Institut für Lebensmittel- und Ressourcenökonomik der
Rheinischen Friedrich-Wilhelms-Universität zu Bonn

Econometric specification of constrained optimization models

I n a u g u r a l - D i s s e r t a t i o n

zur

Erlangung des Grades

Doktor der Agrarwissenschaften
(Dr.agr.)

der

Hohen Lanwirtschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität

zu Bonn

vorgelegt am 3. Mai 2007

von Torbjörn Jansson

aus Västervik, Schweden

Referent: Prof. Dr. Thomas Heckelei

Korreferent: Prof. Dr. Klaus Frohberg

Tag der mündlichen Prüfung: Freitag 29. Juni, 2007

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn elektronisch publiziert (http://hss.ulb.uni-bonn.de/diss_online)

Erscheinungsjahr: 2007

Zusammenfassung

Gegenstand dieser Untersuchung ist die ökonometrische Spezifikation von Parametern beschränkter Optimierungsmodelle mit Schwerpunkt auf Fragen, welche dann auftreten (i) wenn Ungleichheitsbeschränkungen involviert sind und/oder (ii) wenn das Schätzproblem unterbestimmt ist oder die Daten aus verschiedenen Quellen stammen. Die übergeordnete Methode, die hier angewendet wird, ist die direkte Schätzung der Optimalitätsbedingungen des jeweiligen Optimierungsmodells zusammen mit zusätzlichen Gleichungen zur Miteinbeziehung von A-Priori-Informationen.

Wenn das Optimierungsmodell Ungleichungen beinhaltet, dann sind Komplementaritätsbedingungen in den Optimalitätsbedingungen enthalten, die zu einem nicht-konvexen Lösungsraum und zu nicht-stetigen Ableitungen des Schätzproblems führen. Ein Problem solcher Art ist zumeist schwer lösbar. In solchen Fällen ist es vorteilhaft, die Schätzung als ein Optimierungsproblem in zwei Ebenen (Bi-Level Programming Problem, BLPP) zu betrachten, wobei das Problem der oberen Ebene darin besteht, das Schätzkriterium zu optimieren, und die untere Ebene das zu schätzende Modell darstellt. Die BLPP-Litteratur bietet mehrere Lösungsalgorithmen an, die dazu geeignet sind, Probleme mit Komplementaritätsbedingungen zu lösen. In dieser Dissertation wird gezeigt, wie durch deren Einsatz die Effizienz von Schätzungen verbessert werden kann.

Wenn das Schätzproblem unterbestimmt ist, dann müssen zusätzliche Informationen zugeführt werden um die Parameter zu identifizieren. Im vergangenen Jahrzehnt wurde dies häufig mittels so genannten Entropie-Schätzern erreicht. Eine allgemeinere und transparentere Methode zur Einbindung der Zusatzinformation, welche zugleich einfacher zu berechnen sein kann, basiert auf der Erfassung des Schätzproblems als Bayes'sche Schätzung. Die Zusatzinformationen werden als A-Priori-Wahrscheinlichkeitsverteilungen der Parameter definiert, und die Punktschätzung wird mit Hilfe der A-Posteriori-Verteilung, abgeleitet vom Satz von Bayes, ermittelt. Der vorgestellte Bayes'sche Ansatz bietet sich auch an, um Daten unterschiedlicher Quellen in einer theoretisch einheitlichen Schätzung zu nutzen.

Die Dissertation hat vier Hauptkapitel (Kapitel zwei bis fünf). In den beiden ersteren wird das Problem, die Handelskosten in einem räumlichen Preis-Gleichgewichtsmodell zu schätzen, aus jeweils methodologischer und empirischer Sicht behandelt, wobei BLPP zum Einsatz gebracht wird, um die Ungleichheitsbedingungen zu handhaben. In den beiden letzten Kapiteln wird eine Bay-

es'sche Methode, unterbestimmte Schätzprobleme zu lösen, vorgestellt und für die Schätzung der Parameter eines landwirtschaftlichen Angebotsmodells eingesetzt.

In Kapitel zwei wird die Schätzung der Parameter eines traditionellen Transportmodells diskutiert, und eine neue Schätzmethode vorgestellt. In Gegensatz zu anderen Methoden werden Beobachtungen von Handelskosten und von regionalen Preisen verwendet. BLPP wird eingesetzt, um ein quadratisches Schätzkriterium zu minimieren, unter der Nebenbedingung, dass die Kuhn-Tucker-Bedingungen des Transportmodells erfüllt sind. Mit Hilfe von Monte-Carlo-Simulationen werden einige Eigenschaften der Schätzer abgeleitet, und mit den einer traditionellen Kalibrierungsmethode dieses bekannten Modells verglichen. Die Analyse ergibt, dass mit der vorgeschlagenen Methode Preise und Handelskosten effizienter geschätzt werden können. Der Ansatz scheint für ein breites Spektrum linearer und quadratischer Modelle anwendbar zu sein.

Kapitel drei behandelt die Schätzung von regionalen Preisen, Überschussnachfragen und Handelskosten homogener Güter in einem räumlichen, partiellen Preis-Gleichgewichtsmodell und ist eine empirische Anwendung des im zweiten Kapitel vorgestellten Konzeptes. Das geschätzte Modell bezieht sich auf zwölf Regionen in Benin, die in sieben Produkten miteinander (Netto-)Handel treiben. Die zur Verfügung stehenden Preis- und Mengendaten sind mit beachtlichen Unsicherheiten behaftet. Wie in Kapitel zwei wird auch in diesem Kapitel die Schätzung als ein Optimierungsproblem in zwei Ebenen aufgestellt und gelöst, wobei auf der oberen Ebene die Summe der gewichteten quadrierten Abweichungen der geschätzten von den beobachteten Werten, und auf der unteren Ebene die Summe der Handelskosten im Transportmodell minimiert werden. Die Handelskosten werden in einer nichtlinearen Funktion, deren Parameter zu bestimmen sind, von einer Entfernungsmatrix abhängig gemacht. Die geschätzten Handelskosten werden mit entsprechenden Ergebnissen anderer Studien verglichen und als akzeptabel eingestuft. Die Handelskostenfunktion impliziert einen linearen Zusammenhang zwischen Entfernung und Handelskosten, mit einer Proportionalitätskonstante von 0.147 FCFA/kg/km (FCFA ist die Beninsche Währung, 1 USD entsprach in 2002 etwa 700 FCFA).

Kapitel vier stellt eine Bayes'sche Alternative zu Generalized Maximum Entropy- (GME) und Generalized Cross Entropy- (GCE) Lösungen unterbestimmter Gleichungssysteme vor. In vielen Fällen bietet der vorgeschlagene Ansatz eine völlig äquivalente Alternative zu GME- und GCE-Techniken. Er liefert aber in seiner allgemeinen Form eine direktere und einfacher zu interpretierende Möglichkeit zur Einbindung von A-Priori-Informationen und kann zudem auch den erforderlichen Rechenaufwand erheblich verringern. Die Methode lässt sich auch auf den Fall einer nicht-informativen (d.h. gleichverteilten) A-Priori-

Information erweitern. In dem Kapitel werden sechs illustrative Rechenbeispiele ausführlich dargestellt.

Kapitel fünf liefert eine umfassende Anwendung der in Kapitel vier vorgestellten Methoden. Der Bayes'sche Schätzer wird eingesetzt, um die im Angebotsteil des Modellsystems CAPRI enthaltenen Verhaltensparameter zu bestimmen. Das Angebotsmodell in CAPRI besteht aus rund 250 nichtlinearen Optimierungsmodellen, wobei jedes Modell den landwirtschaftlichen Sektor einer europäischen Region abbildet. Die Modelle haben eine quadratische Zielfunktion, wobei die quadratischen Terme das Simulationsverhalten des Modells stark beeinflussen. Das Ziel ist, diese Parameter mit Hilfe der in der CAPRI-Datenbank vorhandenen Daten zu schätzen. Insgesamt wurden Parameter für bis zu 23 Produktionsaktivitäten mit zugehörigen Preisen, variablen Kosten und Verhaltensfunktionen in 165 Regionen der EU-15 geschätzt. Es gibt mehrere Studien, die Ergebnisse auf Mitgliedsstaatebene und/oder für aggregierte Produkte (Beispielsweise Getreide und Ölsaaten) präsentieren. Die Ergebnisse in diesem Kapitel werden systematisch mit den Ergebnissen vier solcher Studien hinsichtlich Angebotselastizitäten verglichen, darunter zwei Studien für Frankreich, eine für die Niederlande und eine für Dänemark. Für aggregierte Produkte auf Mitgliedsstaatsebene sind die Ergebnisse gut vergleichbar mit denen anderer Studien und mit gängigen Faustzahlen. Zum Beispiel ergibt die Schätzung eine Angebotselastizität von einem Aggregat von üblichen Getreidearten in Frankreich von 0.508, und eine von Ölsaaten von 0.807. Auf regionaler Ebene und für einzelne Produkte sind die geschätzte Elastizitäten mehr heterogen. In der Französischen Region (NUTS 2) die in der Analyse als Fallbeispiel dient, fallen die Elastizitäten zwischen 0.38 (Kartoffeln) und 7.9 (Hartweizen). Ob diese Spanne realistisch ist, ist schwierig zu beurteilen, da dem Autor keine andere Publikation mit Schätzergebnissen von vergleichbarem Produktspektrum und regionaler Deckung bekannt sind

Summary

The subject of this thesis is econometric specification of parameters of constrained optimization models, with special attention to issues that arise when (i) inequality constraints are involved and/or (ii) when the estimation problem is ill-posed (underdetermined) or data come from diverse sources. The general approach followed here is to estimate directly the optimality conditions of the optimization model, together with additional equations for including prior information.

If inequality constraints are involved, the optimality conditions will contain complementary slackness conditions, making the space of solutions non-convex with discontinuous derivatives. The extremum estimation problem may then be very difficult to solve. In such cases, the estimation profits from being viewed as a bilevel programming problem (BLPP), where the leader’s problem is to optimize the estimation criterion function, whereas the follower’s problem is the optimization model whose parameters are to be estimated. The BLPP literature offers several solution algorithms capable of handling the complementary slackness conditions, and in this thesis it is shown how the efficiency of estimators may be increased if they are used.

Obtaining a point estimate of the unidentifiable parameters in an ill-posed problem requires additional information. Common practice among applied modellers during the last decade has been to introduce this information using entropy estimators. A more general, more transparent and potentially computationally simpler means to the same end is to cast the estimation in a Bayesian form. The required additional information is defined in terms of a prior probability distribution of the parameters, and the estimation is based on the posterior probability density function which can be found using Bayes's rule. The Bayesian approach also proves useful for utilizing data from heterogeneous data sources in a theoretically sound way.

The thesis has four main chapters (chapters two to five). The chapters two and three are a methodological and an empirical approach respectively to the problem of estimating trade costs in a spatial price equilibrium model using bilevel programming techniques to handle inequality constraints. Chapters four and five introduce a Bayesian approach to estimation of ill-posed problems, and apply a Bayesian estimator to the problem of estimating parameters of an agricultural supply model.

Chapter two discusses the estimation of parameters of a traditional transportation model, and proposes a new estimation method. In contrast to previously used methods, observations of regional prices *as well as* of trade costs are used. The proposed method uses bilevel programming to minimize a weighted least squares criterion under the restriction that the estimated parameters satisfy the Kuhn-Tucker conditions for an optimal solution of the transport model. Monte-Carlo simulations are used to trace out some properties of the estimator and compare it with a traditional calibration method. The analysis shows that the proposed technique estimates prices as well as trade costs more efficiently. The approach appears to be applicable to a wide range of linear and quadratic models.

Chapter three also treats the estimation of regional prices, excess demand and trade costs, for homogeneous products in a spatial partial price equilibrium model, and serves as an empirical application of the theoretical approach developed in chapter two. The estimation is applied to a model for Benin, where twelve market regions are bilaterally trading (net trade) in seven products. The available data is subject to considerable uncertainty both regarding prices and quantities. The estimation is again formulated as a bilevel program, with the upper level objective to minimize the weighted sum of squared deviations of estimated from observed values of prices and excess demand. The estimation is restricted to optimal solutions of the transport cost minimization problem, parametrized by a trade cost function, the parameters of which are also to be determined. The resulting trade cost estimates are compared to those of empirical studies, and are found to be within an acceptable range. Trade costs are found to have a distance elasticity of unity, and are thus linear, with a coefficient of proportionality of 0.147 FCFA/kg/km (FCFA is the currency in Benin, 1 USD ≈ 700 FCFA in 2002)

Chapter four presents a Bayesian alternative to Generalized Maximum Entropy (GME) and Generalized Cross Entropy (GCE) solutions to underdetermined systems of equations. For certain types of economic model specifications, this approach provides fully equivalent results to GME-GCE techniques, but in its general form allows a more direct and straightforwardly interpretable formulation of available prior information and can reduce significantly the computational effort involved in finding solutions. The technique can also be extended to situations with uninformative (uniform distributed) prior information. The chapter provides six fully worked out illustrative numerical examples of the proposed estimator.

Chapter five provides a large-scale application of the methods proposed in chapter four. The estimator is applied to the supply part of the agricultural sector model CAPRI in order to estimate the behavioural parameters embedded there. The supply model in CAPRI consists of around 250 constrained quadratic programming models, where each model represents the agricultural sector in a region in the EU. The models have a quadratic objective function, where the quadratic term influences the simulation behaviour of the model. It is the objective to esti-

mate the parameters of that quadratic term, using the time series data in the CAPRI database. After discarding regions with insufficient data, parameters for up to 23 crop production activities with related inputs, outputs, prices and behavioural functions were estimated for 165 regions in EU-15. There are several studies available that publish supply elasticities for individual countries and major crop aggregates. The results are systematically compared to the outcomes of four such studies, of which two for France, one for the Netherlands and one for Denmark. For crop aggregates (e.g. cereals, oilseeds etc.) on the level of nations, the estimated own price elasticities of supply are found to be in a plausible range. For example, the supply elasticity of common cereals in France is found to be 0.508 and that of oilseeds 0.807. On a regional level and for individual crops, the picture is much more diverse. In a French case-study region (NUTS 2), the supply elasticities ranged between 0.38 (potatoes) and 7.9 (durum wheat). Whether this span is plausible or not is difficult to judge, since no other study of similar regional and product coverage is known to the author.

Table of contents

Chapter 1 Introduction

1. Aim and motivation

The aim of this dissertation is to develop and demonstrate a general framework for the econometric specification of constrained optimization models. Focus is on issues that arise when inequality constraints are involved and/or when the estimation problem is ill-posed (underdetermined) or data come from diverse sources.

There are at least two good reasons for working on a general approach to this subject. One reason is that traditionally, parameters of larger optimization models are gathered from different sources. Take as an example a stylized economic simulation model that contains two types of parameters that we call *supply elasticities* (behavioural parameters) and *technical coefficients*. Assume that supply elasticities are estimated separately or taken from other studies, whereas technical coefficients come from engineering information or external estimations. Thus, the parameters are determined in two steps, each of which may be performed following the state of the art in statistics. The error model used when estimating the behavioural parameters does not include errors on the technical coefficients, and vice versa, so that taken together, the two estimations are likely to be inefficient and inconsistent. Furthermore, in each step, the parameters were estimated using a different model than the one in which the parameters are to be used, and are thus not likely to be consistent with that model. One aim of this dissertation is to point out that such problems do arise and to propose a general way of avoiding them.

The second reason is that the full specification of large scale optimization models typically require some amount of information that is not contained in any data source. Examples of such information can be that certain coefficients have certain signs, or that they should be within some specified range that is based on heuristics. Some parameters of the desired model may not be identifiable at all in the underlying data, implying that there exist many different (generally a continuum of) parameter vectors that all result in models consistent with data but possibly different simulation behaviour. If the researcher wishes to proceed with the chosen model formulation in underdetermined cases, she faces the problem of selecting one of the possible parameter vectors for her model; a task frequently solved using out-of-sample information. Inclusion of such prior information is another core subject of this dissertation.

The first reason is related to heterogeneity of data sources and consistency between the estimating equations and the final model, whereas the second refers to

prior information about the *parameters* of the model. Depending on the estimation approach, however, that distinction blurs, and it is indeed one proposition of this thesis that a proper estimation approach can address both issues simultaneously.

2. Overview of methodological approach

The general approach to the estimation of constrained optimization models proposed in this dissertation is found in the intersection of the three different disciplines (i) Bilevel programming (ii) Errors in variables modelling and (iii) Bayesian estimation. In this section, a general constrained optimization model is constructed, and it is shown how each of the three disciplines contributes to the estimation of the model. This section is brief and only serves as an overview, whereas the methodological approach is developed at length in the subsequent chapters.

2.1. A constrained optimization model

The problem is to estimate the parameter vector $\boldsymbol{\psi}$ of the following general constrained optimization problem:

$$\max_{\mathbf{x}|\boldsymbol{\psi}} \; f(\mathbf{x},\boldsymbol{\psi}): \mathbf{g}(\mathbf{x},\boldsymbol{\psi}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x},\boldsymbol{\psi}) = \mathbf{0} \qquad (1)$$

Here, and throughout the text if not stated otherwise, a bold face letter generally indicates a vector or matrix, whereas scalars are denoted by regular italic letters. We could thus write $\mathbf{x} = \{x_j: j = 1,\ldots,J\}$. The symbols $\mathbf{x}|\boldsymbol{\psi}$ beneath the directive "max" means that the function $f(\mathbf{x},\boldsymbol{\psi})$ is being maximised with respect to $\mathbf{x}$ while treating $\boldsymbol{\psi}$ as exogenous. The general problem can have inequality constraints $\mathbf{g}$ as well as equality constraints $\mathbf{h}$. Further notation is introduced along the way.

We constrain the study to cases where f is twice continuously differentiable and the feasible space S defined by the constraints for any $\boldsymbol{\psi}$, i.e. $S(\boldsymbol{\psi}) = \{\mathbf{x}: \mathbf{g}(\mathbf{x},\boldsymbol{\psi}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x},\boldsymbol{\psi}) = \mathbf{0}\}$ is compact. Furthermore, all elements of $\mathbf{x}$ and $\boldsymbol{\psi}$ are required to be real continuous variables, thus excluding integer programming problems.

The vector $\mathbf{x}$ is generally referred to as the *variables* or *endogenous*, whereas $\boldsymbol{\psi}$ is called the *parameters* or *exogenous*. Those terms, however, depend upon the context, forcing us to leave that convention at times. Take for example the estimation of the vector $\boldsymbol{\Psi}$ using the equation system $\mathbf{F}(\mathbf{x},\boldsymbol{\Psi}) = \mathbf{0}$. When solving the equation system, $\boldsymbol{\Psi}$ is generally termed "parameter" and $\mathbf{x}$ "variable". In the context of this thesis, the estimation itself is explicitly considered an optimization problem in the variable $\boldsymbol{\theta} = (\mathbf{x},\boldsymbol{\Psi})$, possibly parametrized by yet another vector (e.g. weights or probability distribution parameters), so that the vector $\boldsymbol{\Psi}$ sometimes must be referred to as variable and sometimes as parameter.

2.2. A bilevel programming perspective

A bilevel programming problem (BLPP) is an optimization model that has another optimization model in its constraints. A common instance is the case of a *leader* and one or many *followers*. Without loss of generality, we assume that there is only one follower. The leader is in control of some instrument vector $\boldsymbol{\psi}$ which the follower takes as given. The leader seeks to maximise some function F that depends on $\boldsymbol{\psi}$ and the vector of choices $\mathbf{x}$ of the follower, and possibly on a vector of parameters $\boldsymbol{\Theta}$. The leader may be constrained by vectors of inequalities $\mathbf{G}(\mathbf{x},\boldsymbol{\psi},\boldsymbol{\Theta}) \leq \mathbf{0}$ and equalities $\mathbf{H}(\mathbf{x},\boldsymbol{\psi},\boldsymbol{\Theta}) = \mathbf{0}$. The follower seeks to maximise a function f in which the instrument vector $\boldsymbol{\psi}$ of the leader is exogenous, subject to the vectors of inequality constraints $\mathbf{g}(\mathbf{x},\boldsymbol{\psi}) \leq \mathbf{0}$ and equality constraints $\mathbf{h}(\mathbf{x},\boldsymbol{\psi}) \leq \mathbf{0}$. The problem of the leader can then be written

$$
\begin{aligned}
&\max_{\mathbf{x},\boldsymbol{\psi}|\boldsymbol{\Theta}} && F(\mathbf{x},\boldsymbol{\psi},\boldsymbol{\Theta}) && \\
&\text{subject to} && \mathbf{G}(\mathbf{x},\boldsymbol{\psi},\boldsymbol{\Theta}) \leq \mathbf{0} && \\
& && \mathbf{H}(\mathbf{x},\boldsymbol{\psi},\boldsymbol{\Theta}) = \mathbf{0} && (2)\\
&\text{and } \mathbf{x} \text{ solves} && \max_{\mathbf{x}|\boldsymbol{\psi}} \quad f(\mathbf{x},\boldsymbol{\psi}) && \\
& && \text{subject to} \quad \mathbf{g}(\mathbf{x},\boldsymbol{\psi}) \leq \mathbf{0} && \\
& && \qquad\qquad\ \ \mathbf{h}(\mathbf{x},\boldsymbol{\psi}) = \mathbf{0} &&
\end{aligned}
$$

The similarities with an estimation of $\boldsymbol{\psi}$ can be readily seen: Let the leader be a researcher that seeks to estimate the parameter vector $\boldsymbol{\psi}$, and let the function $F(\mathbf{x},\boldsymbol{\psi})$ be the estimation criterion function. Then the BLPP (2) is interpreted as to find the parameters that, when inserted into the model, gives model outcomes that maximise the criterion function. Observations that are somehow related to $\mathbf{x}$ or $\boldsymbol{\psi}$ by some data sampling process may be entered as constants in the objective function F or in the constraints $\mathbf{G}$ or $\mathbf{H}$ via the parameter vector $\boldsymbol{\Theta}$.

The model (2) belongs to a special class of BLPPs that is sometimes called the *weak* or *optimistic* case (Dempe, 1997). That class is characterised by the property that if there are several optimal solutions to the followers problem for a given $\boldsymbol{\psi}$, then the leader may choose the one he prefers and the follower is indifferent to that choice. Or, in terms of an estimation, if the followers problem has several solutions for a given parameter vector, we may choose any of those.

The follower's problem is also referred to as the *inner problem* and the leader's problem as the *outer problem*. In this work, those terms are sometimes used instead of the terms leader/follower, because they are perceived to be more neutral. The terms leader and followers are useful too, however, because they show the relation to the vast literature on leader-follower games originating with Stackelberg (1934).

In general, the model (2) is very difficult to solve (Luo, Pang and Ralph, 1996), and gradient based solution methods are bound to find only local optima. When such models arise in econometrics, common practice is to either apply ordinary solution techniques and ignore the possible existence of "better" solutions[1] (e.g. Heckelei and Wolff, 2003 pp. 44), or to determine the binding status of the complementary slackness conditions by some tailor-made algorithm (e.g. Fischer et al. 1988) or heuristic that effectively turns inequality constraints into equalities (e.g. the budget constraint of the consumer is always exhausted). That such solutions may be appropriate in the cases where they are applied is beside the point. An advantage of viewing the estimation problem as bilevel programming problem is that much research has been done on solution methods for such problems. In this study we make use of the easy to implement techniques proposed in Ferris, Dirkse and Meeraus (2002). A comprehensive but now somewhat dated literature review of bilevel programming is found in Vicente and Calamai (1994).

When the inner problem is replaced by its first order conditions, we obtain (in general) a single level programming problem that contains complementary slackness conditions among its constraints. Such problems are studied within the field of mathematical programming with equilibrium constraints (MPEC), see for example the textbook by Luo, Pang and Ralph (1996).

The proposition to use bilevel programming techniques to solve estimation problems does not *per se* imply doing something entirely different or new. Many econometric problems solved in literature are implicitly BLPPs. Especially, all extremum estimators of equations derived from neoclassic models should belong to that class. Consider for example the estimation of parameters of a demand system. A common way of proceeding includes setting up a microeconomic model of the agent (the inner problem or the follower), and from that model deriving a reduced form system of demand equations. The demand equations are then related to observations of prices and quantities by some error model, and a parameter vector is selected that minimizes (or maximizes) some estimation criterion. The whole process is then equivalent to the solution of a BLPP.

The value of viewing estimations of parameters of constrained programming models as BLPPs is that it provides an alternative perspective of the problem. BLPP provides a toolbox for handling situations where the road appears closed to ordinary statistical methods.

[1] The phrase "better solutions" means that there may exist another local optimum that is better than the one found. In the work of Heckelei and Wolff (2003) referred to in the text, the authors do not report any numerical problems in the estimations.

2.3. A measurement errors perspective

A common characteristic of estimations of economic models is that there are generally few controlled variables, i.e. variables known with certainty, for instance by experimental design. For an economic model such controlled variables can in some cases be political intervention measures, like subsidies and tariffs. Nevertheless, important items like prices, supply, demand, technical coefficients *et cetera* can in general only be inferred from error prone measurements. In fact, the parameters of interest can frequently not be measured *at all*, but we can instead measure some quantity that we believe is *correlated* with the parameter of interest via an error model.

Take for instance prices in some production model. The producer is generally assumed to base his production decision on an *expected* price. So even if we actually could measure the price with very high precision, we do not necessarily know what price the agent was *expecting* to receive, but can only assume that there is some statistical relation between our measurements and the expected price, just as we may assume that there is one between observed quantities and true planned production.

In the situation described above, which should be quite common in the estimation of constrained optimization models, there are no "left hand side" and "right hand side" variables, but the model is regarded as an equation system where measurement errors can enter in several places. Such models are referred to as errors-in-variables models (EVM) or measurement error models. It is well known that failure to incorporate errors in the "explanatory variables" when they are really error prone leads to biased parameter estimates (e.g. Fuller, 1987). Despite the likely bias, estimations in literature frequently use error prone explanatory variables, e.g. prices to explain demanded quantities.

The EIV models that arise in the context of economic models generally contain *nuisance parameters*[2] i.e. items that enter the estimating equations but are of no direct interest to the researcher, and that furthermore tend to increase in number as the number of observations grows. For an explicit example, consider a model of a single economic price taking agent that is assumed to choose some optimal quantity x given a price p and some technical parameter β that is believed to be constant. In the notation of model (1), $\boldsymbol{\Psi} = (p,\beta)$. Assume for simplicity that there are no inequality constraints involved and that the functional forms are such that curvature is no issue (second order conditions are always satisfied), and furthermore that there is a time series containing n observations of prices and quantities. The producer is assumed to choose the optimal x in each period as a function of (p,β) so that we can write $x(p,\beta)$ as the (possibly implicit) solution to the first

[2] Nuisance parameters is the term used by Carroll et al. (1995). Zellner (1971) adopts the term "incidental parameters" introduced by Neyman and Scott (1948).

order conditions of (1) for x. Observations of prices and quantities are assumed to be generated as

$$x^{obs} = x(p,\beta) + u_1$$

$$p^{obs} = p + u_2$$

where u_1 and u_2 are measurement errors with standard deviations σ_1 and σ_2.

From an econometric perspective, p and β are parameters to estimate, and we see that for each new observation, another "true price" p must be estimated, albeit the parameter β may be the only parameter of interest for the modeler. Thus, regardless of sample size n, there are always $n + 1$ parameters to estimate, hence the term "nuisance parameters". Without further information, the parameters are not identifiable. A maximum likelihood estimation of the problem requires knowledge of the ratio of variances σ_1/σ_2 (e.g. Zellner, 1971, Fuller, 1987). If (p,β) are estimated with a Bayesian approach, exact knowledge of σ_1/σ_2 is not required, but other prior information may be used instead (Zellner, 1971, chapter V).

Taking the reasoning one step further, we may also acknowledge that not only the researcher (upon measuring the quantities of the model) can commit errors, but also the agent that is modelled. Such errors may result in non-optimal choices of the producer, which then are measured by the researcher and enter into the estimation. McElroy, (1987) terms a model that considers optimization as well as measurement errors a general error model, and points to the consequences of neglecting optimization errors.

By designing a measurement error model, a more realistic data sampling process can be obtained, and biases and inefficiencies reduced. Unfortunately, the more sophisticated the error model, the greater the data requirements, because the number of endogenous variables in the estimation is increased. The greater number of "unknowns" may even result in an underdetermined, or ill-posed estimation. The measurement error model also requires the researcher to determine the relative variances of the errors involved, either by assumption or using external datasets. Those potential obstacles raised by the measurement error model lead to the last methodological field, Bayesian estimation.

2.4. A Bayesian perspective

There are many situations in which applied modellers need to include prior information into the specification of model parameters: Applied modellers frequently face problems of scarce or missing data (e.g. Howitt and Reynaud, 2003), or want to include non-data information when determining parameters of their models in order to account for sound plausibility considerations (e.g. Fischer et al. 1988 p. 93). Furthermore, modellers may wish to specify a richer model structure than is supported by the data (e.g. Oude Lansink, 1999, Paris and Howitt, 1998), thus

facing ill-posed problems. In the previous section it was also mentioned that the measurement error model generally requires prior information of some kind.

The prior information required to solve ill-posed problems have frequently been introduced using entropy based estimators, as in Howitt and Reynaud (2003), Oude Lansink (1999) and Paris and Howitt (1998) referred to above (more references are found in chapter 4). The Bayesian approach to econometrics is directly aimed at handling prior information, and may thus prove a useful alternative to maximum entropy and cross entropy techniques for the applied modeller.

Let $\phi(\mathbf{z}|\boldsymbol{\theta})$ denote the conditional probability density for observing the outcome $\mathbf{z}$ of some random vector $\mathbf{Z}$ given the parameter vector $\boldsymbol{\theta} \in \Omega$, and let $\xi(\boldsymbol{\theta})$ be the unconditional probability density of $\boldsymbol{\theta}$. Using Bayes's rule (e.g. DeGroot, 1970), the following function (actually a family of functions) for the density of $\boldsymbol{\theta}$ conditional on observing $\mathbf{z}$ can be derived:

$$\xi(\boldsymbol{\theta}|\mathbf{z}) \propto \phi(\mathbf{z}|\boldsymbol{\theta})\xi(\boldsymbol{\theta})$$

$\xi(\boldsymbol{\theta}|\mathbf{z})$ is called the *posterior density*, and $\xi(\boldsymbol{\theta})$ the *prior density* of $\boldsymbol{\theta}$. Note that the parameter vector $\boldsymbol{\theta}$ is not to be confused with the parameter vector $\boldsymbol{\Psi}$. The latter are parameters of the inner problem, whereas the former are parameters of the data sampling model.

The Bayesian approach lets the researcher introduce prior information regarding the parameters, and can contribute to resolving the problems of transparent and consistent inclusion of diverse data sources, a common problem in the estimation of parameters of constrained optimization models. Different data sources and assumptions provide information about the prior density function ξ (e.g. prior mean and variance of $\boldsymbol{\theta}$), and confrontation of the model with data provides information about the outcomes $\mathbf{z}$. As an estimator we may want to choose the mode or mean of the conditional density function $\xi(\boldsymbol{\theta}|\mathbf{z})$. As is shown in chapter 4, the posterior mode estimator entails many other familiar estimators as special cases.

3. Outline of thesis

The methodological approach outlined above is implemented in four steps in the following four chapters, as is illustrated in the two-by-two matrix below. The first two chapters show how the bilevel programming perspective is useful for estimating inequality constrained models. In particular, the two chapters deal with the estimation of regional prices and transportation costs in a transportation model, first from a theoretical point of view and then applied to a real data set. The chapters four and five treat Bayesian methods for inclusion of prior information. In chapter four, a Bayesian estimator for general underdetermined models is proposed, and chapter five applies the methodology developed in chapter four to a

large scale nonlinear model. The models in all chapters can be regarded a kind of measurement error models. The chapters are based on individual papers, and are thus written in such a way that they may be read independent of each other.

Figure 1: Outline of body of thesis.

	Bilevel Programming	**Bayesian estimation**
Theoretical	Chapter 2	Chapter 4
Applied	Chapter 3	Chapter 5

Chapter two discusses the estimation of parameters of a traditional transportation model, as it is typically present in so-called Takayama-Judge type spatial price equilibrium (SPE) models (e.g. Takayama and Judge, 1964). In contrast to previously used estimation methods for this problem, observations of regional prices as well as of trade costs are used. The proposed method uses bilevel programming techniques to minimize a weighted least squares criterion under the restriction that the estimated parameters satisfy the Kuhn-Tucker conditions for an optimal solution of the transport model. A penalty function as proposed in Ferris et al. (2002) is used to iteratively approximate the complementary slackness conditions. Monte-Carlo simulations are used to trace out some properties of the estimator and compare it with a traditional calibration method. The analysis shows that the proposed technique estimates prices as well as trade costs more efficiently. It is suggested to apply the same method to a range of linear and quadratic models.

Chapter three also treats the estimation of regional prices, excess demand and trade costs, for homogeneous products in a spatial price equilibrium model, and serves to illustrate the application of the theoretical approach developed in chapter two. The estimation is restricted to optimal solutions of the transport cost minimization problem, parametrized by a trade cost function, the parameters of which are also to be determined. The data come from an agricultural model for Benin, where twelve market regions are bilaterally trading (net trade) in seven primary crop products. The resulting trade cost estimates are compared to those of empirical studies.

Chapter four presents a Bayesian alternative to Generalized Maximum Entropy (GME) and Generalized Cross Entropy (GCE) solutions to underdetermined systems of equations. For certain types of economic model specifications, this approach provides fully equivalent results to GME-GCE techniques, but in its general form allows a more direct and straightforwardly interpretable formulation of

available prior information and can reduce significantly the computational effort involved in finding solutions. The technique can also be extended to situations with non-informative prior information. Six fully worked out numerical illustrations of the estimator are supplied. The explicit comparison with the entropy-based methods is motivated by the frequent use of GME or GCE for solving underdetermined problems and introducing prior information into model calibration problems (e.g. Paris and Howitt 1998, Witzke and Britz 1998, Paris 2001).

In chapter five, the estimator developed in chapter four is applied to the supply part of the agricultural sector model CAPRI in order to estimate the behavioural parameters embedded there. The supply model in CAPRI consists of a number of constrained quadratic programming models, where each model represents the agricultural sector in a region in the EU. The models have a quadratic objective function, where the quadratic term influences the simulation behaviour of the model. It is the objective to estimate the parameters of that quadratic term, using the time series data in the CAPRI database. Along the way, two alternative formulations of the regional programming models are considered but discarded.

Chapter six summarizes the main findings, discusses them relative to the aim of the study, and identifies some potentially fruitful fields for further research.

References

Carroll, R., Ruppert, D., and Stefanski, L. (1995). *Measurement error in nonlinear models*. Suffolk, Great Britain: Chapman and Hall.

DeGroot, M. H. (1970). *Optimal statistical decisions*. New York: McGraw-Hill.

Dempe, S. (1997). First-order necessary optimality conditions for general bilevel programming problems: Technical comment. *Journal of Optimization Theory and Applications*, 95: 735-739.

Ferris, M.C., Dirkse, S.P., and A. Meeraus (2002). Mathematical programs with equilibrium constraints: Automatic reformulation and solution via constrained optimization. Oxford university, computing laboratory report no. 02/11. Oxford, England.

Fischer, G., Frohberg, K., Keyzer, M.A., and Parikh, K.S. (1988). *Linked national models: A tool for food policy analysis.* Dordrecht: Kluwer.Fuller, W. A. (1987). *Measurement error models.* John Wiley and Sons, New York.

Heckelei, T., and Wolff, H. (2003). Estimation of constrained optimisation models for agricultural supply analysis based on generalised maximum entropy. *European Review of Agricultural Economics*, 30: 27-50.

Howitt, R., and Reynaud, A. (2003). Spatial disaggregation of agricultural production data using maximum entropy. *European Review of Agricultural Economics*, 30(3): 359-387.

Luo, Z., Pang, J., and D. Ralph (1996). *Mathematical programs with equilibrium constraints*. Cambridge University Press, Cambridge, UK.

McElroy, M. B. (1987). Additive general error models for production, cost, and derived demand or share systems. *The journal of political economy, 95* (4), 737-757.

Neyman, J., and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1): 1-32.

Oude Lansink, A. (1999). Generalised maximum entropy estimation and heterogeneous technologies. *European Review of Agricultural Economics*, 26(1): 101-115.

Paris, Q. (2001): Symmetric Positive Equilibrium Problem: A Framework for Rationalizing Economic Behavior with Limited Information. *American Journal of Agricultural Economics*, 83(4): 1049-1061.

Paris, Q., and Howitt, R.E. (1998). An analysis of ill-posed production problems using maximum entropy. *American Journal of Agricultural Economics*, 80: 124-138.

Stackelberg, H. (1934). *Marktform und Gleichgewicht*. J. Springer, Wien.

Vicente, L. N. and Calamai, P. H. (1994). Bilevel and multilevel programming: a bibliography review. *Journal of Global Optimization*, 5 (3), 291-306.

Witzke, H.P and W. Britz (1998): A Maximum Entropy Approach to the Calibration of Highly Differentiated Demand Systems, Capri Working Paper 98-06, University of Bonn.

Zellner, A. (1971): *Introduction to Bayesian Inference in Econometrics*. New York: Wiley.

Chapter 2 An estimator for trade costs and its small sample properties[3]

1. Introduction

The transport model treated in this chapter is a common component of spatial price equilibrium models. It has been analyzed in several central articles in linear programming, for example in Koopmans's original article from 1947, Enke's ingenious "solution by electric analogue" (1951) and Samuelson's formalized treatment (1952). In Dantzig's work on linear programming (1966), the transport problem is referred to as "The classical transport problem." This chapter is not concerned with the *solution* of the transport model, which has been thoroughly treated for more than fifty years, but turns instead to the empirical specification of the model.

In fact, during the long history of this established model, little attention has been paid to the estimation of its parameters. For example, the only article in the edited volume by Labys et al. (1989) that explicitly mentions the estimation (calibration) issues is McCarl et al. (1989, p. 289-290). They describe a process where the model specification is iteratively updated in order to reproduce first observed quantities, then observed prices. Thompson (1989, the same edited volume) proposes to use statistical measures as mean absolute deviation or mean squared error to evaluate goodness of fit of the resulting model. The reader should, however, be aware that already with a modest number of regions in the SPE model, the large number of possible bilateral trade flows result in an equally large number of zero arbitrage conditions, that render the selection of a basis that fits the base data a difficult problem.

The transport cost minimization problem can be written as

$$\begin{aligned} \min_{x} \quad & \sum_{ij} c_{ij} x_{ij} \\ \text{s.t.} \quad & e_i + \sum_{j} (x_{ij} - x_{ji}) = 0 \qquad [p_i] \\ & x_{ij} \geq 0 \qquad [v_{ij}] \end{aligned} \tag{1}$$

[3] Part of the material in this chapter has previously been presented and discussed at conferences, see Jansson and Heckelei (2004) and Jansson (2005).

where c_{ij} is the trade cost from region i to region j, x_{ij} is the traded quantity and e_i is excess demand in i. Letters in square brackets after the restrictions symbolize the dual values of the constraints.

We consider the trade of a single homogeneous good, and assume that reliable (error free) data on regional excess demand is available. Furthermore, we assume that observations of trade costs between regions as well as regional prices are available, but associated with measurement errors. Observed prices are likely to be inconsistent with observed trade costs and excess demand under the assumption that they constitute an equilibrium solution to model 1.

Spatial price equilibrium models frequently contain a similar transport cost minimization model as component. Examples range from the early publications of Judge and Wallace (1958) and Takayama and Judge (1964) to the more recent contributions of Litzenberg, McCarl and Polito (1982), Peeters (1990) and Guijardo and Elizondo (2003)—to name just a few.

In the cases known to this author, including the publications just cited, there was either no calibration at all[4], or the models were calibrated by in three steps by

1. solving the trade cost minimization problem using the observed trade costs,

2. taking the dual values of the market clearing restrictions p_i as prices and

3. shifting the prices so that some important price is matched precisely.

Step 3 is possible because the first order conditions only contain pair-wise *price differences*. Indeed, one of the market clearing restrictions is redundant, because we know that for a solution to the transportation problem to exist, the sum of all regional net demands must be zero (Dantzig 1966), implying that if there are k markets, then if $k - 1$ of them clear, all of them must clear. Because only price differences are identified, one *numerator price* can be chosen arbitrarily and the remaining prices are determined by those price differences.

Obviously, this method for determining regional prices for a transport model does not use any direct observations of regional prices except for the numerator price. The remaining regional price information is extracted from trade costs and excess demand. This procedure is henceforth referred to as “traditional” and abbreviated TRAD.

The purpose of this chapter is (1) to demonstrate an alternative method, a bilevel estimation program (BLEP), for calibrating the input data for a transport

[4] The edited volume by Labys, Takayama and Uri (1989) is entirely devoted to spatial (and temporal) price equilibrium volume. The paper therein by McCarl et al. mention attempts to calibrate the model by trial and error, and the paper by Thompson suggests to use statistical measures to quantify the deviation of the specified model from observations.

model that uses observations of regional prices, and (2) to show that BLEP estimates regional prices more efficiently than TRAD and that this increased efficiency in estimating prices does not come at the expense of a less efficient estimation of trade costs.

The outline is as follows: In the next section, the BLEP is presented in detail, and it is given a geometric interpretation. Then, hypotheses are deduced about the behaviour of the two estimators by treating them as implicit functions, and the ideas are illustrated in a three-region example model. The hypotheses are analyzed using Monte-Carlo simulations, where the performance of the two estimators is evaluated using generated data. The results of the simulations are analyzed and compared to the hypotheses formed. A final section summarizes and discusses the results.

2. A bilevel estimation program

2.1. The estimator

The idea to determine parameters of optimization models by estimating the first order conditions (FOC) is not new—it is standard procedure. Nevertheless, it has not been much used for *inequality constrained* programming models. Fischer et al. mention an algorithm for computing parameters of a linear program so that a statistical measure of deviation from observations is minimized, thus (implicitly) solving a bilevel programming problem, but do not provide any details of the algorithm. Heckelei and Wolff (2003) propose estimating parameters of agricultural supply models by using optimality conditions as estimating equations. Jansson and Heckelei (2004) show how a similar technique can be applied to the estimation of a transport model, where a large number of inequalities renders the estimation numerically difficult. The current chapter contributes to this strand of research by calibrating the parameters of a transport model by direct estimation of the optimality conditions of problem 1, using a least squares objective, and analyzing the finite sample properties of the estimator. The resulting optimization problem, given by equations 2-7, belongs to the class *bilevel programming problems* (BLPP). The term "bilevel" refers to the fact that it is one programming problem (the estimation) that has another programming problem, in this case the transport problem represented by its optimality conditions, in the constraints (Candler and Norton 1977).

$$\min_{p,c} \quad \sum_{ij}\left(c_{ij} - c_{ij}^{obs}\right)^2 + \sum_{i}\left(p_i - p_i^{obs}\right)^2 \tag{2}$$

$$\text{s.t.} \quad e_i + \sum_{j}\left(x_{ji} - x_{ij}\right) = 0 \tag{3}$$

$$c_{ij} - p_j + p_i = v_{ij} \quad (4)$$

$$x_{ij} v_{ij} = 0 \quad (5)$$

$$c_{ij} = c_{ji} \quad (6)$$

$$x_{ij} \geq 0,\ v_{ij} \geq 0 \quad (7)$$

The general BLPP is difficult to solve, so a few words about solution techniques are appropriate, although a substantial treatment of that subject is beyond the scope of this text. Several different solution methods were tested, including approximation by smooth reformulations as suggested by Facchinei, Jiang and Qi (1999), by Ferris et al. (2002, also found in the NLPEC solver for GAMS), a branch-and-reduce algorithm called BARON implemented as solver for the modelling language GAMS and the method proposed by Jansson and Heckelei (2004). Repeated simulation experiments were performed in order to select the most appropriate solution algorithm for the case at hand. In the experiments, normally distributed errors were added to randomly generated true models, and the parameters re-estimated with least squares. It turned out that two algorithms (A1 and A2) based on smooth approximations in Ferris et al. (2002) were performing similarly well.

A1: The algorithm that *most frequently* obtained the smallest sum of squared errors was based on a product reformulation. The idea behind the reformulation can be illustrated for a complementary slackness condition consisting of a slack variable $x \geq 0$ and a dual value $v \geq 0$. The complementary slackness condition can then be written $xv = 0$. With A1, we instead write $xv \leq \mu$ for some positive number μ. The estimation problem with the approximated complementary slackness is a smooth NLP problem that can be solved with gradient based techniques (this paper uses the solver CONOPT for GAMS,). The problem is solved repeatedly, starting with a large μ and finishing with $\mu = 0$, and each time using the previous solution as a starting point. It was found that performance was improved if before each step a new feasible starting point was found by solving the inner problem for the parameter values of the previous step (or the observations in the first step).

A2: The algorithm that *on average* obtained the smallest sum of squared errors was based on a penalty function. Here the complementary slackness conditions were completely removed, and instead the estimation criterion was augmented with a penalty function of the form μxv (with x, v and μ as before). This modified optimization problem was solved iteratively, each time with a larger μ, with μ initially set to a small positive number. The algorithm terminates when the complementarity gap xv was zero. A2 did not find the best solution most frequently, but it always found a solution that was close to the best solution found by any algorithm.

All other methods were either taking considerably more time to solve the problem or were less reliable in finding a good solution. It was thus decided to use both A1 and A2, and for each simulation experiment use the results of A1 except if the sum of squares obtained by A2 was lower, in which case the results of A2 were used.

2.2. A geometric interpretation

In the case where the criterion function to be minimized is the sum of squared deviations and the model to be estimated is a linear model, the BLPP has an intuitive geometric interpretation. Consider the following problem, estimating a parameter x of a linear programming model in one variable y, and restrictions as follows:

$$\begin{aligned} &\min_{x,y} (x - x_o)^2 + (y - y_o)^2 \\ &\text{s.t.} \quad \min_{y|x} y \\ &\qquad \text{s.t.} \quad -y - x \leq -3 \\ &\qquad\qquad -y + x \leq 2 \\ &\qquad\qquad y - x \leq 2 \\ &\qquad\qquad y + x \leq 8 \end{aligned} \tag{8}$$

Here x_o and y_o are observations, and we want to pick x and y that minimize the upper level objective and where y solves the inner problem treating x as given. We note that (x,y) that minimize $(x - x_o)^2 + (y - y_o)^2$ also minimize $[(x - x_o)^2 + (y - y_o)^2]^{½}$, which is the *Euclidean metric*, i.e. the ordinary distance, between the estimated point (x,y) and the observation (x_o,y_o).

In figure 1 the restrictions of problem (8) are drawn as lines, the observed point (x_o,y_o) as a plus sign, and level curves[5] for the criterion function as concentric circles around the observation. All points on a circle have the same distance from the plus sign and hence the same objective values in the criterion function. Following Bard (1998), we call the area enclosed by the restrictions (where the circles are not dashed) the *constraint region S* of the bilevel programming problem.

[5] If the problem is interpreted as a Bayesian problem as in chapter four, with error terms iid normal distributed, the concentric circles are the iso-probability curves of the posterior density function.

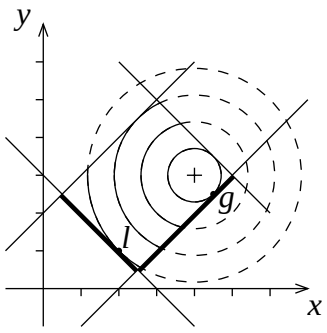


Figure 1. A simple BLEP with OLS criterion and linear inner problem.

The projection of S onto the x-axis is denoted by $S(X)$, and is a convex subset of X with the property that for each $x \in X$ at least one, but possibly several, solutions to the LP exists. If we form the set of all pairs (x,y) where x is in $S(X)$ and y solves the LP, we have the so-called *inducible region*. It is marked with heavy lines in the figure. We seek the point in the inducible region that is closest to the observation.

When the inner problem is an LP, the inducible region is a piecewise linear function derived from the faces of S (Bard 1998). In the general case, it is non-convex, so there may be several local optima. In figure 1, there is a local optimum at the point l and the global optimum is found at g. The non-convexity of the inducible region makes the problem difficult to solve, and is one important reason that special solution methods frequently are needed for BLEPs—and why BLEPs are rarely used.

