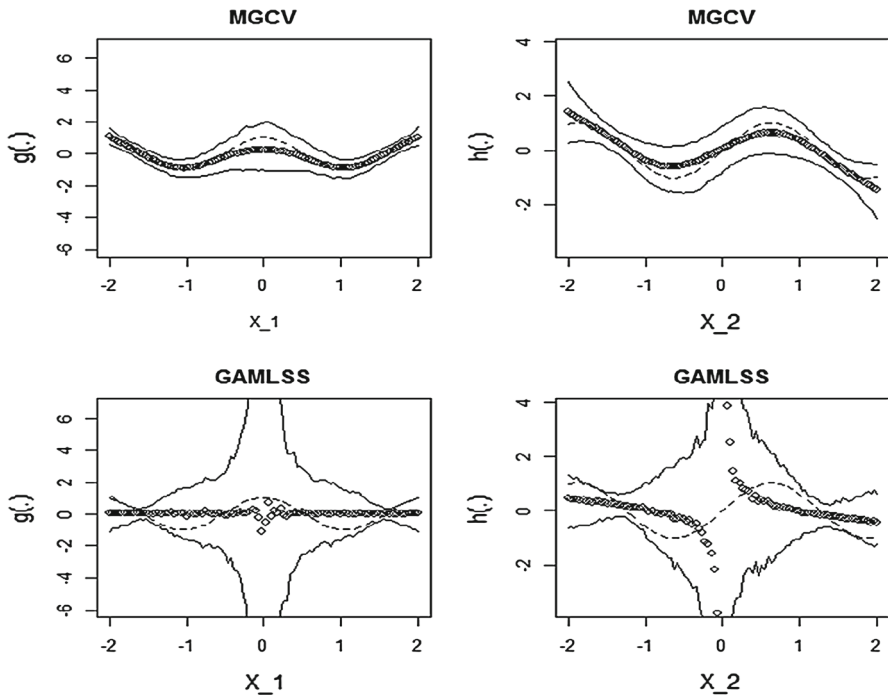
Fig. 5 Design 2, 90 % simulation intervals for sample size $n = 100$ for g on the *left*, h on the *right*

to different effect modifiers. For this reason np is no longer included. Ignoring np allows us to now consider random variables Z that are independently drawn in each simulation study, because both, gamlss and mgcv are able to predict g, h on given fixed grids, no matter where the effect modifier was originally observed. The simulated design is

$$Y = 6 + g(X_1)X_1 + h(X_2)X_2 + \varepsilon$$
$$g(X_1) = \cos(3X_1), \ h(X_2) = \sin(2.5X_2)$$
$$X_1 \sim U[-2, 2], \ X_2 = 0.25X_1 + 0.75U[-2, 2], \ \varepsilon \sim N(0, 1).$$

