

# Essays on structured additive regression models with applications in development economics

Juan Armando Torres Munguía

## Abstract

Structured additive regression models are a particular class of models that provide a flexible framework to deal with a wide class of effects, including linear, nonlinear, random, spatial, and interaction effects, which enables the specification of more complex but more realistic models.

The goal of this dissertation is to use these models to address practical issues in three relevant topics in the field of development economics. First, a Gaussian model is used to study gendered inequalities in time allocation to unpaid housework among partnered women and men. In the second study, we are confronted with the problem of identifying the risk factors associated with emotional intimate partner violence, for which a probit model is used. In the third study, quantile models are applied to examine heterogeneous gendered effects of a set of risk factors associated with the income-to-poverty ratio of the poor and extremely poor families.

Given the complex structure of the models used in the three abovementioned cases, an estimation cannot be computed by traditional inference techniques. To overcome this issue, it is implemented a three-step strategy consisting on the use of the boosting algorithm, complementary pairs stability selection with per-family error rate control, and the calculation of pointwise bootstrap confidence intervals.

From a statistical standpoint, the methodology helps to overcome common issues in regression in development economics, such as dealing with different types of response variables, the inclusion of potential nonlinear (or even a priori unknown) effects of continuous covariates on the response, select the relevant variables at their most suitable functional form, dealing with hierarchical data, to account for spatially correlated observations, to introduce complex interaction effects, and to avoid multicollinearity.

From an empirical perspective, the method applied allows to illustrate how the utilization of the structured additive models contributes to enhancing knowledge on these phenomena by providing new relevant insights on the matter. Findings in the three studies not only yield evidence about significant covariates that were either hitherto unknown, understudied, or that have not yet been tested empirically, but they are also relevant for the design of public policies, such as the identification of the relevance of the individual, household, communities, and regional factors in these studies, the existence of age-varying effects, the determination of the circumstances in which women and men face particular disadvantages, and the identification of some specific risk subgroups of the population that are generally overlooked.

## Zusammenfassung

Strukturiert additive Regressionsmodelle sind eine bestimmte Klasse von Modellen, die einen flexible Struktur für den Umgang mit verschiedenartigen Kovariableneffekten bietet, einschließlich linearer, nichtlinearer, zufälliger, räumlicher und Interaktionseffekte, was die Spezifikation von komplexerer, aber wirklichkeitsgetreuer Modelle ermöglicht.

Das Ziel dieser Dissertation ist es, diese Modelle zu nutzen, um praktische Fragestellungen in drei relevanten Themenfeldern der Entwicklungsökonomie zu untersuchen. Zunächst wird ein Normalverteilungsmodell verwendet, um geschlechtsspezifische Ungleichheiten bei der Zeitverwendung von Frauen und Männer in Partnerschaft für unbezahlte Hausarbeit. In der zweiten Studie werden wir mit dem Problem konfrontiert, die Risikofaktoren emotionaler Partnergewalt zu identifizieren, für die ein Probit-Modell verwendet wird. In der dritten Studie werden Quantilmodelle angewendet, um heterogene geschlechtsspezifische Auswirkungen einer Reihe von Risikofaktoren zu untersuchen, die mit dem Verhältnis von Einkommen zu Armut in armen und extrem armen Familien verbunden sind.

Angesichts der komplexen Struktur der Modelle, die in den drei oben genannten Fällen verwendet werden, kann eine Schätzung nicht durch herkömmliche Inferenztechniken berechnet werden. Um dieses Problem zu lösen, wird eine dreistufige Strategie implementiert, die aus der Verwendung des Boosting-Algorithmus, der complementary pairs stability selection mit per-family error rate control und der Berechnung von punktweisen Bootstrap-Konfidenzintervallen besteht.

Aus statistischer Sicht hilft die Methodik dabei, häufige Probleme bei der Regression in der Entwicklungsökonomie zu überwinden, z.B. verschiedene Arten von Zielgrößen, Auswahl der relevanten Variablen in ihrer am besten geeigneten funktionalen Form, Umgang mit hierarchischen Daten und räumlich

korrelierte Beobachtungen zur Berücksichtigung, komplexe Interaktionseffekte einzuführen und Multikollinearität zu vermeiden.

Aus empirischer Sicht ermöglicht die angewandte Methode darzustellen, wie die Nutzung der strukturierten additiven Modelle dazu beiträgt, das Wissen über diese Phänomene zu erweitern, indem sie neue relevante Erkenntnisse zu diesem Thema liefern. Die Ergebnisse der drei Studien geben nicht nur Hinweise auf signifikante Kovarianzen, die entweder bisher unbekannt, zu wenig untersucht oder noch nicht empirisch getestet wurden und gleichzeitig für die Gestaltung öffentlicher Maßnahmen relevant sind, beispielsweise für die Ermittlung der Relevanz von Einzel-, Haushalts-, Gemeinde- und regionale Faktoren in diesen Studien, das Vorhandensein von altersabhängigen Effekten, die Bestimmung der Umstände, unter denen Frauen und Männer besonderen Benachteiligungen ausgesetzt sind, und die Identifizierung einiger spezifischer Risikountergruppen der Bevölkerung, die im Allgemeinen übersehen werden.