3. Analysis of estimator properties

It is desirable that an estimator on average is close to the true parameter. We call this *efficiency* and measure it by the *mean squared error* (MSE) (Greene 2003). MSE is the mean squared deviation of an estimate from the true parameter value. Efficiency is a relative measure, so what we would like to know is if one of the estimators is more efficient than the other. To our aid, we use the fact that MSE can be split into a variance and a bias component using

$$MSE[\hat{\theta} \mid \theta] = E[(\hat{\theta} - \theta)^2] = Var[\hat{\theta}] + (Bias[\hat{\theta} \mid \theta])^2 \qquad (9)$$

where θ is the true parameter value and $\hat{\theta}$ the estimator.

In this section, we proceed analytically to deduct hypotheses about variances and biases of the estimates of the two methods. If one estimator turns out to be less biased as well as having less variance than the other, we conclude that it is more efficient. If, in contrast, one estimator is less biased but has higher variance

than the other (or vice versa), the qualitative reasoning in this section does not allow us to say that one estimator is more efficient than the other. The last two sections in this chapter report setup and results of simulation experiments designed to investigate the properties numerically.

We consider the case when there is only one observation available and the errors are known to be additive and drawn from a symmetric distribution. Then TRAD and BLEP constitute two different *implicit functions* from price-cost space $\mathbf{\Omega}$ *into* itself (*onto* the inducible region, which is a subspace of $\mathbf{\Omega}$). That is, each vector of observations is mapped into a vector of estimates. If this mapping were linear, the probability distribution of the estimates would be only a scaled and shifted version of the probability distribution of the errors. The mapping is, however, not always linear, due to the complementary slackness conditions, and thus the distribution of the estimates is asymmetric.

Let us analyse the shape of the graph of the implicit function by a few examples. If we pick one component of the implicit function, say the estimated price in some market, and compute its graph against one of the arguments, say the trade cost into that region from some neighbouring region, we could obtain something like the left pane in figure 2. In the figure, the trade route (i,j) is not used if the trade cost is greater than c^*, so observing a trade cost greater than that does not influence the price estimate p' in region j at all neither with TRAD nor BLEP.

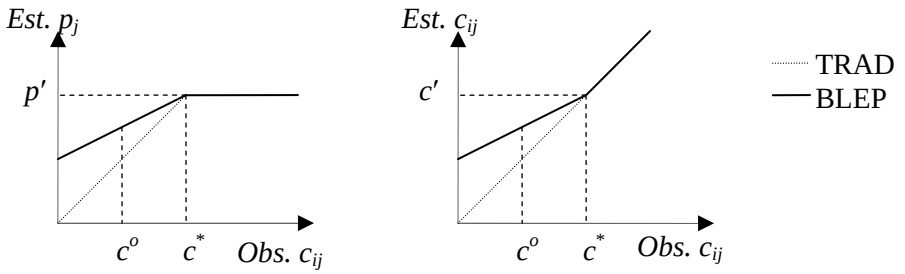


Figure 2. Stylized graph of implicit function component. The break point is due to the complementary slackness conditions.

If in contrast the observed cost is lower than c^*, say c^o, then the zero arbitrage conditions suggest that the price in j is lower than p'. TRAD undertakes the necessary adjustment solely on the price position, so the graph of TRAD slopes 45° downward to the left from (c^*,p') as the dotted line shows. BLEP undertakes the adjustment for the estimated trade cost, shown in the right pane, as well as for the estimated price, so the adjustment of the estimated price is smaller, along the more gently sloping solid line. Both graphs in the left pane do, however, have a kink at (c^*,p'). As a consequence, the density function of the estimated price in j will be asymmetric, and more so for TRAD than for BLEP. For the estimated

trade cost though, as shown in the right pane, only the graph of the implicit function for BLEP has a kink.

What are the consequences of the kinked implicit functions on the efficiency of the estimators? Analysis suggests that a kink causes the estimates to be biased. To see this, we sketch a histogram for a stylized density function for the BLEP estimates in the right pane of figure 2. This is done in figure 3. The quantity on the horizontal axis is, as before, an observation. It is presumed to have an additive error component from a symmetric density function, so that the density function for the observation could give rise to the histogram that is standing on the horizontal axis. Denote the class width by s, and the probabilities of the four classes by a, b, d, and e. Assume, without loss of generality, that c^* is the true trade cost.

On the vertical axis is, as before, the estimate, in this case the BLEP estimate of trade cost. We have put the resulting histogram for the estimates, rotated 90° anti clockwise, along that axis. We see that an observed costs in the interval $(c^*, c^*+s]$ that occurs with probability d will be mapped to an estimate in the interval $(c', c'+s]$ with probability d, and an observation in $(c^*+s, c^*+2s]$ that occurs with probability e maps to $(c'+s, c'+2s]$ with probability e. However, any observation in $[c^*-2s, c^*)$, occurring with probability $a + b$ maps to the smaller interval $[c'-s, c')$ with probability $a + b$. Thus, the distribution of the estimates is skewed.

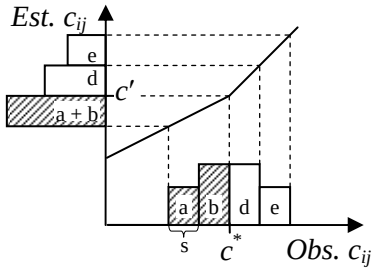


Figure 3. Density functions for a component of the implicit function BLEP.

The mean of the estimates will not be c', the value mapped by the true parameter c^*. We see this by computing the moments around c' and c^* respectively:

$$\text{Moment around } c' = -0.5s(a + b) + 0.5sd + 1.5se$$

$$= -0.5sa - 0.5sb + 0.5sd + 1.5se$$

$$> -1.5sa - 0.5sb + 0.5sd + 1.5se$$

$$= \text{moment around } c^* = 0.$$

Since the moment around c^* is zero, we can compute the bias easily by cancelling terms to obtain moment around $c' = sa > 0$. Thus, the estimates of trade costs

are biased by the amount *sa*, and they must be biased whenever the implicit function is kinked.

Figure 3 also shows that the variance of the estimates decreases, because the width of the histogram decreases. In that way, some of the efficiency lost due to the bias is regained.

Note that the number of kinks and the slopes of the lines depend on the trade pattern. If a region is connected by trade flows to many other regions, then a changed price in that region has implications for many other prices, and conversely when there are few connections. This makes an analytical deduction of the graph of the implicit function practically impossible.

The above reasoning is illustrated in a three-region, single commodity model, for which we assume that there exist true parameter values that represent an equilibrium solution. Figure 4 shows three regions *A*, *B* and *C*, with *B* being a net importing region and *A* and *C* net exporters. The left and right panels of the figure show two of the three possible trade flows that would clear all markets. To be specific, let the true regional prices be $p_A = 100$, $p_B = 109$ and $p_C, = 104$ and the trade costs $c_{AB} = 9$, $c_{AC} = 5$ and $c_{CB} = 5$ and symmetric as in equation (6). In this case, trade will flow as in the left hand panel.

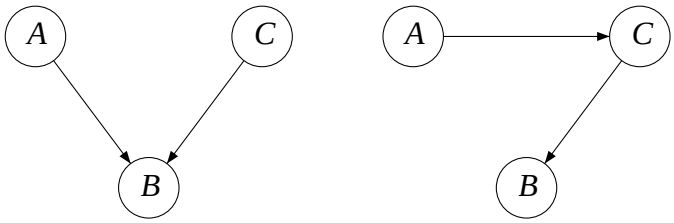


Figure 4. Three region model with *A* and *C* net exporters and *B* a net importer, and two possible market clearing solutions.

To start with, we make the trivial observation that if the all disturbances are degenerate “0”, i.e. no disturbances, then TRAD and BLEP map an “observation” to itself, i.e. both methods return the true parameters. Then, we add random disturbances from a *symmetric* distribution with mean of zero but nonzero variance to only the trade cost c_{AB} and leave all other observations undisturbed, and see what happens to the estimates of TRAD and BLEP. In mathematical terms, we make point approximations of a single partial derivative of the vector valued implicit functions that map observations to estimates, defined by the estimation methods. The symmetric errors will have a *biasing* effect on price estimates, regardless if they are estimated with TRAD or BLEP, as illustrated by the following numerical example:

Example: *We measure prices and costs of the model in figure 4 twice, and after each measurement we use the observation to estimate the true parameters with TRAD and BLEP. Only* c_{AC} *is measured with errors, all other trade costs and prices "happen to be" observed at their true values (but we do not know that). The observations of* c_{AC} *are*

Case 1: $c_{AC} = 10$

Case 2: $c_{AC} = 0$

Estimates with TRAD:

Case 1: The trade cost minimizing solution is the same as that without the error, so trade will still flow as in the left panel of the figure. Conclusion: The dual values of the markets with the numerator price p_A added will equal the true prices, because the flow *AC* is still not used. The costs will, as always with TRAD, be the observed ones: $c_{AB} = 9$, $c_{AC} = 10$ and $c_{CB} = 5$.

Case 2: It is cheaper to transport via *ACB* than via *AB*, so trade will divert from *AB* to *ACB* as in the right panel of figure 4, the prices will be $p_A = 100$, $p_B = 105$ and $p_C = 100$, and costs $c_{AB} = 9$, $c_{AC} = 0$ and $c_{CB} = 5$. Conclusion: In this case, only negative errors that are larger than 1.0 influence the price estimates, because the second cheapest trade route is 1.0 unit more expensive than the cheapest one. The price estimates for *B* should systematically turn out lower than the true prices in this setup, as would the price in *C*.

Estimates with BLEP:

Case 1: The observation is a point in the inducible region, so the estimator will accept the observation unaltered and will measure a deviation of zero. In the estimated model, trade will flow as in the left panel of the figure. Conclusion: Nothing will happen to the prices because the flow *AC* is still not used, and the estimated costs will be $c_{AB} = 9$, $c_{AC} = 10$ and $c_{CB} = 5$ as with TRAD.

Case 2: The observation is not in the inducible region, so the estimator will look for the closest point of the inducible region using the least squares criterion. The best solution means using the trade flow *ACB* and not *AB*, choosing the prices $p_A = 101.9$, $p_B = 108.2$ and $p_C = 102.9$, and the trade costs $c_{AB} = 9$, $c_{AC} = 0.952$ and $c_{CB} = 5.381$. Conclusion: As with TRAD, only the negative error with absolute amount greater than 1.0 influences the estimation, thus a symmetric measurement error with a mean of zero causes the estimated prices to deviate from the true values in only one direction (positive for p_A and negative for p_B and p_C), i.e. being estimated with bias, but less biased than with TRAD.

Further analysis along the same lines as above suggests four hypotheses about the estimators TRAD and BLEP, and those hypotheses are further analysed in the next section with aid of numerical techniques.[6]

H1. BLEP is a more efficient estimator of regional prices than TRAD, because the BLEP estimates have both less variance and less bias than the TRAD estimates.

H2. We cannot *a priori* say that either estimator is a more efficient estimator of trade costs. On the one hand, BLEP estimates have a bias that TRAD estimates lack, but on the other hand the variances of the BLEP estimates are lower. The simulation experiments reported below suggest that this hypothesis can be strengthened.

H3. The variances of the price estimates are heterogeneous, in other words the variance is different in different regions. It is more heterogeneous if estimated with TRAD than with BLEP.

H4. The variance of the cost estimates is heterogeneous when estimated with BLEP but not when estimated with TRAD.

4. Simulation experiments

The small sample properties of the estimators are analyzed using simulation techniques. The basic idea is to generate m randomly chosen "true models," and then estimate each model n times (the simulation size is n), each time adding errors to the true prices and costs. We thus obtain m samples that each consists of n observations of estimated trade cost matrices and price vectors. Throughout this paper we use $m = 100$ and $n = 500$.

The m models, each with ten regions, are generated by drawing regional excess demand from the uniform distribution $[-10,10]$ and trade costs from the uniform distribution $[20,100]$. The excess demand of one region is set to the negative of the sum of excess demand in all other regions to make the problem feasible. The transport model (1) is solved, and the dual values of the market balances plus a constant of are 120 taken as true regional prices. In the following, the index denoting the *model* $(1,\ldots,m)$ to which a certain price or trade cost belongs is omitted for readability.

Each of the m models is estimated n times with TRAD and BLEP, each time with errors added to all true prices and trade costs. The errors are sampled from

[6] A rigorous treatment of the topological properties of problems such as this is beyond the capabilities of the author, but can be found in the literature on mathematical programming with equilibrium constraints (MPEC), e.g. the monograph by Luo, Pang and Ralph (1996).

the normal distribution with mean of zero and standard deviation 6. With this standard deviation, we expect that the major part of the errors are in the interval [−18,18], because plus or minus three standard deviations covers 99.9% of the outcomes. By construction, true trade costs are in [20,100], so with a numerator price of 120 the smallest possible true trade cost as well as price is 20. Hence, adding an error of standard deviation 6 and mean zero will rarely result in negative observed values. Still, they may occur, and to prevent that, the sampled errors are censored to lie within the interval [−19,19]. The errors are censored upwards as well as downwards to avoid censoring being a source of biases.

In the next section, we address the hypotheses put forward in the previous section by analysing MSE, variances and biases of prices and trade costs estimated with TRAD and BLEP. Since equation (9) holds for each parameter in each model, we can compute the mean of each term over all prices or costs in each model, obtaining mean MSE (MMSE), mean squared bias (MSBIAS) and mean variance (MVAR), for which it holds that MMSE = MSBIAS + MVAR. The means are computed in order to obtain an overview over the large number of parameters estimated in the simulation exercise.

The GAMS program “generateSample.gms” that was used for performing the numerical experiments is printed in appendix 2.2, and the program that was used for analysing the results, “analyseSample.gms”, is printed in appendix 2.3. The programs require that the software GAMS is installed, and are better executed from the command prompt in order for progress feedback to work properly.

5. Results

This section presents the results of the simulation experiments in relation to the four hypotheses formed in section 3. The section is subdivided into three parts: 5.1 efficiency of price estimates, 5.2 efficiency of trade cost estimates and 5.3 heterogeneity of variances.

5.1. Efficiency of price estimates (hypothesis H1)

Result 1 *BLEP is a more efficient estimator of regional prices than TRAD.*

The simulation experiments confirm the hypothesis that BLEP is a more efficient estimator of regional prices than TRAD. Figure 5 shows MMSE for price estimates in all models. Each point is the average MSE over all regional prices and all 500 estimations in one model.

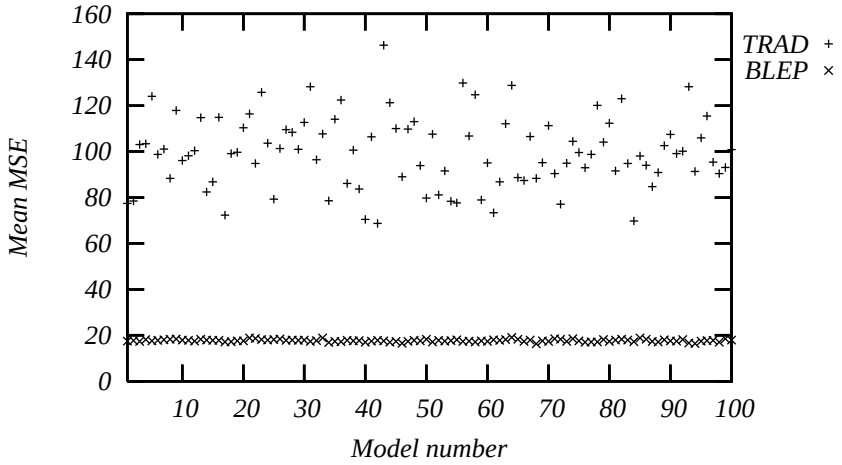


Figure 5. Mean MSE for price estimates for each model.

The results show that BLEP not only delivers more efficient estimates of prices, but also that the efficiency is stable across different models. In other words, it does not depend upon the true data constellation. In contrast, TRAD is less efficient in all cases, and additionally, the efficiency seems to depend on the data constellation. As we will see, the greater efficiency of BLEP regarding price estimates is attributable to less bias as well as less variance, as the qualitative reasoning above suggests.

Result 2 *Both BLEP and TRAD estimate prices with bias, but the bias is smaller for BLEP.*

Figure 6 shows the mean squared bias (MSBIAS)of all price estimates and estimations per model. MSBIAS of prices estimated with TRAD fluctuate strongly between models, whereas the biases of prices estimated with BLEP are much more stable and also smaller. Most of the large biases come from the TRAD estimator.

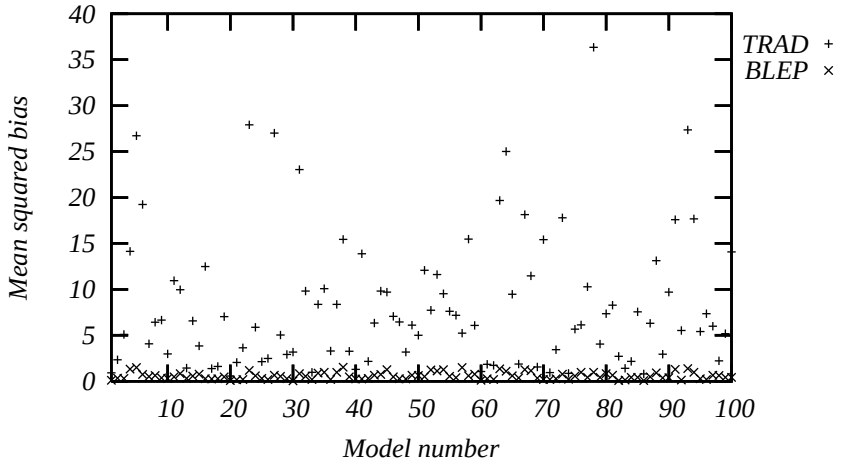


Figure 6. Mean squared bias of price estimates of all models.

It may also be of interest to analyse the biases of the individual regional price estimates (not the mean squared bias). Table 1 shows descriptive statistics of the sample of estimation biases of regional prices (total sample of 100 models × 10 prices × 500 estimations = 500 000 prices). Neither the average nor the median of the biases is far from zero, indicating that there are about as many positive biases as there are negative ones. The larger variance of the biases of TRAD supports the hypothesis that TRAD generally produces price estimates with larger biases. The larger biases also appears in the line "SABIAS", which is the sum of absolute biases. SABIAS of TRAD is more than three times that of BLEP.

Table 1: Descriptive Statistics of Biases of Price Estimates

	TRAD	BLEP
mean	-0.193	0.004
variance	8.242	0.590
median	-0.061	-0.004
SABIAS	2018.280	594.965

According to the reasoning in the previous section, we would expect TRAD to systematically estimate biased prices in some regions in some models, and the BLEP estimates would also be biased. A visual inspection of figure 6 suggests that that the bias of BLEP is much smaller than that of TRAD. It would thus be interesting to *test* if the biases of the price estimates are significant. If we pick at random one of the 1000 prices (10 prices in each of 100 models), we have 500 estimates of that price. Since the estimates are random variables, we may use the Lindberg-Levy central limit theorem (LLCLT) to argue that the mean of the estimates will be asymptotically normally distributed. With 500 observations, the

asymptotic distribution can be considered a fair approximation to the corresponding small sample distribution.

If we apply the LLCLT, using an estimate s^2 of the variance instead of the unknown true variance σ^2, we have that

$$t_{ki} = \frac{n^{1/2}\left(p_{ki} - \overline{p}_{ki}\right)}{s_{ki}} \xrightarrow{d} N(0,1) \quad \text{with} \quad s_{ki} = \frac{\sum_{l=1}^{n}\left(\hat{p}_{kil} - \overline{p}_{ki}\right)^2}{n-1} \tag{10}$$

where $\hat{p}_{kil}$ is the estimated price in model *k*, region *i* for estimation *l* with $l \in \{1,\ldots,500\}$, $\overline{p}_{ki}$ the average of the *n* estimations of the price in model *k*, region *i*, and s_{ki} the sample standard error of the price estimates in model *k*, region *i*. With the null hypothesis that the expectation of the estimates equals the true price, we can compute the probability that the absolute value of the outcome of the test statistic $\left|\hat{t}_{ki}\right|$ in eq. (10) would be as big as it is or bigger. In symbols, that means that we seek the probability $P(\left|t_{ki}\right| \geq \left|\hat{t}_{ki}\right|) = 2(1 - F(\left|\hat{t}_{ki}\right|)$, with *F* the cumulative standard normal distribution. We decide to reject the null hypothesis if the probability is less than 1%.

The result is that for TRAD, the null hypothesis is rejected in 509 cases (of 1000 possible, 100 models with 10 prices per model). For BLEP, the null hypothesis is rejected in 508 cases. The test seems to support the hypothesis that both estimators are biased, but does not make any clear distinction between them. The greater biases of TRAD that are visible in figure 6 are accompanied by greater variances (see figure 7 below), that make the biases less significant.

Result 3 *The variance of prices estimated with TRAD is greater than that of prices estimated with BLEP.*

Figure 7 shows the pooled sample variance of price estimates in each model estimated with TRAD and BLEP. If $k = \{1,\ldots,m\}$ indexes the models, the pooled sample variance[7] s_k^2 of the prices of model *k* is computed as $s_k^2 = \frac{1}{R}\sum_{i=1}^{R}\left(s_{ki}^2\right)$, with s_{ki}^2 indicating the squared sample standard deviation of price *i* in model *k* as defined above, and *R* indicating the number of regions (in this case $R = 10$). As can be clearly seen in the figure, TRAD estimates generally have a higher variance. The variances of prices estimated with TRAD seem to depend more strongly upon the underlying true model than is the case for BLEP. The highest pooled sample variance of the TRAD estimates is about twice the lowest one, whereas the variances of the BLEP estimates are closer together.

[7]Because each price is estimated the same number of times, the pooled variance turns out to be the plain average MVAR.

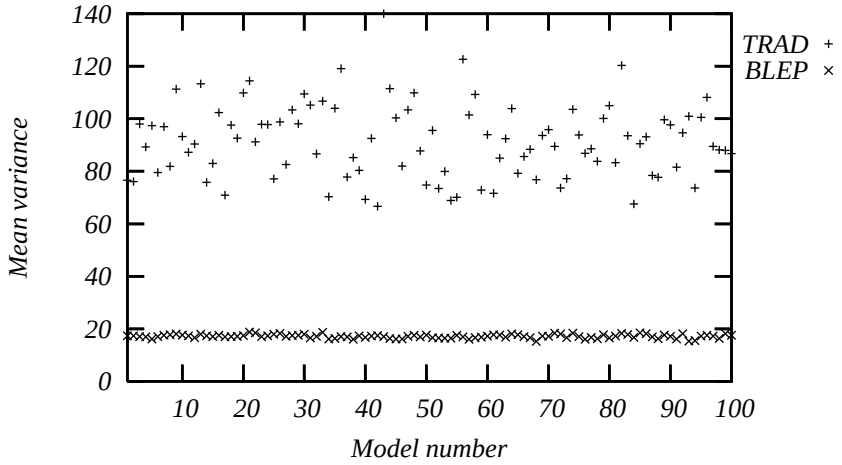


Figure 7. Sample variances of price estimates, pooled together for each model.

It is noteworthy that in all models, the pooled variance for BLEP is clearly smaller than the variance used in the sampling process (36), whereas it is larger than 36 for TRAD. The pooled sample variances reflect the behaviour of the underlying non-pooled variances, described in more detail below under the hypothesis about heterogeneity. Obviously, no statistical test is necessary to see that the variance of the TRAD price estimates is greater than that of the BLEP.

5.2. Efficiency of trade cost estimates (hypothesis H2)

Result 4 *BLEP is a more efficient estimator of trade costs than TRAD.*

Figure 8 shows the mean MSE for all trade cost estimates in each model. MMSE for BLEP is lower than for TRAD in almost all models, but the differences are not as obvious as for the price estimates (figure 5). It also looks as if the efficiency of BLEP is somewhat more sensitive to different data constellations than TRAD, because the BLEP points appear to be vertically more dispersed. Below, MMSE of trade cost estimates is split up into bias and variance components.

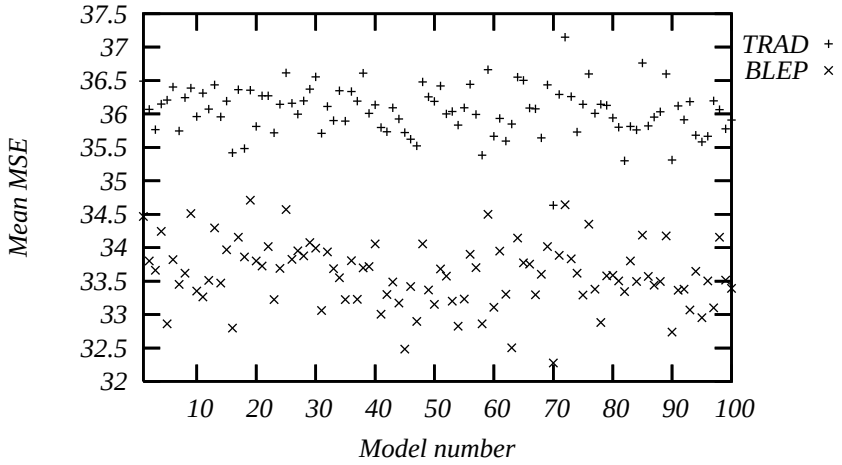


Figure 8. Mean MSE for trade cost estimates for each model.

Result 5 *BLEP but not TRAD produces biased trade cost estimates.*

The TRAD trade cost estimates cannot, by construction, be systematically biased. They are simply the unaltered observations, so their being biased would mean that there is something wrong with our data generating process. However, we have only a finite sample, so the sample mean (the mean of any estimated cost item taken over the *n* repetitions) may very well deviate from the true trade cost.

For BLEP, the qualitative discussion above suggests that the inequalities could cause the trade costs to be systematically biased for some region pairs, but in an unpredictable direction. In figure 9, the MSBIAS of the trade cost estimates in all models are shown. All values are small, and it is not immediately clear whether BLEP is more biased than TRAD, but the tendency is certainly visible, because points further away from zero generally belong to BLEP.

To further investigate the question whether the BLEP cost estimates actually are more biased than the TRAD estimates, we perform a test similar to the one performed for price biases above. As we do not know if a given trade cost will be over- or underestimated, we make the test two-sided as before. For TRAD, the number of rejections of the null hypothesis (that the expectation of each trade cost estimate equals the true trade cost) is close to the number that would be expected, namely 1.31 percent of the cases (59/4500) at the 99 percent level. For BLEP, the number of rejections is higher, with 9.75% of the null hypotheses (439/4500) rejected at the 99 percent level. Thus, there is evidence for the BLEP but not for the TRAD estimates being biased.

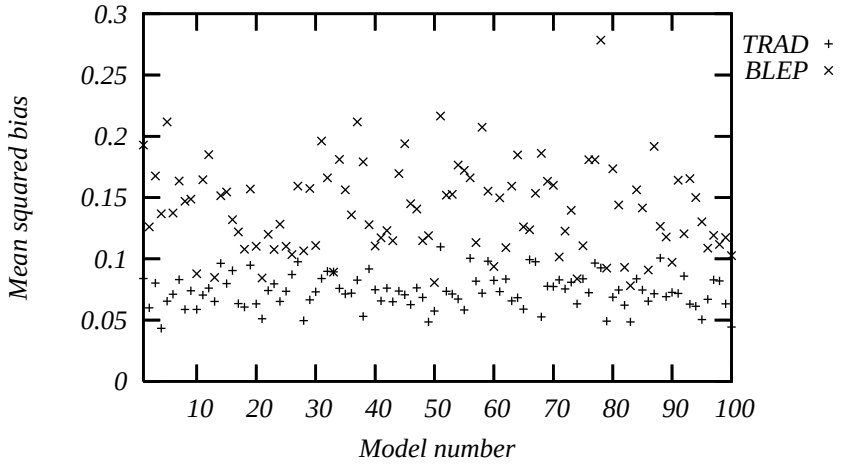


Figure 9. Mean squared bias of trade cost estimates in each model.

If all the computed trade cost biases are considered a sample, we get the sample statistics shown in table 2. The average bias is close to zero for both methods, and so are the variances and the medians of the biases. The sum of absolute biases of all trade costs in all models, SABIAS, is higher for BLEP than for TRAD, also supporting the hypothesis that estimates of BLEP are more biased than TRAD.

Table 2. Descriptive Statistics of Biases of Trade Cost Estimates

	TRAD	BLEP
mean	0.004	0.054
variance	0.073	0.138
median	0.005	0.026
SABIAS	969.585	1271.622

Result 6 *The variance of trade costs estimated with TRAD is greater than that of those estimated with BLEP.*

We expect the pooled sample variance of the estimated trade costs for TRAD to be precisely 36, which is the variance used in the data generation process. Furthermore, the hypothesis states that the pooled sample variance per model (see above) of the estimates performed with BLEP should be lower. Figure 10 shows the pooled sample variances of both methods for each of the 100 models. The data seems to support the hypothesis, because the TRAD data points are nicely dispersed around 36 and all BLEP data points lie below the lowest TRAD data point.

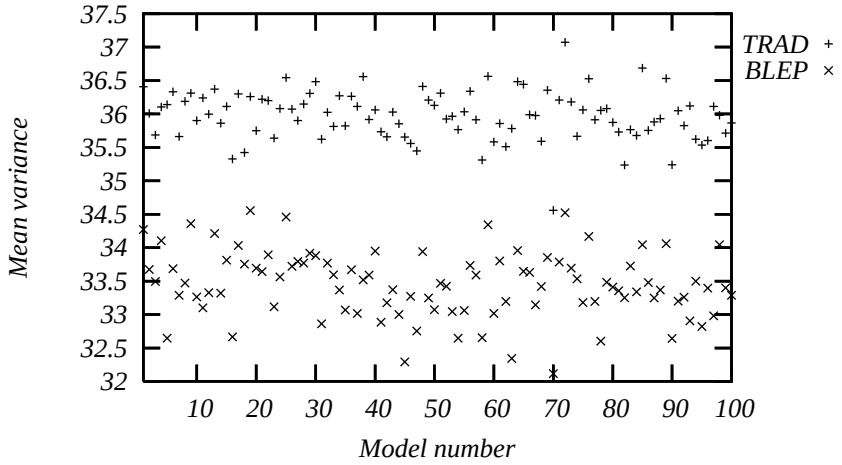


Figure 10. Mean variance of estimated trade costs, pooled together for each model.

The logic of this result is clear in figure 3 above: The kink of the implicit function tends to compress part of the density function of the BLEP cost estimates, making it narrower than the true error distribution, thus reducing variance. If we look at the underlying data in the form of the non-pooled sample variances, the view is more differentiated. It seems that the sample variances of the cost estimates are more dispersed across trade links within each model with BLEP than with TRAD. This observation is further discussed in connection with the hypotheses H3 and H4 regarding heterogeneity of variances below (figure 12).

5.3. Heterogeneity of variances (hypotheses H3 and H4)

Result 7 *The variance of the price estimates is heterogeneous, i.e. the variance is different in different regions. It is more heterogeneous if estimated with TRAD than with BLEP.*

A quick look at the data supports this result. Figure 11 shows the sample variance (not pooled) of the first 200 prices estimated, i.e. the variance over all 500 estimations of each of the ten prices in the first 20 models. It can be seen that the variances of the different regional price estimates fluctuate strongly between the TRAD estimates, whereas the variances seem much more homogeneous for BLEP. All BLEP points are in the thick band at the bottom of the plot. Above that band comes a row of plus signs, which is the TRAD estimates of the numerator prices, all of which have the variance 36 (the sampling variance). Above that row lie all the other TRAD estimates. The higher the variance of a price estimate, the

more trade links are probably separating it from the numerator price. The more trade links, the more measurement errors of trade costs affect the TRAD estimate.

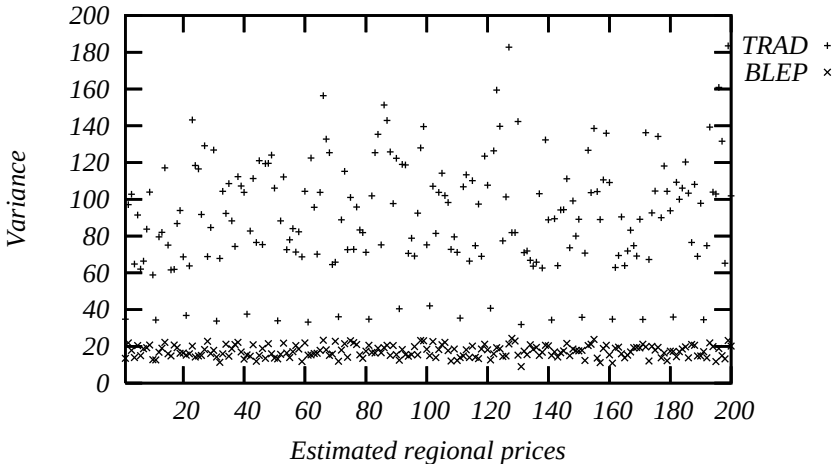


Figure 11. Sample variance of estimates of individual prices

The variances of the TRAD estimates are clearly heterogeneous. However, it is difficult to tell whether the variances of the BLEP estimates are homogeneous or not, that is if the fluctuations observed are random outcomes of the same distribution. If we would estimate the same price item another *n* times, would we then get a similar or different sample variance? We want to test the hypothesis "the variances of prices differ between at least two regions in the estimated model" with the null hypothesis "the variances are equal in all regions of the model." To do this, a *Bartlett's* test (see NIST/SEMATECH 2004) is performed for each model *m*. The results indicate that in 100 models out of 100, TRAD has produced heterogeneous estimates at the 99 percent significance level, whereas BLEP has done so in 99 cases.

However, the Bartlett's test is sensitive to deviations from normality, and we know that the price estimates are biased. Hence, the results may be due to a skewed distribution, not to heterogeneity. To double-check, we perform also a *Levene's* test (NIST/SEMATECH 2004) for heterogeneity, a test that is less sensitive to deviations from normality. The test can be performed using deviations from mean or from the median. Both were tried, with similar results. The following results are for tests with the mean. The result of the test is that the hypothesis that the variances of price estimates equal in all regions is rejected in 99 model for BLEP and in all 100 models for TRAD. So, it seems like the price estimates are likely to be heterogeneous with both methods, albeit the visual impression from figure 11 clearly is that the problem is smaller for BLEP than for TRAD.

Result 8 *The variance of the cost estimates is heterogeneous when estimated with BLEP but not when estimated with TRAD.*

Figure 12 shows the sample variance of the first 200 trade cost estimates with TRAD and BLEP. The first impression is that there is less difference between the methods than was the case for the price estimates. The variances of the TRAD estimates are, as expected due to the data generation method, dispersed around 36. The variances of the BLEP estimates seem to be generally smaller, as previously discussed, and more dispersed, supporting the hypothesis. A lot of the points in the figure coincide. These are trade costs for trade links that are not used regardless of cost, so the observed value need not be modified with either method in order to reach consistency.

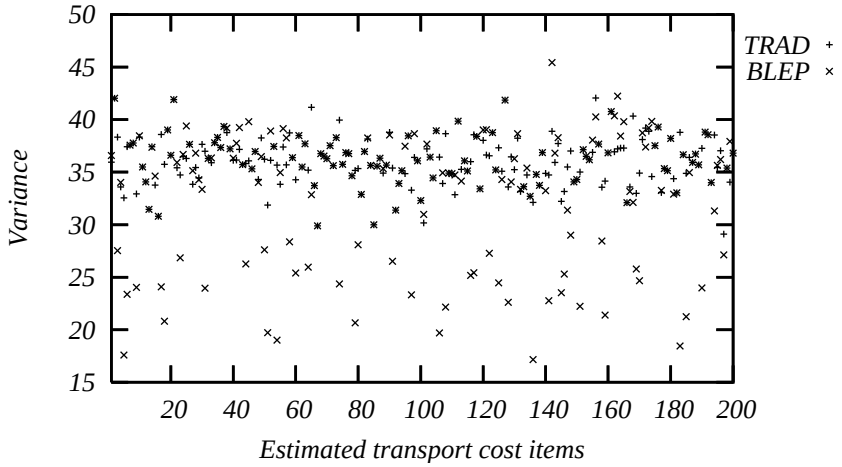


Figure 12. Variance of estimates of individual trade costs.

The tests for heterogeneity detect clear differences between the TRAD and the BLEP estimates: The Bartlett's statistic for the hypothesis "the variances of all cost estimates in each model are not equal," with the null hypothesis "all variances in each model are equal" fully supports the hypothesis. The null hypothesis is rejected at the 99 percent significance level in only a single model of 100 models estimated with TRAD. For BLEP, the null hypothesis is rejected in all 100 models on the 99 percent level.

To double check, the Levene's test was performed also for trade cost estimates with results identical to those of the Bartlett's test (at the 99 percent level). The null hypothesis is rejected in one out of 100 models for TRAD at the 99 percent level, whereas for BLEP, the Levene's test rejects the hypothesis in all models. The result thus seems to be firmly corroborated.

6. Discussion and extensions

We conclude that the traditional way of calibrating a transportation model is inefficient. With the error model studied here, BLEP is a more efficient estimator than TRAD of prices as well as trade costs. For *prices*, BLEP estimates have *smaller biases* as well as *smaller variances* than TRAD. With both methods, the variances of the price estimates depend on the true parameters even when the additive errors come from the same distribution (heterogeneous variances). In other words, also the variances of prices are estimated with biases in both methods.

For *trade costs*, the BLEP estimates are biased whereas the TRAD estimates are not. However, the biases of the BLEP estimates are more than compensated for by lower average variances, obtaining a smaller mean squared error. Variances of trade cost estimates are heterogeneous if estimated with BLEP but not with TRAD.

The BLEP performs better than TRAD in almost all disciplines. Are there no drawbacks? Clearly, one drawback is that BLPPs in general are difficult to solve. However, with increasing computing capacity and the development of new solver software, that argument is losing its strength. And for the incumbent problem—the transport model—existing techniques are able to handle the difficulties.

Hitherto we only considered the estimation of parameters in the classical transportation model. The proposed techniques and main results apply equally well to linear programs and linearly constrained quadratic programs.

The extremum estimation of parameters of a mathematical optimization model can *generally* be formulated as a bilevel programming problem, where the upper level problem is to select the parameters of the optimization model so that the parameters and the solution of the model minimize some estimation criterion. The inner problem is to solve the optimization model that is to be estimated, treating the parameters coming from the upper level as given. Setting up the estimation this way, the parameters are consistently estimated in the sense that the estimating equations are fully equivalent to those in the final simulation model.

In all *linear programs*, the solution of the inner problem will be at a boundary of the constraint region and the solution correspondence, i.e. the implicit function that returns a set of solutions (possibly empty) for each parameter value, will not be continuously differentiable except in trivial cases. This is reflected in for example the stepwise supply response to changing prices in linear supply models, and in the switching of trade flows between different destinations in the model studied in this text. In such cases, bilevel programming techniques as exemplified by the ones employed here are *required* in order to find a solution to the bilevel estimation problem at least close to the global optimum. Neither conventional least squares estimation techniques nor gradient based numerical optimization is feasible due to the discontinuous derivatives of the solution correspondence.

The estimation of parameters of a linear program can also be put in terms of estimation of parameters of a linear complementarity problem, LCP. An LCP is a system of linear equalities and inequalities, where the inequalities are linked via complementary slackness conditions. The system of equations (3-7) is an instance of LCP. We conjecture that the solution method and the analysis of estimator properties applies equally well to the whole class of LCPs, because it only refers to linearity of equations and the existence of complementary slackness conditions. In particular, this class includes first order conditions of *linearly constrained quadratic programs*. In appendix 2.1 we give a definition of a linear complementarity problem and prove that the first order conditions of a linear or quadratic program is an LCP, as claimed above.

The approach described in this chapter would thus be useful also for estimating parameters in such problems. A specific instance that attracted much attention in the last ten years is the quadratic PMP model, introduced to the wider community of modellers by Howitt (1995). Note that with a convex quadratic program, an inner solution is possible, or, in the case of PMP-models, even likely. We may thus expect less problems with "switching complementary slackness constraints" in such models than in linear models. Indeed, Heckelei and Wolff (2003) perform such an estimation and do not report any numerical problems. In a subsequent chapter (chapter five), we estimate parameters of a linear inequality constrained quadratic model, where we are able to *a-priori* determine the status of the single complementary slackness condition, and thus manage to avoid using bilevel solution techniques altogether.

References

Bard, J. (1998). *Practical bilevel optimization*. Kluwer academic publishers.

Candler, Wilfred and Norton, Roger (1977). Multi-level programming and development policy. World Bank Staff Working Paper 258.

Dantzig, G. (1966). *Lineare Programmierung und Erweiterungen*. Springer-Verlag Berlin.

Enke, S. (1951). Equilibrium among spatially separated markets: Solution by electric analogue. *Econometrica 19*, 40–47.

Facchinei, F., H. Jiang, and L. Qi (1999). A smoothing method for mathematical programs with equilibrium constraints. *Mathematical Programming 85*, 107–134.

Ferris, M.C., Dirkse, S.P., and A. Meeraus (2002). Mathematical programs with equilibrium constraints: Automatic reformulation and solution via constrained optimization. Oxford university, computing laboratory report no. 02/11. Oxford, England.

Greene, W. H. (2003). *Econometric analysis* (Fifth ed.). Prentice Hall.

Guajardo, R. G. and H. A. Elizondo (2003). North american tomato market: a spatial equilibrium perspective. *Applied Economics 35*, 315–322.

Heckelei, T. and H. Wolff (2003). Estimation of constrained optimization models for agricultural supply analysis based on generalised maximim entropy. *European Review of Agricultural Economics 30*, 27–50.

Howitt, R. E. (1995). Positive mathematical programming. *American Journal of Agricultural Economics*, *77*, 329-342.

Jansson, T.(2005). Two ways of estimating a transport model. Selected Paper, AAEA Annual Meeting, July 24-27, 2005, Providence, Rhode Island.

Jansson, T. and T. Heckelei (2004). Estimation of a transportation model using a mathematical program with equilibrium constraints. Technical Report 1501, Global Trade Analysis Project.

Judge, G. G. and T. D. Wallace (1958). Estimation of spatial price equilibrium models. *Journal of Farm Economics 15*, 801–820.

Koopmans, T. C. (1947). Optimum utilization of the transportation system. *Econometrica 17*, Supplement.

Labys, W.C., Takayama, T., and Uri, N.D. (1989). *Quantitative methods for market oriented economic analysis over space and time*. Aldershot, UK: Avebury.

Litzenberg, K., B. A. McCarl, and J. Polito (1982). A decomposition solution for spatial equilibrium problems. *American Journal of Agricultural Economics 64*, 590–594.

Luo, Z., Pang, J., and D. Ralph (1996). *Mathematical programs with equilibrium constraints*. Cambridge University Press, Cambridge, UK.

McCarl, B., Brown, D., Adams, R., and Pheasant, J. 1989. Linking farm and sector models in spatial equilibrium analysis: An application to ozone standards as they affect corn belt agriculture. In W. C. Labys, T. Takayama, and N. D. Uri (Ed.), *Quantitative methods for market-oriented economic analysis over space and time*. Aldershot, UK: Avebury.

Murty, K.G. (1997). *Linear complementarity, linear and nonlinear programming.* Berlin: Heldermann Verlag.

NIST/SEMATECH (2004, Dec). NIST/SEMATECH e-handbook of statistical methods. http://www.itl.nist.gov/div898/handbook/, National Institute of Standards and Technology.

Peeters, L. (1990). A spatial equilibrium model of the EU feed grain sector. *European Review of Agricultural Economics 17,* 801–820.

Samuelson, P. A. (1952). Spatial price equilibrium and linear programming. *American Economic Review 42*, 283–303.

Takayama, T. and G. G. Judge (1964). Spatial equilibrium and quadratic programming. *Journal of Farm Economics 46*, 67–93.

Thompson, R.L. (1989). Spatial and temporal price equilibrium agricultural models. In W. C. Labys, T. Takayama, and N. D. Uri (Ed.), *Quantitative methods for market-oriented economic analysis over space and time*. Aldershot, UK: Avebury.

Appendix 2.1. Linear complementarity problems

Definition: Linear Complementarity Problem (LCP)

Let M be an $(n \times n)$ matrix and q, z and w $(n \times 1)$ vectors. (z,w) is said to solve the LCP(M,q) if

$$w - Mz = q$$
$$w \geq 0, z \geq 0$$
$$w'z = 0$$

The necessary first order conditions of linear and quadratic programming problems can be transformed into LCPs.

Theorem: *Quadratic programming and LCP*

Consider the quadratic program (QP)

$$\min_x \mathbf{cx} + \tfrac{1}{2}\mathbf{x'Dx}$$
$$\text{subject to} \quad \mathbf{Ax} \geq \mathbf{b}$$
$$\mathbf{x} \geq 0$$

$\mathbf{x}$ solves the QP iff there exist a vector of dual values $\mathbf{y}$ and slack vectors $\mathbf{u}$ and $\mathbf{v}$ such that $\left(\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}, \begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix}\right)$ solve the LCP$\left(\begin{bmatrix}\mathbf{D} & -\mathbf{A'}\\ \mathbf{A} & \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{c'}\\ -\mathbf{b}\end{bmatrix}\right)$

Proof: (Murty, 1997) The Kuhn-Tucker conditions of the QP are

$$-\mathbf{A'y} + \mathbf{c'} + \mathbf{Dx} \geq 0$$
$$\mathbf{Ax} - \mathbf{b} \geq 0$$
$$\mathbf{x} \geq 0, \mathbf{y} \geq 0$$
$$(\mathbf{c'} - \mathbf{A'y})'\mathbf{x} = 0, (\mathbf{Ax} - \mathbf{b})'\mathbf{y} = 0$$

where $\mathbf{y}$ is the vector of dual values connected to the inequalities. Introducing the (non-negative) slack variables $\mathbf{u} = -\mathbf{A'y} + \mathbf{c'} + \mathbf{Dx}$ and $\mathbf{v} = \mathbf{Ax} - \mathbf{b}$ and rearranging gives

$$\begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix} - \begin{bmatrix}\mathbf{D} & -\mathbf{A}'\\ \mathbf{A} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} = \begin{bmatrix}\mathbf{c}'\\ -\mathbf{b}\end{bmatrix}$$

$$\begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix} \geq 0, \begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} \geq 0$$

$$\begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix}' \begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} = 0$$

which is an LCP according to the definition. □

If the matrix **D** is a matrix of zeros, then the problem is an LP, and the first order conditions are still an LCP. Note that in order to estimate the parameters of the QP, we would in the most common cases like to add the second order condition that **D** be a positive (semi-)definite matrix.