The results are plotted in Figs. 4, 5 and 6, and in Table 3. The mgcv package manages to estimate the coefficient functions but with a bias where the curvature is important. However, this problem decreases with the sample size;so this might simply be a smoothing effect. This is different in the gamlss package that just fails to fit the coefficient function. A direct analysis of ABias and AMSE tends to underestimate these two issues because negative and positive bias compensate each other all along the support of the coefficient functions. In any case, it is obvious that this design should better be estimated by (generalized) additive model estimators. The implementation of the varying-coefficient model in the gamlss package seems to require different (or maybe even independent) variations for the regressors X and the effect modifiers Z.
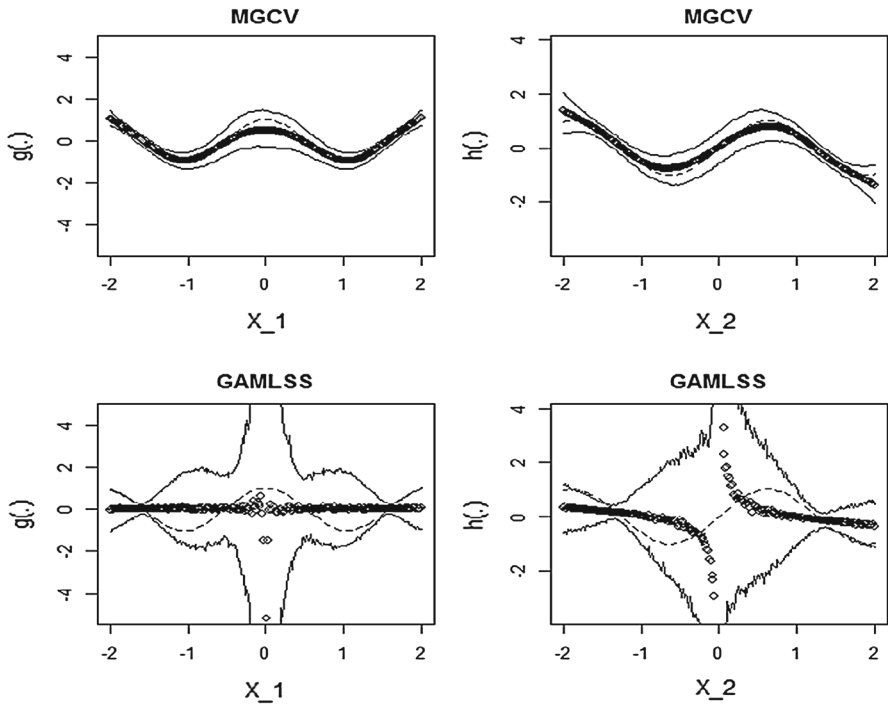
Fig. 6 Design 2, 90 % simulation intervals for sample size $n = 250$ for g on the *left*, h on the *right*

Table 3 Results calculated from samples drawn from Design 2

Sample size	Package	$ABias[\hat{g}(.)]$	$ABias[\hat{h}(.)]$	$AMSE[\hat{g}(.)]$	$AMSE[\hat{h}(.)]$
50	mgcv	−0.2152	0.0259	0.2172	0.1053
	gamlss	−0.0004	0.0003	0.5283	2.3782
100	mgcv	0.1393	0.0106	0.0932	0.0597
	gamlss	0.0430	0.0012	0.5169	3.2682
250	mgcv	−0.0831	0.0038	0.0366	0.0233
	gamlss	0.0647	−0.0043	0.8001	5.3895

We saw already in design 1 that this package has problems where the relation between X and Z are strong.

4.3 Design 3

The design is now crossing the dependence structure, i.e. the effect modifier Z_1 for the coefficient of X_1 is correlated with X_2, and Z_2 with X_1. Else one can speak of a standard bivariate nonparametric varying-coefficient model so that there should be no particular identification problem for gamlss. The simulated design is
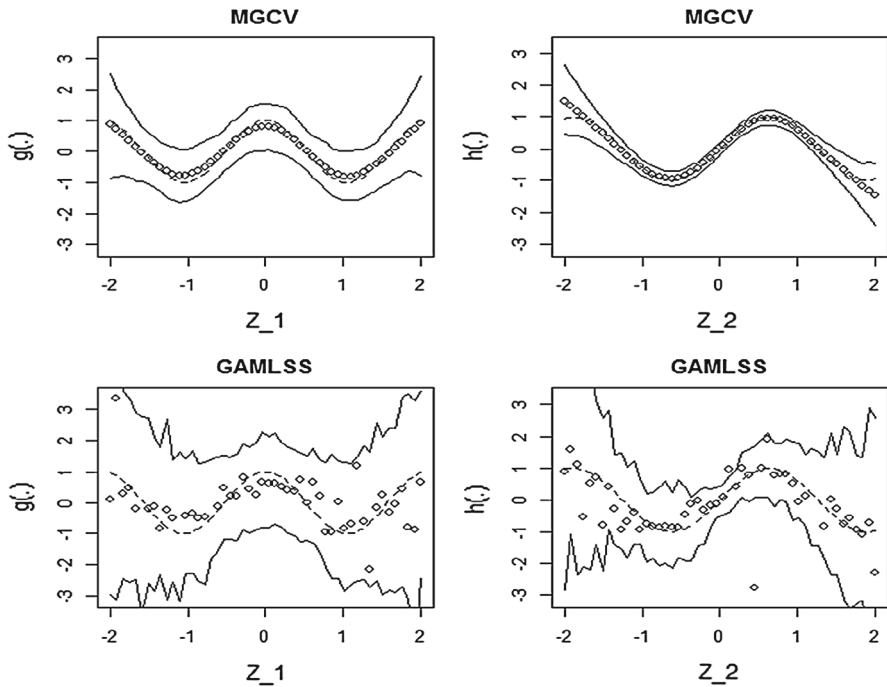
Fig. 7 Design 3, 90 % simulation intervals for sample size $n = 50$ for g on the *left*, h on the *right*