Appendix 2.2. GAMS program for generation of numerical results

```
$ontext
    generateSample.gms
    GAMS code for implementing the bilevel estimation program of chapter 2,
    plus some alternative solution algorithms. Note that storage costs are
    also included in contrast to paper, but with number of periods equal to
    one (the set t contains one element), equivalence is obtained.
    Progress feedback works only when run from DOS prompt. Minimize DOS
    window to improve speed considerably, as screen updating becomes un-
    neccesary. Several solution approaches are included below (a1-a10).
    In the chapter, “A1” corresponds to a6 below, “A2” corresponds to a9,
    and the “traditional method” is a8 below.

    Torbjoern Jansson
    LEI, The Hague, NL
$offtext
$offlisting

* Temporary directory: set to a fast local drive
$setglobal tempdir C:\TEMP

* Create new set of random numbers?
*  execseed=1+gsecond(jnow);

scalar saveGDX "Set to 1 to save results for analyseSample.gms" /1/;

set m Number of models to generate  /m1*m10/;
set n Number of estimation attempts /n1*n10/;

set a Solution approaches implemented/
  a0  Not a solution method just the raw data saved
  a1  Direct solution bound to fail
  a2  Perpendicular starting point working well in symmetric case
  a3  Global optimisation with Baron as NLP
  a4  Global optimisation with Baron as MINLP
  a5  Simpler but not better own method
  a6  Facchinei et al or NLPEC solver smooth approximation
  a7  Yet another enumeration method not working
  a8  Traditional method not using price observations
  a9  Penalty function iterative approximation as in NLPEC solver
  a10 Best of a6 and a9
  /;

set aon(a) Solution approaches to test /a6,a8,a9,a10/;
alias(aon,aon2);

* Declarations for the economic model (the inner problem)

set i Regions /i1*i10/;
alias(i,j,k,ii,jj);

set t Periods /t1*t1/;
alias(t,t1,t2);

set tnext(t,t1);
  tnext(t,t1) $  (ord(t1) = (ord(t)+1))               = yes;
```

```
  tnext(t,t1) $ ((ord(t1) = 1) and (ord(t) = card(t))) = yes;

set im(i) Regions with market balance (one is dropped);
    im(i) = yes$(not sameas(i,"i1"));
*   im(i) = yes;

set AD(i,j) Set of admissible transport flows;
    AD(i,j) $ (not sameas(i,j)) = yes;

parameter c(i,j) Observed (true) transport costs;
parameter so(i)  Observed (true) storage costs;
parameter p(t,i) Observed prices;
parameter e(t,i) Excess demand;

scalar fp Fix price that shifts (the observed) price system /120/;
scalar mu Smoothing parameter of relaxation                /0/;

scalar w1 Weight for transport cost in estimation /1/;
scalar w2 Weight for storage cost in estimation   /1/;
scalar w3 Weight for prices in estimation         /1/;

* *** Declarations belonging to the models ***

free variables
    pe(t,i)   Estimated price of region i
    z         Free objective variable
    zz        Another free objective variable
    zpen      Penalty function value;

positive variables
    x(t,i,j)  Transport stream from i to j
    st(t,i)   Storage
    ce(i,j)   Estimated transport cost
    se(i)     Estimated storage   cost
    pi(t,i,j) Dual value of lower bound on x
    ro(t,i)   Dual value of lower bound on st;

binary variables
    bt(t,i,j) Is there a trade flow
    bs(t,i)   Is there storage;

equations
    F         Objective function definition
    h(t,i)    Market balance i
    dx(t,i,j) First order conditions for optimal transportation
    ds(t,i)   First order conditions for optimal storage
    cx        Complementarity restriction
    cs
    pen       Penalty function for complementary slackness
    FPen      Objective function penalty term

*     Binary tree for Baron
    bdx(t,i,j)    Alternative foc tp
    bds(t,i)      Alternative foc st
    bcx(t,i,j)    Alternateve cmp tp
    bcs(t,i)      Alternateve cmp st

    DUM           Dummy
```

```
    TPC     Transport cost;

*   Weighted least squares deviation

F       ..   z =E=

*                Squared deviations of costs from observations.
*                Symmetry ==> only upper triangle less diagonal

            w1*sum(AD(i,j) $ (ord(i) lt ord(j)), sqr(ce(i,j) - c(i,j)))

*                Squared deviations of prices to observations.

          + w3*sum((t,i), sqr(pe(t,i) - p(t,i)))

*                Squared deviations of storage costs from observations

          + w2*sum(i, sqr(se(i) - so(i)));

*   Penalty function approach

Pen ..  zpen =e= mu*sum((t,i,j) $ AD(i,j), pi(t,i,j) * x(t,i,j))
              + mu*sum((t,i)            , ro(t,i) * st(t,i));

FPen .. zz =e= z + zpen;

*   Market balance

h(t,im)   ..   sum(AD(im,j), x(t,j,im)-x(t,im,j)) - st(t,im)
            + sum(tnext(t1,t), st(t1,im)) =E= e(t,im);

*   First order condition for transport problem

dx(t,i,j) $ AD(i,j) ..   (ce(i,j) $ (ord(i) lt ord(j))
                         +ce(j,i) $ (ord(i) gt ord(j)))
                       + pe(t,i) - pe(t,j) =E= pi(t,i,j);

ds(t,i) ..  se(i) - sum(tnext(t,t1), pe(t1,i)) + pe(t,i) =E= ro(t,i);

*   Complementary slackness condition for transport problem

cx (t,i,j) $ AD(i,j)  ..  pi(t,i,j) * x(t,i,j) =L= mu;

cs (t,i) ..   ro(t,i) * st(t,i) =L= mu;

*   BARONS binary first order condition for transport problem

bdx(t,i,j) $ AD(i,j)  ..  pi(t,i,j) =l= pi.up(t,i,j) * (1-bt(t,i,j));
bds(t,i)              ..  ro(t,i)   =l= ro.up(t,i)   * (1-bs(t,i));

bcx (t,i,j) $ AD(i,j) ..  x(t,i,j)  =l= x.up(t,i,j)  * bt(t,i,j);
bcs (t,i)             ..  st(t,i)   =l= st.up(t,i)   * bs(t,i);

DUM .. z =l= 10;

*   Transport cost minimisation

TPC .. z =E= sum((t,i,j)$AD(i,j), x(t,i,j)*(ce.l(i,j)$ (ord(i) lt ord(j))
```

```
                                          +ce.l(j,i)$ (ord(i) gt ord(j))))
          + sum((t,i), st(t,i)*se.l(i));

model EstimNLP  NLP formulation of MPEC     /F,h,dx,cx,ds,cs/;
model EstimPEN  Penalty formulation of MPEC /F,h,dx,   ds,pen,FPen/;
model TPmin     Transportation model        /TPC,h  /;
model EstimPre  Relaxed version of MPEC     /F,h,dx,ds/;
model BaronNLP  Binary formulation of MPEC  /F,h,dx,bdx,bcx,ds,bds,bcs/;
model DUMMY /DUM/;

EstimNLP.limcol    = 0;
EstimPEN.limcol    = 0;
TPmin.limcol       = 0;
TPmin.limrow       = 0;

TPmin.solprint     = 2;
EstimNLP.solprint  = 2;
EstimPEN.solprint  = 2;
EstimPRE.solprint  = 2;

EstimPRE.solvelink = 2;
EstimNLP.solvelink = 2;
EstimPEN.solvelink = 2;
TPmin.solvelink    = 2;

EstimNLP.workspace = 100;
EstimNLP.optfile   = 1;
EstimPEN.workspace = 100;
EstimPEN.optfile   = 1;

DUMMY.solprint = 2;

*   Declarations of items used to save program output

parameter ptru(m,t,i)     True price;
parameter ctru(m,i,j)     True transportation cost;
parameter stru(m,i)       True storage cost;
parameter etru(m,t,i)     True excess demand;

parameter pest(m,n,t,i,a) Estimated price;
parameter cest(m,n,i,j,a) Estimated transportation cost;
parameter sest(m,n,i,a)   Estimated storage cost;

parameter objes(*,m,n);

set enum /e1*e1000/;
set cenum/s,x,pi,rho/;
parameter zenum(m,n,cenum,enum);
scalar nenum /1/;

*   For the log window

scalar logcount 'Counter for iteration log' /0/;
scalar iflog    'Iteration log frequency'   /0/;
scalar progress 'Progress fraction'         /0/;
file batch /%tempdir%\titlebatch.bat/;
batch.lw = 0;
iflog = (card(m)*card(n))/1000;
```

```
*   Bound prices to help Baron

pe.lo(t,i) = 0;
pe.up(t,i) = fp*2;

$onecho > init.gms
*   Automatically generated include file that initializes problem at a
*   feasible point, using observations.
*   1) Reset all variables with bounds and equations of importance

    option kill=pe;
    option kill=ce;
    option kill=se;
    option kill=x;
    option kill=F;
    option kill=h;
    option kill=dx;
    option kill=ds;
    option kill=cx;
    option kill=cs;
    option kill=ro;
    option kill=pi;
    option kill=z;

*   2) Restore bounds on transport costs

    ce.up(i,j) = 200 $ (AD(i,j) and (ord(i) lt ord(j)));

*   3) Starting point for costs is observation

    ce.l(i,j) $ (AD(i,j) and (ord(i) lt ord(j))) = c(i,j);
    se.l(i) = so(i);

*   4) Find a corresponding feasible price vector by solving TP problem

    solve TPmin using nlp minimising z;

    pe.l(t,i) = fp + h.m(t,i) $ im(i);

*   5) Initialise the first order conditions using duals of TP problem

    pi.l(t,i,j) $ AD(i,j)
        = ce.l(i,j) $ (ord(i) lt ord(j))
        + ce.l(j,i) $ (ord(i) gt ord(j)) + pe.l(t,i) - pe.l(t,j);

    ro.l(t,i) = se.l(i) - sum(tnext(t,t1), pe.l(t1,i)) + pe.l(t,i);
$offecho

$onechov > savesol.gms
* Automatically generated batinclude that saves the outcome of a solution
objes("%1",m,n) = z.l;
pest(m,n,t,i,"%1") = pe.l(t,i);
cest(m,n,i,j,"%1") = ce.l(i,j);
sest(m,n,i,"%1")   = se.l(i);
$offecho

* ---------------------------------------------------------------------------
```

```
*   Generate m models
* ---------------------------------------------------------------------------

loop(m,

*   START: Generate true model with prices consistent with transport costs

    c(i,j) $ (ord(i) lt ord(j)) = uniform(20,100);
    c(i,j) $ (ord(i) gt ord(j)) = c(j,i);

    so(i) = uniform(3,10);

    e(t,i)    = uniform(-10,10);
*   Make sure solution exists by adjusting excess demand on one point
    e("t1","i1") = -sum((t,i) $ (not (    sameas(t,"t1")
                                      and sameas(i,"i1"))), e(t,i));

    ce.l(i,j) =  c(i,j);
    se.l(i)   = so(i);

    solve TPmin using nlp minimising z;

    p(t,i) = fp + h.m(t,i) $ im(i);

    ptru(m,t,i) =  p(t,i);
    ctru(m,i,j) =  c(i,j);
    stru(m,i)   = so(i);
    etru(m,t,i) =  e(t,i);

* ---------------------------------------------------------------------------
*       Add error terms to the model n times and estimate m
* ---------------------------------------------------------------------------

    loop(n,

*       Add error terms to costs and prices, truncate to avoid negatives

        c(i,j) $ (ord(i) lt ord(j)) = ctru(m,i,j)
                                    + min(19, max(-19,normal(0,6)));
        c(i,j) $ (ord(i) gt ord(j)) = c(j,i);

        p(t,i)           = ptru(m,t,i) + min(19, max(-19,normal(0,6)));
        so(i)            = max(stru(m,i)   + normal(0,(3/2)), 0.2);

        pest(m,n,t,i,"a0") =  p(t,i);
        cest(m,n,i,j,"a0") =  c(i,j);
        sest(m,n,i,"a0")   = so(i);

*       Show progress in title bar of DOS window

        logcount=logcount-1;
        if(logcount le 0,
            logcount=iflog;
            progress=100-100*((ord(m)-1)*card(n)+ord(n)-1)/card(n)/card(m);
            putclose batch "title M: ",m.tl,", ",progress:0:1,"%% left";
            execute "%tempdir%\titlebatch";
        );
```

```
*       -------------------- approach 1 ---------------------------------
        if(aon("a1"),

$include "init.gms"

*           a) Solve using NLP from the starting point just obtained

            solve EstimNLP using NLP minimizing z;

$batinclude "savesol.gms" a1
        );
*       -------------------- approach 2 ---------------------------------
        if(aon("a2"),

$include "init.gms"

*        a) Solve the relaxed MPEC

            solve EstimPre using NLP minimizing z;

*        b) Solve TP-problem again

            solve TPmin using nlp minimizing z;

            pe.l(t,i) = fp + h.m(t,i) $ im(i);

            pi.l(t,i,j) $ AD(i,j)
                = ce.l(i,j) $ (ord(i) lt ord(j))
                + ce.l(j,i) $ (ord(i) gt ord(j)) + pe.l(t,i) - pe.l(t,j);

            ro.l(t,i) = se.l(i) - sum(tnext(t,t1), pe.l(t1,i)) + pe.l(t,i);

*        c) Solve the full MPEC from this new point using NLP

            mu = 0;
            solve EstimNLP using NLP minimising z;

$batinclude "savesol.gms" a2
        );
*       -------------------- approach 3 ---------------------------------
        if(aon("a3"),

*           Comment out to let Baron start from previous solution
$include "init.gms"

*           Extra: bounds on all variables, just for the Baron..

            se.up(i)     = 100*smax(t,pe.up(t,i)-pe.lo(t,i));
            pi.up(t,i,j) = ce.up(i,j) + pe.up(t,i)-pe.lo(t,i);
            ro.up(t,i)   = se.up(i)   + pe.up(t,i)-pe.lo(t,i);
            x.up(t,i,j)  = sum((t1,k), ABS(e(t1,k)));
            st.up(t,i)   = sum((t1,j), ABS(e(t1,j)));

*        a) Solve using Baron from the starting point just generated

            option NLP=Baron;
            solve EstimNLP using NLP minimizing z;
            option NLP=Conopt;
```

```
$batinclude "savesol.gms" a3
        );

*       -------------------- approach 4 ---------------------------------
        if(aon("a4"),

*           Comment out to let Baron start from previous solution
$include "init.gms"

*           Extra: bounds on all variables, just for the Baron..

            se.up(i)     = 100*smax(t,pe.up(t,i)-pe.lo(t,i));
            pi.up(t,i,j) = ce.up(i,j) + pe.up(t,i)-pe.lo(t,i);
            ro.up(t,i)   = se.up(i)   + pe.up(t,i)-pe.lo(t,i);
            x.up(t,i,j)  = sum((t1,k), ABS(e(t1,k)));
            st.up(t,i)   = sum((t1,j), ABS(e(t1,j)));

*           Initialise binary variables to feasible point

            bt.l(t,i,j) = 1 $  x.l(t,i,j);
            bs.l(t,i)   = 1 $ st.l(t,i);

*        a) Solve using Baron MINLP from the starting point just generated

            option MINLP=Baron;
            BaronNLP.solprint = 2;
            solve BaronNLP using MINLP minimizing z;
            BaronNLP.solprint = 2;

$batinclude "savesol.gms" a4
        );

*       -------------------- approach 5 ---------------------------------
        if(aon("a5"),

$include "init.gms"

*        a) Solve the MPEC using NLP from the starting point just generated

            solve EstimNLP using NLP minimizing z;

$batinclude "savesol.gms" a5
        );
*       -------------------- approach 6 ---------------------------------
        if(aon("a6"),

$include "init.gms"

*        a) Solve problem without complementarity constraints

            solve EstimPre using NLP minimizing z;

*        b) Iterate over increasingly better approximations

            mu = 1;

            while(mu,
```

```
*               b1) Make feasible

                solve TPmin using nlp minimizing z;

                pe.l(t,i) = fp + h.m(t,i) $ im(i);

                pi.l(t,i,j) $ AD(i,j)
                    = ce.l(i,j) $ (ord(i) lt ord(j))
                    + ce.l(j,i) $ (ord(i) gt ord(j)) + pe.l(t,i)-pe.l(t,j);

                ro.l(t,i) = se.l(i) -sum(tnext(t,t1),pe.l(t1,i))+pe.l(t,i);

*               b2) Solve approximation

                mu $ (mu<0.001) = 0;

                solve EstimNLP using NLP minimising z;

                mu = mu/2;
            );

$batinclude "savesol.gms" a6
        );
*       -------------------- approach 7 ----------------------------------
        if(aon("a7"),

$include "init.gms"

*           Force one column after another into the basis

            nenum=1;

*        a) Spatial price equilibrium

            loop((t2,ii,jj) $ AD(ii,jj),
                pi.up(t2,ii,jj) = 0;
                solve EstimPRE using nlp minimizing z;
                solve TPmin using nlp minimizing z;
                pe.l(t,i) = fp + h.m(t,i) $ im(i);

*               Initialise the first order condition as well

                pi.l(t,i,j) $ AD(i,j)
                    = ce.l(i,j) $ (ord(i) lt ord(j))
                    + ce.l(j,i) $ (ord(i) gt ord(j)) + pe.l(t,i)-pe.l(t,j);

                ro.l(t,i) = se.l(i) -sum(tnext(t,t1),pe.l(t1,i))+pe.l(t,i);
                solve EstimNLP using nlp minimizing z;
                pi.up(t2,ii,jj) = inf;
                zenum(m,n,"pi",enum) $ (ord(enum)=nenum) = z.l;
                nenum=nenum+1;
            );

*        b) Intertemporal price equilibrium

            loop(ii,
                ro.up(t2,ii) = 0;
```

```
                solve EstimPRE using nlp minimizing z;
                solve TPmin using nlp minimizing z;
                pe.l(t,i) = fp + h.m(t,i) $ im(i);

*               Initialise the first order conditions
                pi.l(t,i,j) $ AD(i,j)
                    = ce.l(i,j) $ (ord(i) lt ord(j))
                    + ce.l(j,i) $ (ord(i) gt ord(j)) + pe.l(t,i)-pe.l(t,j);

                ro.l(t,i) = se.l(i) -sum(tnext(t,t1),pe.l(t1,i))+pe.l(t,i);
                solve EstimNLP using nlp minimizing z;
                ro.up(t2,ii) = inf;
                zenum(m,n,"rho",enum) $ (ord(enum)=nenum) = z.l;
                nenum=nenum+1;
            );

            z.l = smin((cenum,enum)$(    zenum(m,n,cenum,enum)
                                             and (ord(enum) lt nenum)),
                                    zenum(m,n,cenum,enum));

$batinclude "savesol.gms" a7
        );
*       -------------------- approach 8 ---------------------------------
        if(aon("a8"),

$include "init.gms"

            z.l =
               sum(AD(i,j) $ (ord(i) lt ord(j)), sqr(ce.l(j,i)-c(i,j)))*w1
             + sum((t,i),                        sqr(pe.l(t,i)-p(t,i)))*w3
             + sum(i,                            sqr(se.l(i)  -so(i))) *w2;

$batinclude "savesol.gms" a8
        );

*       -------------------- approach 9 ---------------------------------
-------
*       A penalty function approach,
*       with ordinary complementary slackness times mu in objective

        if(aon("a9"),

$include "init.gms"

*        b) Solve problem without complementarity constraints

            solve EstimPre using NLP minimizing z;

*        c) Iterate over increasingly better approximations

            mu = 0.1;
            zpen.l = 100;

            while(mu<1001,

*           d) Solve approximation. Stop if comp. gap < 0.01

                mu $ ((mu > 1000) or (zpen.l/mu lt 0.01)) = 100000;
```

```
                solve EstimPEN using NLP minimising zz;

                mu=mu*2;
            );

$batinclude "savesol.gms" a9
        );

*       -------------------- approach 10 ---------------------------------
*       a6 most often produces the best solution, but when it is NOT best,
*       it is much worse than a9 (penalty function). Thus, check which one
*       was the best and use that one.

        if(aon("a10"),

            if((aon("a6") and aon("a9")),
                if((objes("a9",m,n) lt objes("a6",m,n)),
                    objes("a10",m,n) = objes("a9",m,n);
                    pest(m,n,t,i,"a10") = pest(m,n,t,i,"a9");
                    cest(m,n,i,j,"a10") = cest(m,n,i,j,"a9");
                    sest(m,n,i,"a10")   = sest(m,n,i,"a9");
                else
                    objes("a10",m,n) = objes("a6",m,n);
                    pest(m,n,t,i,"a10") = pest(m,n,t,i,"a6");
                    cest(m,n,i,j,"a10") = cest(m,n,i,j,"a6");
                    sest(m,n,i,"a10")   = sest(m,n,i,"a6");
                );
            );
        );
    );
);

parameter objes1;
set objitems /pre2,dds,nds,ddx,ndx/;
objes1(m,n,a) = objes(a,m,n);
objes1(m,n,objitems) = objes(objitems,m,n);
display "All objective values transposed for readability:", objes1;
if(aon("a7"), display zenum);

* Rate solution approaches. Find the best solution, and give one point to
* the approach that found it

scalar bestz /0/;
parameter rating(*);
rating(aon) $ (not sameas(aon,"a0"))
= sum((m,n) $ (objes1(m,n,aon) lt (smin(aon2 $ (not sameas(aon2,"a0")),
    objes1(m,n,aon2)) + 0.0001)),1);

rating("total") = card(n)*card(m);
display rating;

option kill=rating;
option kill=objes;
option kill=objes1;

if(saveGDX,
execute_unload "smp.gdx" m n i t a aon ptru ctru stru etru pest cest sest);
```

Appendix 2.3. GAMS program for analysing numerical results

```
$ontext
    analyseSample.gms
    File for analysing results of estimations carried out with program
    generateSample.gms and stored in GDX-file smp.gdx.

    Torbjorn Jansson
    LEI, The Hague, NL
$offtext
$offlisting
$set TRAD a8
$set BLEP a10
$set OBS  a0

file con /con/;

set m  Models generated ;
set n  Estimation attempts per model;
set a  Solution approaches available in total;
set i  Regions ;
set t  Periods;
set aon(a) Solution approaches actually implemented;

alias(i,j,k);
alias(t,t1);

putclose con / "... Reading sets ..." /;

$gdxin "smp.gdx"
$load m n a i t aon

parameter ptru(m,t,i)     True price;
parameter ctru(m,i,j)     True transportation cost;
parameter stru(m,i)       True storage cost;
parameter etru(m,t,i)     True excess demand;

parameter pest(m,n,t,i,a) Estimated price;
parameter cest(m,n,i,j,a) Estimated transportation cost;
parameter sest(m,n,i,a)   Estimated storage cost;

variable VDUM; equation EDUM; EDUM .. VDUM =e= 10;
model MDUM 'Dummy model clears memory for some odd reason' /EDUM/;
MDUM.solprint = 2;

putclose con / "... Reading data ..." /;

$load ptru ctru stru etru pest cest sest
$gdxin

scalar count /0/;
scalar np Number of prices estimated;
scalar nc Number of transport costs estimated;
scalar ns Number of storage costs estimated;
scalar nn Number of iterations per model;

np = card(i)*card(t);
```

```
nc = card(i)*(card(i)-1)/2;
ns = card(i);
nn = card(n);

parameter iord(i) Ordering of the i set;
parameter nord(n) Ordering of the n set;

count=0;
loop(i, count=count+1; iord(i)=count);
count=0;
loop(n, count=count+1; nord(n)=count);

set AD(i,j) Admissible trade flows;
AD(i,j) = yes $ (iord(j) gt iord(i));

scalar son Storage on /0/;
son = card(t)-1;

set obs/o1*o10000/;
set uobs(obs);

* Some basic reporting to list file

option count:0;
count = card(m);
display "Number of true models in sample:",count;
count = card(n);
display "Number of solutions per model:",count;
count = card(t);
display "Number of time periods in storage model (1=no storage)",count;
count = card(i);
display "Number of regions in model",count;
count = sum((m,n,i,j) $ (AD(i,j) and (1 = cest(m,n,i,j,"%OBS%"))),1);
display "Number of truncated cost items:", count;
count = sum((m,n,t,i) $ ((139 = pest(m,n,t,i,"%OBS%"))
                      or (101 = pest(m,n,t,i,"%OBS%"))),1);
display "Number of truncated price items:", count;


* -------------------------------------------------------------------
*   Declare statistical measures to compute
* -------------------------------------------------------------------

set parvec /preg Regional price, tpc Transport cost, stc Storage cost/;
set meth /TRAD Traditional method , BLEP Bilevel Estimation Program/;

parameter pmean(m,t,i,a)    Sample mean;
parameter cmean(m,i,j,a);
parameter smean(m,i,a);

parameter psvar(m,t,i,a)    Sample variance;
parameter csvar(m,i,j,a);
parameter ssvar(m,i,a);

parameter ptval(m,t,i,a)    T-values for test estimated equals true;
parameter ctval(m,i,j,a);
parameter stval(m,i,a);
```

```
set alpha 'Levels of significanse' /alpha950,alpha990,alpha999/;
parameter pcrit(alpha) 'Critical test statistic value' /
    alpha950    0.05
    alpha990    0.01
    alpha999    0.001 /;

parameter tsummary(a,*);

set mse Items used for analysis of mean squared error /
    msetot  Mean squared error
    msebias Bias
    msevar  Variance    /;

set totmse Items used for analysis of mean squared error /
    MMSE    Mean MSE
    MSBIAS  Mean Squared Bias
    MVAR    Mean (pooled) Variance    /;

parameter pmse(m,t,i,a,mse)  Mean squared error per item;
parameter cmse(m,i,j,a,mse);
parameter smse(m,i,a,mse);

parameter psrmse(*,a,totmse)    Sum of root mean squared error;
parameter csrmse(*,a,totmse);
parameter ssrmse(*,a,totmse);

option pmse:3:3:2;
option cmse:3:3:2;
option smse:3:2:2;

option psrmse:3:1:2;
option csrmse:3:1:2;
option ssrmse:3:1:2;

parameter bartlett(m,a,parvec) Bartlett statistic equality of variances;
parameter levene(m,a,parvec)   Levene statistic equality of variances;

option bartlett:2:1:2;
option levene:2:1:2;

*   Items for GDX-rank

parameter pct(*) /median 50.0/;
parameter someSort1(*);
parameter someSort2(*);
parameter someRank(*);

* ---------------------------------------------------------------------
*   Compute mean and sample variance of estimated parameters
* ---------------------------------------------------------------------

putclose con / "... computing means ..." /;

pmean(m,t,i,aon)           = sum(n, pest(m,n,t,i,aon))/card(n);
cmean(m,i,j,aon) $ AD(i,j) = sum(n, cest(m,n,i,j,aon))/card(n);
smean(m,i,aon)   $ son     = sum(n, sest(m,n,i,aon))  /card(n);

putclose con / "... computing variances ..." /;
```

```
psvar(m,t,i,aon)           = sum(n, sqr(pest(m,n,t,i,aon)
                                    - pmean(m,t,i,aon)))/(card(n)-1);
csvar(m,i,j,aon) $ AD(i,j) = sum(n, sqr(cest(m,n,i,j,aon)
                                    - cmean(m,i,j,aon)))/(card(n)-1);
ssvar(m,i,aon)   $ son     = sum(n, sqr(sest(m,n,i,aon)
                                    - smean(m,i,aon)))  /(card(n)-1);

putclose con / "... computing t-values ..." /;

ptval(m,t,i,aon)            = (ptru(m,t,i)-pmean(m,t,i,aon))
                              /sqrt(psvar(m,t,i,aon)/card(n));
ctval(m,i,j,aon) $ AD(i,j)  = (ctru(m,i,j)-cmean(m,i,j,aon))
                              /sqrt(csvar(m,i,j,aon)/card(n));
stval(m,i,aon)   $ son      = (stru(m,i)  -smean(m,i,aon))
                              /sqrt(ssvar(m,i,aon)/card(n));


* ---------------------------------------------------------------------
*   Compute mean squared error and bias of estimators
* ---------------------------------------------------------------------

putclose con / "... computing MSE ..." /;

pmse(m,t,i,aon,"msebias")           = pmean(m,t,i,aon)-ptru(m,t,i);
cmse(m,i,j,aon,"msebias") $ AD(i,j) = cmean(m,i,j,aon)-ctru(m,i,j);
smse(m,i,aon,"msebias")   $ son     = smean(m,i,aon)  -stru(m,i);

option kill = ptru; option kill = ctru;

pmse(m,t,i,aon,"msevar")            = psvar(m,t,i,aon);
cmse(m,i,j,aon,"msevar") $ AD(i,j)  = csvar(m,i,j,aon);
smse(m,i,aon,"msevar")   $ son      = ssvar(m,i,aon);

option kill = psvar; option kill = csvar; option kill = ssvar;

pmse(m,t,i,aon,"msetot")            = pmse(m,t,i,aon,"msevar")
                                    + sqr(pmse(m,t,i,aon,"msebias")) ;
cmse(m,i,j,aon,"msetot") $ AD(i,j)  = cmse(m,i,j,aon,"msevar")
                                    + sqr(cmse(m,i,j,aon,"msebias")) ;
smse(m,i,aon,"msetot")   $ son      = smse(m,i,aon,"msevar")
                                    + sqr(smse(m,i,aon,"msebias"))   ;

putclose con / "... computing sum of root(MSE) ..." /;

psrmse(m,aon,"msbias") = sum((t,i), sqr(pmse(m,t,i,aon,"msebias")))
                         /(card(i)*card(t));
csrmse(m,aon,"msbias") = sum((i,j)$ AD(i,j),sqr(cmse(m,i,j,aon,"msebias")))
                         /(card(i)*(card(i)-1)/2);
ssrmse(m,aon,"msbias") = sum(i$ son, sqr(smse(m,i,aon,"msebias")))
                         /card(i);;

*   Compute pooled variance

psrmse(m,aon,"mvar") = sum((t,i),           pmse(m,t,i,aon,"msevar"))
                       /(card(i)* card(t));
csrmse(m,aon,"mvar") = sum((i,j) $ AD(i,j), cmse(m,i,j,aon,"msevar"))
                       /(card(i)*(card(i)-1)/2);
```

```
ssrmse(m,aon,"mvar") = sum( i    $ son,     smse(m,i,aon,"msevar"))
                       /card(i);

*   Compute sum of Root Mean Squared Error, SRMSE

psrmse(m,aon,"mmse")  = sum((t,i),           pmse(m,t,i,aon,"msetot"))
                        /(card(i)* card(t));
csrmse(m,aon,"mmse")  = sum((i,j) $ AD(i,j), cmse(m,i,j,aon,"msetot"))
                        /(card(i)*(card(i)-1)/2);
ssrmse(m,aon,"mmse")  = sum( i    $ son,     smse(m,i,aon,"msetot"))
                        /card(i);

*   Average over all models

psrmse("tot",aon,totmse)   = sum(m, psrmse(m,aon,totmse));
csrmse("tot",aon,totmse)   = sum(m, csrmse(m,aon,totmse));
ssrmse("tot",aon,totmse)   = sum(m, ssrmse(m,aon,totmse));


* ---------------------------------------------------------------------
*   Declare some parameters to hold values to plot
* ---------------------------------------------------------------------

parameter plotpmse(m,meth)  Plot of SRMSE for prices for Poster;
parameter plotcmse(m,meth)  Plot of SRMSE for costs for Poster;

parameter plotpvar(m,meth)    Plot of pooled price variance for Poster;
parameter plotcvar(m,meth)    Plot of pooled price variance for Poster;

parameter plotpbias(m,meth)    Plot of price ABIAS for Poster;
parameter plotcbias(m,meth)    Plot of cost ABIAS for Poster;

parameter plotphomo(obs,*)  Plot of individual price variance estimates;
parameter plotchomo(obs,*)  Plot of individual cost variance estimates;

plotpmse(m,"TRAD") = psrmse(m,"%TRAD%","mmse");
plotcmse(m,"TRAD") = csrmse(m,"%TRAD%","mmse");
plotpmse(m,"BLEP") = psrmse(m,"%BLEP%","mmse");
plotcmse(m,"BLEP") = csrmse(m,"%BLEP%","mmse");

plotpvar(m,"TRAD") = psrmse(m,"%TRAD%","mvar");
plotcvar(m,"TRAD") = csrmse(m,"%TRAD%","mvar");
plotpvar(m,"BLEP") = psrmse(m,"%BLEP%","mvar");
plotcvar(m,"BLEP") = csrmse(m,"%BLEP%","mvar");

plotpbias(m,"TRAD") = psrmse(m,"%TRAD%","msbias");
plotcbias(m,"TRAD") = csrmse(m,"%TRAD%","msbias");
plotpbias(m,"BLEP") = psrmse(m,"%BLEP%","msbias");
plotcbias(m,"BLEP") = csrmse(m,"%BLEP%","msbias");


* ---------------------------------------------------------------------
*   Analysis of biases
* ---------------------------------------------------------------------

putclose con / "... Price biases ..." /;

*   Compute descriptive statistics for biases
```

```
parameter biasstat(*,a,parvec) Sample mean and variances of biases;

*   ... average bias ...
biasstat("mean",aon,"preg") =
    sum((m,t,i), pmse(m,t,i,aon,"msebias"))
    /(card(m)*card(i)*card(t));

biasstat("mean",aon,"tpc") =
    sum((m,i,j)$AD(i,j), cmse(m,i,j,aon,"msebias"))
    /(card(m)*card(i)*(card(i)-1)/2);

*   ... sum of absolute biases ...
biasstat("SABIAS",aon,"preg") =
    sum((m,t,i), abs(pmse(m,t,i,aon,"msebias")));

biasstat("SABIAS",aon,"tpc")  =
    sum((m,i,j)$AD(i,j), abs(cmse(m,i,j,aon,"msebias")));

*   ... variance of biases ...
biasstat("variance",aon,"preg") =
    sum((m,t,i), sqr(pmse(m,t,i,aon,"msebias")
                     -biasstat("mean",aon,"preg")))
    /(card(m)*card(i)*card(t)-1);

biasstat("variance",aon,"tpc") =
    sum((m,i,j)$AD(i,j), sqr(cmse(m,i,j,aon,"msebias")
                             -biasstat("mean",aon,"tpc")))
    /(card(m)*card(i)*(card(i)-1)/2-1);

putclose con / "... computing medians ..." /;

*   ... median bias over k times m observations for prices (use gdx-rank)

count = 0;
loop(obs$sameas(obs,"o1"),
    loop((m,t,i),
        someSort1(obs+count) = pmse(m,t,i,"%TRAD%","msebias");
        someSort2(obs+count) = pmse(m,t,i,"%BLEP%","msebias");
        count=count+1;
    );
);
uobs(obs) = yes $ (ord(obs) lt count);

pct("median") = 50.0;
$libinclude rank.gms someSort1 uobs someRank pct
biasstat("median","%TRAD%","preg") = pct("median");
pct("median") = 50.0;
$libinclude rank.gms someSort2 uobs someRank pct
biasstat("median","%BLEP%","preg")  =pct("median");

*   ... median bias over k times m observations for costs (use gdx-rank)

count = 0;
loop(obs$sameas(obs,"o1"),
    loop((m,i,j)$AD(i,j),
        someSort1(obs+count) = cmse(m,i,j,"%TRAD%","msebias");
        someSort2(obs+count) = cmse(m,i,j,"%BLEP%","msebias");
```

```
        count=count+1;
    );
);
uobs(obs) = yes $ (ord(obs) lt count);

pct("median") = 50.0;
$libinclude rank.gms someSort1 uobs someRank pct
biasstat("median","%TRAD%","tpc") = pct("median");
pct("median") = 50.0;
$libinclude rank.gms someSort2 uobs someRank pct
biasstat("median","%BLEP%","tpc")  =pct("median");

display biasstat;


*   Summarize t-statistics
putclose con / "... computing t-tests for biases ..." /;

*   Double sided test if average price differs from zero, based on
*   asymptotic normality of averages (Lindberg-Levy Central Limit Theorem)

tsummary(aon,alpha) =
    sum((m,t,i) $ [2*(1-errorf(abs[ptval(m,t,i,aon)])) le pcrit(alpha)],1);
tsummary(aon,'neg') =
    sum((m,t,i) $ (ptval(m,t,i,aon) lt 0),1);

display "Number of rejections of null hypothesis for price estimates:"
        ,tsummary;

*   Test for average of cost estimates equals true cost

tsummary(aon,alpha) =
    sum((m,i,j)$ (AD(i,j)
             and [2*(1-errorf(abs[ctval(m,i,j,aon)])) le pcrit(alpha)]),1);
tsummary(aon,"neg") =
    sum((m,i,j)  $ (AD(i,j) and (ctval(m,i,j,aon) lt 0)),1);

display "Number of rejections of null hypothesis for cost estimates:"
        , tsummary;

option kill = tsummary;


* ---------------------------------------------------------------------
*   Analysis of variances
* ---------------------------------------------------------------------

putclose con / "... pooling variances ..." /;

parameter msestat(totmse,*,aon,parvec);

msestat(totmse,"mean",aon,"preg") = sum(m, psrmse(m,aon,totmse)) / card(m);
msestat(totmse,"mean",aon,"tpc")  = sum(m, csrmse(m,aon,totmse)) / card(m);

msestat(totmse,"variance",aon,"preg") =
    sum(m, sqr(psrmse(m,aon,totmse)-msestat(totmse,"mean",aon,"preg")))
    / (card(m)-1);
msestat(totmse,"variance",aon,"tpc")  =
```

```
    sum(m, sqr(csrmse(m,aon,totmse)-msestat(totmse,"mean",aon,"tpc")))
    / (card(m)-1);

display msestat;


* ---------------------------------------------------------------------
*   Tests:  Variances of prices are heterogeneous for TRAD and BLEP
*           Variances of costs are heterogeneous for BLEP but not TRAD
*       - Plot of variances
*       - Bartlett test
*       - Levene test (if normality does not hold)
* ---------------------------------------------------------------------

putclose con / "... making vectors suitable for plotting ..." /;

*   Put variances on parameter suitable for plotting

count = 0;
loop(obs$sameas(obs,"o1"),
    loop((m,t,i),
        plotphomo(obs+count,"TRAD") = pmse(m,t,i,"%TRAD%","msevar");
        plotphomo(obs+count,"BLEP") = pmse(m,t,i,"%BLEP%","msevar");
        count=count+1;
    );
);

count = 0;
loop(obs$sameas(obs,"o1"),
    loop((m,i,j)$AD(i,j),
        plotchomo(obs+count,"TRAD") = cmse(m,i,j,"%TRAD%","msevar");
        plotchomo(obs+count,"BLEP") = cmse(m,i,j,"%BLEP%","msevar");
        count=count+1;
    );
);


putclose con / "... testing for equality of variances (bartlett) ..." /;

bartlett(m,aon,"preg") =
    ((card(i)*card(t)*(card(n)-1))*log(psrmse(m,aon,"mvar"))
     -sum((t,i), (card(n) - 1)*log(pmse(m,t,i,aon,"msevar"))))
    /(1 + (1/(3*(card(i)*card(t)-1)))
        * (sum((t,i), 1/(card(n) - 1))
                    - 1/(card(i)*card(t)*(card(n)-1))));

bartlett(m,aon,"tpc")  =
    ((card(i)*(card(i)-1)/2*(card(n)-1))*log(csrmse(m,aon,"mvar"))
      -sum((i,j)$AD(i,j), (card(n) - 1)*log(cmse(m,i,j,aon,"msevar"))))
    /(1 + (1/(3*(card(i)*(card(i)-1)/2-1)))
        * (sum((i,j)$AD(i,j), 1/(card(n) - 1))
                            - 1/(card(i)*(card(i)-1)/2*(card(n)-1))));

parameter plevYi(m,t,i,a)    The Yi hat parameter in the levene statistic;
parameter plevZij(m,n,t,i,a) The Zij parameter in the levene statistic;
parameter plevZi(m,t,i,a)    The Zi hat parameter in the levene statistic;
parameter plevZ(m,a)         The Z hat parameter in the levene statistic;
```

```
parameter clevYi(m,i,j,a)    The Yi hat parameter in the levene statistic;
parameter clevZij(m,n,i,j,a) The Zij parameter in the levene statistic;
parameter clevZi(m,i,j,a)    The Zi hat parameter in the levene statistic;
parameter clevZ(m,a)         The Z hat parameter in the levene statistic;

*   Mean (1) or median (0) for levene measure? Median should be better for
*   skew distributions. Trial show no difference.

scalar levenemean "Use mean (1) or median (0)" /1/;
plevYi(m,t,i,aon) $ levenemean    = pmean(m,t,i,aon);
clevYi(m,i,j,aon) $ levenemean    = cmean(m,i,j,aon);

parameter levSort(n);
parameter levRank(n);

if((not levenemean), putclose con / "... computing medians ..." /);


loop((m,aon) $ (not levenemean),
    putclose con / "... medians for prices in model ", m.tl /;
    loop((t,i),
        pct("median") = 50.0;
        levSort(n) = pest(m,n,t,i,aon);
$libinclude rank.gms levSort n levRank pct
        plevYi(m,t,i,aon) = pct("median");
    );
    putclose con / "... medians for transport costs in model ", m.tl /;

    loop((i,j)$AD(i,j),
        pct("median") = 50.0;
        levSort(n) = cest(m,n,i,j,aon);
$libinclude rank.gms levSort n levRank pct
        clevYi(m,i,j,aon) = pct("median");
    );
);

putclose con / "... free some memory ..." /;
solve MDUM using NLP maximising VDUM;
putclose con / "... computing Levene statistics for prices ..." /;

plevZij(m,n,t,i,aon)  = abs(pest(m,n,t,i,aon)-plevYi(m,t,i,aon));
plevZi(m,t,i,aon)     = sum(n, plevZij(m,n,t,i,aon)) / nn;
plevZ(m,aon)          = sum((t,i), plevZi(m,t,i,aon)) / np;

putclose con / "... free some memory ..." /;
option kill = pest;
solve MDUM using NLP maximising VDUM;
putclose con / "... computing Levene statistics for costs ..." /;

clevZij(m,n,i,j,aon)  = abs(cest(m,n,i,j,aon)-clevYi(m,i,j,aon));
clevZi(m,i,j,aon)     = sum(n, clevZij(m,n,i,j,aon)) / nn;
clevZ(m,aon)          = sum((i,j)$AD(i,j), clevZi(m,i,j,aon)) / nc;

putclose con / "... free some memory ..." /;
option kill = cest;
solve MDUM using NLP maximising VDUM;
putclose con / "... finalize Levene statistics ..." /;
```

```
levene(m,aon,"preg") =
    ((np*(card(n)-1))
     *sum((t,i), card(n)*sqr(plevZi(m,t,i,aon)-plevZ(m,aon))))
    /((np-1)*sum((t,i,n), sqr(plevZij(m,n,t,i,aon)-plevZi(m,t,i,aon))));

levene(m,aon,"tpc")  =
    ((nc*(card(n)-1))
     *sum((i,j)$AD(i,j), card(n)*sqr(clevZi(m,i,j,aon)-clevZ(m,aon))))
    /((nc-1)*sum((i,j,n)$AD(i,j), sqr(clevZij(m,n,i,j,aon)
                                     -clevZi(m,i,j,aon))));

display bartlett,levene;

set homtest /bartlett,levene/;
parameter heterostats(*,alpha,parvec,aon) "Summary of Levene and Bartlett";

table critval(parvec,alpha,homtest)
                bartlett    levene
preg.alpha950   16.92       1.88
preg.alpha990   21.67       2.43
tpc.alpha950    61.66       1.39
tpc.alpha990    69.96       1.61    ;

heterostats("bartlett",alpha,parvec,aon) $ (not sameas(parvec,"stc")) =
   sum(m $ (bartlett(m,aon,parvec) ge critval(parvec,alpha,"bartlett")),1);

heterostats("levene",alpha,parvec,aon)  $ (not sameas(parvec,"stc"))  =
   sum(m $ (levene(m,aon,parvec) ge critval(parvec,alpha,"levene")),1);

display heterostats;
```

Chapter 3 Estimating prices, excess demand and trade costs in a spatial price equilibrium model[8]

1. Introduction

Spatial price equilibrium (SPE) models with homogeneous goods have been used in agricultural sector analysis at least since the publications of Judge and Wallace (1958) and Takayama and Judge (1964). This chapter is concerned with an SPE model for homogeneous primary agricultural products in Benin. The model has twelve regions (administrative departments) and seven goods (cassava, beans, rice, sorghum including millet, maize, yams and groundnut). For each region and product there is data on annual supply, demand and price, and there is also a table of distances between each pair of regions. The objective of the study is to estimate the regional net supply, net trade flows, prices and trade costs.

The trade cost minimization component of the model for each product k is put as a linear program in standard form,

$$\begin{aligned} \min_{x_k} \quad & c_k x_k \\ \text{subject to} \quad & A x_k \geq q_k \\ & x_k \geq 0 \end{aligned} \tag{1}$$

where c_k is a $1 \times n(n\text{-}1)$ vector of trade costs, x_k an $n(n\text{-}1) \times 1$ vector of trade flows, q_k is an $n \times 1$ vector of excess demand and A is an $n \times n(n\text{-}1)$ matrix of "0", "1" and "-1" arranged in such a way that for the i^{th} row, there is a "1" in all columns corresponding to flows into region i, a "-1" in all columns corresponding to flows out of region i and "0" elsewhere. This is not the usual symbolic representation of the transportation model, though it is mathematically equivalent. The representation is chosen because it points at the general applicability of the estimation method to a wider class of problems.

If a model is to be used in a positive way, it is desirable that it is capable of reproducing real world behaviour ex-post, and consequently we would like to interpret real world observations as model solutions. This requires that the ex-post data satisfies the Kuhn-Tucker (KT) conditions for an optimal solution to the transportation problem. For several reasons, the KT conditions are likely to be violated by

[8] A draft of this chapter was presented on the Ecomod conference on Regional and Urban Modeling, Brussels 2006.

ex-post data: The good may not be quite homogeneous; errors may arise when observations are aggregated over time and space, there can be measurement errors involved etc. Given such errors, and the chosen model, some calibration/estimation procedure is required in order to fit the model to the ex-post data.

Traditionally, the calibration of SPE models has been handled by solving the transportation model with observed or engineered trade costs, subject to market clearing constraints for given regional excess demand quantities, and using the Lagrange multiplicators associated with the market clearing constraints to determine the regional prices (e.g. Judge and Wallace (1958), Litzenberg, McCarl and Polito (1982), Peeters (1990), Kawaguchi, Suzuki and Kaiser (1997) and Guajardo and Elizondo (2003)). This implies that any disturbances of observed trade costs and excess demand are accepted, and that all corrections needed to satisfy the KT conditions are undertaken on the price positions, for which only a single observation is used (the numerator price).

Whereas this certainly may be a defensible way of proceeding in some instances, it is equally easy to imagine situations where there are observations of regional prices available and the observations of trade costs and excess demand are associated with errors. Then the traditional procedure described above is inefficient, because the price observations are ignored. It is also unable of identifying autarky regions; an observed nonzero regional excess demand, however tiny, enforces a fixed price difference (equal trade cost) to some other region.

A general approach to this type of estimation problem is to recognize that *given* a set of trade costs and regional excess demand (parameters), trade flows and regional price differences (variables) result from observing a solution of the transportation model. Thus, the estimation problem at hand is to select the parameters so that they, together with the solution variables of the transportation problem, minimize (or maximize) some estimation criterion. Viewed that way, the problem falls within the class of *bilevel programs*, prominently exemplified by the Stackelberg game. In terms of a leader-follower problem, the leader is the person conducting the estimation, the leader's cost function is the estimation criterion, his decision variables the parameters of the transportation problem. The follower's problem is the transportation model with parameters given by the leader.

In terms of a bilevel programming problem, the situation at hand is sometimes (e.g. in Dempe, 1997) described as the *optimistic* or *weak* approach. In economic terms that would mean that if the follower is indifferent regarding two solutions, he chooses the one preferred by the leader. In mathematical terms it means that if the solution of the inner problem is not a singleton but a set with several points, the leader is allowed to choose that value from the *set* of solutions of the inner problem that minimizes the estimation criterion. This property simplifies the solution of the bilevel estimation problem compared to the general bilevel program,

where the weak approach cannot be assumed *a priori*. In the case at hand, it has special implications for the prices, which are not fully identified in a solution to the transportation problem; only price differences are. The weak approach means that we, among all sets of prices satisfying those price differences, may choose those that score highest on our estimation criterion, e.g. are closest to the observed prices or have the highest probability density.

The outline is as follows: Section two formulates the bilevel estimation program, section three discusses alternative solution methods, and in section four the required weights are computed. Section five presents and analyses the data used in the estimation, and section six reviews other relevant studies of trade costs in Benin. In section seven, the results of the estimation are described, and section eight concludes with a discussion.

2. A bilevel estimation program

The mathematical representation of the bilevel programming problem in this application is based on a representation of the transportation problem by its first order conditions, with a weighted least squares objective function penalizing deviations from observations of prices and excess demand. The first order conditions here are cast as a *linear complementarity problem* (LCP), thus formulating the estimation problem explicitly as a *mathematical program with equilibrium constraints*—the branch of literature from which the solution method is borrowed.

$$\min_{q,p,c,x,\beta} \sum_k \left(w_k^p \left(p_k - p_k^o\right)' \left(p_k - p_k^o\right) + w_k^q \left(q_k - q_k^o\right)' \left(q_k - q_k^o\right) \right) \tag{2}$$

$$\text{subject to} \quad \begin{bmatrix} u_k \\ v_k \end{bmatrix} - \begin{bmatrix} 0 & -A' \\ A & 0 \end{bmatrix} \begin{bmatrix} x_k \\ p_k \end{bmatrix} = \begin{bmatrix} c_k' \\ -q_k \end{bmatrix} \tag{3}$$

$$\begin{bmatrix} u_k \\ v_k \end{bmatrix} \geq 0, \begin{bmatrix} x_k \\ p_k \end{bmatrix} \geq 0 \tag{4}$$

$$\begin{bmatrix} u_k \\ v_k \end{bmatrix}' \begin{bmatrix} x_k \\ p_k \end{bmatrix} = 0 \tag{5}$$

$$c_k' = \beta_1 \delta^{\beta_2} \tag{6}$$

The objective function (2) minimizes the weighted sum of squared deviations of estimated prices and excess demand from observations. If observations of trade flows and costs were available, those could be similarly included into the objective. Equations 3-6 form an LCP that is equivalent to the Kuhn-Tucker conditions for the LP (1), with p_k the dual vector of the constraints in the LP, and u_k and v_k slack vectors. Equation 6 is a function relating the trade cost between any two regions to the distance δ between them, parametrized by β_1 and β_2.

Trade costs are expressed per weight unit, and in order to economize on degrees of freedom, the trade costs per weight unit were assumed to be equal for all products. This would be reasonable if all products were equally perishable and with similar prices, which is not perfectly true for the set of products at hand: cassava, yams, maize, rice, sorghum, beans and peanuts. On the other hand, IFPRI (2004) does not find that traders in Benin discriminate between different agricultural products when setting transportation rates, supporting the use of a single trade cost function for all products.

3. Solution method

An optimization problem constrained by an LCP falls in the class of *mathematical programs with equilibrium constraints* (MPEC), that started to attract attention in the literature in the 1990's, evidenced by the publication of two books on the subject (Luo, Pang and Ralph (1996), Outrata, Kocvara and Zowe (1998)). The solver NLPEC (see NLPEC solver manual) for GAMS (General Algebraic Modelling System) solves MPECs via smooth reformulation of the complementarity constraints. Several of those reformulations were tested, and the method finally selected is one of the reformulations implemented in NLPEC. The intuition behind the smooth reformulations is the following:

The complementary slackness constraint (5) is the equation causing trouble when attempting to solve the problem (2-6), because it makes the feasible space non-convex and it has "corners". The key idea of the smooth reformulations is to replace (5) by a sequence of increasingly accurate approximations. Several such reformulations are available, and after extensive testing with synthetic data, a method where a penalty function minimizes the *complementarity gap* was chosen.

Before proceeding, we note that data is unlikely to support solutions with zero price for any product. Thus, the slack vectors v_k can be fixed to zero, reducing the problem somewhat. Next, the remaining complementary slackness condition $u_k' x_k = 0$ is removed, and instead a penalty term $\mu(u_k' x_k)$ is added to the objective function with μ a nonnegative real number. The resulting system is solved repeatedly, with μ initially set to a small number and then stepwise increased, each time using the previous solution as starting point, until the complementarity gap $u_k' x_k$ is zero. The estimation problem then is:

$$\min_{q,p,c,x,\beta} \sum_k \left(w_k^p \left(p_k - p_k^o\right)' \left(p_k - p_k^o\right) + w_k^q \left(q_k - q_k^o\right)' \left(q_k - q_k^o\right) + \mu\left(u_k' x_k\right) \right) \tag{7}$$

$$\text{subject to} \quad \begin{bmatrix} u_k \\ 0 \end{bmatrix} - \begin{bmatrix} 0 & -A' \\ A & 0 \end{bmatrix} \begin{bmatrix} x_k \\ p_k \end{bmatrix} = \begin{bmatrix} c_k' \\ -q_k \end{bmatrix} \tag{8}$$

$$u_k \geq 0, x_k \geq 0 \tag{9}$$

$$c_k' = \beta_1 \delta^{\beta_2} \tag{10}$$

with w_k^p and w_k^q weights to be defined below. Note that when $\mu \to \infty$, the factor $u'x$ in the penalty term eventually becomes zero (if feasible), so that the original problem is recovered. Testing with different sequences of μ and different synthetic data constellations revealed that this method is *not guaranteed* to find the global minimum for the problem size at hand[9]. Nevertheless, of the methods tested it performed on average and in median best, measured by the sum of squared deviations obtained, on a large number of similarly structured, randomly generated problems.

In an attempt to verify that the iterative approximation method finds the unique global minimum, or at least a point close to it, for the incumbent data, the problem was also reformulated as a mixed integer programming problem, with binary variables in a so-called "big M" construct switching the complementary slackness conditions. To reduce the size of the problem, only one product (maize) was included, and the problem initialised with the solution obtained by the iterative approximation described above. The so obtained problem in 132 binary variables was solved with a branch-and-bound algorithm (the solver SBB in GAMS on the NEOS server). The solver terminated after 38 minutes and 1.6 million iterations without any significant improvement of the objective, though still with a possible gap (between best found and best possible) of round 10% of the objective function value. For the entire problem (around 900 binary variables) the solver terminated due to limited system resources (memory). As a comparison, the iterative smooth approximation solution of the entire problem solves in about 20 seconds on a standard workstation. So, even if a better solution may exist, it is difficult to find.

The estimation was programmed in GAMS and the extremum optimisation problem solved numerically using the non-linear programming software CONOPT.

4. Assigning weights

The objective function of the problem (2-6) literally compares apples to pears. It actually does more than that, because it also weighs an error in the *price* of one commodity against the error in *quantity* of another. In order to make the estimator more efficient, the error terms need to be weighed by the inverse of their variances, which in this case are unknown.

In other circumstances, one approach would be to estimate the variances simultaneous with the parameters, either using maximum likelihood or by iteratively computing the sample variances from the residuals of previous estimation steps. To this end, one could assume that prices and quantities of each commodity

[9] Several studies formulates conditions for the existence of a global optimum for bilevel programs. ,Following Dempe (1997) a global minimum exists in the case at hand.

constitute two homoscedastic groups with variances σ_p^2 and σ_q^2 (or some more complex matrix function of those variances). Endogenously determined σ_p^2 and σ_q^2, would likely result in one group having variance close to zero and the other a very high variance. The reason for this is that either observed prices or quantities always can be matched perfectly by the estimates in this model. If the ratio of the variances σ_p^2/σ_q^2 is shifted towards zero, prices will be matched perfectly and the objective value be depending only on the inverse of σ_q^2, and vice versa.

Thus, some external source or assumption must be used to assign weights. In this analysis, we *assumed* that variances are proportional to the absolute size of the related variables. More specifically we assumed that the variances of prices are proportional to the observed price and to the inverse market share of the current region. The variances of excess demand were assumed to be proportional to the sum of regional supply and demand, the sum being motivated by the fact that the variance of a difference is the sum of the variances. The weights were computed as

$$w_{ik}^p = \frac{1}{p_{ik}} \frac{nd_{ik}}{\sum_j d_{jk}}, \quad w_{ik}^q = \frac{1}{d_{ik} + s_{ik}},$$

and the resulting weights are shown in tables 1 for prices and 2 for excess demand.

Table 1: Weights for price disturbances

	CASS	SORM	PULS	MAIZ	PEAN	RICE	YAMS
ALI	0.446	1.155	0.769	0.554	0.29	0.258	2.834
ATA	0.136	1.966	0.628	0.171	0.744	0.267	3.634
ATL	1.293	0.003	0.231	2.244	0.268	0.474	0.079
BOR	0.373	2.644	1.031	0.522	0.431	0.328	3.832
COL	0.589	0.215	0.56	1.497	0.383	0.335	0.065
COU	1.449	0.325	0.211	0.611	0.271	0.299	1.335
DON	0.059	0.813	0.359	0.118	0.818	0.227	15.644
LIT	0.924	0.002	0.129	0.683	0.223	0.998	0.067
MON	0.631	0.284	0.129	0.451	0.116	0.186	0.815
OUE	0.966	0.986	0.377	2.527	0.588	0.517	0.078
PLA	0.437	0.681	0.207	1.511	0.344	0.245	0.056
ZOU	0.666	0.232	0.614	1.463	0.441	0.373	0.048

Table 2: Weights for excess demand disturbances

	CASS	SORM	PULS	MAIZ	PEAN	RICE	YAMS
ALI	1.161	1.877	7.438	1.97	4.476	7.898	0.5
ATA	1.533	1.574	8.602	5.481	4.959	7.452	0.261
ATL	0.204	1230.172	21.374	0.74	12.518	5.863	13.186
BOR	0.668	1.484	5.19	1.282	7.614	6.748	0.179
COL	0.325	11.913	5.505	1.305	4.259	5.211	0.729
COU	0.411	13.756	11.086	2.404	6.704	9.198	0.874
DON	3.837	5.235	20.935	11.328	7.289	13.243	0.491
LIT	0.56	1399.573	49.265	3.394	18.692	2.68	15.357
MON	0.595	12.846	50.817	2.164	36.331	15.597	1.32
OUE	0.452	3.528	14.49	0.916	7.048	5.275	10.906
PLA	0.278	6.575	15.385	0.747	9.709	12.695	6.174
ZOU	0.567	20.176	6.503	1.401	3.718	8.555	3.402

5. Data

The data used in the estimation is the data in the BenImpact model data base for 2001. Aggregates of the data sufficient for reproducing the results of the estimations are presented in this section. In the BenImpact database, regional demand stems from the Benin statistic agency ONASA (several publications), as do regional prices. Regional supply is based on ONASA-data on yields and acreage on the level of the administrative units sub-prefectures (SP), of which there are 77 in Benin.

Both yields and acreages fluctuate strongly between adjacent time periods as well as regions. In order to arrive at reasonable estimates of yields at least on aggregate levels, the data for the individual SPs were scaled to fit yield and acreage trends estimated for “agri-ecological zones” (AEZ) based on survey data from van den Akker (2000). AEZ are eight agronomically homogeneous but spatially discontinuous geographical units, to which the SPs can be mapped. After fitting it to the AEZ data, excess demand was computed and aggregated to the administrative regional level of departments, which was the regional level used in the BenImpact model.

Prices are collected by ONASA at the level of selected market places, sometimes several times per year due to the seasonality of production. Since no data on market place or seasonal *turnover* was available, the prices for the departments were computed as the arithmetic average of the prices on market places that fall within that department. Having all this in mind, we conclude that data on prices as well as excess demand are subject to considerable uncertainty, and thus that the traditional calibration approach is likely to make inefficient use of the available information.

Figure 1 shows the regions used in the estimation and lists the abbreviations used for the twelve departments (Littoral, containing only the city Cotonou, was considered one department). Table 3 lists the abbreviations used for the seven products. A major crop that is missing is cotton, which was omitted from the estimations due to special regulated internal trading conditions. Data on prices and excess demand are shown in tables 4 and 5.

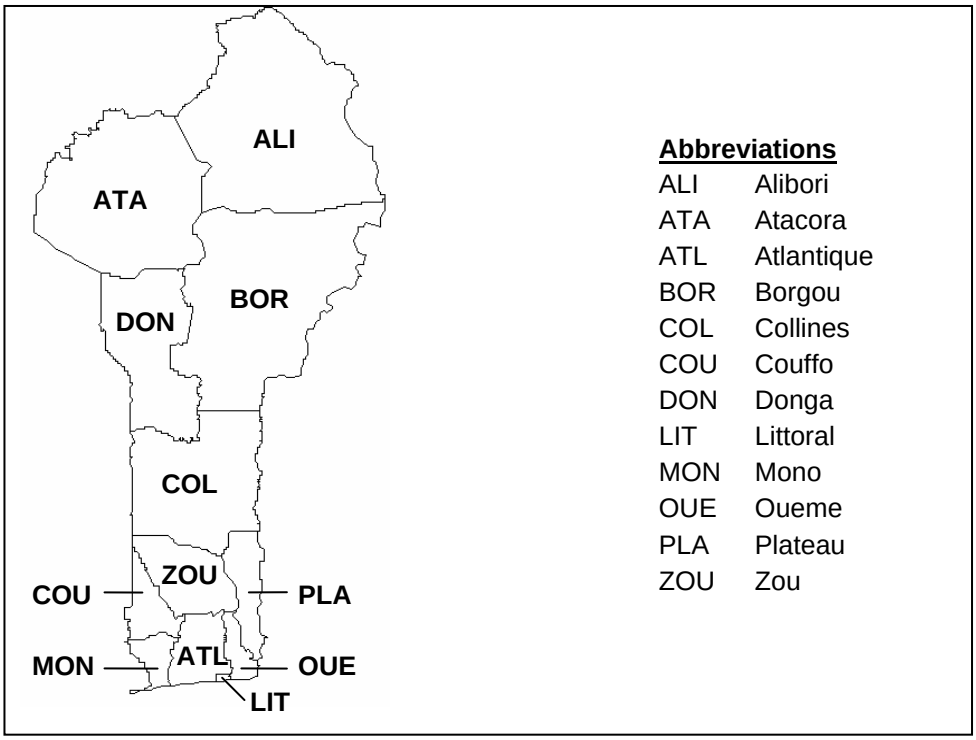


Figure 1: Regions (departments) of Benin in estimation. Source: The BenImpact mapping tool built by Wolfgang Britz (Bonn University), with own graphical modifications.

Table 3: Products in the estimation and their abbreviations.

CASS	Cassava	PULS	Beans	RICE	Rice
SORM	Sorghum and millet	MAIZ	Maize	YAMS	Yams
		PEAN	Groundnut		

Table 4: Price observations for regions and products

	CASS	SORM	PULS	MAIZ	PEAN	RICE	YAMS
ALI	172.42	107.06	219.74	88.52	250.45	266.31	73.93
ATA	171.70	113.21	182.53	96.45	225.54	266.36	72.83
ATL	156.99	186.65	269.81	104.10	286.02	277.70	136.28
BOR	182.98	112.39	231.28	91.68	236.80	291.55	59.56
COL	153.40	111.35	219.01	86.83	233.85	246.89	79.29
COU	117.67	163.74	224.43	85.91	197.73	273.59	119.74
DON	166.81	106.76	201.69	87.70	175.50	235.24	n.a.
LIT	159.55	205.29	288.96	127.37	364.50	288.62	138.55
MON	165.71	204.19	253.03	110.65	294.84	264.46	127.13
OUE	142.66	191.01	285.67	92.06	324.33	248.11	126.91
PLA	156.02	168.16	292.00	94.35	264.97	248.43	96.18
ZOU	149.85	132.42	219.66	101.41	220.39	237.59	118.71

Source: BenImpact database for 2001, based on data from ONASA.

Table 5: Regional excess demand observations

	CASS	SORM	PULS	MAIZ	PEAN	RICE	YAMS
ALI	60.76	-20.45	4.94	-17.54	-12.88	5.18	91.95
ATA	-20.77	-4.45	0.85	-7.10	1.65	5.04	-13.77
ATL	-102.96	0.08	2.10	23.07	1.97	17.04	7.43
BOR	-19.50	11.51	6.68	-45.59	0.13	9.99	-242.02
COL	-134.79	-2.04	-4.84	11.41	-11.84	2.25	-129.94
COU	82.68	6.86	-3.87	-6.08	-7.95	10.31	108.25
DON	-7.14	3.93	3.10	-1.85	4.95	6.29	57.79
LIT	103.32	0.06	2.03	29.46	5.22	37.31	6.51
MON	31.75	7.62	1.58	-12.44	1.68	6.36	68.67
OUE	42.39	21.64	4.81	48.30	10.63	14.26	4.56
PLA	-229.93	15.21	0.08	-37.34	1.55	7.88	-8.64
ZOU	14.57	3.20	-0.71	29.08	-14.27	11.27	-21.41

Source: BenImpact database for 2001, based on data from ONASA and van den Akker (2000).

The distances between departments required by (6) were computed using a table of line-of-sight-distances between the principal market places in each department. In praxis, sometimes different market places are important for different products, so that the selection of principal market places had to be a compromise if not one unique distance matrix was to be used for each product. The distance matrix used is shown in table 6.

Table 6: Distance matrix (kilometres)

	ALI	ATA	ATL	BOR	COL	COU	DON	LIT	MON	OUE	PLA	ZOU
ALI		300	574	213	425	592	347	628	631	658	570	502
ATA			485	212	336	503	78	539	542	569	481	413
ATL				361	149	102	407	54	63	84	140	72
BOR					212	379	134	415	418	445	357	289
COL						167	258	203	206	233	145	77
COU							425	106	39	136	158	90
DON								461	464	491	403	335
LIT									67	30	138	126
MON										97	197	129
OUE											108	156
PLA												68

Source: Own measurement on a map of Benin.

6. Results of other studies

There are other sources of trade cost estimates for Benin. Thus, before proceeding with a presentation of the estimation results, a brief survey is of interest. One recent study is IFPRI 2004, performing a survey of traders in Benin. They find that on distances of 160 km, large trucks are used, and that motorized transport on average costs 0.28 USD/ton/km. Converted to FCFA using an exchange rate of 700 FCFA/USD this corresponds to 31 FCFA per kg for 160 km. It is not clear to this author if those rates also contain mark-ups other than transportation costs.

Van den Akker (2000) surveys production and trade in Benin and finds transport costs for maize that, when fitted to the trade cost function used in this article, correspond to a distance elasticity of transport costs of 0.37, 0.71 and 0.41 for southern, central and northern Benin respectively (own computations). For a typical truck operated distance of 160 km, this amounts to transportation costs of 9.60, 17.92 and 10.51 FCFA per kg. These numbers are supposed to contain only transportation costs and not other costs connected to trade, whence we expect our estimated trade costs to be somewhat higher. For maize, van den Akker finds that marketing costs and profit each amount to approximately as much as the transportation costs.

Finally, there are estimates of distance elasticities of trade costs from other studies, prominently in the gravity literature. Hummels (1999) estimates a trade cost function similar to ours but with *ad valorem* trade costs and finds a distance elasticity of 0.27 (all products), and commodity specific elasticities “tightly clustered in the 0.2 to 0.3 range” (ibid p.11). Elasticities smaller than one indicate concave trade costs, which would be easily explained:

The main reason for expecting a concave trade cost function with a distance elasticity of less than unity is that trade takes place with a multitude of means,

ranging from transportation by foot over bicycles, motorcycles, modified ordinary automobiles, small trucks up to large trucks (IFPRI 2004), all with different fix charges and costs per km. If always the cheapest available means of transportation were used for a given haul, this would result in concave trade costs as illustrated in figure 2, where the heavy grey line shows the graph of a trade cost function as of equation 6. Most distances in the model are relatively long and therefore could be operated by a more homogeneous class of transportation means, allowing the function to be closer to linear, but never convex. Having all this in mind, we would expect our estimated trade cost function to be such that the elasticity is between 0.2 (the lowest value in Hummels) and 1.0 (the upper limit for concavity), and the function value for 160 km to be around 30 FCFA per kg (indicated by the IFPRI survey).

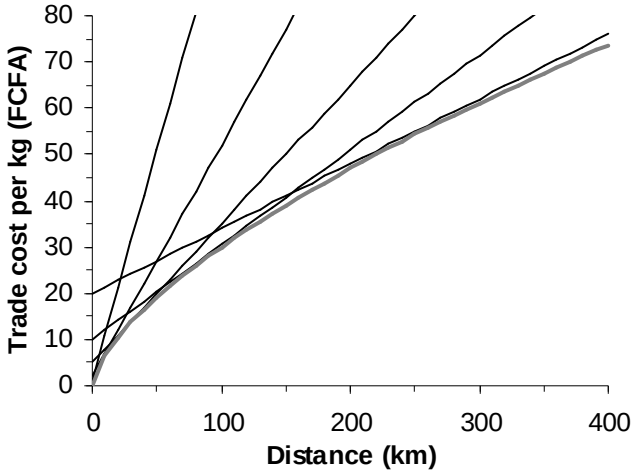


Figure 2: Concavity of trade costs resulting from a heterogeneous class of means of transportation.

7. Results of estimation

The fit of the estimates were evaluated using the R-squared measure, computed for each product separately according to the formula

$$R^2 = \frac{SSR}{SST} \text{ with } SST = \sum_r (\hat{p}_r - \bar{p}_r)^2 \text{ , } SSE = \sum_r (\hat{p}_r - p_r)^2$$

and *SSR* = *SST* – *SSE*, and the same for excess demand. Thus, no correction for degrees of freedom was undertaken. A bar denotes the sample mean and a hat denotes the estimated value. Tables 8 and 9 show the estimated prices and excess demand, table 7 the computed R-squared values.

For sorghum and millet, peanuts and yams, the R-squared measures indicate a fair fit, with errors distributed over both quantities and prices. This means that for those products, the price was *in general* higher in regions with positive excess demand and lower in regions with excess supply. A check with the observations of prices and net supplies supports this conclusion: Take for example sorghum and millet. Here, the R-squared measures are 0.71 for prices and 0.82 for quantities (table 7). The greatest net supplier of sorghum and millet was Alibori in the far north of Benin with a surplus of 20 450 tons (table 5). As expected, that region had a low price (table 4), a relation which remains in the estimates. The price or sorghum and millet in Donga was also low, but the excess demand positive and smaller than in Alibori. In that case, the estimator determined that Donga should be an autarky region. The largest net importers of sorghum and millet—Mono, Oueme and Plateau—also had among the highest observed prices.

Another example is yams. The majority of the yams trade according to observations originates in Borgou and Collines (table 5), where the prices also are lowest. The low prices enable them to export yams to the higher priced net demanding regions. Only in the northern region of Alibori a low observed price of yams is in contrast to a significant excess demand. In that case, the estimator chooses to adjust the price upwards in order to be able to import the required quantities.

Table 7: Measures of determination with all products included (P = price, Q = excess demand).

		mean	sse	sst	ssr	R2
CASS	P	157.98	5031	3113	-1918	-0.62
CASS	Q	-14.97	26	104029	104003	1.00
SORM	P	150.18	5163	17822	12658	0.71
SORM	Q	3.60	227	1250	1023	0.82
PULS	P	240.65	5588	14594	9006	0.62
PULS	Q	1.40	111	129	18	0.14
MAIZ	P	97.25	2217	1623	-594	-0.37
MAIZ	Q	1.12	464	8721	8256	0.95
PEAN	P	256.24	17654	31893	14238	0.45
PEAN	Q	-1.60	200	719	519	0.72
RICE	P	262.07	5306	3924	-1382	-0.35
RICE	Q	11.10	336	940	605	0.64
YAMS	P	104.46	2249	8280	6030	0.73
YAMS	Q	-5.89	234	104109	103875	1.00

Source: Own estimations.

The R-squares for the three products cassava, maize and rice the R-squares for prices turn out negative, indicating a poor support in the data for the assumed model. Indeed, a look at the data reveals that for those products, excess demand is sometimes positive where the price is low and vice versa. One example where

data is contradicting the assumption of arbitraging traders is the case of cassava, especially in the region of Borgou. In that region, the data claim an excess supply of 19 500 tons (table 5), which has to be exported, but at the same time, Borgou has the highest cassava price of all regions (182.98 FCFA/ton, table 4). This is clearly not consistent with the idea that the cassava surplus is traded to some other region by profit maximizing traders facing a positive unit trade cost and homogeneous goods. The estimator thus chooses the "smallest possible" correction that brings quantities and prices in line with the model. This turns out to be a strong correction of the price and little correction of the excess demand. The estimated price is 157.95 FCFA/ton (table 8), implying an error of about 25 FCFA/ton, whereas the estimated excess demand (-20 420 tons) is closer to the data. In general, the estimator favoured price corrections rather than quantity corrections for cassava, as evidenced by the R-squared measures in table 7 (1.00 for quantities, -0.62 for prices).

Table 8: Price estimates with all products included (FCFA/kg)

	CASS	SORM	PULS	MAIZ	PEAN	RICE	YAMS
ALI	189.21	88.06	219.74	87.42	250.45	266.31	87.24
ATA	145.18	113.21	192.91	87.27	218.92	262.73	72.83
ATL	148.70	172.30	256.64	109.14	267.20	263.46	108.96
BOR	157.95	119.32	224.03	56.16	227.14	293.84	55.99
COL	126.84	150.43	234.77	87.27	245.33	262.73	87.10
COU	151.34	174.94	241.67	94.17	259.57	271.09	111.61
DON	138.28	112.57	204.36	75.83	207.47	274.17	75.65
LIT	156.63	170.83	248.71	109.73	275.13	255.54	116.89
MON	157.07	180.66	247.39	99.89	265.29	265.37	117.33
OUE	152.22	175.23	253.12	105.32	279.53	259.94	112.49
PLA	136.37	171.71	256.05	89.47	266.61	275.79	96.64
ZOU	138.14	161.73	246.07	98.57	256.63	274.03	98.40

Table 9: Excess demand estimates (1000 ton, empty = autarky)

	CASS	SORM	PULS	MAIZ	PEAN	RICE	YAMS
ALI	60.22	-25.34					89.93
ATA	-21.17		-2.62			-2.21	
ATL	-99.99	0.07	4.79	18.69	1.50	21.49	7.35
BOR	-20.42	5.33	0.94	-48.13		4.66	-247.69
COL	-136.70	-2.81	-2.36	8.92	-13.21	-4.66	-131.33
COU	81.18	6.19	-1.99	-7.43	-8.81	13.15	107.09
DON	-7.30		1.68			2.21	55.73
LIT	104.40	0.09	3.20	28.51	4.91	47.04	6.45
MON	30.71	10.59	1.99	-13.94	1.52	8.03	67.90
OUE	43.73	32.48	8.77	43.38	9.80	19.21	4.46
PLA	-227.74	13.81	0.96	-43.38	0.95	9.93	-8.80
ZOU	13.48	2.74	1.39	26.76	-15.83	14.32	-21.71

The estimated parameters of the trade cost function turn out (β_1,β_2) = (0.147,1.000) corresponding to a trade cost at 160 km of 23.5 FCFA/kg. The cost for 160 km is of plausible magnitude, judged by the results of van den Akker and IFPRI discussed in the previous section. The distance elasticity β_2 of unity, implying a linear trade cost function, appears high compared to the considerations expressed in the previous section, implying linear trade costs. However, one should keep in mind that most distances involved are of similar magnitude since all regions are relatively large. In other words, if there would also have been a number of very small regions, close by one another, involved, then the observations would have contained more information about how the trade costs vary with varying distance. Indeed, most distances in table 6 are large enough to be covered by similar means of motorized transport.

Finally, in order to investigate the sensitivity of the estimates to changes in data, an additional estimation was performed, using only products for which R-squared was positive, thus not contradicting the assumed model. That meant omitting cassava, maize and rice. That results in parameter estimates of (β_1,β_2) = (1.558,0.643), corresponding to a trade cost for 160 km of 40.7 FCFA per kg, the estimated prices and excess demand shown in tables 10 and 11, and the measures of determination shown in table 12. Those estimates, as far as the trade cost function is concerned, is more in line with other results. A conclusion must be that the estimation is not very robust, and that there possibly are problems with the data, the model or both.

Table 10: Price estimates with product subset (FCFA/kg, n.a. = not available)

	CASS	SORM	PULS	MAIZ	PEAN	RICE	YAMS
ALI	n.a.	90.20	219.74	n.a.	250.45	n.a.	95.93
ATA	n.a.	113.21	182.56	n.a.	225.54	n.a.	72.83
ATL	n.a.	182.73	265.60	n.a.	270.65	n.a.	112.77
BOR	n.a.	114.13	231.34	n.a.	236.80	n.a.	47.01
COL	n.a.	144.11	226.72	n.a.	233.77	n.a.	74.69
COU	n.a.	184.59	235.12	n.a.	249.96	n.a.	116.53
DON	n.a.	108.61	208.21	n.a.	201.90	n.a.	83.32
LIT	n.a.	165.28	245.35	n.a.	281.20	n.a.	122.12
MON	n.a.	188.54	251.55	n.a.	266.39	n.a.	122.47
OUE	n.a.	179.16	259.23	n.a.	285.59	n.a.	116.73
PLA	n.a.	182.32	264.92	n.a.	269.77	n.a.	85.11
ZOU	n.a.	158.84	241.44	n.a.	246.29	n.a.	88.41

Table 11: Excess demand estimates, product subset (1000 ton, empty = autarky)

	CASS	SORM	PULS	MAIZ	PEAN	RICE	YAMS
ALI	n.a.	-20.24		n.a.		n.a.	89.93
ATA	n.a.		-2.62	n.a.		n.a.	
ATL	n.a.	0.08	4.79	n.a.	1.50	n.a.	7.35
BOR	n.a.		0.94	n.a.		n.a.	-247.69
COL	n.a.	-2.00	-1.33	n.a.	-13.21	n.a.	-131.33
COU	n.a.	6.89	-1.99	n.a.	-8.81	n.a.	107.09
DON	n.a.		1.68	n.a.		n.a.	55.73
LIT	n.a.	0.09	3.20	n.a.	4.91	n.a.	6.45
MON	n.a.	10.59	1.99	n.a.	1.52	n.a.	67.90
OUE	n.a.	32.48	8.77	n.a.	9.80	n.a.	4.46
PLA	n.a.	15.27	1.33	n.a.	0.95	n.a.	-8.80
ZOU	n.a.			n.a.	-15.83	n.a.	-21.71

Table 12: Measure of determination with product subset

		mean	sse	sst	ssr	R2
SORM	P	150.18	4698	17822	13123	0.74
SORM	Q	3.60	304	1250	945	0.76
PULS	P	240.65	4045	14594	10549	0.72
PULS	Q	1.40	114	129	15	0.12
PEAN	P	256.24	13605	31893	18288	0.57
PEAN	Q	-1.60	200	719	519	0.72
YAMS	P	104.46	2661	8280	5618	0.68
YAMS	Q	-5.89	234	104109	103875	1.00

The data and GAMS-code required for reproducing the results presented here can be obtained for test purposes from the author upon request.

8. Discussion

The estimations above show that the proposed method for estimating trade costs is computationally feasible, and theoretical considerations indicate that it is superior to traditional calibration methods from an efficiency point of view. The estimation results were, however, not stable when the composition of the sample was modified, as demonstrated by the rather different results obtained when using a subset of the products. This signals that there are problems with the model specification, with the data or both. The list of potential specification errors is long:

(i) Lack of temporal disaggregation. In the tropical country of Benin, there are two production seasons, with somewhat different time windows in the south and in the north. Thus, production, demand and trade takes place within shorter time frames and trade flows may even reverse within a year. Van den Akker uses four time periods. This setup was tried but discarded, as it requires the estimation of a storage function and urges considerations of uncertainty.

(ii) Products may not be homogeneous. For some products there are local as well as commercial varieties, which may sell at different prices.

(iii) Great circle distances neither reflect the state of maintenance and other qualities of the road network nor actual distances of road to travel.

(iv) Congestion effects are not considered. According to IFPRI, congestion in the transportation system sometimes occur during cotton harvest.

(vi) External trade, occurring via the harbour in Cotonou or across the borders to neighbour countries, is not considered due to lack of data on prices and quantities. The only external trade allowed is by a constant, derived from the sample, and attributed entirely to Cotonou.

The data problems for a country suffering under deficits in all kinds of infrastructure and low literacy (33.6% according to CIA 2006) are obvious. Official statistics frequently appear to be more "guesstimates" than the results of actual measurements, and utterly sparse. The data used in the estimations has gone through a gap-filling process already before entering the estimation, and is still not complete (one price is missing for yams). For some products and regions there are obvious problems. As an example, the cassava price in the department Couffo is the lowest of all regions even though Couffo has the second to largest excess demand, clearly contradicting the assumption of spatial arbitraging. Indeed, the coefficient of correlation between prices and excess demand is negative (-0.138) for cassava, though positive for all other products. However, trade does occur and is not likely to take place at a loss, so this is more likely to be a data than a specification problem.

To conclude, the estimation is on the one hand connected with some severe difficulties. Not only is the available data barely supporting the assumed model, but furthermore, the bilevel program is difficult to solve due to non-convexities. On the other hand, the method is workable, delivering reasonable estimates compared to expert knowledge and other trade cost studies. Furthermore, for the given model, it is difficult to see how the available data could be used more efficiently. The estimation uses all available information, and, by the proper use of weights, attaches more confidence to items that for some reason are believed to be more certain (have less variance).

The bilevel programming approach to the estimation of constrained programming models can, as noted above, be extended to include observations also of trade flows and trade costs in a similar manner, as well as a time series of observations. One could also attempt similar (bilevel) techniques to estimate parameters of general LP models, or, with additional second order conditions to NLP models, and could thus be of interest to a wider range of modellers as an alternative to separate estimations or calibration methods.

References

van den Akker, E. (2000). Makroökonomische Bewertung der Auswirkungen von technischen und institutionellen Innovationen in der Landwirtschaft in Benin. Hohenheim University Dissertation. Grauer, Stuttgart.

CIA 2006. The world factbook online, (downloaded in March 2006) (http://www.cia.gov/cia/publications/factbook/index.html).

Dempe, S. (1997). First-order necessary optimality conditions for general bilevel programming problems: Technical comment. *Journal of Optimization Theory and Applications*, 95: 735-739.

Guajardo, R.G., and H.A. Elizondo (2003). North American tomato market: a spatial equilibrium perspective. *Applied Economics*, 35: 315-322.

Hummels, D. (1999). Toward a geography of trade costs. Work. pap. University of Chicago.

IFPRI (2004). Benin: small farmer survey, 1998. Final report, Volume 2, Results of the trader and market surveys. International Food Policy Research Institute, Washington D.C.

Judge, G.G., and T.D. Wallace (1958). Estimation of spatial price equilibrium models. *Journal of Farm Economics*, 15: 801-820.

Kawaguchi, T., Suzuki, N., and H.M. Kaiser (1997). A spatial equilibrium model for imperfectly competitive milk markets. *American Journal of Agricultural Economics*, 79: 851-859.

Litzenberg, K., McCarl, B.A., and J. Polito (1982). A decomposition solution for spatial equilibrium problems. *American Journal of Agricultural Economics*, 64: 590-594.

Luo, Z., Pang, J., and D. Ralph (1996). *Mathematical programs with equilibrium constraints*. Cambridge University Press, Cambridge, UK.

NLPEC solver manual. GAMS Corporation, Washington D.C. http://www.gams.com/dd/docs/solvers/nlpec.pdf

Office National d'Appui à la Sécurité Alimentaire (ONASA). Rapport d'evaluation de la campagne agricole et les perspectives alimentaires pour 2001 au Benin, Situation par departement, Volume II, several editions.

Outrata, J., Kocvara, M., and J. Zowe (1998). *Nonsmooth approach to optimization problems with equilibrium constraints*. Kluwer academic publishers, Dordrecht.

Peeters, L. (1990). A spatial equilibrium model of the EU feed grain sector. *European Review of Agricultural Economics*, 17: 801-820.

Takayama, T., and G.G. Judge (1964). Spatial equilibrium and quadratic programming. *Journal of Farm Economics*, 46: 67-93.

Chapter 4 A Bayesian alternative to generalized cross entropy solutions for underdetermined models [10]

1. Introduction

In 1996, Golan, Judge and Miller published a book on "Maximum Entropy Econometrics" introducing Generalized Maximum Entropy (GME) and Generalized Cross Entropy techniques (GCE) to a wider range of applied econometricians. These estimation approaches were attractive to empirical modelers mainly for two reasons: First, they allow empirical specification and estimation of underdetermined models, i.e. models where the number of unknowns is larger than the number of equations, a capability not provided by classical solution techniques. Second, prior information on model unknowns can be included in a technically straightforward way, making estimates potentially more efficient in a mean square error sense, or at least more "plausible" for model simulation, interpretation, and analysis subsequent to estimation..

Since their introduction, a notable number of applications of GME and GCE have appeared in the empirical economics literature. A significant area of application relates to balancing large raw data sets using accounting identities and prior information to fill gaps and reconcile conflicting data sources. The techniques allow setting ranges for missing data values and provide a means of differentiating the reliability of various sources in the balancing process (e.g. Robinson, Cattanbo, and El-Said 2000; Britz and Wieck 2002, Robilliard and Robinson 2003). A related line of work deals with allocating input quantities to outputs from data on total input use and prior information on the input-output relationships (e.g. Lence and Miller 1998a and b, Léon et al. 1999). Calibration of simulation models to base year quantities and theory-consistent parameter sets is often done using entropy methods (e.g. Paris and Howitt 1998; Witzke and Britz (1998); Paris 2001) and a fairly new but increasingly important area is the spatial disaggregation of technological and economic data (Howitt and Reynaud 2003). However, GME and GCE applications are not reserved for data recovery and calibration issues, and have been employed in attempts to better solve traditional estimation problems or analyze new ones (e.g. Golan, Judge and Perloff 1996; Oude Lansink 1999; Zhang and Fan 2001; Arndt, Robinson, and Tarp (2002), Heckelei and

[10] Authors: Thomas Heckelei (Bonn University), Ron Mittelhammer (Washington State University), Torbjörn Jansson (Bonn University).

Wolff 2003). In essence, any economic model characterized by a vector of M equations in K > M unknowns, say $\mathbf{g}(\mathbf{z}) = \mathbf{0}$, is an underdetermined model that can be solved through the use of GME or GCE techniques.

Despite the growing number of applications, GME and GCE techniques are arguably subject to at least three difficulties, the first being the specification and interpretation of prior information imposed via the use of discrete support points and a corresponding reference prior probability distribution on that support. In fact, the actual prior information ultimately imposed is a rather complicated composite of the choice of support points, the choice of reference prior probabilities on support points, and their interaction with the criterion of maximum entropy in determining the final estimated subject probabilities on the support points. A second issue relates to challenges in characterizing the nature of the estimation objective that is actually being used to combine prior and data information, with attendant difficulties in evaluation of the estimation results by the scientific community. Thirdly, the entropy approach introduces additional variables (the probabilities linked to the supports) and equations (adding up constraints for the probabilities) to the estimation process, which leads to a potential computational challenge especially for large data balancing applications. We elucidate as well as address these issues in the sections ahead.

The overall objective of this chapter is to introduce a Bayesian alternative to GME and GCE techniques that allows for a direct and straightforwardly interpretable formulation of prior information and a clearly defined estimation objective while also reducing computational demands considerably when estimating an underdetermined economic model. The specific objectives are reflected in the organization of the remaining sections of the paper, which is as follows. Section 2 reviews the GME-GCE approach in the context of estimating an underdetermined linear model without noise. We clarify the interpretation of the effective prior information imposed as being a combined effect of supports, reference probabilities on supports, and the solution for the subject probabilities via the maximum entropy criterion. Section 3 introduces a formulation of the underdetermined linear model estimation problem using a Bayesian approach that is fully equivalent to GME-GCE, where the underdetermined model equations and the data together represent the "Likelihood" information and all prior information is represented in terms of a prior density on model unknowns. This approach is then extended to solving general systems of underdetermined equations. In section 4, the approach is extended to accommodate the situation where the prior information is completely uninformative relative to the unknowns in the equation system. Section 5 provides illustrative applications, followed by concluding remarks.

2. Prior information in GME-GCE approaches

The principles of GME (later extended to GCE) estimation as introduced by Golan, Judge and Miller (1996) and discussed further in Mittelhammer, Judge, and Miller (2000) are briefly reviewed here in the linear model context without noise to provide a conceptual foundation and identify notation for use in later sections. Within this basic model context, we elucidate the actual nature of the prior information that is implicitly used in the GME and GCE approaches.

Consider the underdetermined linear regression model, without noise, given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} \tag{1}$$

where **y** is a T-dimensional column vector of observations on the dependent variable, **X** is a T×K matrix of observations on independent regressors with T<K, and **β is** a K-dimensional column vector of unknown parameters. The values of **β** cannot be uniquely identified with classical estimation techniques, such as ordinary least squares, because the number of observations is smaller than the number of parameters. The basic GME approach is to "reparameterize" the vector of parameters **β** such that each element is expressed as an expectation of a discrete probability distribution. Let **S** be a block-diagonal K×KL matrix of *support points*, where L is the number of support points associated with each parameters, and let **p** be a corresponding KL×1 vector of weights that have the properties of probabilities. The vector **β** can then be represented as

$$\boldsymbol{\beta} = \mathbf{S}\mathbf{p} = \begin{bmatrix} \mathbf{s}_1' & 0 & \cdots & 0 \\ 0 & \mathbf{s}_2' & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{s}_K' \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_K \end{bmatrix} \tag{2}$$

with $\mathbf{s}_k' = \begin{bmatrix} \mathbf{s}_{k1} & \mathbf{s}_{k2} & \cdots & \mathbf{s}_{kL} \end{bmatrix}$ such that $s_{k1} < s_{k2} < \ldots < s_{kL}$. A reparameterized version of (2) is then given by

$$\mathbf{y} = \mathbf{X}\mathbf{S}\mathbf{p} \tag{3}$$

which corresponds to the 'data constraints' of the GME approach. Realizing that the elements of each $\mathbf{p}_k$, k = 1,...,K sum to 1 consistent with their interpretation as having the properties of probabilities, equation (2) defines the *admissible values* for the elements of **β** as convex combinations of the corresponding support points $\mathbf{s}_k$, k = 1,...,K. This implies that the range of possible values for $\boldsymbol{\beta}_k$ is given by the interval $[s_{k1}, s_{kL}]$. The GME approach chooses among the infinite number of vectors **p** satisfying (3) so as to maximize the entropy criterion[11]

[11] The value of $p_{ij} \ln\left(p_{ij}\right)$ is defined to equal its limiting value of 0 when $p_{ij} = 0$

$$H[\mathbf{p}] = -\mathbf{p}' \ln \mathbf{p} \tag{4}$$

The objective function (4) attains an unconstrained maximum when all elements of **p** have the value 1/L, i.e. when the probabilities are uniform. Since the uniform distribution treats each outcome as equally likely one can view this distribution as the maximally uninformative distribution with respect to anticipating outcomes of a random variable. 'Thus, the *maximum value of entropy* is uniquely associated with the *maximally uninformative weight- probability distribution*" (Mittelhammer et al. 2000, E3: 8). However, the notion of "uninformative" probabilities has caused some confusion in some applications of GME in that it has been incorrectly interpreted as characterizing the prior probabilities associated with various possible values of the parameters in the GME problem formulation. We will address this issue in more detail shortly.

The complete estimation problem can now be stated as

$$\begin{aligned} \max_{\mathbf{p}} \quad & H(\mathbf{p}) = -\mathbf{p} \ln \mathbf{p} \\ \text{subject to} \quad & \mathbf{y} = \mathbf{XSp} \\ & \boldsymbol{\iota}'\mathbf{p}_k = 1 \quad \forall\, k \end{aligned} \tag{5}$$

where the last constraint ensures that the probabilities appropriately sum to one, with $\boldsymbol{\iota}$ being a L×1 'summation vector', i.e. a conformable vector of ones. The values of $\boldsymbol{\beta}$ can be recovered after optimization by the definition given in (2).

A crucial question for interpreting the results of the GME estimation approach is how one can interpret the notion of "uninformative" claimed above for the entropy criterion in the GME context. Of principal interest is the interpretation of the *expectation* of the probability distribution over the support points, since it is this expectation that represents the final estimate of the parameter vector $\boldsymbol{\beta}$, as defined in (4). The probability distributions inherent in the solved value of **p** merely serves as a vehicle for the entropy criterion to choose particular values of the expectation that maximize entropy. Or as Preckel (2001, p. 375) states: "*Thus, the role of the distribution is simply to serve as intermediary in expressing the desirability of the value of a parameter*…".

Preckel reinterprets GME as minimizing a penalty function on these expectations subject to the data constraints., and compares the approach to the case of the penalty function implied by a least squares criterion. We instead conceptualize the GME-implied weighting on expectations as the prior probability distribution in a Bayesian context. This prior density turns out to be a reflection of Preckel's penalty function (see his equation (5), p. 368).

For an explicit illustration of the implied prior, consider just one parameter β_k from the linear model in (1) and suppose that only two support points s_{k1} and s_{k2} are used, i.e. L=2. Recalling that $p_{k1} + p_{k2} = 1$ we write the expectation of β_k as

$$E\beta_k = p_{k1}s_{k1} + (1-p_{k1})s_{k2} \tag{6}$$

Solving for the probability as a function of $E\beta_k$ obtains

$$p_{k1}(E\beta_k) = (E\beta_k - s_{k2})/(s_{k1} - s_{k2}) \tag{7}$$

The component of the entropy criterion in (5) relating to the expectation of β_k can then be expressed as

$$\begin{aligned} H(E\beta_k) &= -p_{k1}\ln(p_{k1}) - (1-p_{k1})\ln(1-p_{k1}) \\ &= -(E\beta_k - s_{k2})/(s_{k1} - s_{k2})\ln\left((E\beta_k - s_{k2})/(s_{k1} - s_{k2})\right) \\ &\quad -(s_{k1} - E\beta_k)/(s_{k1} - s_{k2})\ln\left((s_{k1} - E\beta_k)/(s_{k1} - s_{k2})\right) \end{aligned} \tag{8}$$

which defines the prior weight that the entropy criterion assigns to each possible value of the expectation of β_k. The criterion is maximized if the distance of $E\beta_k$ from the lower support point s_{k1} is equal to the distance of $E\beta_k$ from the upper support point s_{k2}, which coincides with $p_{k1} = p_{k2} = 0.5$, i.e. a uniform distribution over the supports, and a value for $\beta_k = (s_{k1} + s_{k2})/2$. All other values of $E\beta_k$ are assigned lower prior weights via $H(E\beta_k)$. A graphical illustration of the weight distribution is given in Figure 1, where we chose $s_{k1} = 0$ and $s_{k2} = 10$.

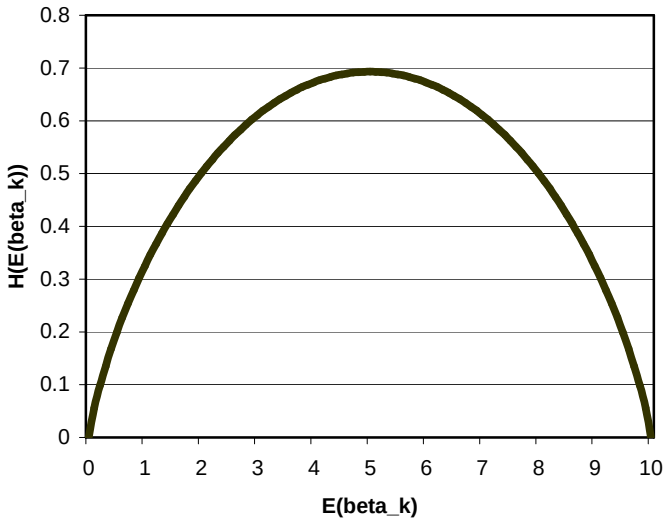


Figure 1: Prior weighting of parameter expectations based on the entropy criterion. Source: Maximum entropy calculations with $s_{k1} = 0$, $s_{k2} = 10$.

The mathematical and graphical illustration above demonstrates that the use of the maximum entropy criterion implies different prior weights on the different

possible outcomes, $E\beta_k$, of the GME estimator. These are prior weights in that they are independent of any the data information. The highest weight is given to the parameter expectation that would be generated by a uniform probability distribution over the supports.

The GME approach is a special case of GCE, where the latter method allows defining a reference probability distribution over the support points. Denoting the vector of reference distribution probabilities as **q**, the cross entropy criterion can be written as

$$I(\mathbf{p},\mathbf{q}) = \mathbf{p}'\ln(\mathbf{p}/\mathbf{q}) \tag{9}$$

where **p**/**q** is to be interpreted as a vector with elements p_{sk}/q_{sk}. The value of I(**p**,**q**) is smallest if all elements of the vector **p** are equal to the corresponding elements of the vector **q**. Consequently, an unconstrained *minimization* of the cross entropy measure over **p** will result in a probability distribution equal to **q**, and provides estimates of parameters according to expectations implied by the probabilities in **q**. The GME approach considered above is equivalent to an application of the GCE approach with a uniform *reference* distribution.

The use of a non-uniform reference distribution leads to modifications in the implicit prior weighting on parameter expectations under the GCE approach. Without repeating what amounts to a similar mathematical derivation to that in (6)-(8), we illustrate in Figure 2 the impact on the prior weights for the two support points example above. The reference probabilities were chosen such that q_{k1} = 0.3 and q_{k2} = 0.7. Note that we reflected the cross entropy value—which is minimized rather than maximized as in the GME case—around 0.6 to make the graph more easily comparable to Figure 1. In this case the highest cross-entropy weight is given to $E\beta_k$ = 7, which would be the parameter estimate chosen by the GCE approach if data constraints render the value β_k = 7 feasible. A general principle of GCE is illustrated by the two examples—the prior that is actually implied by the method places the highest prior weight on the expectation that is implied by the *reference* probability distribution.

In summary, the GME/GCE approaches implies the use of informative prior information on parameters to be estimated. This is true, even if the reference distribution employed is uniform over the set of support points because the actual GME/GCE estimates are defined as expectations with respect to the discrete probability distribution used to reparameterize the parameters of interest. To solve underdetermined systems of equations, the use of prior information is unavoidable and by itself is not a caveat regarding the use of GME techniques. It is in fact this specific feature, i.e. the flexibility in formulating prior information, that makes the GCE/GME framework of analysis so interesting to applied modelers who seek plausible simulation models and consistent data sets. The prior information actually employed is, however, a result of interactions between chosen support points

and the reference distributions on the chosen supports as well as the final weighting on support points implied by the maximum entropy criterion. The total effect of this interaction—especially for applications with many parameters and more than two support points—is not transparent. Furthermore, the introduction of a set of probabilities for each parameter to be estimated increases the computational demand on solving complex problems, which renders some very complex data reconciliation and estimation exercises intractable with currently available hardware and optimization solvers. In the next section we develop a Bayesian alternative to the GME approach which allows a direct and transparent formulation of prior information and potentially reduces the computational demand significantly.