$$Y = 6 + g(Z_1)X_1 + h(Z_2)X_2 + \varepsilon$$
$$g(Z_1) = \cos(3Z_1), h(Z_2) = \sin(2.5Z_2)$$
$$Z_1 \sim U[-2, 2], Z_2 = 0.25Z_1 + 0.75U[-2, 2]$$
$$X_1 = 0.25W_1 + 0.4W_2 + Z_2, X_2 = 2W_1 - 0.5W_2^2 + 2Z_1$$
$$W_1 \sim N(0, 1), W_2 \sim N(0, 1), \varepsilon \sim N(0, 1).$$

The numerical results are shown in Figs. 7, 8 and 9, and Table 4. Both packages, mgcv and mgcv manage to predict the shape of the coefficient functions. One can say that the mgcv clearly preforms better. Similar conclusion can be drawn from Table 4. Note that the identification issue we detected in design 2 has disappeared, and the smoothing effect mentioned in the simulations with design 1 seems to be reduced a lot.

4.4 Design 4

This design studies the control function approach introduced in Sect. 3.2. Note that in this design, our function $h(.)$ is no longer a standard varying coefficient but plays the role of the control function. The simulated design is

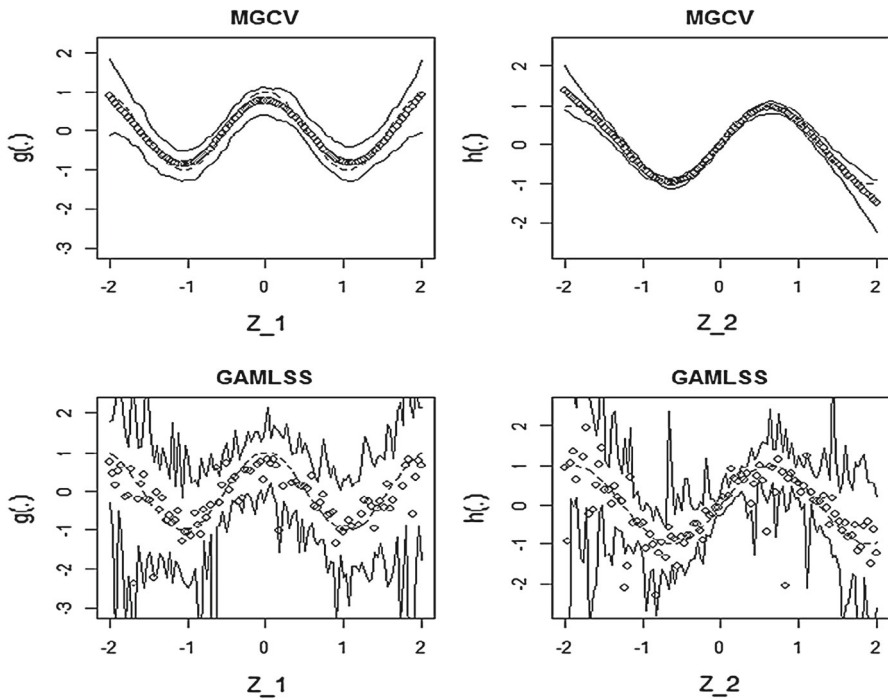**Fig. 8** Design 3, 90 % simulation intervals for sample size $n = 100$ for g on the *left*, h on the *right*