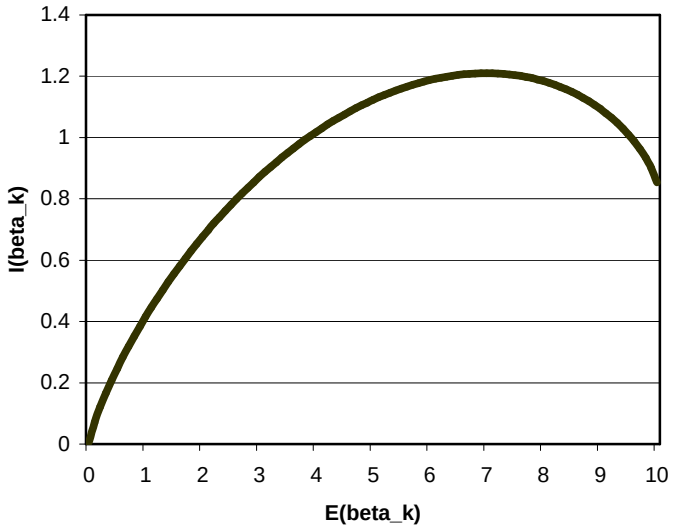


Figure 2: Prior weighting of parameter expectations with the cross-entropy criterion. Source: Maximum cross-entropy calculations with $s_{k1} = 0$, $s_{k2} = 10$, and a reference distribution where $q_{k1} = 0.3$, $q_{k2} = 0.7$.

3. A Bayesian approach to the solution of underdetermined systems

To motivate the general concepts underlying the Bayesian alternative to GCE/GME we first reconsider the linear model without noise used in the previous section. We then extend the approach to a general system of underdetermined structural equations.

3.1. The linear model revisited

The Bayesian approach to parameter estimation treats model parameters as stochastic variables. In this context the method distinguishes between the prior den-

sity, $p(\boldsymbol{\beta})$, summarizing prior information on parameters, the Likelihood function, $L(\boldsymbol{\beta}|\mathbf{y})$, representing information obtained from the data in conjunction with the assumed model, and the posterior density, $h(\boldsymbol{\beta}|\mathbf{y})$, which is the result of combining prior and data information based on Bayes's theorem. The relationship between these three elements can be expressed as (e.g. Zellner 1971, p.14)

$$h(\boldsymbol{\beta}|\mathbf{y}) \propto p(\boldsymbol{\beta}) L(\boldsymbol{\beta}|\mathbf{y}), \tag{10}$$

where the posterior density is proportional to the prior density multiplied by the Likelihood function. The posterior density allows drawing statistical inference about $\boldsymbol{\beta}$ using probability statements or by deriving point estimates that are optimal with respect to some loss criterion. For example, the mean of the posterior (density) is the value which minimizes quadratic loss.

Through appropriate interpretation of its components, the GME approach to estimating the parameters of the underdetermined linear model given in the previous section can be subsumed within the Bayesian formalism. For the case of two support points, using (8) and suppressing the GCE/GME expectation operator henceforth by simply representing the resultant estimator by $\boldsymbol{\beta}$, the GME optimization problem can be represented as

$$\max_{\boldsymbol{\beta}} \left\{ h(\boldsymbol{\beta}|\mathbf{y}) \propto p(\boldsymbol{\beta}) L(\boldsymbol{\beta}|\mathbf{y}) \propto \left[\sum_{k=1}^{K} H(\beta_k) \right] I_{\{\boldsymbol{\beta}: \mathbf{y}=\mathbf{X}\boldsymbol{\beta}\}}(\boldsymbol{\beta}) \right\} \tag{11}$$

where $I_A(\boldsymbol{\beta})$ is the standard indicator function that takes the value 1 when $\boldsymbol{\beta} \in \mathbf{A}$ and equals 0 otherwise. If $H(\beta_k)$ is chosen according to (8), the optimal value for $\boldsymbol{\beta}$ will be equal to the optimal $E\boldsymbol{\beta} = \mathbf{Sp}$ obtained in the GME solution, with an analogous result holding for GCE with $H(\beta_k)$ defined appropriately. In the Bayesian context, the objective function can be interpreted as the joint posterior density of the model parameters, $h(\boldsymbol{\beta}|\mathbf{y})$, defined via a prior density defined by $p(\boldsymbol{\beta}) \propto \sum_k H(\beta_k)$ [12] that is multiplied by a likelihood function that assigns zero weights to values of $\boldsymbol{\beta}$ that do not satisfy the linear model constraints $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$ and a positive constant weight to the values of $\boldsymbol{\beta}$ that are compatible with the data and the linear model relationship.[13] This implies zero *posterior* density weights for the values of $\boldsymbol{\beta}$ not satisfying the constraints and differential *posterior* weighting ac-

[12] The function $\sum_{k=1}^{K} H(\beta_k)$ would need to be scaled appropriately to integrate to unity in order to be interpreted as a proper density, but this scaling is irrelevant for the outcome of the maximization.

[13] In the classical case of a linear model with noise, the Likelihood function would also have the error variance as an argument and would imply some continuous differential weighting according to the assumed error distribution. All that can be learned from the model (the underdetermined data constraints, without noise) in this case is which parameter vectors satisfy the data constraints and which do not.

cording to the prior (density) for all other values of $\boldsymbol{\beta}$. The value of $\boldsymbol{\beta}$ that maximizes $h(\boldsymbol{\beta}|\mathbf{y})$ is the mode of the posterior distribution of $\boldsymbol{\beta}$, which leads to the *Highest Posterior Density* (HPD)-estimate of $\boldsymbol{\beta}$.[14]

The preceding interpretation of GCE/GME within the Bayesian formalism suggests a general Bayesian alternative to the entropy approach that has three useful characteristics: (1) it can be formulated such that it is fully equivalent to the GCE/GME approach if support point choice and implicit weighting by the entropy criterion are appropriately represented, (2) the prior information on unknowns can be transparently formulated by assigning any appropriate prior density $p(\boldsymbol{\beta})$ *directly* to the unknowns, and (3) the optimization model has a smaller number of variables and, for an appropriate choice of the prior density functions, can be less computationally demanding.

Having motivated the Bayesian alternative with a basic underdetermined linear model example, we now turn to a more general treatment of the Bayesian solution to underdetermined systems and the connection to entropy-based approaches.

3.2. General structural equation system

The general mathematical problem now being addressed is one where there are M equations, represented in vector function form as $\mathbf{g}(\mathbf{z}) = \mathbf{0}$, involving an unknown (K×1) vector argument $\mathbf{z}$, with M < K, so that the system of equations *underdetermines* the unknown vector $\mathbf{z}$.[15] Thus, in the absence of any additional information, and assuming the original equation system $\mathbf{g}(\mathbf{z}) = \mathbf{0}$ is consistent so that at least some solution actually exists, then indeterminacy implies that there is generally an infinite number of solution vectors that solve the system of equations.

One method of obtaining a unique solution to the system of equations is to choose $\mathbf{z}$ so as to optimize an extremum metric $v(\mathbf{z})$, subject to the constraints that $\mathbf{g}(\mathbf{z}) = \mathbf{0}$. So long as there exists a unique optimum of $v(\mathbf{z})$ within the feasible space of $\mathbf{z}$ values determined by $\mathbf{z} \in \Psi = \{\mathbf{z}: g(\mathbf{z}) = \mathbf{0}\}$, a unique solution to the original equation system can be identified. In general terms, such a solution could be represented as

$$\mathbf{z}^* = \arg\max_{\mathbf{z}} \left\{ v(\mathbf{z}) \text{ s.t. } g(\mathbf{z}) = \mathbf{0} \right\} \tag{12}$$

where it is assumed without loss of generality that *maximization* is the type of optimization pursued.

---

[14] Using the mode of the posterior for estimation was suggested before in the context of well-posed estimation problems, for example by DeGroot (1970), who called the estimator "generalized maximum likelihood". More frequently used terms are "maximum a-posteriori estimator" and "posterior mode estimator". In accordance with the Bayesian confidence intervals we prefer HPD-estimator.

[15] The elements of $\mathbf{z}$ are not restricted to model parameters. They could also represent unknown variable values in a data reconciliation exercise where data are measured with errors or not observed at all.

In fact *any* extremum metric v(**z**) that exhibits an optimum within the feasible space $\mathbf{z} \in \Psi$ defines a possible solution to the equation system. There is thus a problem of deciding *which* metric to optimize, which in turn determines *which* solution from among a generally infinite number will be chosen as *the* solution to the original equation system. In general, any of the solutions in Ψ can be obtained given an appropriate corresponding choice of extremum metric v(**z**). Thus, the solution obtained to a system of equations in this way is only defensible to the extent that the extremum metric used to obtain that solution is defensible. Before returning to this issue we discuss some necessary conditions for the solution.

Assume that the equation system of $\mathbf{g}(\mathbf{z}) = \mathbf{0}$ is a collection of functionally independent equations, so that the equations effectively determine M of the z_i's as a function of the remaining K-M z_i-values. It is not necessary, conceptually, that *explicit* solutions exist for M of the variables in terms of the other K-M variables, but only that solutions exists. The solution might only be implicitly defined (which would then require numerical solution techniques). It is apparent that a general *necessary* condition for an extremum solution to exist is that v(**z**) for $\mathbf{z} \in \Psi$ be informative, i.e. non-constant, in at least K-M of the variables in the vector **z**. Among other things, this means that v(**z**) cannot be uniform (or "uninformative" in prior distribution parlance) in more than M of the z_i arguments.[16] We note that there are other conditions that might be necessary in any given application, because depending on the nature of the equations in the system, it may be that informative information would have to exist on a specific as opposed to an arbitrary subset of **z** arguments given the solution space to $\mathbf{g}(\mathbf{z}) = \mathbf{0}$. It should also be noted that if v(**z**) is informative on precisely K-M variables in the **z** vector, then the solution can be trivial in the sense that unconstrained optimization of the v(**z**) metric in these K-M dimensions could be pursued independent of the equation system $\mathbf{g}(\mathbf{z}) = \mathbf{0}$ to determine K-M of the unknowns. The remaining arguments in the **z** vector could then be solved based on the relationships among the z_i's determined by the equation system.

Given that the data information serves only to narrow the feasible space of solutions for the unknowns and is otherwise uninformative, a useful and defensible choice for the extremum metric, v(**z**), is the additional prior information held by the analyst, which summarizes the available non-data information on **z**. If $p_i(z_i)$ represents general prior distribution weights on the possible solution values for the i^{th} component of the **z** vector, and if the prior weightings of the different components are considered to be independent, then the optimization metric used to obtain a solution to the equation system could be specified as

[16] Given this observation, it is clear that the GME approach to solving underdetermined systems works because it "automatically" implies a non-uniform prior weighting with respect to the variable of interest.

$$v(\mathbf{z}) = p(\mathbf{z}) = \prod_{i=1}^{K} p_i(z_i) \tag{13}$$

as example of which was given in (11). In the absence of independence, p(**z**) can represent any joint prior distribution on potential solution values **z**.

Now consider Bayes's rule applied to the problem of solving the equation system for **z**. In the absence of any information that would link **z** values to data and allow a likelihood function to be specified, the likelihood function would be considered undetermined or undefined. In this case, the Bayesian posterior and prior on the **z** vector would be identical and the maximization of the prior v(**z**) = p(**z**) would yield the maximum of the posterior. But in the current problem context the system of equations **g**(**z**) = **0** in effect constrains the support of the posterior h(**z**) to $\mathbf{z} \in \Psi = \{\mathbf{z}: \mathbf{g}(\mathbf{z}) = \mathbf{0}\}$. The Likelihood function in this case can be interpreted as an indicator function $I_{\psi}(\mathbf{z})$ that assigns weights of 1 to admissible values of **z** and 0 otherwise. The posterior is then in the form

$$h(\mathbf{z}) \propto p(\mathbf{z}) I_{\psi}(\mathbf{z}). \tag{14}$$

Consequently, the argument that maximizes the prior probability p(**z**) subject to the constraint $\mathbf{z} \in \Psi$ (or **g**(**z**) = **0**) will provide a *Bayesian highest posterior density (HPD) solution* to the equation system.

4. Solutions for uninformative priors

For reasons discussed earlier, the HPD approach to solving the system of equations cannot be applied in cases where the prior weighting on solution values is not sufficiently informative, i.e. p(**z**) cannot be uniform in more than M of the z_i arguments as the optimum will not be unique. However, in this case, solving for the posterior *mean*, which is the posterior risk-minimizing Bayesian estimate under quadratic loss, will generally be possible as long as the uniform distribution is proper in the sense of integrating to 1. This will follow naturally if the prior support space is a priori compact, so that there is indifference among values of **z** within a hyperrectangle of values having finite boundaries. In the extreme case of no informative prior information at all, the values in the support space defined by the equation system. $\Psi = \{\mathbf{z}: \mathbf{g}(\mathbf{z}) = \mathbf{0}\}$, are all equally likely, so that the Bayes's posterior mean solution would be the mean of **z** from among all equally likely values in this support space. A computational method of finding such a solution would be to draw uniform random outcomes of **z** from Ψ, forming their sample mean, and for large enough simulated sample sizes, the sample mean would converge in probability (or almost surely) to the true mean by the weak (strong) law of large numbers.

In some cases, the posterior mean solution might be identifiable analytically. For example, consider again the underdetermined linear model without noise,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$$

where $\mathbf{X}$ is a T × K matrix, with T < K and rank ($\mathbf{X}$) = T. Since rank ($\mathbf{X}$) is smaller than the number of columns, an infinite number of solutions exist for $\boldsymbol{\beta}$. These solutions will form a hyperplane in $\Re^K$, which can be described by a linear function of the form

$$\boldsymbol{\beta} = \boldsymbol{\beta}_0 + \mathbf{B}\boldsymbol{\xi} \tag{15}$$

where $\mathbf{B}$ is a K × (K – T) matrix that is a basis for the subspace of solutions to the *homogeneous model* $\mathbf{0} = \mathbf{X}\boldsymbol{\beta}$, and $\boldsymbol{\xi}$ is an arbitrary (K – T) vector. This follows from the following results:

Lemma 1: *Any* solution $\boldsymbol{\beta}^*$ to the inhomogeneous linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$ can be written as the sum of a *particular* solution $\boldsymbol{\beta}^0$ to the inhomogeneous model plus *some* solution $\boldsymbol{\beta}^1$ to the homogeneous linear model $\mathbf{0} = \mathbf{X}\boldsymbol{\beta}$ (e.g. de la Fuente 2000, p. 197).

Lemma 2: $\boldsymbol{\beta}^1$ in lemma 1 can be written as $\mathbf{B}\boldsymbol{\xi}$, for *some* matrix $\mathbf{B}$ and *any* vector $\boldsymbol{\xi}$ of dimension (K – T).

If there are uniform priors for at least K – T of the elements of $\boldsymbol{\beta}$, then those priors constitute a hyperrectangle $\mathbf{U}$ in $\Re^K$, and the posterior mean is the geometrical centre of the intersection between the solution hyperplane and the hyperrectangle $\mathbf{U}$. We can then compute the posterior mean through a sequence of four steps that include first finding $\boldsymbol{\beta}^0$. , then computing the matrix **B,** next finding the intersection between the solution hyperplane and the prior hyper rectangle, and finally finding the center of the intersection. A specific algorithm for accomplishing these steps is as follows:

Step 1. A particular solution $\boldsymbol{\beta}^0$ to the inhomogeneous system can be found by solving $\boldsymbol{\beta}^0 = \mathbf{X}^+\mathbf{y}$, where $\mathbf{X}^+$ is the generalized inverse of $\mathbf{X}$.

Step 2. Since K > T and rank ($\mathbf{X}$) = T, K – T columns of $\mathbf{X}$, together forming the matrix $\mathbf{X}_{(i)}$, can be written as linear combinations of the other T columns, which are kept in the T × T matrix $\mathbf{X}_{(-i)}$. The coefficients of each of the K – T columns in $\mathbf{X}_{(i)}$ can be chosen arbitrarily. If this is repeatedly done for each column in $\mathbf{X}_{(i)}$, the following expression is obtained, where the columns of $\mathbf{B}_{(i)}$ are the arbitrary coefficient vectors for $\mathbf{X}_{(i)}$:

$$-\mathbf{X}_{(i)}\mathbf{B}_{(i)} = \mathbf{X}_{(-i)}\mathbf{B}_{(-i)}$$

Choosing the (T – K) × (T – K) identity matrix for $\mathbf{B}_{(i)}$, the above expression can be solved for $\mathbf{B}_{(-i)} = -(\mathbf{X}_{(i)})^{-1}\mathbf{X}_{(i)}$, and $\mathbf{B}$ can be obtained by vertical concatenation of $\mathbf{B}_{(i)}$ and $\mathbf{B}_{(-i)}$, keeping the rows in proper order.

Step 3. Find the values of $\boldsymbol{\xi}$ for which the resulting $\boldsymbol{\beta}$ is inside the prior hyperrectangle. This can be done by trial and error if the dimension of $\boldsymbol{\xi}$ is low, and

numerically by repeated linear programming (solving $\{\min \mathbf{p}'\xi: \boldsymbol{\beta}_0 + \mathbf{B}\xi = \boldsymbol{\beta} \in \mathbf{U}\}$, with **p** being some permutation of (-1) and 1 of length K – T, for all such permutations) if the dimension is higher. The set of solutions will be the bounds of a hyperrectangle in $\Re^{K-T}$.

Step 4. Since $\boldsymbol{\beta}$ is linearly dependent on ξ, and $\boldsymbol{\beta}$ is uniformly distributed, the expected value of $\boldsymbol{\beta}$ is found by computing the geometrical mid point of the hyperrectangle found in step 3.

Various empirical illustrations of the Bayesian methodology are presented in the next section.

5. Illustrative applications

This section presents two illustrative applications of the HPD-estimator based on underdetermined problem specifications that are typical of applications for entropy estimators: Balancing of a Social Accounting Matrix (SAM) and a linear regression problem. A GAMS program for the balancing of the SAM is printed in appendix 4.1, and a GAMS program for the illustrative applications for the underdetermined linear regression is printed in appendix 4.2. An installation of the GAMS software is required in order to run the applications.

5.1. Balancing a Social Accounting Matrix

In 1994, Golan, Judge and Robinson (GJR) used entropy based estimators to create a consistent SAM. Variants of their approaches can be found in the empirical Computable General Equilibrium literature to prepare complete databases out of incomplete and uncertain data information.

The basic problem of balancing a SAM can be formulated as follows: find a square matrix of coefficients **A** and vectors **x** and **y** satisfying the equations

$$\mathbf{Ax} = \mathbf{y} \qquad (16)$$

$$\mathbf{A}'\boldsymbol{\iota} = \boldsymbol{\iota}. \qquad (17)$$

with $\boldsymbol{\iota}$ the vector of ones of appropriate dimension. In general, information about the accounts **x** and **y** are available from observable data, whereas the coefficient matrix **A** is difficult to obtain. A common situation is thus that **x** and **y** are given, and **A** needs to be determined subject to the restrictions (16) and (17), possibly given some prior information about **A**, perhaps in the form of the same matrix for another region or for the same region for a different period. We take the example studied by GJR and provide a Bayesian alternative.

Table 1 in their paper provides the “true parameters”,

$$\mathbf{A} = \begin{bmatrix} 0.726 & 0.000 & 0.165 & 0.301 \\ 0.161 & 0.268 & 0.000 & 0.451 \\ 0.113 & 0.678 & 0.714 & 0.000 \\ 0.000 & 0.054 & 0.121 & 0.248 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 62 \\ 56 \\ 91 \\ 266 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 140 \\ 145 \\ 110 \\ 80 \end{bmatrix}.$$

The authors proceed to construct a (synthetic) prior for **A** by multiplying each entry in **A** by a random number drawn from a normal distribution, N(1, .05). They present the outcome

$$\mathbf{A}^o = \begin{bmatrix} 0.730 & 0.000 & 0.172 & 0.278 \\ 0.159 & 0.259 & 0.000 & 0.480 \\ 0.111 & 0.688 & 0.694 & 0.000 \\ 0.000 & 0.053 & 0.135 & 0.243 \end{bmatrix},$$

and estimate **A** with GCE using $\mathbf{A}^o$ as a prior.

The GCE problem is

$$\begin{aligned} &\min && \mathbf{p}'\ln(\mathbf{p}/\mathbf{q}) \\ &\text{such that} && \mathbf{p} \geq \mathbf{0},\ \mathbf{Sp} = vec(\mathbf{A}),\ \mathbf{y} = \mathbf{Ax},\ \mathbf{A}\iota = \mathbf{1} \end{aligned} \qquad (18)$$

where the prior probabilities $\boldsymbol{q}$ of the support point matrix **S** are selected so that $\mathbf{Sq} = \mathbf{A}^o$, with vec(**A**) being the operator that reshapes the matrix **A** to a column vector by vertically concatenating respective columns, and **p**/**q** as in section 2 the vector whose i^{th} element is p_i/q_i, Note that this approach requires the researcher to define a set of at least two (GJR use five) support points for each parameter, and also to define a corresponding set of prior probabilities such that the prior SAM is recovered. GJR use the same support points for all elements of **A**, and choose **q** using an initial GME estimation of $\mathbf{Sq} = \mathbf{A}^o$, which effectively doubles the computational effort needed to produce the final estimates of the **A** matrix.

Now construct an alternative Bayesian estimator for the same problem. The HPD framework allows the use of any prior distribution. Assume, for example, that the researcher had a-priori knowledge that the observed matrix $\mathbf{A}^o$ was generated as in GJR. Taking $\mathbf{A}^o$ as prior mean, and continuing to follow GJR, the corresponding prior density function would be vec(**A**) ~ N(vec($\mathbf{A}^{o)}$,**Σ**), The covariance matrix **Σ** is set equal to a diagonal matrix with elements $\left(\text{vec}(\mathbf{A}^o)0.05\right)^2$, the square taken element-wise.

Formulating the HPD estimator as discussed previously, taking natural logs, and restricting the objective function to the terms that are relevant for optimization leads to the following extremum estimation problem:

$$\begin{aligned} \max_{\mathbf{A}} \quad & \left[vec(\mathbf{A}) - vec(\mathbf{A}^o) \right]' \boldsymbol{\Omega}^{-1} \left[vec(\mathbf{A}) - vec(\mathbf{A}^o) \right] \\ \text{subject to} \quad & \mathbf{Ax} = \mathbf{y} \\ & \mathbf{A}\boldsymbol{\iota} = \mathbf{1} \end{aligned} \tag{19}$$

For the synthetic data provided in GJR, GCE and HPD give very similar results, shown below (results for GCE as printed in GJR). Note that the HPD estimation tacitly assumed degenerate priors for **x** and **y**. The estimation is easily extended to encompass the fact that **x** and **y** are not known with certainty.

$$\mathbf{A}^{GCE} = \begin{bmatrix} 0.732 & 0.000 & 0.168 & 0.298 \\ 0.155 & 0.251 & 0.000 & 0.456 \\ 0.114 & 0.697 & 0.702 & 0.000 \\ 0.000 & 0.052 & 0.129 & 0.246 \end{bmatrix}, \mathbf{A}^{HPD} = \begin{bmatrix} 0.731 & 0.000 & 0.167 & 0.299 \\ 0.157 & 0.248 & 0.000 & 0.456 \\ 0.112 & 0.699 & 0.702 & 0.000 \\ 0.000 & 0.053 & 0.131 & 0.245 \end{bmatrix}$$

As can be seen from (19), the choice of a normal prior distribution results in a weighted least squares approach implying numerically desirable properties for large scale problems. Compared to GME or GCE approaches, explicit accounting for support points and adding up constraints for probabilities are unnecessary and infeasibilities are less likely to lead to numerical problems. Other prior distributions can be flexibly accommodated and will be considered in the next example.

5.2. Regression models

In this section we consider an ill-posed linear regression model with and without noise, and characterized by three equations and four parameters. For two of the parameters there is prior information available. In total five cases are studied which are distinguished by the prior information used for the parameters, and the type of estimation objective applied:

1. Uniform priors given bounds [**u**,**v**], with parameters estimated by posterior means;

2. Symmetric triangular distributed priors, with parameters estimated by the posterior mode;

3. GME estimation with [**u**,**v**] as supports, represented and solved equivalently as a Bayesian HPD-estimator;

4. Priors are distributed as beta(2,2) between the bounds [**u**,**v**] and estimated by a Bayesian HPD estimator;

5. As previous, but also including additive white noise;

True parameters for the model without noise, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$, were chosen arbitrarily and the columns 2-4 of $\mathbf{X}$ drawn from normal distributions with means 20, 8 and 12 and variances equal to ¼ of the means. By multiplication with the selected true parameters, the true $\mathbf{y}$ was obtained. In all five cases, prior information was available for β_2 and β_3, based on a lower bound of zero and an upper bound found by adding errors from N(0,0.1) to the true parameters and multiplying the result by two. The procedure resulted in the following numbers. Note that the mid points between the bounds ($\mathbf{u}$,$\mathbf{v}$) used in some cases are *not* equal to the true parameter value.

$$\boldsymbol{\beta} = \begin{bmatrix} 10.0 \\ 0.5 \\ 1.5 \\ 1.0 \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 20.733 & 8.656 & 8.830 \\ 1 & 17.827 & 7.443 & 13.619 \\ 1 & 20.001 & 6.715 & 12.596 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 42.180 \\ 43.697 \\ 42.668 \end{bmatrix}, (\mathbf{u}, \mathbf{v}) = \begin{bmatrix} 0 & 0.868 \\ 0 & 2.903 \end{bmatrix}$$

Case 1: Since we are dealing with a linear system with (-1) degrees of freedom, the vector $\boldsymbol{\xi}$ in equation (15) is a scalar, and all feasible $\boldsymbol{\beta}$ lie on a line segment limited by ($\mathbf{u}$,$\mathbf{v}$). Following the steps indicated in section 4, choosing the second column of $\mathbf{X}$ for $\mathbf{X}_{(i)}$, we obtain

$$\boldsymbol{\beta} = \begin{bmatrix} 0.1132 \\ 0.7284 \\ 1.8604 \\ 1.2300 \end{bmatrix} + \xi \begin{bmatrix} -43.2306 \\ 1.0000 \\ 1.5735 \\ 1.0054 \end{bmatrix} \tag{20}$$

for arbitrary ξ. In order for $\boldsymbol{\beta}$ to be within ($\mathbf{u}$,$\mathbf{v}$), it is required that $\xi \in$ (-0.7284, 0.1396). Since the uniform density indicates the same posterior density weight for all values for ξ in that interval and zero elsewhere, we can compute the posterior mean as the mid point of the interval, or $\hat{\xi} = -0.2944$. Inserting that value into the expression (20) gives us the point estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 12.842 \\ 0.434 \\ 1.397 \\ 0.934 \end{bmatrix}.$$

Case 2: Let the prior density for β_2 and β_3 have the same bounds as before, but now follow a symmetric triangular distribution, i.e. the mid point of the interval is favored. Now a unique posterior mode exists, and we may apply the HPD estimator. Since we strive to maximize the posterior, the piecewise linear formulation of the triangular density can be relaxed to three linear inequalities, each representing a side of the triangle. For ease of notation, we first introduce the subvector $\boldsymbol{\beta}_p$

consisting of the elements (β_2,β_3) for which there are priors, the corresponding subset of probability densities $\mathbf{p}_p$, and the linear mapping $\mathbf{g}:[\mathbf{u},\mathbf{v}] \rightarrow [0,1]\times[0,1]$. Thus $g_k(\beta_k)$ expresses β_k in terms of the *share* of the distance from u_k to v_k, for k = 2,3.

The HPD estimator is then

$$\max_{\mathbf{p}_p,\boldsymbol{\beta}_p} p_2 p_3$$

subject to $\boldsymbol{\beta}_p$ being triangular distributed, i.e.

$$\mathbf{p}_p \le 4\mathbf{g}(\boldsymbol{\beta}_p),\ \mathbf{p}_p \le 4-4\mathbf{g}(\boldsymbol{\beta}_p),\ \mathbf{p}_p \ge 0,$$

and the data constraints,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$$

The outcome of the estimator is identical to the outcome of the posterior mean estimator and is not repeated here.

Case 3: Let (**u**,**v**) be support points for a GME estimation. Using the normalization $\mathbf{g}(\boldsymbol{\beta}_p)$ as before, the support points become {0,1}, the probabilities of the supports become $\mathbf{p} = \mathrm{vec}\left(\left[\mathbf{g}(\boldsymbol{\beta}_p) \quad \mathbf{1}-\mathbf{g}(\boldsymbol{\beta}_p)\right]'\right)$ and we may write the GME estimator as

$$\max_{\boldsymbol{\beta}_p} H = -\left[\mathbf{g}(\boldsymbol{\beta}_p)'\ln\left(\mathbf{g}(\boldsymbol{\beta}_p)\right)+\left(1-\mathbf{g}(\boldsymbol{\beta}_p)\right)'\ln\left(\mathbf{1}-\mathbf{g}(\boldsymbol{\beta}_p)\right)\right]$$

subject to $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$

$$\mathbf{u} \le \boldsymbol{\beta}_p \le \mathbf{v}$$

For the sake of illustration, we re-write this as a fully equivalent HPD estimation problem. Note that the GME problem is equivalent to maximizing e^H (the maximum is maintained under monotonic transformation). Substitution and some algebra lead to the equivalent HPD problem

max $\mathbf{f}_2(\boldsymbol{\beta}_2)\mathbf{f}_3(\boldsymbol{\beta}_3)$

subject to $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$

$$\mathbf{u} \le \boldsymbol{\beta}_p \le \mathbf{v}$$

where $f_k(\beta_k) = c g_k(\beta_k)^{-g_k(\beta_k)}\left(1-g_k(\beta_k)\right)^{(g_k(\beta_k)-1)}$ (for k = 2,3) is a probability density function if the constant *c* is chosen properly (c ≈ 0.6 makes *f* integrate to unity, non-negative values are prevented by the mapping ***g*** and the bounds

$\mathbf{u} \leq \boldsymbol{\beta}_p \leq \mathbf{v}$). We see that, interpreted in this way, the GME estimator is an instance of a HPD estimator. The GME estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 12.586 \\ 0.440 \\ 1.406 \\ 0.940 \end{bmatrix}$$

Case 4: The upper and lower bounds on two of the parameters make it natural to describe the estimates in terms of a fraction of the distance between the bounds (as expressed in the mapping **g**). In such cases the beta distribution is sometimes used. Let the distribution of β_k for k = 2,3 be such that $g_k(\beta_k) \sim$ beta(2,2). The HPD estimate of $\boldsymbol{\beta}$ with beta-distributed priors is identical to the GME estimate at least up to three decimal places in this case, and not repeated here.

Case 5: In the previous cases we assumed that **X** and **y** were observable without noise. We now introduce white noise for **y** by adding iid errors drawn from N(0,1) whereas **X** is still assumed to be known with certainty. The resulting stochastic vector of left hand side variables is denoted by $\mathbf{y}_s$, and the outcome of a draw was

$$\mathbf{y}_s = \mathbf{y} + \mathbf{e} = \begin{bmatrix} 44.064 & 42.976 & 41.369 \end{bmatrix}'$$

where **e** is an outcome of the error $\boldsymbol{\varepsilon}$, and the system to estimate is $\mathbf{y}_s = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

If we consider $\boldsymbol{\varepsilon}$ yet another parameter to determine, and introduce the prior information that errors were drawn from N(0,1) and still assume that β_2 and β_3 belong to the same beta distributions as in the previous example, the HPD estimator for $\boldsymbol{\beta}$ is found by solving the problem

$$\max_{\boldsymbol{\beta}_p, \boldsymbol{\varepsilon}} \; h = \prod_{k \in \{2,3\}} p_p\left(g_k(\beta_k)\right) \prod_{i=1}^{4} p_e\left(\varepsilon_i\right)$$

$$\text{subject to} \quad \mathbf{u} \leq \boldsymbol{\beta}_p \leq \mathbf{v}$$

$$\mathbf{y}_s = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

with $p_b(\cdot)$ being the beta density function as in the previous example and $p_e(\cdot)$ being the standard normal univariate density. A form more easily computed is obtained by recognizing that $p_b(x) = 6(x - x^2)I_{(0,1)}(x)$ and taking the logarithm of the objective function, which then becomes

$$\max \; \ln(h) = \sum_{k \in \{2,3\}} \ln\left(g_k(\beta_k) - g_k(\beta_k)^2\right) - \tfrac{1}{2} \sum_{i=1}^{4} \varepsilon_i^2$$

The resulting estimate of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = \begin{bmatrix} 16.668 & 0.379 & 1.820 & 0.419 \end{bmatrix}'$.

6. Summary and conclusions

This paper presents a Bayesian alternative to the solution of underdetermined systems of equations. First, we reviewed the GME-GCE approach in the context of estimating an underdetermined linear model without noise and identified the effective prior information as a combined effect between supports, reference probabilities, and the entropy criterion. It was indicated that a "uniform distributions over supports" does not imply a "non-informative" prior on the parameters of interest, but rather a clear prior preferential weighting on estimation outcomes. In the suggested Bayesian alternative the underdetermined model equations and the data represent the "Likelihood" information. Deviating from standard Likelihood functions of conventional models with a predefined family of distributions, the Likelihood implies a constant positive weight for all possible solutions of the model equations and a zero weight for infeasible values. All prior information is represented in a standard Bayesian way via prior probability densities on model parameters. Highest Posterior Density (HPD) estimates are obtained using an optimization algorithm.

The Bayesian approach can be formulated to mimic the behavior of GME-GCE models perfectly. However, more interesting is its general structure allowing full flexibility in formulating directly and transparently the prior information held by the analyst. For a unique solution to exist, a certain amount of informative prior information is necessary. However, if this is not the case, a solution based on the posterior mean can—at least conceptually—still be provided.

The suggested approach lends itself easily to the type of problems currently solved with GME or GCE techniques. It has been successfully applied to large scale estimation and calibration exercises (Britz et al. 2004, Jansson 2007). It facilitates the peer review of methodology and underlying assumptions by making the employed prior information directly visible. Further research should examine computational approaches for generating posterior mean estimates under insufficient identifying prior information.

References

Arndt, C., Robinson, S., and F. Tarp (2002): Parameter Estimation for a Computable General Equilibrium Model: a Maximum Entropy Approach. *Economic Modelling*, 19:375-398.

Britz, W. and C. Wieck (2002): Completeness and Consistency in a Multidimensional Data Base using Constrained Simultaneous Estimation Techniques. Poster presented at the "42. Jahrestagung der GEWISOLA" (German Agricultural Economics Society).

Britz, W. Witzke, H.P., and T. Heckelei (2004): Estimating Trade Matrices and Supply Utilization Accounts using a Bayesian Estimator, Final Report of EuroCARE to FAO, Bonn.

DeGroot, M. H. (1970). *Optimal statistical decisions*. New York: McGraw-Hill.

Fuente, A. de la (2000). *Mathematical methods and models for economists*. Cambridge University Press, New York.

Golan, A., G. Judge, and S. Robinson (1994): Recovering Information from Incomplete of Partial Multisectoral Economic Data. *The Review of Economics and Statistics*, 76(3):541-549.

Golan, A., G. Judge, and D. Miller (1996): *Maximum Entropy Econometrics*, Chichester UK: Wiley.

Golan, A., G. Judge and J.M. Perloff (1996): A Maximum Entropy Approach to Recovering Information From Multinomial Response Data. *Journal of the American Statistical Association* 91, No. 434, Theory and Methods: 841-853.

Howitt, R.E. and A. Reynaud (2003): Spatial Disaggregation of Agricultural Production Data using Maximum Entropy. *European Review of Agricultural Economics*, 30:359-387.

Jansson, T. (2007). Econometric specification of constrained optimization problems. Dissertation, Bonn University (forthcoming).

Lence, H.L., and D. Miller (1998a): Estimation of Multioutput Production Functions with Incomplete Data: A Generalized Maximum Entropy Approach. *European Review of Agricultural Economics*, 25:188-209.

Lence, H.L., and D. Miller (1998b): Recovering Output Specific Inputs from Aggregate Input Data: A Generalized Cross-Entropy Approach. *American Journal of Agricultural Economics*, 80(4):852-867.

Léon Y., L. Peeters, M. Quinqu and Y. Surry (1999): The use of maximum entropy to estimate input-output coefficients from regional farm accounting data. *Journal of Agricultural Economics,* 50: 425-439.

Mittelhammer, R.C., G.G. Judge, and D.J. Miller (2000): Econometric Foundations. Cambridge: Cambridge University Press.

Oude Lansink, A.G.J.M.(1999b): Generalized Maximum Entropy and heterogeneous technologies. *European Review of Agricultural Economics* 26:101-115.

Paris, Q. (2001): Symmetric Positive Equilibrium Problem: A Framework for Rationalizing Economic Behavior with Limited Information. *American Journal of Agricultural Economics*, 83(4): 1049-1061.

Paris, Q., and R.E. Howitt (1998): An Analysis of Ill-Posed Production Problems Using Maximum Entropy. *American Journal of Agricultural Economics*, 80(1): 124-138.

Paris, Q., and R.E. Howitt (2001): The Multi-Output and Multi-Input Symmetric Positive Equilibrium Problem. In: Heckelei, T., H.P. Witzke, and W. Henrichsmeyer (Eds.*): Agricultural Sector Modelling and Policy Information Systems*. Proceedings of the 65th EAAE Seminar, March 29-31, 2000 at Bonn University, Vauk Verlag Kiel: 88-100.

Preckel, P.V. (2001): Least Squares and Entropy: A Penalty Function Perspective. *American Journal of Agricultural Economics,* 83(2): 366-377.

Robilliard, A.-S., and S, Robinson (2003): Reconciling Household Surveys and National Accounts Data using a Cross Entropy Estimation Method. *Review of Income and Wealth*, 49:395-406.

Robinson, S., Cattanbo, A., and M. El-Said (2000): Updating and Estimating a Social Accounting Matrix using Cross Entropy Methods. Economic Systems Research, 13:47-67.

Witzke, H.P and W. Britz (1998): A Maximum Entropy Approach to the Calibration of Highly Differentiated Demand Systems, Capri Working Paper 98-06, University of Bonn.

Zellner, A. (1971): *Introduction to Bayesian Inference in Econometrics*. New York: Wiley.

Zhang, X. and S. Fan (2001): Estimating Crop-Specific Production Technologies in Chinese Agriculture: A Generalized Maximum Entropy Approach. *American Journal of Agricultural Economics*, 83(2): 378-388

Appendix 4.1. GAMS code for section 5.1

```
$title Balance a SAM with HPD estimator
$ontext
    This program illustrates the use of highest posterior density estimator
    to balance a SAM.
    The SAM is 4 by 4, taken from Golan, Judge, Robinson (1994) (GJR94).
$offtext
$offlisting
option limrow=0; option limcol=0;


*---------------------------------------------------------------------------
* Declarations and true sample data. We start with the SAM,
* and compute the coefficient matrix.
*---------------------------------------------------------------------------

set i 'rows in SAM' /1*4/;
set j 'columns in SAM' /1*4/;
alias(i,ii); alias(j,jj);

set fun 'different datasets or extremum metrics ' /
    ori 'original data'
    me  'maximum entropy solution'
    ce  'cross entropy solution'
    gme 'generalized maximum entropy solution'
    gce 'generalized cross entropy solution'
    hpd 'highest posterior density estimator outcome' /;

table FLOWDATA(i,j) 'true SAM'
    1   2   3   4
1   45  0   15  80
2   10  15  0   120
3   7   38  65  0
4   0   3   11  66  ;

table AO(i,j) 'Stochastic coefficient matrix (observation)'
    1       2       3       4
1   0.730   0       0.172   0.278
2   0.159   0.259   0       0.480
3   0.111   0.688   0.694   0
4   0       0.053   0.135   0.243 ;

parameter x(j)              'column sum of SAM';
parameter y(i)              'row sum of SAM';
parameter A(i,j)            'coefficient matrix';
parameter AFUN(fun,i,j)     'comparison of different A-matrices';
parameter dist(fun)         'distance of estimated A from original';

x(j) = sum(i, FLOWDATA(i,j));
y(i) = sum(j, FLOWDATA(i,j));
A(i,j) = FLOWDATA(i,j)/x(j);
AFUN('ori',i,j) = A(i,j);

* Declare fundamental SAM identities

variable o          'objective variable';
```

```
variable va(i,j)    'estimated coefficients';
equation e1(i)      'balance condition';
equation e2(j)      'addup condition';

e1(i) ..    sum(j, va(i,j)*x(j)) =e= y(i);
e2(j) ..    sum(i, va(i,j)) =e= 1;

va.lo(i,j) = 0.00001 $ AO(i,j);
va.up(i,j) = 1       $ AO(i,j);

*-------------------------------------------------------------------------
* Implement balancing solutions by GJR94 and the HPD estimator.
*
* A: Maximum entropy procedure
*-------------------------------------------------------------------------

equation eentropy   'ordinary entropy';
eentropy .. o =e= -sum((i,j)$AO(i,j), va(i,j)*log(va(i,j)));
model mme 'maximum entropy model' /eentropy,e1,e2/;
va.l(i,j) = 0.25;
solve mme using nlp maximising o;
AFUN('ME',i,j) = va.l(i,j);


*-------------------------------------------------------------------------
* B: Cross entropy with the same priors as GJR94
*-------------------------------------------------------------------------

scalar s 'standard deviation' /0.05/;

equation ece    'cross entropy';
ece .. o =e= -sum((i,j)$AO(i,j), va(i,j)*log(va(i,j)/AO(i,j)))
model mce 'maximum cross entropy model' /ece,e1,e2/;
solve mce using nlp maximising o;
AFUN('CE',i,j) = va.l(i,j);


*-------------------------------------------------------------------------
* C: Generalized maximum entropy (GME) with the same priors as GJR94,
*    i.e. "uninformative" with (0,0.25,0.5,0.75,1)
*-------------------------------------------------------------------------

set k 'support points' /1,2,3,4,5/;
parameter ps(k) 'support values' /
    1   0
    2   0.25
    3   0.5
    4   0.75
    5   1 /;

variable vp(i,j,k) 'probability of supports';
equation eGMEcrit       'GME criterion function';
equation eReParam(i,j)  'Reparametrization of A in terms of supports';
equation eProbSum(i,j)  'Adding up criterion for probabilities';

eGMEcrit .. o =e= -sum((i,j,k), vp(i,j,k)*log(vp(i,j,k)));
eReParam(i,j) .. va(i,j) =e= sum(k, vp(i,j,k)*ps(k));
eProbSum(i,j) .. sum(k, vp(i,j,k)) =e= 1;
```

```
model MGME 'GME model' /eGMEcrit,eReParam,eProbSum,e1,e2/;

vp.lo(i,j,k) = 1E-12;
vp.l(i,j,k) = 1/card(k);

solve MGME using nlp maximizing o;
AFUN('GME',i,j) = va.l(i,j);


*-------------------------------------------------------------------------
* D: Generalized cross entropy (GCE) with the same priors as GJR94, i.e.
*    Step 1: assign probabilities to the supports (0,0.25,0.5,0.75,1)
*            so that the prior SAM is recovered.
*    Step 2: use the probabilities of step 1 as priors in the GCE
*-------------------------------------------------------------------------

parameter q(i,j,k)  'Prior probabilities in GCE';
variable vS(i,j)    'Generalized cross-entropy per point';
equation eGCEpre(i,j)'Definition of generalized cross-entropy';
equation eGCE       'Negative of sum of generalized cross entropies';

eGCEpre(i,j) .. vS(i,j) =e=  sum(k,     vp(i,j,k)*log(vp(i,j,k)/q(i,j,k)));
eGCE         .. o       =e= -sum((i,j), vS(i,j));

* Step 1

va.fx(i,j) = AO(i,j);
model MGCE1 'First step in GCE as GME' /eGMEcrit,eReParam,eProbSum/;
solve MGCE1 using NLP maximizing o;

* Step 2

va.lo(i,j) = 0.00001 $ AO(i,j);
va.up(i,j) = 1       $ AO(i,j);
q(i,j,k) = vp.l(i,j,k);
model MGCE2 'Second step in GCE' /eGCE,eGCEpre,eReParam,eProbSum,e1,e2/;
solve MGCE2 using NLP maximizing o;
AFUN('GCE',i,j) = va.l(i,j);

option q:3:2:1;
display q;


*-------------------------------------------------------------------------
* Estimate with HPD using the error density used by GJR94 for generating
* the data.
*-------------------------------------------------------------------------

equation epd    'posterior density equals prior for feasible values';
epd ..  o =e= -sum((i,j)$AO(i,j), 0.5*sqr((va(i,j)-AO(i,j))/(s*AO(i,j))));
model mhpd 'highest posterior density model' /epd,e1,e2/;
solve mhpd using nlp maximising o;
AFUN('HPD',i,j) = va.l(i,j);

dist(fun) = sum((i,j), sqr(AFUN('ori',i,j)-AFUN(fun,i,j)));
display AFUN,dist;
```

Appendix 4.2. GAMS code for section 5.2

```
$ontext
    This GAMS program illustrates the estimation of the underdetermined
    linear regression in chapter 4. It implements, using an NLP solver:
    1. Posterior mode estimator with triangular priors for two parameters
    2. Posterior mean estimator with uniform priors for two parameter
    3. Generalized maximum entropy with supports equal to uniform bounds
    4. Beta distributed normalised priors
    5. Beta distributed normalised priors, normally distributed noise

    Torbjorn Jansson
    LEI, The Hague, NL
$offtext
$offlisting
$eolcom #
option limrow=0; option limcol=0;


*---------------------------------------------------------------------------
* SEC 0     Declarations and basic problem setup
*---------------------------------------------------------------------------

* Declarations of some basic items

set i 'Rows in equation system' /i1*i3/;
set k 'Columns in equation system' /k1*k4/;
set stochk(k) 'All k except constant'; stochk(k) = yes$(ord(k)-1);
alias(i,j); alias(k,l);
set fun 'Different estimation settings' /
    f1 'HPD with triangular prior'
    f2 'posterior mean with uniform distribution'
    f3 'GME with supports = uniform bounds'
    f4 'beta(1.75,1.75) distributed priors for k2,k3 normalised to (0,1)'
    f5 'with normal(0,1) noise in likelihood function'
    /;
parameter b(k) 'True parameter vector';
parameter estb(k,fun) 'Estimated parameter vector';
parameter x(i,k) 'Exogenous';
parameter y(i) 'Endogenous';
parameter xmean(k);
parameter xvar(k,k);

* Re-initialise pseudo random number generator if needed
* Default seed is 3141, but I type it anyway just to make sure...
execseed = 3141;
*execseed = gmillisec(jnow);

* Data generation, manually entered for now

b('k1') = 10.0;
b('k2') = 0.5;
b('k3') = 1.5;
b('k4') = 1.0;
xmean('k1') = 1;
xmean('k2') = 20;
xmean('k3') = 8;
```

```
xmean('k4') = 12;
xvar('k2','k2') = 5;
xvar('k3','k3') = 2;
xvar('k4','k4') = 3;
* Replace random number generator with outcome in paper
*x(i,k) = normal(xmean(k),sqrt(xvar(k,k)));
x(i,'k1') = 1;
x('i1','k2') = 20.733;
x('i2','k2') = 17.827;
x('i3','k2') = 20.001;
x('i1','k3') = 8.656;
x('i2','k3') = 7.443;
x('i3','k3') = 6.715;
x('i1','k4') = 8.830;
x('i2','k4') = 13.619;
x('i3','k4') = 12.596;

y(i) = sum(k, b(k)*x(i,k));

* Prior distributions: triangular for k2 and k3, and uniform for k1 and k4

set pri(k)      'parameters with prior distribution' /k2,k3/;
set uml         'upper, mid and low points' /upp,mid,low/;
set uplo(uml)   'upper and lower bounds' /upp,low/;
parameter bound(k,uplo) 'bounds';

* Declaration of model variables and equations used in all five examples

variable p          'posterior density';
variable f(k)       'prior density of parameter value';
variable vb(k)      'estimated parameter vector';
variable vnb(k)     'b normalised to [0,1]';

equation eqpd       'definition of posterior density';
equation eql(i)     'likelihood function for parameters as restrictions';
equation eqnb(k)    'normalization of b to the interval [0,1]';


eqpd   .. p    =e= prod(pri, f(pri));
eql(i) .. y(i) =e= sum(k, vb(k)*x(i,k));

eqnb(pri) .. vnb(pri) =e= (vb(pri) - (bound(pri,'low')+bound(pri,'upp'))/2)
                         /(bound(pri,'upp') - bound(pri,'low')) + 0.5;


*---------------------------------------------------------------------------
* SEC 1     Posterior mode estimator with triangular priors for two params
*---------------------------------------------------------------------------

* Implementation of triangle distribution:
*   To avoid abs and max, the triangle is put as three linear inequalities,
*   which is equivalent to the standard piecewise linear formulation. The
*   first two equations are the legs of the triangle, the third is the base
*   plus a tiny amount to avoid log of zero.

parameter primode(k) 'mode of prior dist';
parameter prispan(k) 'span of prior dist';
* Replace random number generator with outcome in paper
```

```
*primode(pri) = b(pri) + normal(0,0.1);
primode('k2') = 0.434;
primode('k3') = 1.4515;
bound(pri,'upp') = primode(pri)*2;
bound(pri,'low') = 0;
prispan(pri) = abs(bound(pri,'upp') - bound(pri,'low'));

display x,y,b,bound,primode;

equation eqf1(k)    'definition of triangular density left line';
equation eqf2(k)    'definition of triangular density right line';
equation eqf3(k)    'definition of prior density non-negativity';

eqf1(pri) .. f(pri) =l=     4*vnb(pri);
eqf2(pri) .. f(pri) =l= 4 - 4*vnb(pri);
eqf3(pri) .. f(pri) =g= 0.00001;


model hpd1 /eqf1,eqf2,eqf3,eql,eqpd,eqnb/;
solve hpd1 using nlp maximizing p;
estb(k,'f1') = vb.l(k);
display primode;


*-------------------------------------------------------------------------
* SEC 2     Posterior mean estimator with uniform priors for two parameter
*-------------------------------------------------------------------------

* The same span is assumed as with the triangular distribution.

parameter line(k,uml) 'end and mid points of line';
alias(pri,pri2);

* The feasible space for B is a line in 4-space, limited by hyper rectangle
* defined by the bounds (low,upp) for each of the four parameters.
* Find the line by trial and error; there are only four points to examine.

model mlin /eql/; mlin.holdfixed = 1; mlin.solprint = 2;
loop(pri2,              # loop over parameters
    loop(uplo,          # loop over bounds
        vb.lo(pri) = bound(pri,'low');
        vb.up(pri) = bound(pri,'upp');
        vb.fx(pri2) = bound(pri2,uplo);
        solve mlin using cns;
        line(k,uplo)$(mlin.modelstat = 15) = vb.l(k);
        ));

* The posterior mean is the midpoint of the line just identified
line(k,'mid') = (line(k,'upp') + line(k,'low'))/2;
estb(k,'f2') = line(k,'mid');

display line,bound,estb;


*-------------------------------------------------------------------------
* SEC 3     Generalized maximum entropy with supports equal to bounds
*-------------------------------------------------------------------------
```

```
variable prob(k,uplo);
equation eentropy, egme(k),eqaddup(k);

eentropy .. p =e= -sum(pri,    vnb(pri) *log(  vnb(pri))
                        + (1-vnb(pri))*log(1-vnb(pri)));
egme(pri) .. vb(pri) =e= sum(uplo, prob(pri,uplo)*bound(pri,uplo));
eqaddup(pri) .. sum(uplo, prob(pri,uplo)) =e= 1;

model mgme /eentropy,eql,eqnb/;
prob.lo(pri,uplo) = 0.00001; prob.l(pri,uplo) = 0.5;
vb.lo(pri) = bound(pri,'low');
vb.up(pri) = bound(pri,'upp');
solve mgme using nlp maximising p;

estb(k,'f3') = vb.l(k);


*-------------------------------------------------------------------------
* SEC 4     Beta distributed normalised priors
*-------------------------------------------------------------------------

scalar   t          'parameter of symmetric beta' /2/;
variable ff(k)      'intermediary variable'
equation eqbp(k)    'definition of beta density';
equation eqff(k)    'definition of special density';

eqbp(pri) .. ff(pri) =e= 1/beta(t,t)*vnb(pri)**(t-1)*(1-vnb(pri))**(t-1);
eqff(pri) .. f(pri) =e= ff(pri);            t = 2;

model mhpdbeta /eql,eqnb,eqbp,eqpd,eqff/;
vnb.lo(pri) = 0.00001;
solve mhpdbeta using nlp maximising p;
estb(k,'f4') = vb.l(k);


*-------------------------------------------------------------------------
* SEC 5     Beta distributed normalised priors, normally distributed noise
*-------------------------------------------------------------------------

parameter e(i)      'additive error';
parameter ys(i)     'Y plus error';
variable ve(i)      'estimated error';
equation eqls(i)    'stochastic version of linear model';
equation eqlogp     'sum of logged probabilities';

eqls(i) .. ys(i) =e= sum(k, x(i,k)*vb(k)) + ve(i);
eqlogp  .. p =e= sum(pri, log(vnb(pri) -sqr(vnb(pri))))
              - sum(i, 0.5*sqr(ve(i)));

model mNoiseHPD /eqls,eqlogp,eqnb/;
ys(i) = y(i) + normal(0,1);
solve mNoiseHPD using nlp maximising p;
estb(k,'f5') = vb.l(k);

* Save everything to a GDX file for later analysis

display estb,ys; execute_unload 'hpd1.gdx';
```

Chapter 5 Estimation of supply response in CAPRI

1. Introduction

The primary objective of this research is to develop a robust method for estimating the behavioural parameters of the supply module in the regionalised European agricultural sector model CAPRI, utilizing the time series of observations available in the CAPRI database and the optimality conditions of the model. As a secondary objective, the current model assumption of constant yields will be reviewed and, if feasible, revised.

The CAPRI model is a constrained quadratic programming model for NUTS2 regions in 34 European countries, where agriculture in each region is represented by an instance of a template programming model.

In this context we only consider the arable annual crop producing part of the representative regional farm, keeping other parts (husbandry, permanent grassland and permanent crops) fixed when necessary or leaving them out altogether when possible. We also ignore the fertilization constraints of the full model, working with Leontieff fertilizer input coefficients. With those restrictions, we need to estimate parameters for a maximum of 23 land use activities using ten inputs in 172 regions in EU-15 (thus excluding new member states).

Since most regions produce only a subset of the 23 crops, and some regions have too short time series of data, the actual extent of the exercise is somewhat smaller. Still, it is a large scale application that requires a method equally applicable to all regions and that is robust to data problems. The full list of crops and crop groups (see following sections) is provided in appendix 1, table 16. The ten inputs are listed in table 17.

Data for the model is provided by the CAPRI database. The part of the dataset that is relevant for this research has been compiled from the *Economic Accounts for Agriculture* (EAA, production values and volumes at national level) and *New Cronos Regio* (acreages and yields on regional level), both databases from Eurostat, completed with policy information from regulations and expert data where necessary. The dataset has been processed by econometric/heuristic software of the CAPRI system to be made complete (no holes in time series) and consistent (with respect to physical and economical interrelations) on member state as well as NUTS2 level.

The estimation is explicitly uses the optimality conditions of the model, together with some parameter restrictions. Prior information is included in a Bayes-

ian estimation approach as outlined in the previous chapter, and the point estimator is computed by numerical maximization of the posterior density. The statistical model estimated resembles the Bayesian analysis of the measurement error model in Zellner (1971), but is more complex since it instead of the linear model in Zellner (eq. 5.31) has a system of equation representing the optimality conditions of CAPRI, and instead of the additive measurement error model for the "exogenous" (Zellner eq. 5.30) it relates some model parameters to observations through a simple expectation model.

The report is outlined as follows: In section two we describe the structure of the template regional representative farm model that is used for all regions. The existing model has *fixed* input and output coefficients. In order to check whether that is a good specification, two sections follow that investigate two different extensions of the model to endogenous yields. In section three we test for all regions of the model whether yield significantly depends on inputs. Section four analyses in greater depth for one single region, selected for its good data quality (long time series, many crops produced) whether changed acreages lead to changed yields. Since it is concluded that none of the extensions in section three and four is statistically reliable, we return in section five to the primary objective to estimation of the model with fixed yield coefficients. In section six, results are presented for selected regions, and compared to the results of other studies.

2. A regional supply model

The regional representative farm is assumed to act as if solving a linearly constrained quadratic programming problem (1) in every time period *t*. Throughout this chapter we generally use lower case bold face letters to represent items that are column vectors for each *t*, upper case bold face letters to represent matrices and italic letters to represent scalars. The dimensions of vectors and matrices are denoted by upper case letters, where a lower case version of the same letter denotes the indices of the elements in that dimension, so that for instance the "*J*-vector of acreages **x**" means a vector of length *J*, with elements x_j, $j = 1 \ldots J$. The *prime* character (′) denotes the ordinary transpose of a vector or a matrix.

All regional models have identical structure, and no cross-regional constraints or relationships are assumed, in order to keep down the complexity of the estimation. Thus, indices for regions can be omitted. The producer is assumed to solve the optimization problem in each period independently of other periods, thus all items that change across periods obtain an index *t*, so that for example $\mathbf{x}_t$ denotes the vector **x** in period *t*. This implies that **x** also can be considered a 3-dimensional array with dimensions with only one column, or $\dim(\mathbf{x}) = (J,1,T)$. At some occasions it is convenient to denote the time series for some element *j* of **x**,

and this is done somewhat sloppy as $\mathbf{x}_j$, where the reader is assumed to remember that $\mathbf{x}$ also has another dimension T, which is now in the rows[17].

The model can then be written for each period as

$$\begin{aligned} &\max_{\mathbf{x}_t} \mathbf{x}_t'\left[\mathbf{Y}_t\mathbf{p}_t + \mathbf{s}_t - \mathbf{A}_t\mathbf{w}_t\right] - \mathbf{x}_t'\left[q_t\mathbf{c} - \tfrac{1}{2}l_t\left[\mathbf{D} + \mathbf{GBG}'\right]\mathbf{x}_t\right] \\ &\text{s.t.} \quad \mathbf{R}_t\mathbf{x}_t = \mathbf{v}_t \end{aligned} \tag{1}$$

where for each t,

$\mathbf{x}_t$ vector of acreages for each of J land uses

$\mathbf{Y}_t$ $J \times J$ diagonal matrix with yields on the diagonal

$\mathbf{p}_t$ J vector of prices

$\mathbf{s}_t$ J vector of direct subsidies

$\mathbf{A}_t$ $J \times I$ matrix of input coefficients for I inputs

$\mathbf{w}_t$ I vector of input prices

q_t price index

$\mathbf{c}$ J vector of parameters

l_t a land availability index (described further below)

$\mathbf{D}$ $J \times J$ diagonal matrix of parameters

$\mathbf{G}$ $J \times M$ matrix that sums up land use by each of M = 6 crop groups, i.e. with $g_{jm} = 1$ if crop j belongs to group m, else $g_{jm} = 0$

$\mathbf{B}$ 6×6 matrix of parameters

$\mathbf{R}_t$ $2 \times J$ matrix of constraint coefficients, where $r_{1j} = 1$ for $j = 1 \ldots J$ and r_{2j} is the net set-aside contribution of crop j

$\mathbf{v}_t$ 2 vector with v_1 total land available, $v_2 = 0$.

The model implies that the producer maximises the sum of gross margins (the first term) minus a quadratic function (the second term), subject to a land constraint and set-aside requirement. The quadratic function in the objective function is a behavioural term in the tradition of *positive mathematical programming* (PMP, see e.g. Horner et al. 1992 or Howitt 1995) that is intended to capture the aggregated influence of economic factors that are not explicitly included in the model, like land heterogeneity and additional resource constraints (Heckelei 2002). The function is in what follows sometimes referred to as the *PMP-function*, and the parameters $\mathbf{c}$, $\mathbf{D}$ and $\mathbf{B}$ as the *PMP parameters*, or the behavioural parameters of the model. It is the objective of this work to estimate those parameters.

In order to reduce the number of parameters to estimate, we assume that the quadratic function has a special structure: Cross-crop effects are only permitted between *groups of crops*, so that for instance an increase in the area of potatoes

[17] I.e. we perform a generalised transpose of the 3-D array $\mathbf{x}$ where the first and last dimensions are swapped, and signal this only by a switch of indices. In general, symbols are better thought of as 3-D arrays where the index denotes the 3rd dimension.

plus sugar beet may influence the cost of producing cereals and increase the cost of producing both sugar beet and potatoes. In order to provide each individual crop with increasing marginal costs[18], we also admit a quadratic term that depends only on the individual crop. The structure described is implemented using a vector **c** of linear effects, a diagonal matrix **D** of quadratic own-crop effects, and a matrix **B** of cross-group effects. The $J \times M$ matrix **G** is used to sum the acreages within each group, substantially reducing the number of parameters compared to estimation of a full $J \times J$ matrix.

The prices **p** and **w** in the model are nominal, and since the quadratic function is assumed to capture, among other things, the opportunity cost of resources not explicitly modelled, it should be inflated. This is obtained by multiplication of **c** by the general price index q_t.

The total amount of land fluctuates slightly between years, in general with a downward trend due to migration of land into other sectors (fallow land is modelled explicitly as a land use activity). We do not know if it is productive or unproductive land that migrates, so to avoid that land migration strongly influences land rent (the dual value of the first constraint), we use land shares instead of absolute land use in the quadratic term of the PMP-function. This is equivalent to scaling the matrix [**D** + **GBG**′] by the square inverse of total land available in each period. In order to obtain values approximately interpretable as "marginal cost change in euro per hectare" it is also multiplied by ½ times square of total land available in year 2000, or $(v_1)_{2000}$. Thus, the $l_t = ((v_1)_{2000}/(v_1)_t)^2$.

The optimization model (1) can be equivalently described by the following first- and second order conditions for optimal **x**

$$\mathbf{Y}_t\mathbf{p}_t + \mathbf{s}_t - \mathbf{A}_t\mathbf{w}_t - q_t\mathbf{c} - l_t\left[\mathbf{D} + \mathbf{GBG}'\right]\mathbf{x}_t - \mathbf{R}_t'\boldsymbol{\lambda}_t = \mathbf{0} \tag{2}$$

$$\mathbf{R}_t\mathbf{x}_t = \mathbf{v}_t \tag{3}$$

$$\mathbf{B} = \mathbf{U}'\mathbf{U} \tag{4}$$

$$d_{jj} \geq 0 \text{ for } j = 1\ldots J \text{ (and } d_{ij} = 0 \text{ for } i \neq j) \tag{5}$$

$\boldsymbol{\lambda}_t$ is the 2×1 vector of dual values for the constraints. Note that for positive semi-definiteness of the Hessian matrix it is sufficient that **B** is positive semi-definite, which is satisfied by the Cholesky factorisation with the upper triangular matrix **U**, and that all elements of **D** are non-negative[19].

[18] More precisely, to ensure a strictly definite Hessian matrix.

[19] In fact, we will use a stronger restriction of $d_{jj} \geq \delta_{ij} > 0$ in estimations to avoid numerical problems when estimating elasticities.

The primary objective of the chapter can now be more precisely formulated as estimating the PMP parameters by using the optimality conditions as estimating equations.

The secondary objective of evaluating the assumption of constant yields can be restated as an attempt to lift some of the non-linearity out of the PMP-function and explicit it in the form of a non-constant marginal gross value added, i.e. to estimate the relationship between yields and input use.. The first such extension is a variant of the model where yield depends endogenously on input use (land counting as an input). A second extension is the lesser modification that yields depend on allocated acreage.

3. Should yield depend on input use?

3.1. Motivation

The purpose of this section is to determine if prices of outputs and inputs are important determinants of yields of major agricultural crops in the EU. If a significant relationship between prices and yields can be identified, yields should be an endogenous function of input use in the CAPRI model, else input use and yields should be treated as exogenous to the model. The underlying idea is that perhaps some of the nonlinearity of the model, which is currently modelled only by the quadratic cost component, can be explained more explicitly (cf. Heckelei 2002). To decide which of those two alternative formulations to use, we estimate a yield function.

We start from the microeconomic model (1), and augment it with yields endogenously depending on **x** and **A** as in equation (6). We thus assume that yield Y_{jt} of crop j in period t can be approximated by a function that is quadratic in inputs $\mathbf{A} = (a_{ij})$, linearly dependent on planned number of hectares **x** and on trend T and with a random term $\boldsymbol{\varepsilon}$:

$$Y_{jt} = \gamma_{0j} + \gamma_{1j} T_t + \gamma_{2j} x_{jt} + \alpha_{1ij} a_{ijt} + \alpha_{2ij} a_{ijt}^2 + \varepsilon_{jt} \qquad (6)$$

In this estimation, it is assumed that the acreage allocation **x** is the optimal solution to the maximization problem at some expected prices and yield. We may then use the envelope theorem to obtain the optimality conditions for input use.

The first order condition for profit maximum of the extended model with respect to **A** at the expected output prices **p** and input prices **w** can be written

$$\frac{\partial Y_{jt}}{\partial a_{ijt}} p_{jt} = \left(\alpha_{1ij} + 2\alpha_{ij} a_{2ijt}\right) p_{jt} = w_{it}$$

Solving for the optimal input quantities gives $a_{ijt}^{*} = (w_{it}/p_{jt} - \alpha_{1ij})/(2\alpha_{2ij})$. Substituting that expression into the yield function (6) and defining

$$\beta_{0j} = \gamma_{0j} - \sum_i \frac{\alpha_{1ij}^2}{4\alpha_{2ij}}$$

$\beta_{1j} = \gamma_{1j}$

$\beta_{2j} = \gamma_{2j}$

$\beta_{3ij} = \frac{1}{4\alpha_{2ij}}$ and

$r_{ijt} = w_{it}/p_{jt}$

gives us an expression for yields that depends on the square price ratio r_{ijt}:

$$Y_{jt} = \beta_{0j} + \beta_{1j} T_t + \beta_{2j} x_{jt} + \sum_i \beta_{3ij} r_{ijt}^2 + \varepsilon_{it} \qquad (7)$$

The second-order condition for a profit maximum is that $\alpha_{2ij} < 0$, so we expect β_3 to be negative. Without that condition holding true, we will not obtain useable estimates, and we would better choose exogenous yields. Thus, we want to test the hypothesis that $\beta_3 < 0$ versus $\beta_3 \geq 0$.

3.2. Data

The estimation is carried out on NUTSII level for the EU15 member states. All input prices have been aggregated to a single input price index by first computing the Laspeyres price index of the aggregates "plant protection" (PLAP) and "all other inputs" (REST), with the average total input quantities 2001-2003 as weights, and then merging them into a single input price index for each crop by computing the Laspeyres price index using the average 2001-2003 crop specific input coefficients as weights (input coefficients coming from the CAPRI database). Expected output prices were observed prices lagged one year (naïve price expectation), whereas input prices entered without lag.

It is crucial to be able to separate the effect of *trend* from that of the other explanatory variables., The squared price ratio is, however, likely to contain a trend component as well, which we will not be able to separate from the pure trend. To be on the safe side, i.e. not to find a significant influence of prices that is really only the influence of the trend in prices, we subtract linear trends from the explanatory variables **x** and **r**. This is done by fitting and subtracting a simple trend from each variable $\boldsymbol{\xi}$ using the equation

$$\boldsymbol{\xi}^* = \boldsymbol{\xi} - \mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\boldsymbol{\xi}$$

where **C** is the $n \times 2$ matrix with ones in the first column and the sequence from 1 to n in the second column, and $\boldsymbol{\xi}$ a time series for some exogenous variable to clear of trend.

3.3. Estimation method

The principal estimation method used is Least Squares. Three problems are likely to be present in the data set, so that some modifications of the ordinary least squares (OLS) seem appropriate. Firstly, there may be problems with endogeneity, because the lagged price ratio is likely to influence acreage. To avoid obtaining biased estimates, we try an alternative estimation where the trend free acreage is instrumentalized by lagged acreage, lagged output price, lagged price index of substitutes, direct subsidies and the other explanatory variables in (7) except for lagged squared price ratio. Denoting, for now, the $T \times K$ matrix of K explanatory variables T years for each crop j by $\mathbf{X}_j$ (not to confuse with acreage $\mathbf{x}_j$) the instrumental variables matrix by $\mathbf{Z}_j$, and the vector of coefficients by $\boldsymbol{\beta}_{IV}$, we estimate

$$\hat{\boldsymbol{\beta}}_{IVj} = \left(\hat{\mathbf{X}}'_j \mathbf{X}_j\right)^{-1} \mathbf{X}'_j \mathbf{y}_j, \;\; \hat{\mathbf{X}}_j = \mathbf{Z}_j \left(\mathbf{Z}'_j \mathbf{Z}_j\right)^{-1} \mathbf{Z}'_j \mathbf{X}_j \;\; \text{for } j = 1 \ldots J$$

The correlation between acreage and instrumentalised acreage should be rather high for the instrumentation to make sense. The coefficients of correlation are shown for all relevant crops in table 1. Albeit there are some cases with low correlation, the general impression is that the instrumentation is good, with 55% of the correlations greater than 0.80.

Table 1: Correlation between acreage and instrumentalized acreage

	BL	DK	DE	EL	ES	FR	IR	IT	NL	AT	PT	SE	FI	UK
SWHE	0.72	0.84	0.86	0.87	0.87	0.70	0.54	0.61	0.39	0.83	0.79	0.78	0.90	0.54
DWHE			0.84	0.67	0.89	0.92		0.85		0.90	0.84			0.95
RYEM	0.89	0.55	0.87	0.61	0.83	0.96		0.42	0.56	0.89	0.79	0.77	0.61	0.83
BARL	0.92	0.74	0.87	0.94	0.78	0.88	0.96	0.89	0.70	0.85	0.52	0.90	0.79	0.91
OATS	0.79	0.65	0.96	0.46	0.97	0.99	0.63	0.91	0.87	0.77	0.85	0.74	0.88	0.66
MAIZ			0.78	0.97	0.78	0.53		0.86	0.83	0.85	0.74			
OCER	0.60		0.63	0.92	0.78	0.90		0.81			0.95			0.90
RAPE		0.86	0.87		0.93	0.79	0.84	0.94		0.96		0.87	0.26	0.41
SUNF			0.91	0.97	0.96	0.82		0.86		0.88	0.77			
SOYA					0.86	0.71		0.74		0.69				
PULS	0.78	0.40	0.96	0.98	0.80	0.95		0.93	0.89	0.93	0.79	0.53	0.84	0.73
POTA	0.88	0.88	0.99	0.69	0.94	0.74	0.53	0.87	0.68	0.89	0.43	0.64	0.73	0.62
SUGB	0.77	0.94	0.76	0.54	0.94	0.68	0.82	0.72	0.77	0.99	0.54	0.68	0.88	0.90
MAIF		0.98	0.80	0.64	0.91	0.53		0.97	0.80	0.84	0.85	0.58		
OFAR	0.72	0.69	0.95	0.99		0.72	0.82	0.62	0.71	0.93	0.63	0.85	0.81	

Secondly, a strong correlation between error terms of certain crops should be expected due to the similar influence of weather on similar crops. For example, one should expect a positive correlation between the yields of barley and rye, because their vegetative periods are similar and they have similar requirements

regarding weather and soil. Thus, a seemingly unrelated regression (SUR) would be appropriate. Such an estimator would be more efficient than OLS would the covariance matrix be known. In the current case, the covariance matrix is not known, but has to be estimated, which may hamper efficiency considerably. This was tried out using iterated SUR with and without the instrumentation above. The estimation was carried out in three steps: (i) instrumentation of $\mathbf{X}$ by $\hat{\mathbf{X}}$ as above, (ii) iterated SUR of $\mathbf{Y}$ on $\hat{\mathbf{X}}$ to obtain stable weights matrix $\mathbf{W}$, which was computed from the inverse covariance matrix $\boldsymbol{\Sigma}$ of the error terms of the regression of $\mathbf{Y}$ on $\hat{\mathbf{X}}$, weighting each element of the covariance matrix by the harmonic mean of the degrees of freedom of the relevant equations, $DF_{ij} = \sqrt{(N_i - K_i)(N_j - K_j)}$, and (iii) computation of estimator $\boldsymbol{\beta}_{SUR} = \left(\hat{\mathbf{X}}'\mathbf{W}\mathbf{X}\right)^{-1}\hat{\mathbf{X}}'\mathbf{W}\mathbf{y}$. The index free matrices represent the stacked system as in Greene (2003, p. 342). $\mathbf{X}$ is the $(JT) \times (JK)$ partitioned matrix with matrix $\mathbf{X}_j$ on the j^{th} diagonal position and zeros elsewhere, and similar for $\hat{\mathbf{X}}$. $\mathbf{W} = \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}$, and $\mathbf{y}$ the vertically concatenated vectors $\mathbf{y}_j$.

Thirdly, there could be an aggregation bias. It may well be that for example a price increase has a greater production response in a sub region with generally low yields. The weight of the low yield region in the aggregate would increase, leading to reduced aggregate yield although the yield in each sub-region increased as response to the higher price. To investigate this effect to the extent possible by the available data, the regressions were re-run on sub national level (NUTS2 where possible, UK NUTS1). Prices are only available on national level. They were mapped down to the respective sub regions. Acreages and yields, on the other hand, are also available for NUTS2 regions.

Table 2: Different estimation methods tried

Estimation nr.	Regional resolution	Acreage instrumentation	SUR
1	national	no	no
2	sub regions	no	no
3	national	yes	no
4	sub regions	yes	no
5	national	no	yes
6	sub regions	no	yes
7	national	yes	yes
8	sub regions	yes	yes

Alltogether, eight different estimations were run to account for each of the three problems. The estimation setups are shown in table 2. The estimations were evaluated based on the number of significant coefficients using t-tests on the 5% level. The t-tests were computed for the test $\beta_{3j} = 0$ using standard deviations of the vector of estimators computed as the square root of the diagonal elements of

$\mathrm{Cov}(\boldsymbol{\beta}) = \left(\hat{\mathbf{X}}'\mathbf{W}\hat{\mathbf{X}}\right)^{-1}$. Note that in the case of no instrumentation, $\hat{\mathbf{X}}$ = **X**, and $\mathbf{W}^{-1}$ becomes the standard variance estimator with degrees of freedom correction.

3.4. Results

The results indicate that there is a relationship between yields and input prices and also between acreages and yields in some regions for some crops., The relationships ,however, can not be statistically detected for all crops in all regions. For the *major share* of all crops no significant influence at all of neither input prices nor acreages on yields is found. The results also show that the sophistication of the estimation method by the instrumentation of acreages, use of sub regions and SUR covariance structure is worthwhile, because the number of significant coefficients increase by their introduction, and the signs of the price influence tend to be more conform with theory (which suggests a negative influence of the output-input price ratio). Table 3 shows the number of estimated equations, the number of coefficients with positive and negative signs and the number of coefficients significantly different from zero with each sign.

Since a rather large number of t-tests were carried out at the 5% level, one would expect 5% of the tests to show a significant $\beta_3 \neq 0$ even if the true $\beta_3 = 0$. For example, in the estimations with sub regions, 1858 t-tests were carried out. We would then expect 2.5% of 1858 = 93 tests to show *b* significantly different from zero in each direction even if the true *b* = 0. Even with this in mind, it seems that the number of significant coefficients is too large to be a pure random outcome (e.g. 184 negative significant to 96 positive significant out of 1858 tests for regionalised iterated SUR estimation with instrumentation). Therefore, we conclude that there is indeed a general influence of prices on yields, but that the influence is so hard to detect statistically that it does not seem worthwhile to try to estimate an economic model with endogenous yields.

Table 3: Summary of results for different estimation setups.

Est. nr.	Eq.	b_3<0	b_3>0	b_3<0*	b_3>0*	b_2<0	b_2>0	b_2<0*	b_2>0*
1	163	90	73	12	8	95	68	20	6
2	1858	1097	745	137	63	1051	807	204	114
3	163	95	68	9	7	96	67	14	4
4	1858	1112	730	125	61	1024	834	117	103
5	163	98	65	24	9	84	79	29	11
6	1858	1032	810	189	117	995	863	273	192
7	163	98	65	28	6	90	73	31	21
8	1858	1062	780	184	96	986	871	176	178

Est. nr. refers to estimation number in table2, b_3 *is the coefficient of price ratio,* b_2 *the coefficient of acreage, and a star refers to significance of 5 % level double sided t-test.*

Why is there no statistically reliable influence of prices on yields? It is well known that yield of most crops is a concave function of inputs. Given profit

maximizing behaviour of producers, a relationship similar to that estimated here should result. There are, however, at least five major obstacles involved.

(1) The quadratic yield model implied here may be wrong. In reality, yield also depends on a lot of other factors that are all collected in the error term. Crop rotation is certainly a significant determinant of yield that is not controlled for in these estimations. This could be introduced by a careful selection of substitute acreages. A share of this influence should already be represented by the inclusion of own acreage, and introducing further explanatory variables would reduce the degrees of freedom and aggravate the problems with endogeneity (acreages depending on prices)

(2) The producers may not be rational in the way assumed here. Output price expectations may not be naïve, and the decision on input use may have to be taken with some time lag so that an input price expectation is required as well. It may also be the case that the yield function is largely unknown to the producer, so that rational behaviour as in the conceptualized model is impossible. Producers are perhaps more likely to choose input amounts from a table or heuristic with very few, if any, alternative levels of inputs. As an alternative price expectation, the formula $0.67P_{t-1} + 0.33P_{t-2}$ was tried, but without improvement in fit.

(3) The yield function may have a shape that implies almost the same input use and yield for a wide range of price ratios, so that there are almost only two different profit maximizing solutions: either “zero” or “full” input use. That would be the case if the graph of yield to inputs has an almost linear initial part and then bends sharply downwards at some point. Then the influence of the price ratio would be “almost” discontinuous, with almost no change in yield for moderate price ratio changes and a big leap at some point. Then, for most price ratios, the optimal yield choice is almost the same.

(4) It may well be the case that the sub regional level used in the estimations 2, 4, 6 and 8 is still too aggregated so that an aggregation bias remains.

(5) The data sampling model underlying the estimations is inappropriate. Actually, observed acreages and prices are only indicators of the true (latent) *planned acreage* and *expected price*. Because the errors on acreages and prices now (erroneously) are attributed to measurement errors in yields, the estimated variance is too large, and thus the tests likely weaker. The coefficients are also likely to be biased in unknown directions (Fuller 1987). In addition, the observed yield is the *average* yield, whereas if yield really is endogenous the decision is based on the *expected marginal* yield. Actually, a model including measurement errors and marginal yield expectation together with the full optimality conditions (2-5) was the starting point of the estimation, but proved too complex to handle efficiently. Thus, the estimations of yield functions were performed in this separate step to determine whether endogenous yield should be part of the final model.

We return to the measurement error model and yield expectations below, though without endogenous input coefficients.

4. Should yields depend on land allocation?

4.1. Problematic marginal cost curves

If the zero-profit condition (2) is solved for **x** we find that acreages are linearly depending on prices according to the relation

$$\mathbf{x}_t = l_t^{-1}\left[\mathbf{D} + \mathbf{G}\mathbf{B}\mathbf{G}'\right]^{-1}\left[\mathbf{Y}_t\mathbf{p}_t + \mathbf{s}_t - \mathbf{A}_t\mathbf{w}_t - q_t\mathbf{c} - \mathbf{R}_t'\boldsymbol{\lambda}_t\right] \qquad (8)$$

Because the matrix $l_t^{-1}\left[\mathbf{D} + \mathbf{G}\mathbf{B}\mathbf{G}'\right]^{-1}$ is required to be positive semi-definite by the second order conditions, we expect the graph of $\mathbf{x}_t$ to gross margin $\mathbf{m}_t = \mathbf{Y}_t\mathbf{p}_t + \mathbf{s}_t - \mathbf{A}_t\mathbf{w}_t$, to be an upward sloping curve, so that increasing gross margin leads to increased acreage. Figure 1 shows the development of rye acreage and gross margin (nominal prices) between 1985 and 2003 for one of the most important cereals producing regions in France, the Nuts 2 region with code FR24 (Centre). Obviously, it would be difficult to fit acreage to gross margin with a positive slope if no other information is included, because the gross margin has increased whereas production decreased. In fact, the coefficients in an OLS regression of acreage on constant and gives the slope coefficient -0.0122 with a p-value of 0.0152. The points and the fitted line are shown in figure 2.

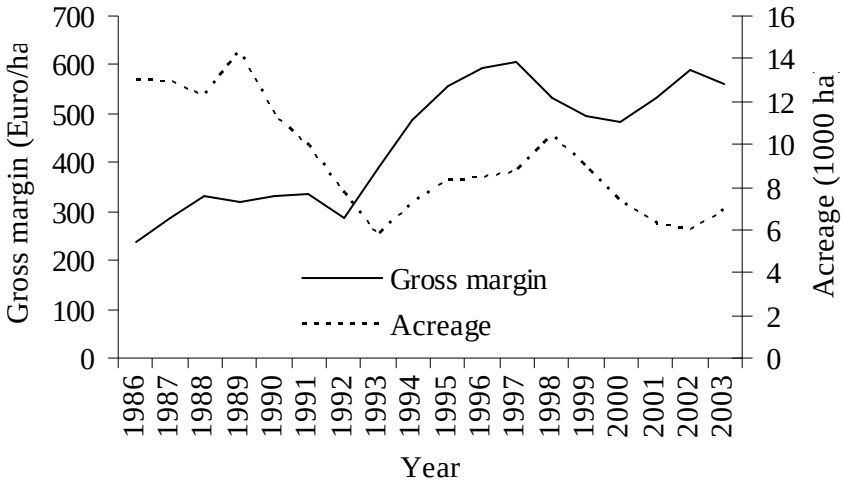


Figure 1. Gross margin and acreage of rye in FR24.

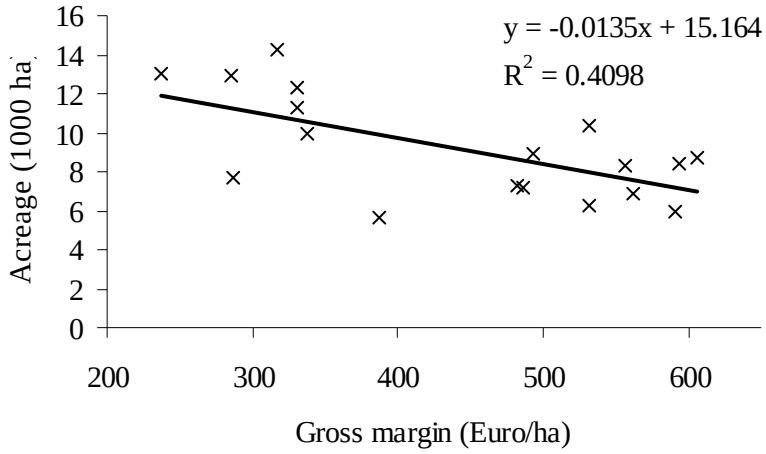


Figure 2. Acreage linearly fitted to gross margin for rye in FR24.

Thus, something more is influencing the producer decision to decrease rye production despite apparently increasing gross margin. Several auxiliary hypotheses come to mind. For instance, we tacitly assumed that the dual vector $\boldsymbol{\lambda}$ was constant, whereas it in fact $\boldsymbol{\lambda}$ depends on the gross margins of all other crops. Perhaps gross margins in, say the most important crop soft wheat, has increased enough to increase land price enough to force back rye. Figure 3 shows acreage and gross margin in soft wheat in the same region and time period. As can be seen in the figure, the gross margin in soft wheat has decreased slightly during the time period, which is not favourable for that hypothesis (though it is not enough to reject it; soft wheat may have been the wrong crop).

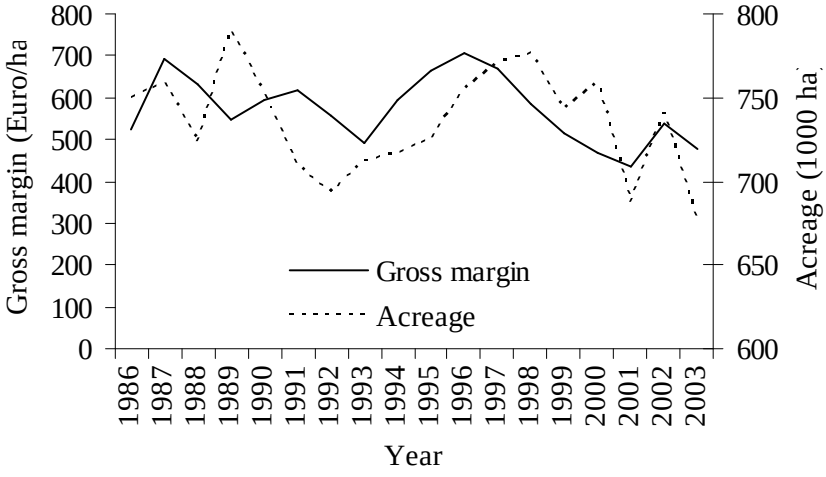


Figure 3. Gross margin and acreage of soft wheat in FR24.

A second assumption in the simple regression of acreage on gross margin was that the coefficient is constant over time. Comparison of the regression model with the equation (8) derived from the first order conditions reveals that the coefficient contains the parameters **c**, **D** and **B** which change over time with price index and total area. Thus the cost component **c** actually increases in nominal terms over time, which also helps alleviate the problem of reverse reaction of rye. A proper analysis should thus include at least the full first order conditions.

Estimation of (2-5) for all crops simultaneously, with a measurement error approach[20] allowing for errors on **x**, **Y**, **p**, **A** and **w**, and endogenous dual values with prior information for identification, did however result in a boundary solution for **D** and/or **B**. The boundary solution is such that rye obtains as small a coefficient as possible, still yielding a positive definite matrix. That implies an elasticity of supply of rye of close to infinity in the resulting simulation model, which is simply not plausible. That model is further discussed in the next section.

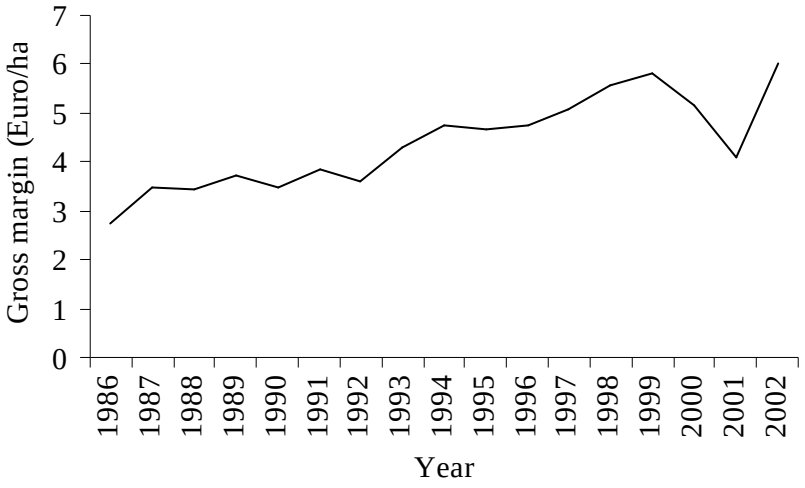


Figure 4. Average yields of rye in FR24

The rest of this section discusses a third auxiliary hypothesis that is sufficient to estimate rye parameters with the expected sign. The hypothesis is based on the fact that our yield data are really *average* yields, whereas the producer is assumed to base his production decision on *expected marginal* yield. Then gross margins **m** were computed in the wrong way above, using average yields. In fact, a closer look on the components **p**, **Y**, **s**, **A** and **w** of gross margins reveals that output prices have dropped steadily, and that the main reason for the increasing gross

[20] The estimation also uses linear trends for expected yield and expected input requirements to remove stochastic weather influences, and uses prior information of 0.5 times gross margin of soft wheat for land price dual value and similar for set-aside for identification of the model.

margin is that rye yields have risen sharply from about three to about five tons per hectare (figure 4). If the marginal yield is actually depending on **x**, then the development of marginal yields may be a qualitatively different from that of averages.

4.2. Motivation for endogenous yield

It could be the case that rye, which is grown on a considerably smaller area than soft wheat, is treated as an inferior alternative of many producers, and is thus grown on soil less suitable for cereals production. If prices increase, rye becomes an increasingly competitive alternative to soft wheat on the better soils. In that case, the marginal yield of rye with respect to acreage would be an *upward* sloping function.

One could also motivate a *downward* sloping yield function (of acreage) by assuming that first the soil that is best suited for rye is used, or that there is some rotational effect favouring smaller land use for rye. To investigate which is the case, we attempt to estimate the relation between yield and acreage.

Assume that marginal yield is approximated by the linear model

$$f_{jt}(x_{jt}) = \beta_{0j} + T_t\beta_{1j} + 2x_{jt}\beta_{2j}$$

with T a linear trend, and that observations of average yields arise according to

$$y_{jt} = \frac{1}{x_{jt}} \int_0^{x_{jt}} f_{jt}(z)dz + \varepsilon_{jt} \ .$$

Integration gives the model to estimate,

$$y_{jt} = \beta_{0j} + T_t\beta_{1j} + x_{jt}\beta_{2j} + \varepsilon_{jt}. \qquad (9)$$

Note that the coefficient β_2 in the expression for the marginal yield enters with twice its estimated value. Thus, if β_{2j} is positive and x_{jt} decreases over time, then the marginal yield decreases over time compared to average yield. If the β_{2j} is big enough, this may be enough to turn the apparent positive gross margin development for example in rye in the case study region FR24 into a negative one.

4.3. Pitfalls when estimating the expected marginal yield

A straightforward least squares estimation of (9) gives a β_2 for rye of 0.04577, which is supporting the hypothesis that gross margin actually has been increasing less rapidly than indicated by the average yields. The t-test for $\beta_2 = 0$ gives a poor p-value of 0.544. The estimation, however, has at least two pitfalls that potentially make the estimation less efficient and reduces the power of the t-test of $\beta_2 = 0$.

(i) The yields of all crops tend to be correlated.

(ii) We ignore that acreage is measured with errors.

The first pitfall makes the LS estimation inefficient, because a more efficient estimator would recognise that if, say, all cereals have a low yield in 2003 (which was the actual case), then error terms in that year should have less weight in the estimation. That is, a *seemingly unrelated regression* (SUR) could be a more appropriate model (as in the previous section).

The second pitfall must be further explained. Above it was briefly mentioned that the ultimate goal is to perform an estimation with errors on the acreages **x**. So we should not now ignore that our observations of acreages may not be the true planned acreages, but acknowledge that a measurement error may be involved. If we assume the simple model that observed acreages **X** relate to true planned acreages **x** with a simple additive error model,

$$\mathbf{X}_j = \mathbf{x}_j + \mathbf{u}_j$$

then the estimates of β_2 are likely to be biased and the variances of the estimates are likely to be biased too (see Fuller 1987 for a thorough treatment of the linear measurement error model). In a simple linear model with a single explanatory variable, the coefficient is biased towards null by a factor $\kappa = \sigma_{xx}(\sigma_{xx} + \sigma_{uu})^{-1}$, and the estimated variance of the coefficient is biased by κ^2. (but t-test $\beta = 0$ is not weakened). Unfortunately the situation becomes more complicated when there are two explanatory variables (TREND and ACREAGE), one of which is measured without error (TREND). To correct for these biases, a measurement error model seems to be the appropriate method.

4.4. A seemingly unrelated regression

The SUR estimator requires knowledge of the covariance matrix of yields. If that is not available, it can be estimated in a feasible generalised least squares estimation (FGLS). In this analysis we use an iterated SUR. In the first step, we estimate the model with independent error terms (identity matrix as weighing matrix). The residuals are used to estimate the yield covariance matrix $\mathbf{\Sigma}_e$. The inverse covariance matrix $\mathbf{\Sigma}_e^{-1}$ is used in the second step to estimate the FGLS model by minimising the generalised sum of squares

$$\min \sum_{jkt} \left(Y_{jt} - \beta_{0j} - \beta_{1j} T_t - \beta_{2j} X_{jt}\right)\left(\Sigma_e^{-1}\right)_{jk}\left(Y_{kt} - \beta_{0k} - \beta_{1k} T_t - \beta_{2k} X_{kt}\right)$$

where *j*,*k* are alternative indices for *J* crops and *t* the index for time.

In order for the coefficient vector to converge, certain limitations are required to bring down the number of elements in $\mathbf{\Sigma}_e$. This was done by subdividing the crops into five groups that were conjectured to react similarly or perform a similar

function in the rotation. This is equivalent to a separate SUR estimation for each group. The groups are the ones shown in table 16 in the appendix to this chapter, except of course for the group NOCR (crops with no physical yield) which was not included. In FR24 there was sufficient data for 15 cropping activities.

The SUR estimator $\hat{\beta}_2$ for β_2 in rye is considerably smaller than the OLS estimator, 0.01878 instead of 0.04577, and the t-statistic indicates an even less significant coefficient, $P(abs(\hat{\beta}_2) \geq 0.01878 | \beta_2 = 0) = 0.598$. The block wise covariance matrix and the estimated coefficients are shown in the following tables (4-9). One can see in the table that the assumption of covariation of yields across crops within the groups is reasonable, because all items except for the covariation PULS.POTA in table 8 are positive. Nevertheless, the estimated $\hat{\beta}_2$ are significantly different from zero only in 4 out of 15 cases (determined by Student's t-test), rye not being one of them. So even if the coefficient on rye tends to have the right sign, the effect could just as well be coincidence in most cases.

Table 4. Coefficients for TREND and ACREAGE in SUR estimation

	$\hat{\beta}_1$ .value	$\hat{\beta}_1$ .p	$\hat{\beta}_2$ .value	$\hat{\beta}_2$ .p	Significance of $\hat{\beta}_2$
SWHE	0.0688	0.0090	-0.0011	0.5870	
DWHE	0.0564	0.0250	-0.0022	0.1210	
RYEM	0.1492	0.0000	0.0188	0.5980	
BARL	0.0689	0.0050	0.0001	0.9770	
OATS	0.0534	0.0140	0.0177	0.0010	***
MAIZ	0.1785	0.0001	-0.0009	0.7880	
OCER	-0.0318	0.2620	0.0511	0.0190	*
RAPE	-0.0013	0.9680	0.0012	0.7070	
SUNF	0.0150	0.4100	-0.0020	0.3220	
PULS	0.0214	0.2760	0.0066	0.1460	
POTA	0.7869	0.0030	-0.6038	0.4310	
SUGB	0.9813	0.0000	-2.1111	0.0008	***
MAIF	-0.0790	0.7810	-0.2972	0.0440	
OFAR	-0.4271	0.1190	-0.1502	0.0000	***
NONF	0.0731	0.0020	0.0072	0.3830	

Table 5. Covariance matrix of SUR residuals for Cereals

	SWHE	DWHE	RYEM	BARL	OATS
SWHE	0.305	0.244	0.202	0.210	0.151
DWHE	0.244	0.274	0.183	0.184	0.133
RYEM	0.202	0.183	0.228	0.147	0.146
BARL	0.210	0.184	0.147	0.233	0.149
OATS	0.151	0.133	0.146	0.149	0.147

Table 6. Covariance matrix of SUR residuals for Cereals2

	MAIZ	OCER
MAIZ	0.413	0.256
OCER	0.256	0.263

Table 7. Covariance matrix of SUR residuals for Oilseeds

	RAPE	SUNF	NONF
RAPE	0.129	0.018	0.038
SUNF	0.018	0.067	0.021
NONF	0.038	0.021	0.061

Table 8. Covariance matrix of SUR residuals for Other Arable Crops

	PULS	POTA	SUGB
PULS	0.205	-0.034	0.835
POTA	-0.034	9.426	2.167
SUGB	0.835	2.167	11.051

Table 9. Covariance matrix of SUR residuals for Fodder

	MAIF	OFAR
MAIF	22.212	12.228
OFAR	12.228	15.024

4.5. A measurement error model

To include the assumption that $\mathbf{X}_j = \mathbf{x}_j + \mathbf{u}_j$ into the estimation, a total least squares estimation is performed by minimising the following extremum estimation criterion, scaled by the inverse of the number of observations $n = JT$ (for J crops and T periods):

minimize

$$\frac{1}{n}\sum_{jkt}\left(Y_{jt}-\beta_{0j}-\beta_{1j}T_t-\beta_{2j}x_{jt}\right)\left(\Sigma_e^{-1}\right)_{jk}\left(Y_{kt}-\beta_{0k}-\beta_{1k}T_t-\beta_{2k}x_{kt}\right) + \frac{1}{n}\sum_{jkt}\left(X_{jt}-x_{jt}\right)\left(\Sigma_u^{-1}\right)_{jk}\left(X_{kt}-x_{kt}\right) \tag{10}$$

Here $\mathbf{\Sigma}_e$ denotes the covariance matrix between the residuals obtained from the SUR estimation mentioned previously, whereas $\mathbf{\Sigma}_u$ is a prior covariance matrix of acreages. $\mathbf{\Sigma}_u$ only contains diagonal entries that are constructed following the principle that the standard deviation always is 6⅔ percent of the sample mean

(over time for each crop), implying that virtually all outcomes are within ± 20% of the observations. That is, for σ_{jj} diagonal element of $\mathbf{\Sigma}_u$,

$$\sigma_{jj} = \left(\overline{X}_{j\bullet} \frac{0.20}{3} \right)^2 .$$

The model 10 with errors in the explanatory variables is referred to as a measurement error model (Fuller 1987), or sometimes *Errors-In-Variables-model* (EIV). The coefficients of the EIV estimation are solved for using a non-linear programming (NLP) solver software, and the results shown in the following table (10). The signs and sizes of the coefficients are generally similar to those of the SUR estimators.