$$Y = 6 + g(Z)X + uX + \varepsilon$$
$$g(Z) = \cos(3Z), h(v) = \sin(2.5v)$$
$$Z \sim U[-2, 2], X = 0.25Z + 0.4W_2 + v$$
$$W_1 \sim N(0, 1), W_2 \sim N(0, 1)$$
$$u = h(v) + N(0, 1), v \sim U[-2, 2], \varepsilon \sim N(0, 1).$$

At the first step, v can be predicted by regressing X on Z and W_2 with any non-parametric method. At the second step, $h(.)$ is estimated on $\widehat{v}$ jointly with $g(.)$ on X. For a fair comparison we do the first step for all simulations and packages applying gam no matter whether the second step is then performed with mgcv or gamlss. So if (only) one of the method fails, then this is due to the second step; if both methods fail it might be due to the first and/or second step.

The results are given in Figs. 10, 11 and 12 and Table 5. We see that both packages manage to predict $g(.)$ with again a better fit for mgcv. Clearly, gamlss barely fits $h(.)$ exhibiting important errors and a very wide confidence band that makes the estimation unreliable, whereas mgcv gives a much better fit. However, when the sample size increases, the problems encountered by the gamlss package tend to be mitigated.

It should be remarked that there exists also the possibility to estimate varying-coefficient models with the BayesX R package. This is a software developed by
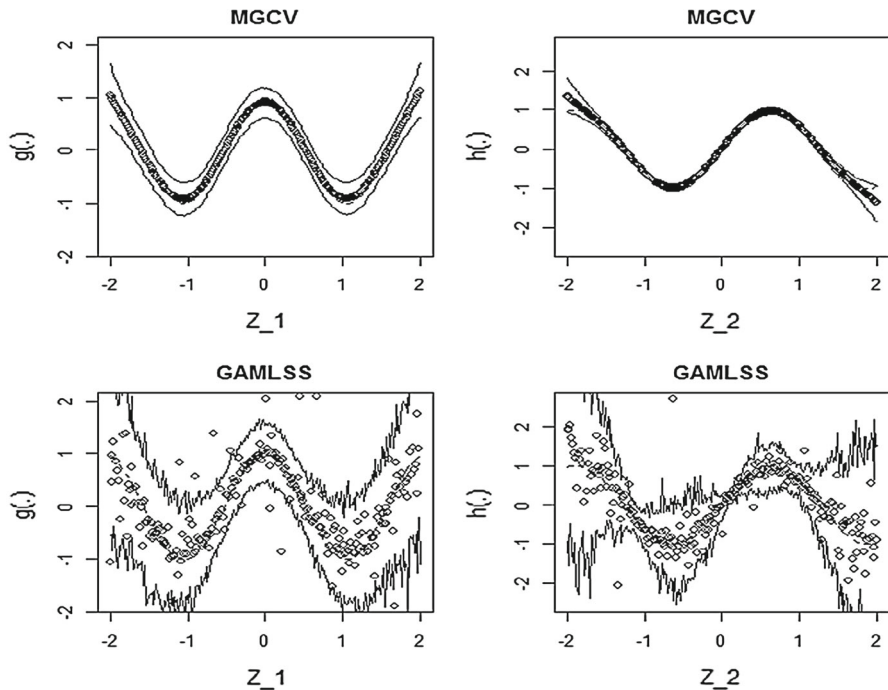
Fig. 9 Design 3, 90 % simulation intervals for sample size $n = 250$ for g on the *left*, h on the *right*