Table 10. Coefficients in EIV estimation

	B0	B1	B2
SWHE	37.21941	0.00644	-0.04172
DWHE	5.58560	0.05500	-0.00250
RYEM	2.82318	0.14463	0.00699
BARL	4.91312	0.06313	0.00195
OATS	2.60144	0.05768	0.01952
MAIZ	6.17032	0.18124	-0.00051
OCER	3.05974	-0.04094	0.06173
RAPE	2.90948	-0.00335	0.00137
SUNF	2.74305	0.01259	-0.00229
PULS	3.68404	0.02101	0.00965
POTA	32.54321	0.85206	-0.86423
SUGB	166.77830	0.67172	-3.90225
MAIF	55.23074	-0.29928	-0.44824
OFAR	69.80859	-0.68218	-0.18152
NONF	0.67922	0.07145	0.00797

It would be desirable to obtain an estimator of the standard deviations of the EIV coefficient estimators. Fuller (1987) finds that he is unable to establish the exact distribution of the estimators even in the simple case with one explanatory variable. He instead derives an approximate (normal) distribution for the coefficient vector in large samples.

Here we follow another approach using asymptotic properties of extremum estimators as described in Mittelhammer, Judge, Miller (2000, ch. 7).

4.6. Asymptotic properties of the estimators in the EIV model

We start off by putting the model (10) in matrix form. Rewrite it separating the exogenous variable “acreages” that is measured with errors from the matrix of

exogenous **Z** that is known with certainty; constant and trend. Denote the coefficients of **x** by $\boldsymbol{\gamma}$ and the coefficients for **Z** by $\boldsymbol{\beta}$. We denote the true planned acreages by lower case **x** and the observed values from the statistics by the random variable upper case **X**. Then the model can be written in matrix form as

$$\begin{aligned} &\min_{\beta,\gamma,x} n^{-1}\left(\mathbf{Y}-\mathbf{z}_b\boldsymbol{\beta}-\mathbf{x}_b\boldsymbol{\gamma}\right)'\boldsymbol{\Omega}_e^{-1}\left(\mathbf{Y}-\mathbf{z}_b\boldsymbol{\beta}-\mathbf{x}_b\boldsymbol{\gamma}\right)+\left(\mathbf{X}-\mathbf{x}\right)'\boldsymbol{\Omega}_u^{-1}\left(\mathbf{X}-\mathbf{x}\right) \\ &\Leftrightarrow \min_{\beta,\gamma,x} m(\boldsymbol{\beta},\boldsymbol{\gamma},\mathbf{x}\,|\,\mathbf{Y},\mathbf{z},\mathbf{X}) \end{aligned} \tag{11}$$

where, for $\mathbf{I}_T$ of size T, and $\otimes$ the Kronecker product,

$$\boldsymbol{\Omega}_e^{-1}=\boldsymbol{\Sigma}_e^{-1}\otimes\mathbf{I}_T,\boldsymbol{\Omega}_u^{-1}=\boldsymbol{\Sigma}_u^{-1}\otimes\mathbf{I}_T$$

The vectors/matrices **Y**, **x**, **z**, **X**, and $\boldsymbol{\beta}$ are the vertically concatenated vectors/matrices $\mathbf{Y}_j$, $\mathbf{x}_j$, $\mathbf{z}_j$, $\mathbf{X}_j$, and $\boldsymbol{\beta}_j$. $\boldsymbol{\gamma}$ is the vector of γ_j for crops $j = 1\ldots J$, the subscript b denotes the block-wise diagonalisation where the j^{th} diagonal block of the $JT \times JK$ (for K columns in **z**) matrix $\mathbf{z}_b$ is $\mathbf{z}_j$ (and similar for $\mathbf{x}_b$), so that the function m can be written alternatively as

$$m=\frac{1}{n}\left(\begin{bmatrix}\mathbf{Y}_1\\ \vdots\\ \mathbf{Y}_J\end{bmatrix}-\begin{bmatrix}\mathbf{z}_1 & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \mathbf{z}_J\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_1\\ \vdots\\ \boldsymbol{\beta}_J\end{bmatrix}-\begin{bmatrix}\mathbf{x}_1 & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \mathbf{x}_J\end{bmatrix}\begin{bmatrix}\gamma_1\\ \vdots\\ \gamma_J\end{bmatrix}\right)'\begin{bmatrix}\mathbf{I}_T\sigma_e^{11} & \cdots & \mathbf{I}_T\sigma_e^{1J}\\ \vdots & & \vdots\\ \mathbf{I}_T\sigma_e^{J1} & \cdots & \mathbf{I}_T\sigma_e^{JJ}\end{bmatrix}\begin{pmatrix}\\ \cdots\\ \\ \end{pmatrix}$$

$$+\frac{1}{n}\left(\begin{bmatrix}\mathbf{X}_1\\ \vdots\\ \mathbf{X}_J\end{bmatrix}-\begin{bmatrix}\mathbf{x}_1\\ \vdots\\ \mathbf{x}_J\end{bmatrix}\right)'\begin{bmatrix}\mathbf{I}_T\sigma_u^{11} & \cdots & \mathbf{I}_T\sigma_u^{1J}\\ \vdots & & \vdots\\ \mathbf{I}_T\sigma_u^{J1} & \cdots & \mathbf{I}_T\sigma_u^{JJ}\end{bmatrix}\left(\begin{bmatrix}\mathbf{X}_1\\ \vdots\\ \mathbf{X}_J\end{bmatrix}-\begin{bmatrix}\mathbf{x}_1\\ \vdots\\ \mathbf{x}_J\end{bmatrix}\right)$$

The rightmost vector (…) in the first term is the same as the first bracketed expression, omitted to save space. The extremum estimator defined in (11) is equivalent to an element-wise weighted total least squares estimator, shown to be consistent in Kukush and Huffel (2004).

We will now attempt to obtain a Lagrange Multiplier (LM) test of the hypothesis that $\boldsymbol{\gamma} = \mathbf{0}$, following the procedure described in Mittelhammer, Judge and Miller (2000) (section 7.6). One can show that the conditions in theorem 7.3.3 are satisfied, so that $\hat{\boldsymbol{\theta}}$ is asymptotically normally distributed, with $n^{½}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{true}) \xrightarrow{d} \mathrm{N}(\mathbf{0},\mathbf{H}^{-1}\boldsymbol{\Sigma}\mathbf{H}^{-1})$, with $\mathbf{H}^{-1}$ the Hessian of m and $\boldsymbol{\Sigma}$ the covariance matrix of $n^{½}$ times the Jacobian of m, both w.r.t. $\boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta}_{true}$.

We may then use the operational Lagrange Multiplier (LM) test with the test statistic

$$LM = n\boldsymbol{\Gamma}_r'\left[\mathbf{c}\hat{\mathbf{H}}^{-1}\mathbf{c}'\right]\left[\mathbf{c}\hat{\mathbf{H}}^{-1}\hat{\boldsymbol{\Sigma}}\hat{\mathbf{H}}^{-1}\mathbf{c}'\right]^{-1}\left[\mathbf{c}\hat{\mathbf{H}}^{-1}\mathbf{c}'\right]\boldsymbol{\Gamma}_r \sim \chi^2(J,0)$$

for $\boldsymbol{\Gamma}_r$ the Lagrange Multipliers associated with the model (11) restricted by the J restrictions $\gamma_j = 0$, or in matrix form as a linear restriction of the entire parameter vector, $\mathbf{c}\boldsymbol{\theta} = \mathbf{0}$, where $\mathbf{c}$ is a $J \times (3J + JT)$ matrix of zeros and ones constructed by horizontal concatenation of $(J \times 2)$ zeros, $\mathbf{I}_J$ and $(J \times JT)$ zeros. Differentiation of m gives the Jacobian $\mathbf{J}(m)$ as column vector as

$$\mathbf{J}(m) = \begin{bmatrix} \dfrac{\partial m}{\partial \boldsymbol{\beta}} \\ \dfrac{\partial m}{\partial \boldsymbol{\gamma}} \\ \dfrac{\partial m}{\partial \mathbf{z}} \end{bmatrix} = n^{-1} 2 \begin{bmatrix} \mathbf{z}_b' \boldsymbol{\Omega}_e^{-1} (\mathbf{Y} - \mathbf{z}_b \boldsymbol{\beta} - \mathbf{x}_b \boldsymbol{\gamma}) \\ \mathbf{z}_b' \boldsymbol{\Omega}_e^{-1} (\mathbf{Y} - \mathbf{z}_b \boldsymbol{\beta} - x_b \boldsymbol{\gamma}) \\ \boldsymbol{\gamma}_d \boldsymbol{\Omega}_e^{-1} (\mathbf{Y} - \mathbf{z}_b \boldsymbol{\beta} - \mathbf{x}_b \boldsymbol{\gamma}) + \boldsymbol{\Omega}_u^{-1} (\mathbf{x} - \mathbf{X}) \end{bmatrix}$$

$$= n^{-1} 2 \begin{bmatrix} \mathbf{z}_b' \boldsymbol{\Omega}_e^{-1} \mathbf{e} \\ \mathbf{x}_b' \boldsymbol{\Omega}_e^{-1} \mathbf{e} \\ \boldsymbol{\gamma}_d \boldsymbol{\Omega}_e^{-1} e - \boldsymbol{\Omega}_u^{-1} \mathbf{u} \end{bmatrix} = n^{-1} 2 \begin{bmatrix} \mathbf{z}_b' \boldsymbol{\Omega}_e^{-1} & 0 \\ \mathbf{x}_b' \boldsymbol{\Omega}_e^{-1} & 0 \\ \boldsymbol{\gamma}_d \boldsymbol{\Omega}_e^{-1} & -\boldsymbol{\Omega}_u^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{e} \\ \mathbf{u} \end{bmatrix}$$

where $\boldsymbol{\gamma}_d$ denotes the $JT \times JT$ diagonal vector with $\boldsymbol{\gamma} \otimes \boldsymbol{\iota}_T$ (for $\boldsymbol{\iota}$ vector of "1") on the diagonal. Since $\mathrm{E}(\mathbf{e}) = \mathrm{E}(\mathbf{u}) = \mathbf{0}$, we have that $\mathrm{E}(\mathbf{J}(m)) = \mathbf{0}$. The Jacobian is a linear combination of the (assumed) normally distributed random variables in $[\mathbf{e}' \; \mathbf{u}']'$, the covariance matrix

$$\mathrm{est.cov}\left(\begin{bmatrix} \mathbf{e} \\ \mathbf{u} \end{bmatrix} \right) = \hat{\sigma}^2 \boldsymbol{\Omega} = n^{-1} \left(\mathbf{e}' \boldsymbol{\Omega}_e^{-1} \mathbf{e} + \mathbf{u}' \boldsymbol{\Omega}_u^{-1} \mathbf{u} \right) \begin{bmatrix} \boldsymbol{\Omega}_e & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_u \end{bmatrix} = m(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{x} \mid \mathbf{Y}, \mathbf{z}, \mathbf{X}) \begin{bmatrix} \boldsymbol{\Omega}_e & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_u \end{bmatrix}$$

so the covariance $\boldsymbol{\Sigma}$ of $n^{½}\mathbf{J}(m(\boldsymbol{\theta}))$ evaluated at the estimated $\boldsymbol{\theta}$ computed is given by

$$\boldsymbol{\Sigma} = m(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{x} \mid \mathbf{Y}, \mathbf{z}, \mathbf{X}) n^{-½} 2 \begin{bmatrix} \mathbf{z}_b' \boldsymbol{\Omega}_e^{-1} & 0 \\ \mathbf{x}_b' \boldsymbol{\Omega}_e^{-1} & 0 \\ \boldsymbol{\gamma}_d \boldsymbol{\Omega}_e^{-1} & -\boldsymbol{\Omega}_u^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Omega}_e & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_u \end{bmatrix} \begin{bmatrix} \mathbf{z}_b' \boldsymbol{\Omega}_e^{-1} & 0 \\ \mathbf{x}_b' \boldsymbol{\Omega}_e^{-1} & 0 \\ \boldsymbol{\gamma}_d \boldsymbol{\Omega}_e^{-1} & -\boldsymbol{\Omega}_u^{-1} \end{bmatrix}' n^{-½} 2$$

$$= m(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{x} \mid \mathbf{Y}, \mathbf{z}, \mathbf{X}) n^{-1} 4 \begin{bmatrix} \mathbf{z}_b' \boldsymbol{\Omega}_e^{-1} \mathbf{z}_b & \mathbf{z}_b' \boldsymbol{\Omega}_e^{-1} \mathbf{x}_b & \mathbf{z}_b' \boldsymbol{\Omega}_e^{-1} \boldsymbol{\gamma}_d \\ \mathbf{x}_b' \boldsymbol{\Omega}_e^{-1} \mathbf{z}_b & \mathbf{x}_b' \boldsymbol{\Omega}_e^{-1} \mathbf{x}_b & \mathbf{x}_b' \boldsymbol{\Omega}_e^{-1} \boldsymbol{\gamma}_d \\ \boldsymbol{\gamma}_d \boldsymbol{\Omega}_e^{-1} \mathbf{z}_b & \boldsymbol{\gamma}_d \boldsymbol{\Omega}_e^{-1} \mathbf{x}_b & \boldsymbol{\gamma}_d \boldsymbol{\Omega}_e^{-1} \boldsymbol{\gamma}_d + \boldsymbol{\Omega}_u^{-1} \end{bmatrix}$$

The Hessian matrix is obtained by differentiation of the Jacobian, to obtain

$$\mathbf{H}(m) = \begin{bmatrix} \dfrac{\partial m^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} & \dfrac{\partial m^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}'} & \dfrac{\partial m^2}{\partial \boldsymbol{\beta} \partial \mathbf{z}'} \\ \dfrac{\partial m^2}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}'} & \dfrac{\partial m^2}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} & \dfrac{\partial m^2}{\partial \boldsymbol{\gamma} \partial \mathbf{z}'} \\ \dfrac{\partial m^2}{\partial \mathbf{z} \partial \boldsymbol{\beta}'} & \dfrac{\partial m^2}{\partial \mathbf{z} \partial \boldsymbol{\gamma}'} & \dfrac{\partial m^2}{\partial \mathbf{z} \partial \mathbf{z}'} \end{bmatrix} = -2n^{-1} \begin{bmatrix} \mathbf{z}_b' \boldsymbol{\Omega}_e^{-1} \mathbf{z}_b & \mathbf{z}_b' \boldsymbol{\Omega}_e^{-1} \mathbf{x}_b & \mathbf{z}_b' \boldsymbol{\Omega}_e^{-1} \boldsymbol{\gamma}_d \\ \mathbf{x}_b' \boldsymbol{\Omega}_e^{-1} \mathbf{z}_b & \mathbf{x}_b' \boldsymbol{\Omega}_e^{-1} \mathbf{x}_b & 2\mathbf{x}_b' \boldsymbol{\Omega}_e^{-1} \boldsymbol{\gamma}_d \\ \boldsymbol{\gamma}_d \boldsymbol{\Omega}_e^{-1} \mathbf{z}_b & 2\boldsymbol{\gamma}_d \boldsymbol{\Omega}_e^{-1} \mathbf{x}_b & \boldsymbol{\gamma}_d \boldsymbol{\Omega}_e^{-1} \boldsymbol{\gamma}_d - \boldsymbol{\Omega}_u^{-1} \end{bmatrix}$$

The model (11) is then solved twice, once unconstrained and once constrained by $\gamma_j = 0$ for $j = 1\ldots15$. The LM test statistic is computed using the Lagrange Multipliers obtained in the constrained model and the estimated **H** and **Σ**. The resulting test statistic is 26.7, which is asymptotically distributed as chi-square(15) if the constraints are true. For a test on 5% level we compare LM with the 5% tabular value of the chi-square distribution, which is 25.0, and conclude that the null hypothesis is rejected at the 5% level (the exact p-value is 0.031).

So, yields could depend on acreages. However, a look at the estimated coefficients in table 10 shows that the estimations are not sufficiently robust to use on a large scale: Sugar beet (SUGB) obtains the (significant at 0.1% level in a test using the asymptotic normal distribution for $\hat{\boldsymbol{\theta}}$) coefficient of minus 3.90 tons per thousand hectares. This implies that the marginal yield at the observed acreage (about 25'000 hectares) is negative, which is unacceptable. At the same time, the coefficient on rye is very close to zero and not significant, so the original problem is not solved. Thus, we decide to discard the model with yield depending on acreage *despite* the failure to reject the hypothesis that $\boldsymbol{\gamma} = \mathbf{0}$.

5. A Bayesian estimator based on highest posterior density

5.1. Principles of estimator

After having discussed two different versions of yield endogeneity in sections three and four, we now return to the primary objective and model (1). The basic assumption underlying the data sampling model is that there exists a set of true parameters $\boldsymbol{\psi}$ = (**p**,**Y**,**s**,**A**,**w**,**q**,**l**,**c**,**D**,**B**,**R**,**v**) of the model, satisfying the second order conditions (4-5), a vector of true planned acreages $\mathbf{x}^*$ and a vector of dual values $\boldsymbol{\lambda}^*$ such that $(\mathbf{x}^*,\boldsymbol{\lambda}^*)$ is the unique optimal solution to the model parametrized by $\boldsymbol{\Psi}$. We may thus write $\mathbf{x}^* = \mathbf{x}^*(\boldsymbol{\psi})$ and $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}^*(\boldsymbol{\psi})$. Furthermore, the values $\mathbf{z} = (\mathbf{x}^{obs},\mathbf{p}^{obs},\mathbf{Y}^{obs},\mathbf{s}^{obs},\mathbf{A}^{obs},\mathbf{w}^{obs},\mathbf{q}^{obs},\mathbf{l}^{obs},\mathbf{R}^{obs},\mathbf{v}^{obs})$ in the CAPRI database are considered the outcome of a random variable vector **Z** that is conditional on $\boldsymbol{\Psi}$, i.e. there exists a probability density function $f(\mathbf{z}|\boldsymbol{\Psi})$.

We have prior beliefs regarding the parameter $\boldsymbol{\Psi}$ that are not contained in the CAPRI database. We expect the dual values of the constraints and the price elas-

ticities implied by $\boldsymbol{\psi}$ to be of "reasonable size". If we are express those beliefs as a prior density function $\xi(\boldsymbol{\Psi})$, we may use Bayes's rule to derive the posterior density function of $\boldsymbol{\Psi}$ conditional on the outcome $\mathbf{z}$:

$$\xi(\boldsymbol{\Psi}|\mathbf{z}) \propto f(\mathbf{z}|\boldsymbol{\Psi})\xi(\boldsymbol{\Psi})$$

In the following sections, we first develop an error model that relates $\mathbf{z}$ to $\boldsymbol{\Psi}$ in order to derive the function f. We discuss the chosen error model and compare it to alternatives. Then we formulate our prior beliefs regarding elasticites and dual values in terms of the unconditional density function ξ. Finally, we devise an estimation method that chooses as an estimate the parameter vector $\boldsymbol{\Psi}$ that maximises the conditional density $\xi(\boldsymbol{\Psi}|\mathbf{z})$. DeGroot (1970) calls this estimator the generalised maximum likelihood estimator. Other authors have called it the posterior mode estimator, the maximum a-posteriori estimator or the highest posterior density estimator.

5.2. Data sampling process

The distribution of $\mathbf{Z}$ is based on the following assumptions, which are detailed further below:

(i) All elements in $\mathbf{Z}$ are independent.

(ii) Subsidies, price index, set-aside rate and total land constraint are known with certainty, i.e. are degenerate random variables.

(iii) Errors are additive.

(iv) Producers have naïve price expectations.

(v) Expected yields and input requirements follow linear trends.

(i) The covariance matrix $\boldsymbol{\Sigma}$ only contains diagonal elements. This is discussed further in the following section on prior distributions.

(ii) We assume that set-aside rate, subsidies, price index and total land constraint are known with certainty. Since the outcomes of those items in the random vector $\mathbf{Z}$ will be the corresponding items of $\boldsymbol{\Psi}$ itself, they are from now on removed from $\mathbf{Z}$. An outcome of $\mathbf{Z}$ is thus written $\mathbf{z} = (\mathbf{x}^{obs},\mathbf{p}^{obs},\mathbf{w}^{obs},\mathbf{Y}^{obs},\mathbf{A}^{obs})$.

(iii) We write an outcome of $\mathbf{Z}$ as the sum of its conditional expectation $\boldsymbol{\mu}(\boldsymbol{\psi}) = (\boldsymbol{\mu}_x,\boldsymbol{\mu}_p,\boldsymbol{\mu}_w,\boldsymbol{\mu}_Y,\boldsymbol{\mu}_A)$, (with appropriate dimensions), and the random error vector $\boldsymbol{\varepsilon}$, so that, $\mathbf{Z} = \boldsymbol{\mu}(\boldsymbol{\psi}) + \boldsymbol{\varepsilon}$. For acreages, we have

$$\boldsymbol{\mu}_x = \mathbf{x}^*(\boldsymbol{\psi})$$

(iv) Naïve price expectations imply that the expectation of the price measurement in period *t*-1 equals the producer price in that period, or conversely,

$$\mathbf{p}_t = (\boldsymbol{\mu}_p)_{t-1}$$

$$\mathbf{w}_t = (\boldsymbol{\mu}_w)_{t-1}$$

where the expression on the right hand side denotes the expected value of the output and input prices for all crops in period *t*-1.

(v) The producers expect the yield in each period to equal $(\boldsymbol{\mu}_y)_t$, which increases over time by an exogenous linear trend. The same assumption is made for input coefficients, We thus have that

$$\mathbf{Y}_t = (\boldsymbol{\mu}_y)_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 T_t \tag{12}$$

$$\mathbf{A}_t = (\boldsymbol{\mu}_A)_t = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 T_t$$

with T being a linear trend and $\boldsymbol{\beta} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1)$ new parameters to estimate. Unfortunately, there are no observations available for actual input application. Instead, we use estimated input coefficients (available in the CAPRI database), that are based on total input use in the agricultural sector in combination with farm level data, economic reasoning and engineering knowledge. Those expert coefficients are denoted by $\mathbf{A}^{obs}$. The actual amount of inputs applied in any given year may differ from the expected value due to unexpected climatic conditions, just as the yield may deviate from expected yield, though the hypothesis is that the agricultural production plan is made up with the expected values in mind[21].

5.3. Discussion of alternative error models

The error model developed above is fairly sophisticated in the sense that it attempts to take into account that all measurements are likely to be subject to errors[22]. The sophistication comes at a cost, because it requires information about the covariance matrix of **Z**. Ideally, this information would be supplied by replicate measurements or external datasets (Carroll, Ruppert and Stefanski 1995). In the case at hand, no such replicates are available, and instead, the relative variability of the different errors is based on assumptions.

---

[21] This implies a general error model, but the resulting formulation is indistinguishable from the measurement error model.

[22] Griliches and Ringstad (1970 p. 370) conclude, in relation to measurement errors in nonlinear models, that "*In short, errors in variables are bad enough in linear models. They are likely to be disastrous to any attempts to estimate additional nonlinearity or curvature parameters*."

Although the error model is sophisticated on the side of the researcher (measurement errors), it is very simple on the side of the economic agent. We assume that the agent has perfect information about the true parameters, and that he is able to determine the optimal production decision exactly. That is, no part of the errors enter the model equations, thereby influencing production. A more general error model, as discussed by McElroy (1987) and Pope and Just (2002) would also take into account the possibility that the producer may not correctly appreciate the true parameters and/or is not able to determine exactly the optimal supply decision. Let us look at the implications of neglecting those errors.

The exogenous (in this model) parameters that are subject to considerable uncertainty are prices (**p**,**w**) and I/O coefficients (**A**,**Y**). Saying that the producer does not correctly appreciate those is silly, since they are *defined* as the producer's expectation. It may however be the case that the *expectation model* is not the correct one (the possibility that the producer does not base his expectation on the same observations as the researcher is already included in the error term). In those cases, the producer bases his land allocation decision not on the true parameters (**p**,**w**,**A**,**Y**) (which can then no longer be called "true") but on stochastic ($\mathbf{p} + \boldsymbol{\delta}_p$,$\mathbf{w} + \boldsymbol{\delta}_w$,$\mathbf{A} + \boldsymbol{\delta}_A$,$\mathbf{Y} + \boldsymbol{\delta}_Y$) for some deviations $\boldsymbol{\delta}$. This is a kind of specification error of the model. If we at this point assume that the producer solves the optimization problem correctly, we can substitute the disturbed parameters into the first order conditions and rearrange to obtain

$$\mathbf{Y}_t\mathbf{p}_t + \mathbf{s}_t - \mathbf{A}_t\mathbf{w}_t - q_t\mathbf{c} - l_t\left[\mathbf{D} + \mathbf{GBG}'\right]\mathbf{x}_t - \mathbf{R}_t'\boldsymbol{\lambda}_t = \boldsymbol{\Delta}_t$$

with $\boldsymbol{\Delta} = \mathbf{A}\boldsymbol{\delta}_w + \mathbf{w}\boldsymbol{\delta}_A + \boldsymbol{\delta}_A\boldsymbol{\delta}_w - \mathbf{Y}\boldsymbol{\delta}_p - \mathbf{p}\boldsymbol{\delta}_Y - \boldsymbol{\delta}_p\boldsymbol{\delta}_Y$ (time indices omitted). This makes the relationship between the true parameters stochastic. It is not clear what effect the omission of $\boldsymbol{\Delta}$ has on the estimation of the parameters of interest, (**c**,**B**,**D**).

The producer may also commit an optimization error, i.e. instead of choosing the optimal acreage vector $\mathbf{x}^*$, which would solve the optimization problem, he allocates $\mathbf{x}^* + \boldsymbol{\delta}_x$, which does not solve it, but satisfies the constraints. That kind of error would be impossible to distinguish from a pure measurement error on the side of the researcher, except that we would require $\mathbf{R}\boldsymbol{\delta}_x = \mathbf{0}$ (because $\mathbf{R}\mathbf{x}^* = \mathbf{v} = \mathbf{R}(\mathbf{x}^* + \boldsymbol{\delta}_x)$).

Since the general error model requires an increased amount of prior information and is anyway difficult to distinguish from the measurement error model, we choose to limit ourselves to measurement errors. We now proceed with explicit assumptions regarding the data sampling processes.

5.4. Augmented parameter vector and its prior distribution

In the ex-post perspective, the outcome **e** of the error vector $\boldsymbol{\varepsilon}$ has actually already been determined, but the outcome is not directly observable. We thus choose to

consider **e** yet another unknown parameter. If the density function *f* for the random vector **Z** is conditional also on **e** and the yield and input parameters $\boldsymbol{\beta}$, and $\boldsymbol{\alpha}$ defined above, then there are no random components left, and *f* becomes the degenerate density function

$$f(\mathbf{z} \mid \boldsymbol{\psi}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{e}) = \begin{cases} 1: & \mathbf{z} = \boldsymbol{\mu}(\boldsymbol{\psi}, \boldsymbol{\beta}, \boldsymbol{\alpha}) + \mathbf{e},\ g(\boldsymbol{\psi}, \mathbf{x}^*, \boldsymbol{\lambda}^*) = 0 \\ 0: & \text{else} \end{cases}$$

One can immediately see that there must be a large number of vectors ($\boldsymbol{\psi}$,$\boldsymbol{\beta}$,$\boldsymbol{\alpha}$,**e**) that give the density value "1" for almost any outcome **z** of **Z**. Without further information, there is no way of discriminating between any two such vectors by saying that one is any more likely than the other to be the true parameter vector. This is why we require a prior distribution $\xi(\boldsymbol{\Psi},\mathbf{e},\boldsymbol{\alpha},\boldsymbol{\beta})$. In this section, we define the prior distribution based on the following assumptions, detailed below:

(i) $\xi(\boldsymbol{\Psi},\mathbf{e},\boldsymbol{\alpha},\boldsymbol{\beta}) = \xi(\mathbf{e})\xi(\boldsymbol{\lambda}^*(\boldsymbol{\Psi},\boldsymbol{\alpha},\boldsymbol{\beta}))\xi(\boldsymbol{\eta}(\boldsymbol{\Psi},\boldsymbol{\alpha},\boldsymbol{\beta}))$, with $\boldsymbol{\eta}(\boldsymbol{\Psi},\boldsymbol{\alpha},\boldsymbol{\beta})$ denoting the vector of implied own price supply elasticities. That is, we assign prior distributions to error terms, dual values and implied point price elasticities of supply, and assume that those are functionally independent.

(ii) The errors **e** are independent and normally distributed with standard deviations equal to a fix share of the observed value of the respective parameter.

(iii) The dual values are independent, with means proportional to average observed gross margins over all crops in each region each year, and standard deviations proportional to a fix share of that.

(iv) We believe that the parameter vector is such that the implied point price elasticity of supply matrix $\boldsymbol{\eta}(\boldsymbol{\Psi},\boldsymbol{\alpha},\boldsymbol{\beta})$ is normally distributed with mean depending on the crop mix (rotational shares) and standard deviation independent for each item. For non-diagonal elements of $\boldsymbol{\eta}$, the prior distribution is non-informative (i.e. we have no specific beliefs regarding cross price elasticities).

Regarding (ii): Specifically we assume that $\mathbf{e} \sim N(0,\boldsymbol{\Sigma}_e)$ with $\boldsymbol{\Sigma}_e$ a diagonal matrix with $\sigma_{ei}^2 = (0.20/3z_i)^2$ on the i^{th} position. This means that we assume that errors are independent normally distributed with mean zero covariance matrix such that three standard deviations cover 20% of the observed value of the related parameter.

Regarding (iii): In order for the posterior density to have a unique maximum, we require informative priors also for the dual values $\boldsymbol{\lambda}$ in order to be able to iden-

tify $\boldsymbol{\Psi}$ (since for example $\mathbf{c}$ and $\boldsymbol{\lambda}$ enter the first order conditions additively). We make the assumptions

$$\lambda_{1t} \sim N\left(0.25\overline{m}_t, \left(\frac{0.20}{3}0.25\overline{m}_t\right)^2\right)$$

$$\lambda_{2t} \sim N\left(0.25\left(m_{"OSET"t}(1-\rho_t)-\overline{m}_t\rho_t\right), \left(\frac{0.20}{3}0.25\left(m_{"OSET"t}(1-\rho_t)-\overline{m}_t\rho_t\right)\right)^2\right),$$

where $m_{"OSET"t}$ is the observed gross margin in compulsory set aside, $\overline{m}_t$ the average gross margin over all crops and ρ_t the general set-aside rate in period t. The prior mode (mean of normal distribution) of $\boldsymbol{\lambda}$ is thus based on the assumption that the expected land rent is approximately 25% of the average observed gross margin $\overline{m}_t$ in the respective year taken over all crops except sugar beet (whereas sugar quota rents are missing in the model). For the case study region FR24 this fits reasonably well with data on land rental prices obtained from Eurostat for France, shown in table 11. The priors for dual values of the set-aside constraint were derived in a similar manner, but also including. The variances of $\boldsymbol{\lambda}$ were assumed to be such that 20% of the prior means equal three standard deviations.

Table 11. Land rents in France (Euro per ha)

	Eurostat*	λ_1 prior	λ_2 prior
1986	102	86	
1987	104	132	
1988	106	94	
1989	109	113	
1990	111	134	
1991	113	119	
1992	115	101	-116
1993	117	91	-23
1994	119	124	-5
1995	121	175	-8
1996	122	163	-27
1997	125	148	11
1998	129	155	-98
1999	132	193	-6
2000	132	139	39
2001	131	145	74
2002	131	181	24

**Source: Eurostat (2003)*

Regarding (iv): There are cases when the observations imply a supply elasticity that is far outside any plausible range, e.g. > 1000. One case when this would happen is when the observations imply a downward sloping supply function, as in the case of rye in FR24 discussed in a previous section. Given the second order

conditions for optimality, the best fit is obtained by a horizontal supply curve, implying an infinite elasticity. Such a simulation behaviour of the model is unacceptable, and we firmly believe that the aggregate supply response of regions in reality is much smoother. Put differently, we believe that the parameter vector comes from a distribution that makes such extreme values utterly improbable, but is rather indifferent for elasticities within some plausible range. For this purpose, we choose a very wide normal distribution, with mean and standard deviation derived below.

Most studies (see comparison to other studies below) find supply elasticities in the range of, say, 0.1 to 5. More specifically, we see that the elasticity is typically around unity for major crops, but that it is higher for crops that occupy a small share of the total area. One motivation for such a relation is that if a small crop expands with a certain percentage, that should have less effect on the value of fixed resources, like pushing other crops out of the rotation on the constrained land, compared to if a major crop expands by the same percentage.

Letting r_j denote the share of land allocated to crop j, we believe that the own price supply elasticities have means $0.5r_j^{-1/3}$ and standard deviations such that three standard deviations cover 1000% of the mean (the standard deviation relative to mean is thus fifty times that of the acreages, prices or yields). There are no priors at all for cross price elasticities. In the result section below, the priors are compared to elasticities from literature for the Netherlands, Denmark and France, and found to be in a plausible range.

We will see that the explicit expression for supply elasticities is a nonlinear function of the parameters. That makes its inclusion into the estimation difficult. Jansson (2005) solves a similar model for supply elasticities and includes the expression explicitly in the estimation. His model, however, did not have area constraints, and imposed land constrain only implicitly over curvature constraints on the Hessian matrix, which simplified the expressions for supply elasticities considerable. Heckelei and Wolff (2003) makes a similar estimation but with invented data for a didactic size problem, with a simultaneous incorporation of elasticity priors. Here, we have two constraints in most years and only one constraint in some years (before set-aside regime), which complicates things further. The elasticities of supply in our model can be obtained by solving the first order conditions for $\mathbf{x}_t$ (repeated here for convenience),

$$\mathbf{x}_t^*(\mathbf{p}_t, \boldsymbol{\lambda}_t) = l_t^{-1}\left[\mathbf{D} + \mathbf{G}\mathbf{B}\mathbf{G}'\right]^{-1}\left[\mathbf{Y}_t\mathbf{p}_t + \mathbf{s}_t - \mathbf{A}_t\mathbf{w}_t - q_t\mathbf{c} - \mathbf{R}_t'\boldsymbol{\lambda}_t\right]. \qquad (13)$$

Let $\mathbf{E}_t = l_t\left[\mathbf{D} + \mathbf{G}\mathbf{B}\mathbf{G}'\right]$ and insert that expression into the constraints to obtain a solution for $\boldsymbol{\lambda}$,

$$\boldsymbol{\lambda}_t^*(\mathbf{p}_t) = \left[\mathbf{R}_t\mathbf{E}_t^{-1}\mathbf{R}_t'\right]^{-1}\left[\mathbf{R}_t\mathbf{E}_t^{-1}\left(\mathbf{Y}_t\mathbf{p}_t + \mathbf{s}_t - \mathbf{A}_t\mathbf{w}_t - q_t\mathbf{c}\right) - \mathbf{v}_t\right]. \qquad (14)$$

Computing $\mathbf{x}^*_t(\mathbf{p}_t,\boldsymbol{\lambda}^*_t(\mathbf{p}_t))$ by inserting (14) into (13), taking derivatives and multiplying the result by yield gives us the following expression for marginal production[23]:

$$\frac{\partial(\mathbf{Y}_t\mathbf{x}_t)}{\partial\mathbf{p}_t} = \mathbf{Y}_t\left(\mathbf{E}_t^{-1}\mathbf{Y}_t - \mathbf{E}_t^{-1}\mathbf{R}_t'\left[\mathbf{R}_t\mathbf{E}_t^{-1}\mathbf{R}_t'\right]^{-1}\mathbf{R}_t\mathbf{E}_t^{-1}\mathbf{Y}_t\right) \tag{15}$$

Using the definition of elasticity, we finally obtain the expression

$$\boldsymbol{\eta}_t = \mathbf{X}_t^{-1}\left(\mathbf{E}_t^{-1}\mathbf{Y}_t - \mathbf{E}_t^{-1}\mathbf{R}_t'\left[\mathbf{R}_t\mathbf{E}_t^{-1}\mathbf{R}_t'\right]^{-1}\mathbf{R}_t\mathbf{E}_t^{-1}\mathbf{Y}_t\right)\mathbf{P}_t \tag{16}$$

where upper case $\mathbf{X}_t$ means the square diagonal matrix with $\mathbf{x}_t$ on the diagonal, and similar for upper case $\mathbf{P}_t$.

This expression is strongly non-linear in **D** and **B** (via **E**) and thus difficult to include as constraint in the estimation. In some models, the expression has been simplified by neglecting the second term in the bracket and only computing diagonal elements in **E**. That simplification was previously used in different model to compute only diagonal elements of the quadratic PMP-parameter, e.g. in the CAPRI model (not published), and by Helming (2005) in the DRAM model.

Nevertheless, with appropriate initialisation of the solution algorithm (CONOPT for GAMS) together with reasonable bounds for the variables, equation 16 turns out to be possible to solve simultaneously in the estimation, thus enabling us to include our prior beliefs regarding elasticities of supply in a transparent way.

5.5. Definition of the estimator

Putting all the pieces together, we can now formulate the estimation problem using Bayes's theorem as above and write

$$\hat{\boldsymbol{\Psi}} = \arg\max \xi(\boldsymbol{\psi},\boldsymbol{\beta},\boldsymbol{\alpha},\mathbf{e}|\mathbf{z}) \propto f(\mathbf{z}|\boldsymbol{\psi},\boldsymbol{\beta},\boldsymbol{\alpha},\mathbf{e})\xi(\boldsymbol{\psi},\boldsymbol{\beta},\boldsymbol{\alpha},\mathbf{e})$$

To repeat, the point estimate of $(\boldsymbol{\psi},\boldsymbol{\beta},\boldsymbol{\alpha},\mathbf{e})$ that we are looking for is the value that maximises the posterior density $\xi(\boldsymbol{\psi},\boldsymbol{\beta},\boldsymbol{\alpha},\mathbf{e}|\mathbf{z})$, i.e. the posterior mode. Note that with the degenerate density function this is equivalent to solving

$$\begin{aligned} &\max \quad \xi(\boldsymbol{\psi},\boldsymbol{\beta},\boldsymbol{\alpha},\mathbf{e}) \\ &\text{subject to} \quad \mathbf{z} = \boldsymbol{\mu}(\boldsymbol{\psi},\boldsymbol{\beta},\boldsymbol{\alpha}) + \mathbf{e} \\ &\qquad\quad \mathbf{g}(\boldsymbol{\psi},\mathbf{x}^*,\boldsymbol{\lambda}^*) = \mathbf{0} \end{aligned}$$

[23] In this case, the marginal production could be solved for directly. In the general case with continuous derivatives, the implicit function theorem may be used instead.

Since the value that maximises some function h also maximises log (h), we may take the logarithm of the objective function (which is a multivariate normal density function with covariance matrix $\boldsymbol{\Sigma}$). Doing that and replacing the constraints with the equations derived above, we arrive at the following extremum estimation problem:

minimise

$$\begin{aligned}&\operatorname{vec}\left(\mathbf{e}_x, \mathbf{e}_Y, \mathbf{e}_p, \mathbf{e}_w, \mathbf{e}_A, \left(\boldsymbol{\lambda} - \boldsymbol{\lambda}^{prior}\right), \left(\operatorname{diag}(\mathbf{v}) - \hat{\mathbf{v}}\right)\right)' \\ &\times \boldsymbol{\Sigma}_{total}^{-1} \operatorname{vec}\left(\mathbf{e}_x, \mathbf{e}_Y, \mathbf{e}_p, \mathbf{e}_w, \mathbf{e}_A, \left(\boldsymbol{\lambda} - \boldsymbol{\lambda}^{prior}\right), \left(\operatorname{diag}(\mathbf{v}) - \hat{\mathbf{v}}\right)\right)\end{aligned}$$

subject to

$$\mathbf{Y}_t \mathbf{p}_t + \mathbf{s}_t - \mathbf{A}_t \mathbf{w}_t - q_t \mathbf{c} - l_t \left[\mathbf{D} + \mathbf{G}\mathbf{B}\mathbf{G}'\right]\mathbf{x}_t - \mathbf{R}_t' \boldsymbol{\lambda}_t - MAC_t \delta = \mathbf{0}$$

$$\mathbf{R}_t \mathbf{x}_t = \mathbf{v}_t$$

$$\mathbf{B} = \mathbf{U}'\mathbf{U}$$

$d_{jj} \geq 0$ for $j = 1 \ldots J$ (and $d_{ij} = 0$ for $i \neq j$)

$$\mathbf{x}^{obs} = \mathbf{x} + \mathbf{e}_x$$

$$\mathbf{Y}^{obs} = \mathbf{Y} + \mathbf{e}_Y$$

$$\mathbf{Y}_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 T_t$$

$$\mathbf{p}_t = \mathbf{p}_{t-1}^{obs} + (\mathbf{p}_t^{adm} - \mathbf{p}_{t-1}^{adm}) + \mathbf{e}_{pt}$$

$$\mathbf{A}^{exp} = \mathbf{A} + \mathbf{e}_A$$

$$\mathbf{A} = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 T_t$$

$$\nu_t = \mathbf{X}_t^{-1}\left(\mathbf{E}_t^{-1}\mathbf{Y}_t - \mathbf{E}_t^{-1}\mathbf{R}_t'\left[\mathbf{R}_t \mathbf{E}_t^{-1}\mathbf{R}_t'\right]^{-1}\mathbf{R}_t \mathbf{E}_t^{-1}\mathbf{Y}_t\right)\mathbf{P}_t$$

The dummy variable MAC_t with associated parameter δ in the first order condition was added to control for additional effects of the MacSharry reform. It is equal to 1 for year 1992 and earlier for regions that were member of the EU then, and zero from 1993 and on. This is motivated by an optical inspection of the time series. For example, looking again at the gross margin and acreages of rye if figure 1 suggests that there are two clouds of observations, which correspond to pre- and post MacSharry reform (1993). Thus the reform is likely to have influenced behaviour in some way not captured in the present model (the situation is similar for some other products).

5.6. Data preparations

The time series in the CAPRI database is different long for different crops even within regions. It also contains holes and obvious errors, especially for crops of residual character like "other cereals", or when the area cropped is very small compared to other crops in the region. Thus, the estimations require data to be processed prior to estimation in order to make sure that no obvious data errors corrupt the estimations, we must select a strategy for choosing which regions, crops and years to include in estimation, and we must decide what to do with zeros in the data.

Selection of crops: A potentially different set of crops were estimated in each region. To start with, all acreages smaller than 1000 ha were set to zero. Then, the crops to be estimated were those satisfying all of the following three conditions: (1) There is acreage data in year 2000, (2) there is acreage data in at least five years, and (3) the sum of acreage over all years is at least 10 000 ha.

Selection of years: A year *t* was included in the estimation if the total acreage over all crops just selected was at least 10 000 ha in year *t*-1. The lag is necessary for the lagged prices to work. The longest possible time series was 1986 to 2003.

Selection of regions: A region was included in the estimation if the following three conditions were satisfied: (1) Year 2000 was included in the set of years to estimate for that region, (2) the set of crops to estimate contain at least three elements, and (3) the number of observations over all crops and years is at least 50. The number of regions to estimate determined in this way turned out to be 165.

Treatment of outliers: Outliers for prices, yields and input coefficients were detected and replaced with time series mean using the following procedure:

Do for $i = 1,2$

1. Compute mean $\bar{z}$ using all but the greatest and the smallest value.
2. If not ($ai \le z_t \le b/i$), then replace z_t with $\bar{z}$

where a and b are constants. The replacement was done twice, and with narrower bounds in the second repetition in order to alleviate the problem that the presence of two outliers biases the mean. Trial and error revealed that $(a,b) = (0.1,6.0)$ worked fine for prices, (0.2,4.0) for yields and (0.25,4.0) for input coefficients.

Unbalanced panels: In the cases where some time series were shorter than the others, it was assumed that this was really due to missing data, perhaps truncated by the "1000 ha rule", not that the data truly was zero (except in the case of the "political activities" compulsory set-aside and non-food production on set-aside). Then, the estimator was allowed to choose any value satisfying the equation system as the estimate, but the item did not enter the posterior density function. Since consecutive years are interlinked via the other parameters (yield, input requirement, PMP terms), this does not generally cause any problems. In most regions

where some time series was shorter than the other, it was early years that were missing, which are of lesser importance for the intended use of the estimates.

6. Results

The estimation produced a large number of results: 1917 elements of the key parameters **c** and **D** respectively, and 5457 elements of the cross group effects matrix **B**. Furthermore, 329 092 price elasticities were computed, including the cross price elasticities. To this comes a very large number of fitted values and all other parameters in Ψ. It is impossible to give even an overview of all those results, and in this section we only present estimation results for the French case study region FR24 and for France as a whole. The results are evaluated following two criteria:

1. How well is the prior information recovered? To address this, a kind of R^2 measure is computed as the share of the explained variance observations or prior mode. In an appendix, we also provide a visual presentation of prior and posterior mode for selected items (plots).

2. How is the resulting model behaving in simulation? We discuss our estimated point price supply elasicities and compare them to estimates from literature.

6.1. Measures of fit

Table (12) shows the share of explained variation, R^2, for acreages, prices and yields for all land use activities in FR24. We see that in most cases, the fit of acreage is high, above 0.90. Exceptions are soft wheat, potatoes, sugar beet and voluntary set-aside. Only the last of those crops has an R^2 below 0.50 (0.393). The fit of prices is equally high in general. The fit of yields is lower because here a more restrictive error model is employed: the expected yields have to lie on a straight line (12). In four cases, the fit of yield is even negative. One can see on the plots in the appendix that the yields of those crops are highly variable.

Table 12. R^2 for acreages (X), prices (P) and yields(Y)

Crop	Item	N	R2	Crop	Item	N	R2
SWHE	P	18	0.928	PULS	X	18	0.907
DWHE	P	18	0.820	POTA	X	18	0.649
RYEM	P	18	0.927	SUGB	X	18	0.805
BARL	P	18	0.791	MAIF	X	18	0.987
OATS	P	18	0.915	OFAR	X	18	0.938
MAIZ	P	18	0.794	NONF	X	11	0.999
OCER	P	18	0.935	OSET	X	12	1.000
RAPE	P	18	0.923	VSET	X	14	0.393
SUNF	P	18	0.932	SWHE	Y	18	0.291
PULS	P	18	0.838	DWHE	Y	18	0.235
POTA	P	18	0.964	RYEM	Y	18	0.673
SUGB	P	18	0.455	BARL	Y	18	0.179
MAIF	P	18	0.716	OATS	Y	18	0.012
OFAR	P	18	0.685	MAIZ	Y	18	0.657
NONF	P	18	0.948	OCER	Y	18	-0.030
SWHE	X	18	0.591	RAPE	Y	18	-0.164
DWHE	X	18	0.995	SUNF	Y	18	0.234
RYEM	X	18	0.998	PULS	Y	18	-0.027
BARL	X	18	0.977	POTA	Y	18	0.490
OATS	X	18	0.997	SUGB	Y	18	0.717
MAIZ	X	18	0.909	MAIF	Y	18	-0.086
OCER	X	18	0.988	OFAR	Y	18	0.428
RAPE	X	18	0.979	NONF	Y	11	0.891
SUNF	X	18	0.934				

Source: Own estimations.

6.2. Elasticities

The point price elasticities of supply are computed simultaneous in the estimations by equation (16). In this section we present elasticities for individual crops and for crop groups for one selected subregion, FR24, and for the aggregate France, all in the year 2002. The aggregation from the 22 French regions estimated and whole of France was done by weighing the regional elasticities with the region's share of national crop area, or

$$\sum_r \eta_{rj} x_{rj} \Big/ \sum_r x_{rj}$$

Aggregation to crop groups was done similarly, by weighing with each crop's share in the crop group to which it belongs. The crop groups are the same that were used in the estimation, reported in table (16). Table (18) and (19) shows the

elasticities of individual crops for FR24 and France respectively. Table (20) and (21) show the elasticities of the crop groups. Some of the elasticities, especially for individual crops of minor land share on regional level, are high. This is true for e.g. rye and durum wheat, which both have elasticities above 7 and small rotational shares. In contrast, soft wheat has the moderate elasticity of 0.79 for a land share of 36%; however, there are notable exceptions. Potatoes has a rotational share of only 0.36%, but only an elasticity of 0.38.