Table 4 Results calculated from 500 samples drawn from Design 3

Sample size	Package	$ABias[\hat{g}(.)]$	$ABias[\hat{h}(.)]$	$AMSE[\hat{g}(.)]$	$AMSE[\hat{h}(.)]$
50	mgcv	−0.0043	0.0040	0.0230	0.0193
	gamlss	0.0312	−0.0720	11.2338	2.9439
100	mgcv	−0.0310	−0.0166	0.0190	0.0124
	gamlss	0.2775	−0.0574	24.0567	0.4506
250	mgcv	−0.0007	−0.0022	0.0040	0.0066
	gamlss	0.1755	0.2812	4.6757	5.2134

Belitz et al. (2013). It was originally designed to mainly deal with generalized additive models, generalized mixed additive models, generalized geo-additive mixed models, and geographically weighted regression. But since a while it includes also special cases of varying-coefficient models. BayesX considers three inferential procedures: Full Bayesian (FB) inference based on MCMC simulation techniques (the function is `bayesregobjects`), mixed model based estimation corresponding to an Empirical Bayesian (EB) inference or a penalized likelihood (the function is `remlregobjects`), and inference based on penalized least squares (the function is `bstepwiseregobjects`). In brief, in the case of the FB approach, the prior of the coefficient functions are assumed to be random with appropriate hyper-priors
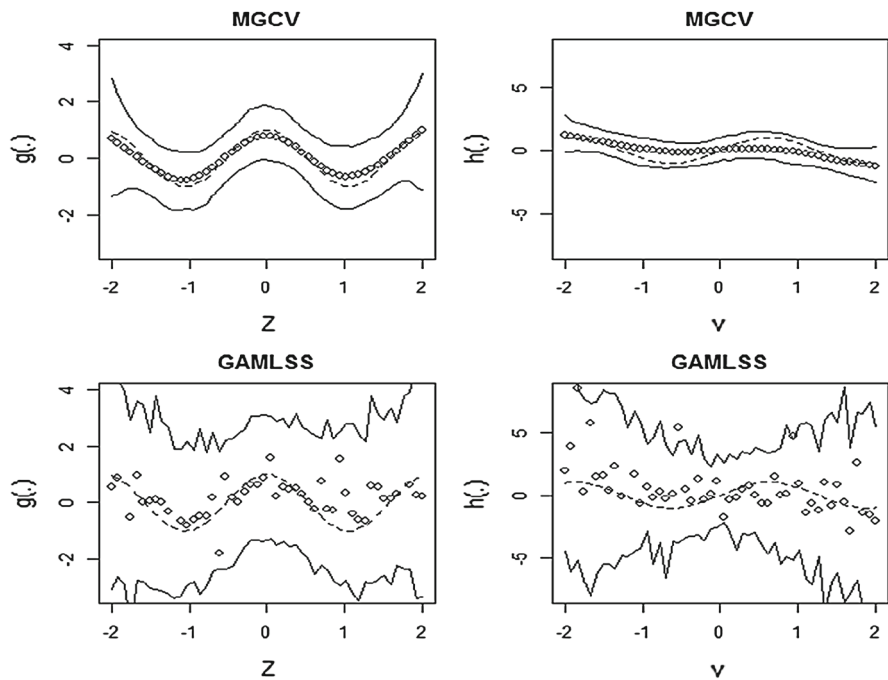
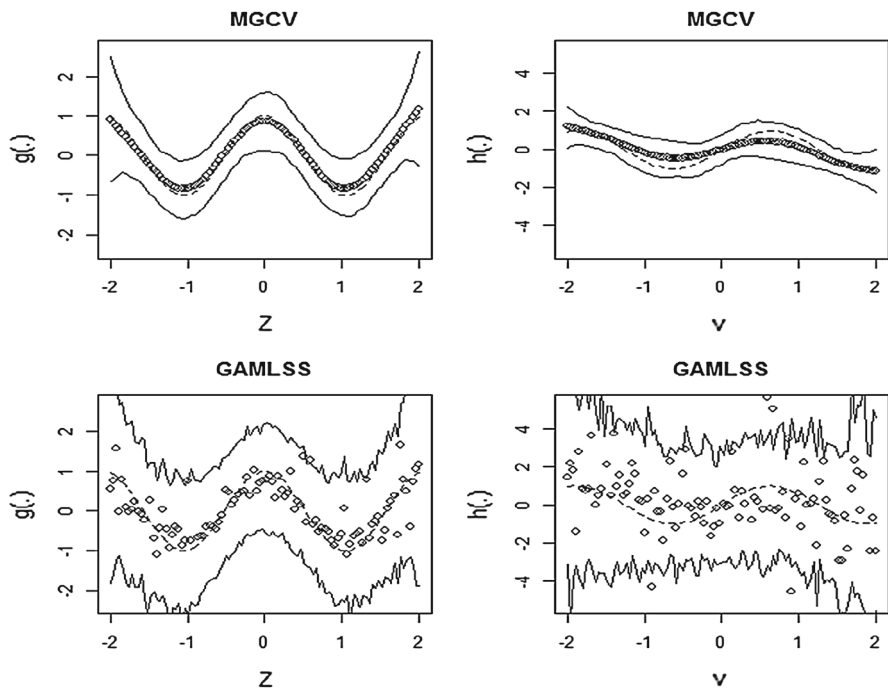**Fig. 10** Design 4, 90 % simulation intervals for sample size $n = 50$ for g on the *left*, h on the *right*

Fig. 11 Design 4, 90 % simulation intervals for sample size $n = 100$ for g on the *left*, h on the *right*
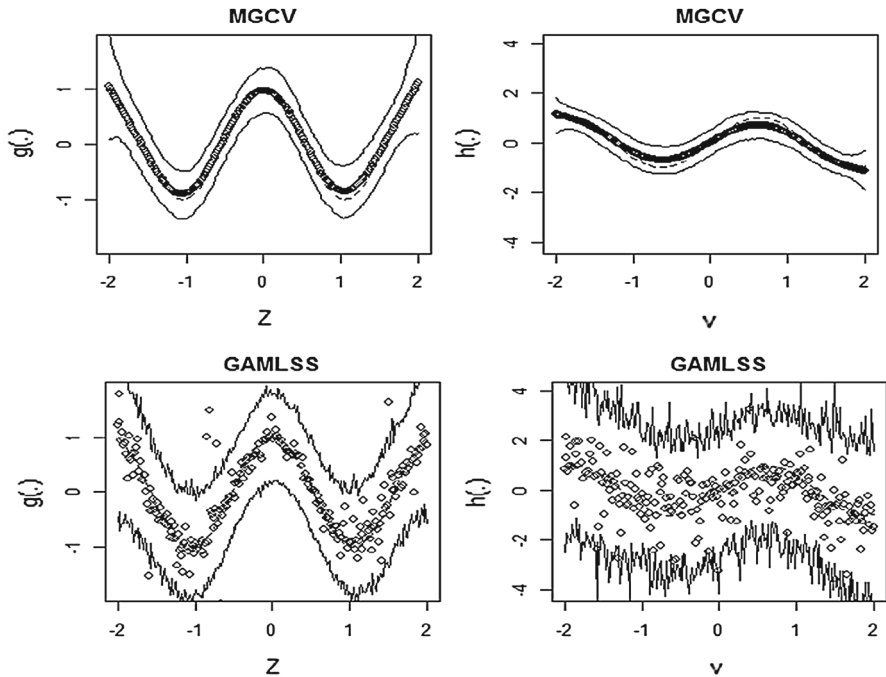
Fig. 12 Design 4, 90 % simulation intervals for sample size $n = 250$ for g on the *left*, h on the *right*

whereas in the EB approach they are assumed to be unknown parameters. Let us consider these procedures in the specific case of varying-coefficient models which can be divided in three subgroups: models with continuous variables as effect modifiers, models with spacial coordinates, and models with unordered group indicators (say, random slopes). For the inferences based on the FB or on the EB approach, the three types are implemented in the same way. However, in our study we are more concerned with continuous variables. This case can be estimated in (at least) three different ways: first `rw1` or second `rw2` order random walk, P-spline with first `psplinerw1` or second `psplinerw2` order random walk penalty—or with a 'seasonal prior'. For the case of doing inference based on a penalized least squares approach, two methods are implemented for continuous regressors: one considering P-splines with first or second order difference penalty, and one with zero degree P-splines. For non-Bayesian statisticians and therefore for almost all empirical economists it is very hard to see how to choose between this impressive set of options, and to set the parameters. Results certainly also depend on the prior choice and simulation method chosen. Furthermore, without going into the source codes, on can only obtain figures with Bayesian confidence intervals but no predictions at given points, nor numerical results on these. All this makes it a bit difficult to compare this package with the above mentioned methods. It would be, however, interesting to study this direction deeper as some small simulation studies (not presented here) seem to indicate a quite competitive performance.

Table 5 Results calculated from 500 samples drawn from Design 4

Sample size	Package	$ABias[\hat{g}(.)]$	$ABias[\hat{h}(.)]$	$AMSE[\hat{g}(.)]$	$AMSE[\hat{h}(.)]$
50	mgcv	0.0262	$-9.2227\mathrm{e}{-17}$	0.0390	0.3151
	gamlss	0.0233	0.4281	2.4170	6.0226
100	mgcv	0.0364	$8.9392\mathrm{e}{-17}$	0.0128	0.1210
	gamlss	-1.0389	-3.1684	105.6606	1072.2730
250	mgcv	0.0425	$-9.7946\mathrm{e}{-17}$	0.0056	0.0321
	gamlss	-0.0598	-0.9288	1.6459	178.4230

5 Conclusion

First the problem of endogenous coefficients is presented. For them we show the failure of the OLS and liner IV approaches when estimating an average effect or return. The use of varying-coefficient models is suggested because of their potential to handle heterogeneous returns. The enable us to directly model the coefficients as functions. This is partly underlined by some examples from empirical economics. We have explicitly shown the inadequacy of traditional econometric inference to estimate an average effect when the coefficients are endogenous. We then propose alternative modeling approaches by using varying-coefficient models. This can be extended to IV estimation to overcome the problem of remaining endogeneity.

Some simulation studies were performed to first show the failure of classical methods, and in Sect. 4 to compare the performance of existing varying-coefficient model estimators. More specifically, we have presented several R packages commonly available to implement the discussed varying-coefficient models.

This article has highlighted the potentiality of varying-coefficient models to estimate individual or average effects of regressors while usual econometric approaches fail because of considering endogenous coefficients. Regarding existing estimation methods, we have provided an overview of the literature on estimators for varying-coefficient models and their available implementations in R. At this point, for the sake of generality, we have proposed solutions to deal with endogenous coefficients in several specifications. Nevertheless, we hope that our contribution will motivate further researches on this track.

References

Allais M (1953) Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. Econometrica 21(4):503–546

Belitz C, Brezger A, Kneib T, Lang S, Umlauf N (2013) `BayesX`: Software for Bayesian Inference in Structured Additive Regression Models. Version 2.1

Branca M (2013) An analysis of implementations of varying-coefficient models. Master Thesis, University of Geneva, Switzerland

Cai Z, Fan J, Li R (2000) Efficient estimation and inferences for varying-coeffcient models. J Am Stat Assoc 95(451):888–902

Card D (1995) Using geographic variation in college proximity to estimate the return to schooling. In: Christofides LN, Vanderkamp J, Grant EK, Swidinsky R (eds) Aspects of labour market behavior: essays in honour of John Vanderkamp. University of Toronto Press, Toronto, pp 201–222

Chiang C-T, Rice JA, Wu CO (2001) Smoothing spline estimation for varying coeffcient models with repeatedly measured dependent variables. J Am Stat Assoc 96(454):605–619

Crossley TF, Pendakur K (2010) The commonscaling social cost-of-living index. J Bus Econ Stat 28(4):523–538

Deaton A, Muellbauer J (1980) Economics and consumer behavior. Cambridge University Press, Cambridge

Diewert WE (1998) Index number issues in the consumer price index. J Econ Perspect 12(1):47–58

Fahrmeir L, Kneib T, Lang S (2004) Penalized structured additive regression for space-time data: a bayesian perspective. Stat Sin 14:715–745

Fan J, Zhang W (1999) Statistical estimation in varying coefficient models. Ann Stat 27(5):1491–1518

Fan J, Zhang W (2000) Simultaneous confidence bands and hypothesis testing in varying-coefficient models. Scand J Stat 27(4):715–731

Fan J, Zhang W (2008) Statistical methods with varying coefficient models. Stat Interface 1(1):179–195

González Manteiga W, Lombardía MJ, Martínez Miranda MD, Sperlich S (2013) Kernel smoothers and bootstrapping for semiparametric mixed effects models. J Multivar Anal 114:288–302

Hastie T, Tibshirani R (1990) Generalized additive models. Chapman and Hall, London

Hastie T, Tibshirani R (1993) Varying-coeffcient models. J R Stat Soc Ser B (Methodol) 55(4):757–796

Hayfield T, Racine JS (2008) Nonparametric econometrics: the np package. J Stat Softw 27(5):1–32

Imbens GW, Angrist JD (1994) Identification and estimation of local average treatment effects. Econometrica 62(2):467–475

Kneib T, Fahrmeir L (2006) Structured additive regression for multicategorical spacet-ime data: a mixed model approach. Biometrics 62:109–118

Mammen E, Nielsen JP (1999) The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. Ann Stat 27:1443–1490

Mammen E, Nielsen JP (2003) Generalised structured models. Biometrika 90(3):551–566

Newey WK, Powell JL, Vella F (1999) Nonparametric estimation of triangular simultaneous equations models. Econometrica 67(3):565–603

Park BU, Mammen E, Lee YK, Lee ER (2015) Varying coefficient regression models: a review and new developments. Int Stat Rev 83(1):36–64

Pendakur K, Sperlich SA (2010) Semiparametric estimation of consumer demand systems in real expenditure with partially linear price effects. J Appl Econom 25:420–457

Pendakur K, Scholz M, Sperlich SA (2010) Semiparametric indirect utility and consumer demand. Comput Stat Data Anal 54:2763–2775

Profit S, Sperlich S (2004) Non-uniformity of jobmatching in a transition economy—a nonparametric analysis for the Czech Republic. Appl Econ 6(7):695–714

R Development Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rigby RA, Stasinopoulos DM (2005) Generalized additive models for location, scale and shape (with discussion). Appl Stat 54:507–554

Roca-Pardinas J, Sperlich SA (2010) Feasible estimation in generalized structured models. Stat Comput 20:367–379

Roy R (1947) La Distribution du Revenu Entre Les Divers Biens. Econometrica 15:205–225

Severance-Lossin E, Sperlich S (1999) Estimation of derivatives for additive separable models. Statistics 33(3):241–265

Sperlich S (2009) A note on nonparametric estimation with predicted variables. Econom J 12:382–395

Sperlich S, Tjøstheim D, Yang L (2002) Nonparametric estimation and testing of interaction in additive models. Econom Theory 18:197–251

Telser L (1964) Iterative estimation of a set of linear regression equations. J Am Stat Assoc 59:845–862

Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J R Stat Soc (B) 73(1):3–36

Wood SN (2006) Generalized additive models: an introduction with R. Chapman and Hall/CRC, London

Wood SN (2004) Stable and efficient multiple smoothing parameter estimation for generalized additive models. J Am Stat Assoc 99(467):673–686

Wood SN (2003) Thin-plate regression splines. J R Stat Soc (B) 65(1):95–114

Wood SN (2000) Modelling and smoothing parameter estimation with multiple quadratic penalties. J R Stat Soc (B) 62(2):413–428

Yang L, Park BU, Xue L, Härdle W (2006) Estimation and testing for varying coefficients in additive models with marginal integration. J Am Stat Assoc 101:1212–1227