As one might expect, the crop groups generally show less elasticity to price changes than the individual crops. This is partly due to the land restriction, but also to the crop group structure of the model, that allows catching substitution effects between related crops. The most notable case for FR24 is perhaps oil seeds. In table (18) we see that rapeseed and sunflower are good substitutes, but table (20) reveals inelastic supply response as a group.

Aggregation from regions to the member state offers no great surprises. Most of the elasticities are of similar size at national as on regional level in the case studied. The greatest difference is for durum wheat, where the elasticity in FR24 is much higher than that in the member state aggregate. One reason for not finding greater differences between the region and the aggregate is probably that the rotational shares in the region are similar to those on national level.

Although there are several studies that present elasticities on national level, no other study that the author is aware of publishes elasticities for individual crops on regional level with this crop coverage. Below we compare our point elasticity estimates as well as our priors with four other studies. Two of the other studies are for France, one study is for the Netherlands and one is for Denmark. In all comparisons, we use our point price elasticities for the year 2002.

For France, we can compare our results to those in Heckelei and Britz (2000) (HB00) and Guyomard et al. (1996) (GBC96). This has been done in table (13), where also the land share and prior mode are printed. GBC96 estimates a model with seven outputs and three inputs based on a restricted profit function, using annual data for France. HB00 estimate a similar model as ours, but they use a cross-section data set of French regions for the year 1994 instead of time series for individual regions as we do.

We see that GBC96 finds considerable smaller elasticities for barley (0.35) and other coarse grains (0.76) than this study (2.24 and 2.53), HB00 (2.65 for barley) or the priors (1.11 and 1.55). For soft wheat the results are much more in line, with the priors (0.77) quite close to GBC96 (0.72) and the estimates (1.01) in between GBC96 and HB00 (1.32). For maize the estimates (1.68) are close to GBC96 (1.63) but much higher than HB00 (0.65), whereas the priors lie in between (1.07). Rapeseed and sunflower occupy small rotational shares, less than 5%, and as a consequence the priors are higher, about 1.5. The elasticity estimates for those crops are also much higher, 1.28 and 2.96, than GBC96, which finds

values of 0.42 and 0.22, and more in line with HB00, which finds elasticities greater than unity. All of the three studies find high elasticities for soya, for which the rotational share is less than 0.5%.

Table 13. Comparison with other studies of own price supply elasticities in France

Crop	Land share[b]	Prior[c]	Own estimate	GBC96[d]	HB00[e]
Other coarse grains[a]	0.034	1.547	2.531	0.758	-.---
Soft wheat	0.273	0.771	1.009	0.715	1.322
Maize	0.102	1.070	1.680	1.630	0.653
Barley	0.092	1.109	2.243	0.351	2.647
Rapeseed	0.045	1.405	1.284	0.418	1.457
Sunflower	0.027	1.664	2.959	0.223	1.126
Soya	0.004	3.276	2.020	3.701	1.861

a: Aggregated from rye, oats and other cereals.
b: Computed from the data in CAPRI for 2002
c: Using the formula for priors reported above
d: Guyomard et al. (1996)
e: Heckelei and Britz (2000)

For the Netherlands, Oude Lansink and Peerlings (1996) (OLP96) estimate twelve farm type models producing three outputs (CO = Cereals and oilseeds, Rootcrops = Potatoes and sugar beet, and Other = all other crops). They estimate the model using panel data on individual farms, and also have a land constraint and a fixed area of rootcrops. In their table A3 they present supply elasticites, of which the own price effects are compared to our estimates for the Netherlands for similar product aggregates in table (14). To make the comparison, our individual crop elasticities have been aggregated with estimated planned rotational shares for 2002. The "other crops" aggregate in OLP96 could not be formed, since we have three crops, (voluntary and compulsory set-aside and fallow land) for which there is no output price.

Our estimates for CO (0.94) are quite close to OLP96 (0.90), but considerably higher for root crops (OLP96 find 0.34, our estimate 0.91). We must then keep in mind that in OLP96, the area used in root crops was fixed, so that the price elasticity can come only from a change in intensity. It then seems reasonable that their estimates for that aggregate turn out lower.

Table 14. Comparison with other own price supply elasticity estimates for the Netherlands

Crop group	Land share	Prior	Own estimate	OLP96[a]
CO	0.266	0.778	0.937	0.90
Root crops	0.342	0.715	0.909	0.24

Oude Lansink and Peerlings (1996)

Jensen (1996) estimates an econometric model of Danish agriculture, and also presents aggregated supply elasticities for three selected crop groups. In table (15) we have reprinted those elasticities and also our implied estimates for the corresponding aggregates. We see that for the first two groups, our elasticities are higher than those in ibid., though our prior for cereals is similar to the estimate in ibid. For the last group, root crops, the elasticities are very similar and more than twice as high as our prior.

Table 15. Comparison with other own price supply elasticity estimates for Denmark

Crop	Land share	Prior	Own estimate	Jensen (1996)
Cereals	0.575	0.601	1.073	0.60
Pulses + rapeseed	0.037	1.498	1.999	0.66
Root crops	0.035	1.522	3.772	3.80

6.3. Complete results and estimation program

The GAMS program and the data used to produce the results in this chapter can be obtained for test purposes from the author upon request.

6.4. Conclusive remarks

No confidence regions for the estimates are established. Exact analytical confidence regions are very difficult to deduce. Approximations would in theory be possible. Reilly and Patino-Leal (1981) compute approximate probability contours of the posterior in a non-linear errors-in-variables model by iterated linearisations. In our case, analytical deduction of approximate confidence regions is more difficult than in ibid. due to the curvature constraints. Numerical computation by Monte Carlo simulations is not feasible because of the amount of computation time required with the present setup (several hours for a single simulation of all regions).

We conclude that the estimated elasticities compare well with estimates in the four cases from literature studied. Nevertheless, only a handful elasticities from three member states could be compared. The vast amount of estimates are for individual crops in NUTS2 regions, and for them, we have nothing to compare to. Some of those elasticities appear high, e.g. rye and durum wheat in FR24 (table 18). Such parameter settings will result in a model that reacts strongly on shocks in simulation compared to the current CAPRI model that in the past had inelastic supply. However, the high elasticities are most often found for crops with small rotational shares, where an elastic response is sensible.

With repeated future applied analyses with the full CAPRI modelling system and the new parameters, experiences will be gained regarding the performance of the estimates.

References

Carroll, R., Ruppert, D.& Stefanski, L. (1995). *Measurement error in nonlinear models* (63). Suffolk, Great Britain: Chapman and Hall.

DeGroot, M. H. (1970). *Optimal statistical decisions*. New York: McGraw-Hill.

Eurostat (2003). Eurostat Agricultural Statistics — Quarterly Bulletin 2003:4

Fuller, W. A. (1987). *Measurement error models*John Wiley and Sons.

Greene, W. H. (2003). *Econometric analysis* (Fifth ed.). Prentice Hall.

Griliches, Z.& Ringstad, V. (1970). Error-in-the-variables bias in nonlinear contexts. *Econometrica, 38* (2), 368-370.

Guyomard, H., Baudry, M.& Carpentier, A. (1996). Estimating crop supply responses in the presence of farm programmes: application to the CAP. *European Review of Agricultural Economics, 23*, 401-420.

Heckelei, T.& Britz, W. (2000). Positive mathematical programming with multiple data points: a cross-sectional estimation procedure. *Cahiers d'economie et sociologie rurales, 57*, 28-50.

Heckelei, T. (2002). *Calibration and estimation of programming models for agricultural supply analysis*. Bonn University, Habilitationsschrift.

Heckelei, T.& Wolff, H. (2003). Estimation of constrained optimization models for agricultural supply analysis based on generalised maximum entropy. *European Review of Agricultural Economics, 30*, 27-50.

Helming, J. F. (2005). *A model of Dutch agriculture based on positive mathematical programming with regional and environmental applications*. PhD thesis Wageningen University.

Horner, G., Corman, J., Howitt, C., Carter, C.& MacGregor, R. (1992). *The Canadian regional agriculture model*.

Howitt, R. E. (1995). Positive mathematical programming. *American Journal of Agricultural Economics, 77*, 329-342.

Jansson, T. (2005). *The economic model BenImpact*. Impetus Technical Paper, Bonn University, May 2005.

Jensen, J. D. (1996). *An applied econometric sector model for Danish agriculture (ESMERALDA)*. SJFI report 90:1996. Copenhagen.

Kukush, A.& Huffel, S. V. (2004). Consistency of elementwise-weighted total least squares estimator in a multivariate errors-in-variables model AX=B. *Metrika, 59*, 75-97.

McElroy, M. B. (1987). Edditive general error models for production, cost, and derived demand or share systems. *The journal of political economy, 95* (4), 737-757.

Mittelhammer, R. C., Judge, G. G.& Miller, D. J. (2000). *Econometric foundations*. Cambridge: Cambridge University Press.

Oude Lansink, A.& Peerlings, J. (1996). Modelling the new EU cereals and oilseeds regime the Netherlands. *European Review of Agricultural Economics*, *23*, 161-178.

Pope, R. D.& Just, R. E. (2002). Random profits and duality. *American journal of agricultural economics*, *84* (1), 1-7.

Reilly, P. M.& Patino-Leal, H. (1981). A Bayesian study of the Error-in-variables model. *Technometrics*, *23* (3), 221-231.

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. New York: Wiley.

Appendix 1. Activities and inputs in estimation

Table 16. Crop groups and activities modelled

Group	Description	Crop	Description
CERE	Cereals	SWHE	Soft wheat
		DWHE	Durum wheat
		RYEM	Rye
		BARL	Barley
		OATS	Oats
CER2	Cereals2	MAIZ	Maize
		OCER	Other cereals
OILS	Oil seeds	RAPE	Rapeseed
		SUNF	Sunflower
		SOYA	Soya
		OOIL	Other oilseeds
		NONF	Ind. rapeseed
OARA	Other arable crops	POTA	Potatoes
		PULS	Pulses
		SUGB	Sugar beet
		TEXT	Fibre crops
FARA	Fodder on arable land	MAIF	Fodder maize
		OFAR	Silage grass
		ROOF	Fodder root crops
NOCR	Non-yield crops	OSET	Obligatory set-aside
		VSET	Voluntary set-aside
		FALL	Fallow land

Table 17: Inputs in estimation

Seed	Repairs buildings	Fuel
Plant protection	Electricity	Lubricants
Fertilize	Gas for drying	Other inputs
Repairs machinery		

Appendix 2: Supply elasticity estimates in France

Table 18. Supply elasticities for FR24 in year 2002 for individual crops.

	Share	SWHE	DWHE	RYEM	BARL	OATS	MAIZ	OCER	RAPE	SUNF	PULS	POTA	SUGB	MAIF	OFAR	NONF
SWHE	36.17%	**0.786**	-0.127	-0.018	-0.280	-0.027	-0.061	-0.007	0.003	0.016	0.032	0.000	0.002	-0.016	-0.055	0.039
DWHE	1.80%	-2.226	**7.913**	-0.134	-2.037	-0.199	-0.441	-0.049	0.022	0.115	0.234	0.001	0.016	-0.118	-0.397	0.283
RYEM	0.29%	-3.366	-1.392	**7.733**	-3.079	-0.301	-0.667	-0.073	0.034	0.174	0.354	0.001	0.024	-0.179	-0.600	0.427
BARL	12.04%	-0.860	-0.356	-0.052	**2.261**	-0.077	-0.171	-0.019	0.009	0.045	0.090	0.000	0.006	-0.046	-0.153	0.109
OATS	1.08%	-2.320	-0.959	-0.140	-2.122	**2.903**	-0.460	-0.051	0.023	0.120	0.244	0.001	0.017	-0.123	-0.413	0.294
MAIZ	7.27%	-0.237	-0.098	-0.014	-0.217	-0.021	**3.168**	-0.261	-0.109	-0.559	-0.965	-0.001	-0.020	0.155	0.742	0.064
OCER	1.93%	-0.233	-0.096	-0.014	-0.213	-0.021	-2.334	**2.074**	-0.107	-0.549	-0.949	-0.001	-0.019	0.152	0.729	0.063
RAPE	9.18%	0.012	0.005	0.001	0.011	0.001	-0.109	-0.012	**1.659**	-1.265	0.043	0.000	0.003	-0.014	-0.033	-0.066
SUNF	4.94%	0.134	0.056	0.008	0.123	0.012	-1.214	-0.134	-2.751	**4.059**	0.480	0.001	0.036	-0.151	-0.366	-0.738
PULS	2.89%	0.465	0.192	0.028	0.425	0.042	-3.568	-0.392	0.159	0.817	**2.225**	-0.040	-1.264	-0.302	-1.434	-0.117
POTA	0.38%	0.001	0.000	0.000	0.001	0.000	-0.002	0.000	0.000	0.001	-0.031	**0.384**	-0.001	0.000	-0.001	-0.001
SUGB	1.22%	0.017	0.007	0.001	0.016	0.002	-0.040	-0.004	0.006	0.033	-0.683	-0.001	**3.083**	-0.003	-0.021	-0.015
MAIF	1.36%	-0.711	-0.294	-0.043	-0.650	-0.064	1.728	0.190	-0.151	-0.775	-0.913	0.000	-0.015	**6.590**	-6.639	0.135
OFAR	6.90%	-0.366	-0.151	-0.022	-0.335	-0.033	1.274	0.140	-0.056	-0.289	-0.666	-0.001	-0.018	-1.020	**2.108**	-0.056
NONF	1.43%	1.658	0.685	0.100	1.516	0.148	0.695	0.076	-0.723	-3.708	-0.344	-0.003	-0.083	0.132	-0.355	**3.944**
OSET	6.29%	-0.289	-0.120	-0.017	-0.265	-0.026	-0.246	-0.027	0.181	0.930	0.139	-0.001	-0.035	0.077	-0.135	-0.917
VSET	0.82%	-3.982	-1.646	-0.240	-3.643	-0.356	-3.266	-0.359	-0.563	-2.886	1.564	0.002	0.073	-0.125	-0.127	0.619
FALL	4.01%	-1.044	-0.432	-0.063	-0.955	-0.093	-0.857	-0.094	-0.129	-0.661	0.412	0.001	0.016	-0.025	-0.044	0.081

Table 19. Supply elasticities for France in year 2002 for individual crops.

	Share	SWHE	DWHE	RYEM	BARL	OATS	MAIZ	OCER	RAPE	SUNF	PULS	POTA	SUGB	MAIF	OFAR	NONF
SWHE	26.84%	**1.009**	-0.056	-0.010	-0.397	-0.048	-0.091	-0.006	-0.003	-0.011	-0.001	-0.010	-0.006	-0.029	-0.020	-0.090
DWHE	1.84%	-0.766	**2.102**	-0.054	-0.480	-0.072	-0.132	-0.013	0.004	0.004	0.001	0.029	-0.002	-0.014	-0.022	-0.230
RYEM	0.16%	-3.276	-1.086	**8.577**	-2.818	-0.970	-0.939	-0.397	-0.018	-0.055	-0.013	0.074	0.092	0.003	-0.115	-1.240
BARL	9.01%	-1.322	-0.112	-0.028	**2.243**	-0.113	-0.199	-0.032	-0.027	-0.035	-0.001	-0.007	-0.004	-0.023	-0.028	-0.144
OATS	1.14%	-2.202	-0.241	-0.133	-1.666	**2.884**	-0.391	-0.126	-0.015	-0.036	-0.005	0.002	0.021	-0.031	-0.059	-0.440
MAIZ	10.04%	-0.195	-0.022	-0.006	-0.129	-0.018	**1.680**	-0.285	-0.020	0.043	0.011	-0.160	-0.024	-0.143	0.020	-0.314
OCER	2.03%	-0.122	-0.018	-0.022	-0.181	-0.050	-2.384	**2.205**	0.016	-0.003	0.009	-0.047	-0.002	-0.043	0.009	-0.566
RAPE	4.44%	-0.011	0.003	0.000	-0.049	-0.001	-0.054	0.006	**1.284**	-0.539	-0.011	0.087	-0.008	0.061	-0.098	-0.404
SUNF	2.67%	-0.076	0.007	-0.001	-0.097	-0.006	0.084	-0.001	-1.016	**2.959**	-0.042	0.358	-0.008	0.087	-0.181	-1.727
PULS	0.35%	-0.058	0.008	-0.003	-0.034	-0.009	0.482	0.047	-0.196	-0.373	**2.020**	0.066	-0.056	0.002	-0.050	-1.443
POTA	2.40%	-0.130	0.044	0.003	-0.014	0.001	-0.993	-0.005	0.180	0.458	0.010	**2.113**	-0.234	-0.712	-0.200	-1.812
SUGB	0.89%	-0.020	0.000	0.002	-0.004	0.004	-0.065	0.000	-0.006	-0.006	-0.003	-0.099	**1.210**	-0.023	0.000	0.059
MAIF	2.40%	-0.114	-0.003	0.000	-0.026	-0.002	-0.205	-0.008	0.030	0.023	0.000	-0.172	-0.018	**2.434**	-0.043	-0.144
OFAR	7.74%	-0.089	-0.009	-0.002	-0.043	-0.006	0.051	0.002	-0.071	-0.071	-0.002	-0.076	-0.001	-0.063	**1.304**	-1.114
NONF	18.08%	-0.195	-0.039	-0.009	-0.102	-0.022	-0.203	-0.059	-0.121	-0.287	-0.036	-0.237	0.033	-0.064	-0.417	**2.059**
OSET	2.02%	1.329	0.146	0.025	0.785	0.092	0.897	0.286	-0.586	-1.196	-0.083	0.381	-0.250	0.209	-0.127	-1.973
VSET	5.17%	-0.442	-0.043	-0.006	-0.256	-0.027	-0.241	-0.092	0.279	0.603	0.048	-0.062	0.076	-0.163	0.206	0.124
FALL	1.15%	-1.355	-0.141	-0.026	-0.824	-0.071	0.012	0.030	-0.208	-0.430	-0.010	0.292	-0.018	-0.001	-0.664	-1.499

Note: Numbers in parentheses from Heckelei and Britz (2000 table 2), in square brackets from Guyomard et al. (1996 table 2).

Table 20. Supply elasticities for FR24 for crop groups in 2002.

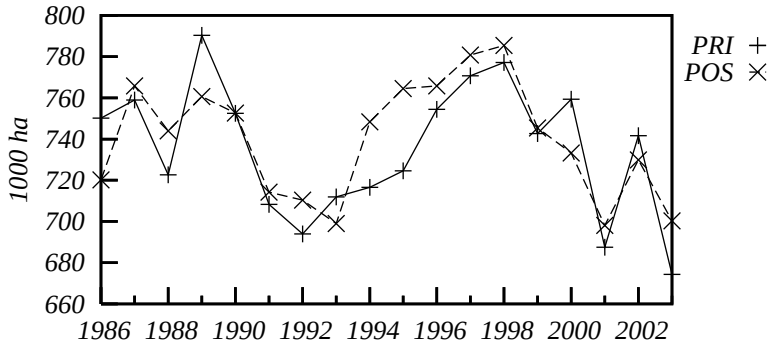
	CERE	CER2	OILS	OARA	FARA
CERE	**0.509**	-0.124	0.107	0.064	-0.131
CER2	-0.666	**2.554**	-0.685	-1.118	1.017
OILS	0.489	-0.418	**0.321**	0.148	-0.207
OARA	0.727	-2.473	0.540	**1.228**	-1.085
FARA	-0.999	1.428	-0.444	-0.691	**0.861**
NOCR	-2.066	-0.765	-0.354	0.334	-0.076

Table 21. Supply elasticities for France for crop groups in 2002.

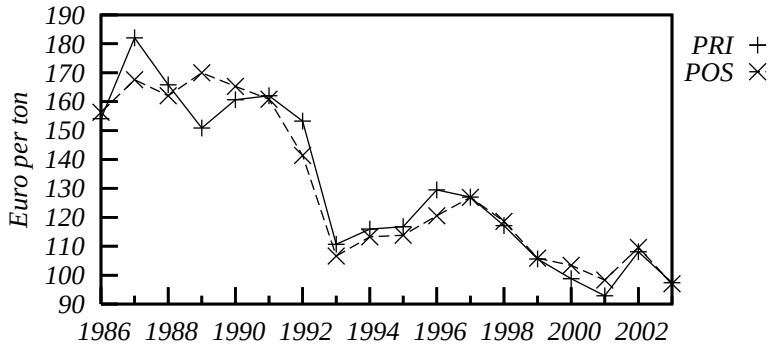
	CERE	CER2	OILS	OARA	FARA
CERE	**0.508**	-0.152	0.046	-0.038	-0.151
CER2	-0.395	**1.220**	0.076	-0.343	-0.343
OILS	0.352	0.209	**0.807**	0.240	-1.042
OARA	-0.138	-0.543	0.402	**1.623**	-0.895
FARA	-0.299	-0.167	-0.428	-0.231	**1.201**
NOCR	-1.273	-0.360	-0.353	-0.127	-0.656

Appendix 3:Plots of prior versus posterior mode for FR24

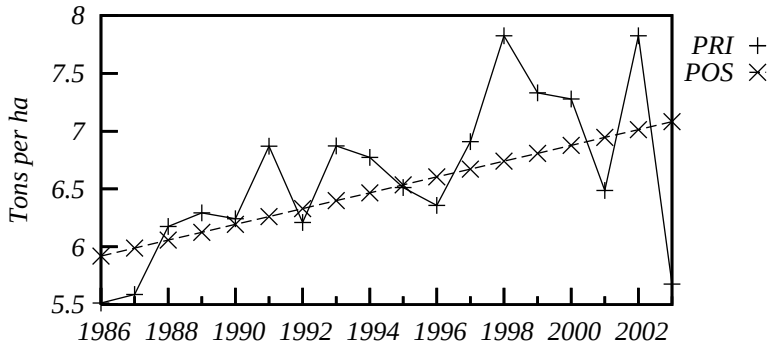
FR240000: SWHE, Acreage



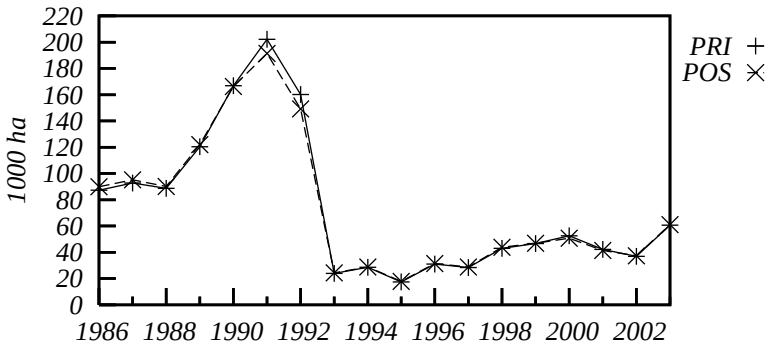
FR240000: SWHE, Output Price



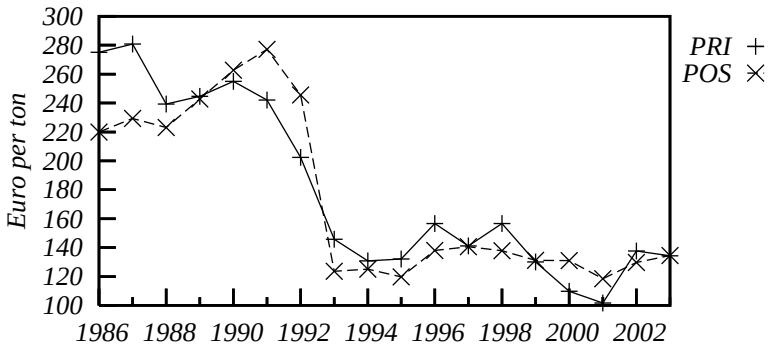
FR240000: SWHE, Yield



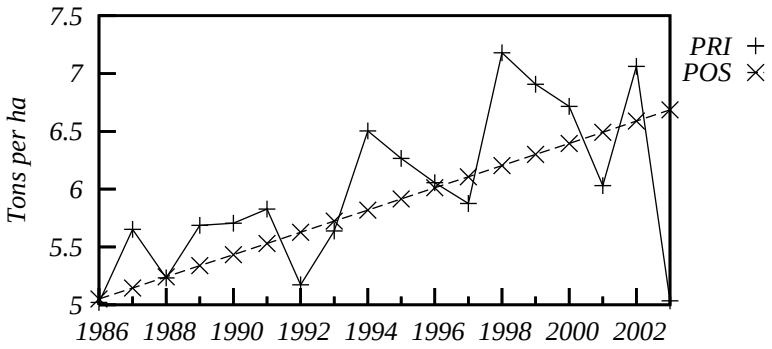
FR240000: DWHE, Acreage



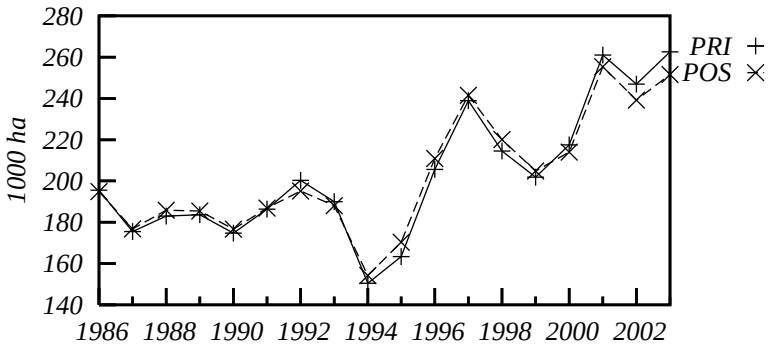
FR240000: DWHE, Output Price



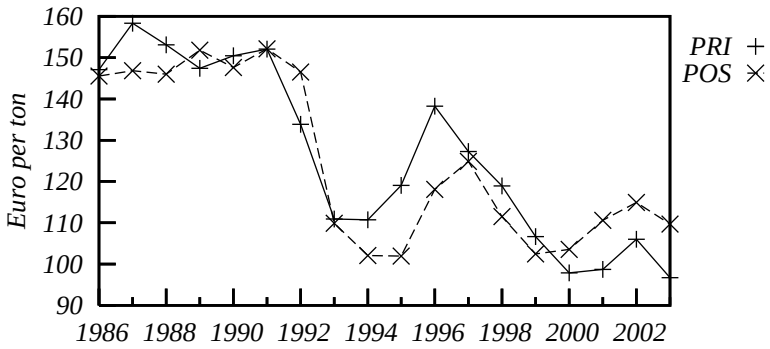
FR240000: DWHE, Yield



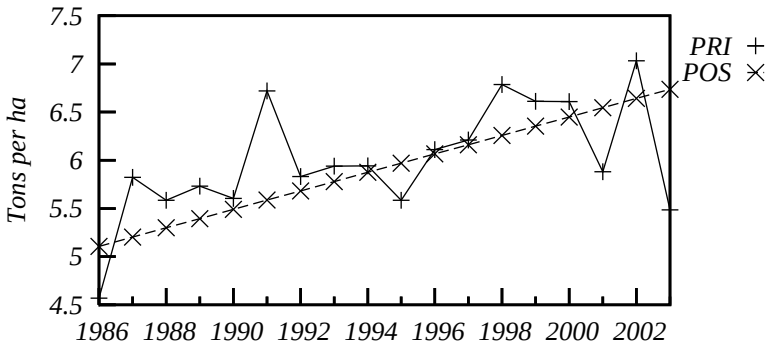
FR240000: BARL, Acreage



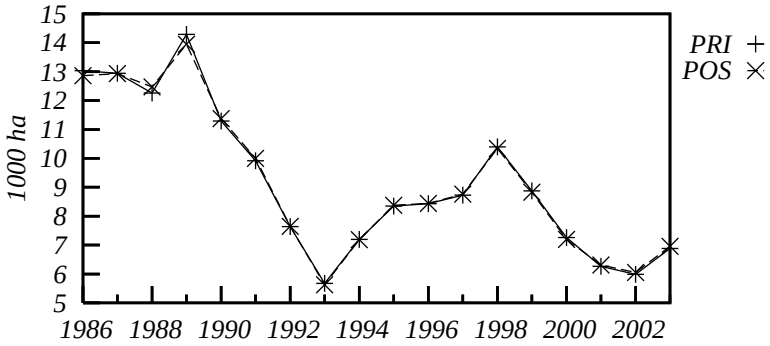
FR240000: BARL, Output Price



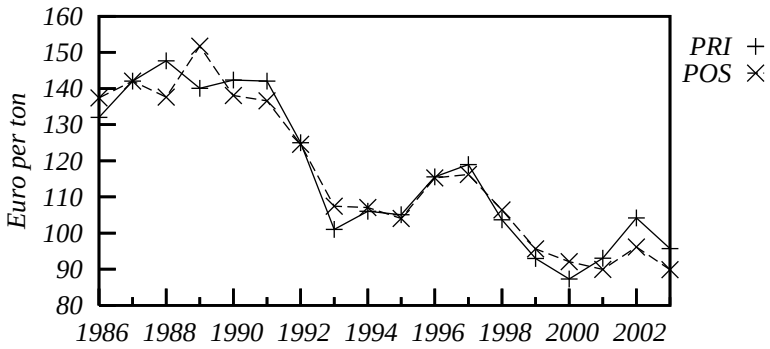
FR240000: BARL, Yield



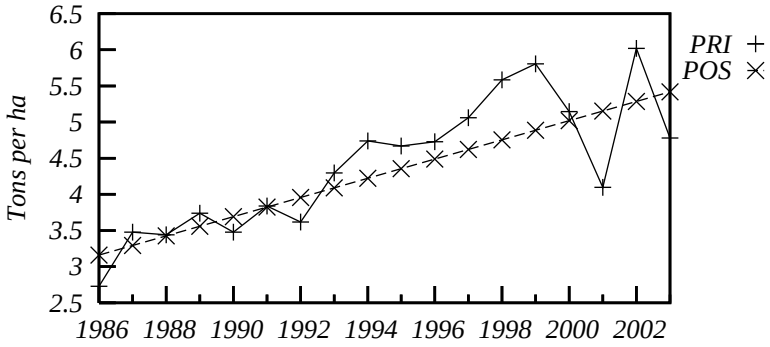
FR240000: RYEM, Acreage



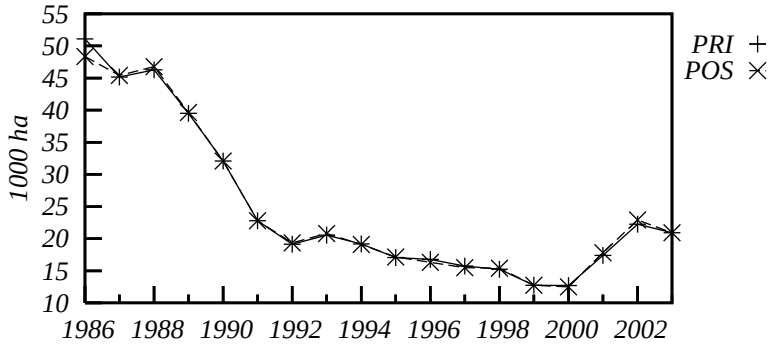
FR240000: RYEM, Output Price



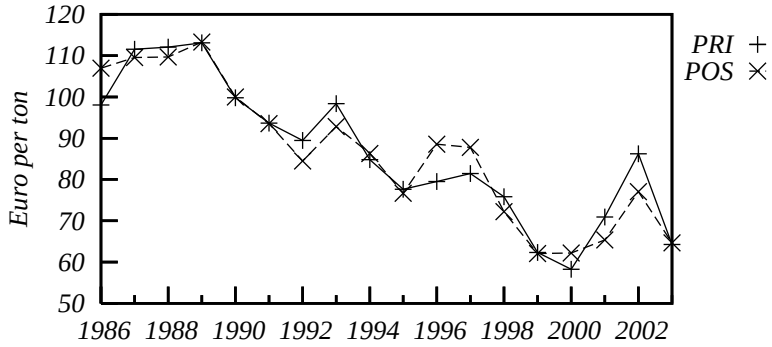
FR240000: RYEM, Yield



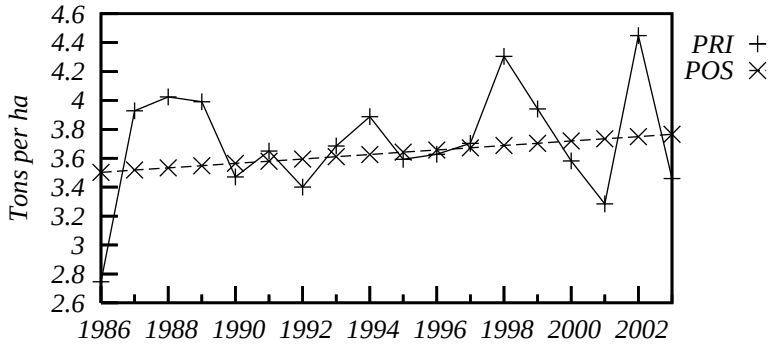
FR240000: OATS, Acreage



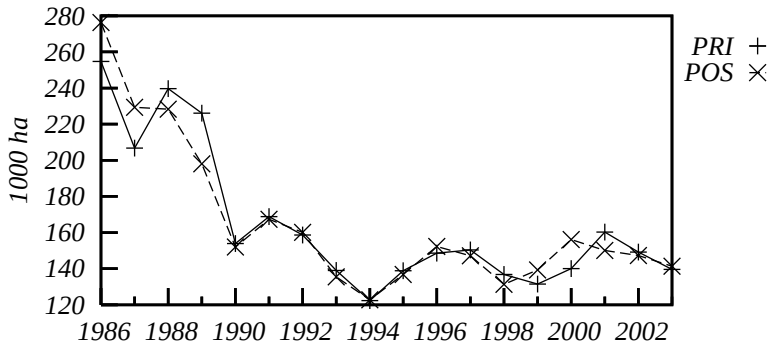
FR240000: OATS, Output Price



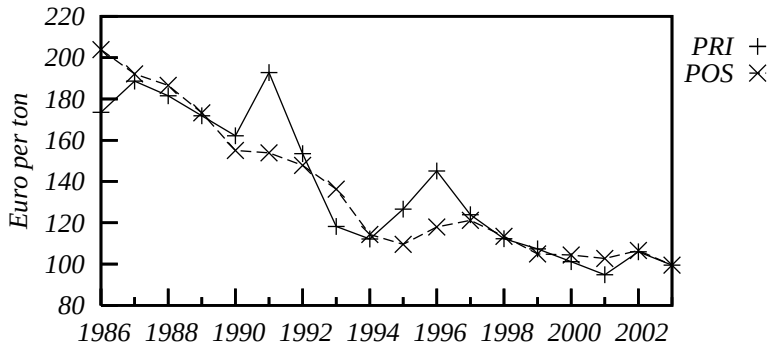
FR240000: OATS, Yield



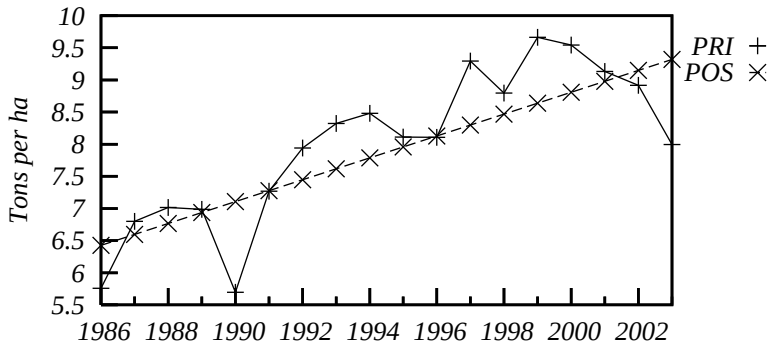
FR240000: MAIZ, Acreage



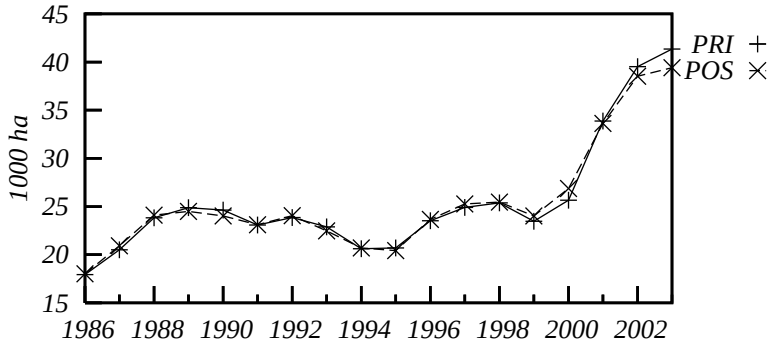
FR240000: MAIZ, Output Price



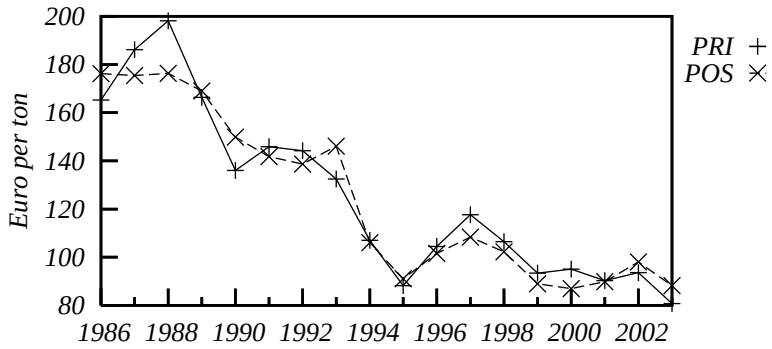
FR240000: MAIZ, Yield



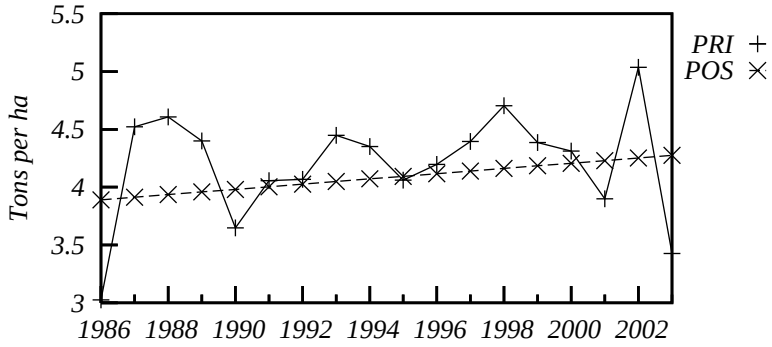
FR240000: OCER, Acreage



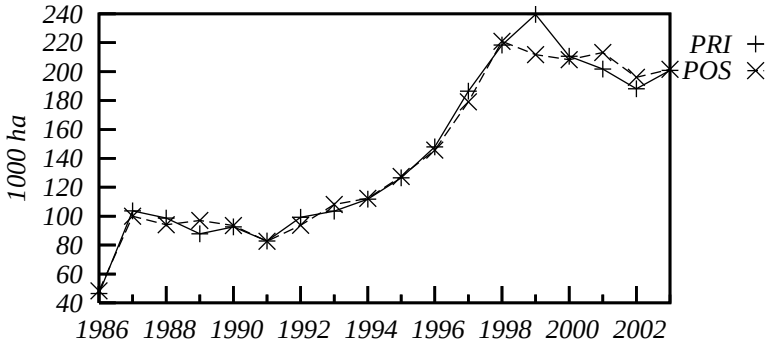
FR240000: OCER, Output Price



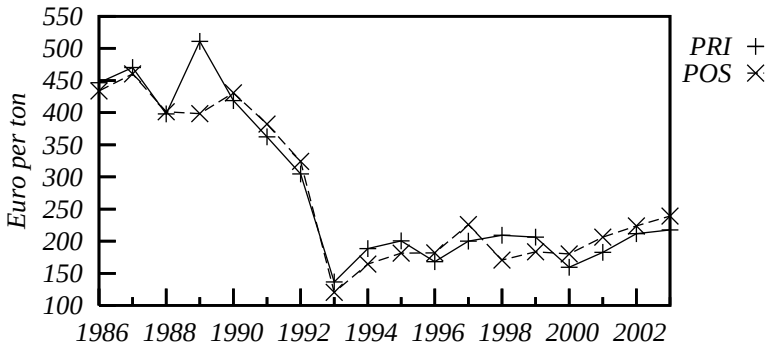
FR240000: OCER, Yield



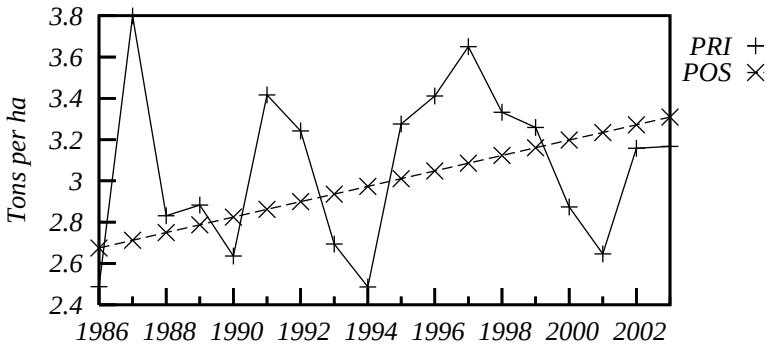
FR240000: RAPE, Acreage



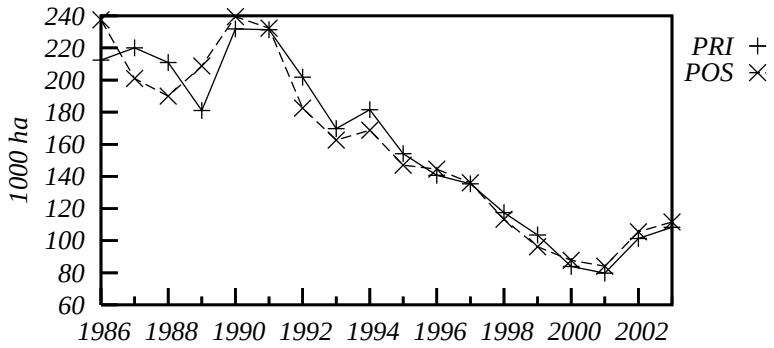
FR240000: RAPE, Output Price



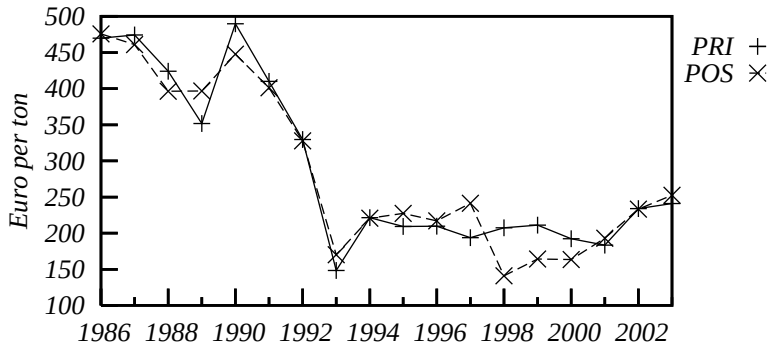
FR240000: RAPE, Yield



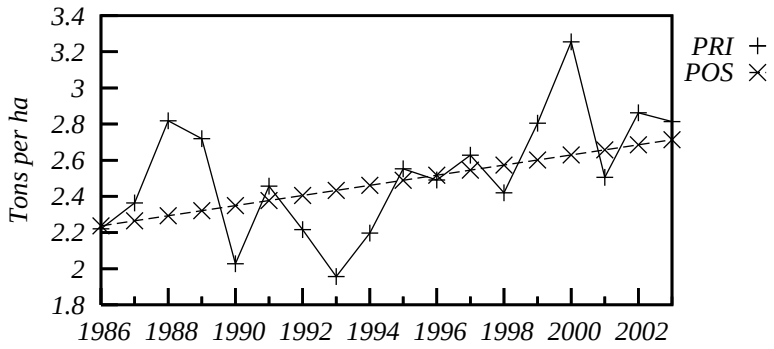
FR240000: SUNF, Acreage



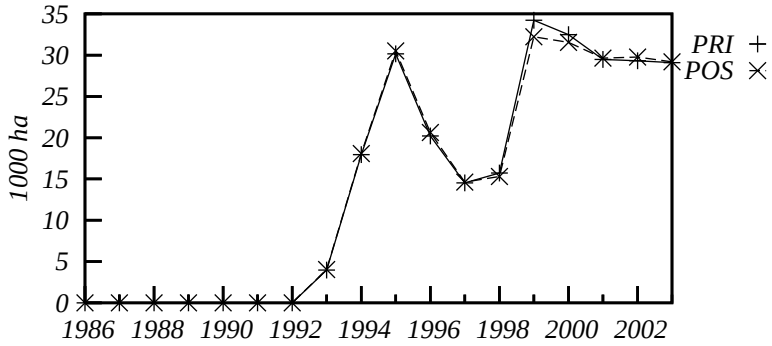
FR240000: SUNF, Output Price



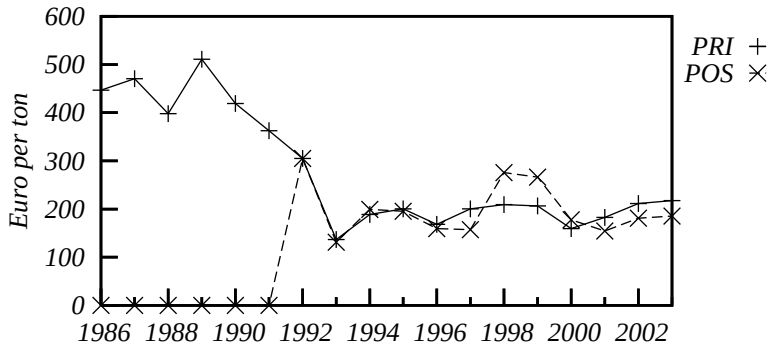
FR240000: SUNF, Yield



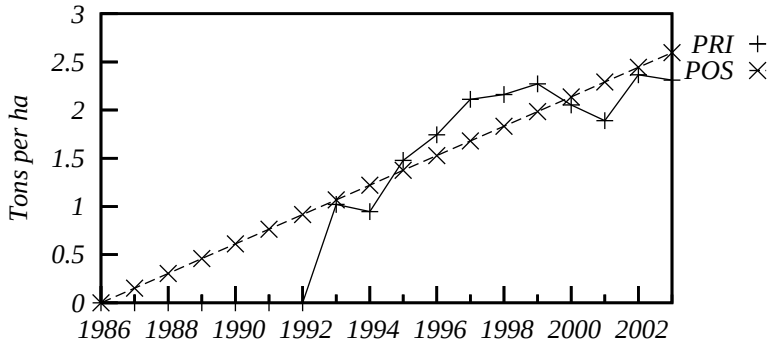
FR240000: NONF, Acreage



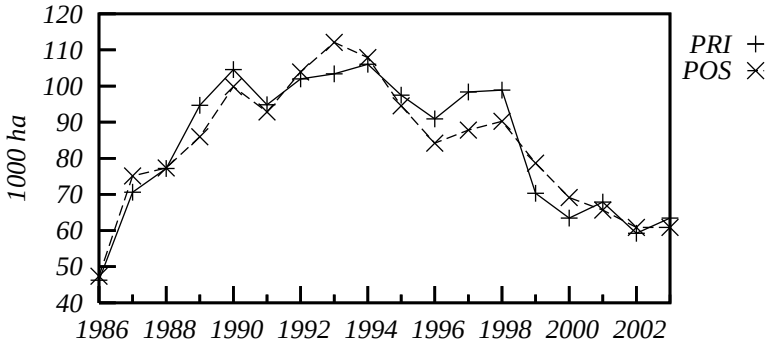
FR240000: NONF, Output Price



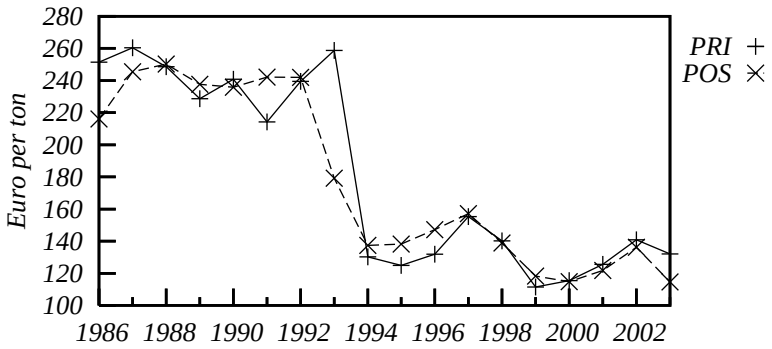
FR240000: NONF, Yield



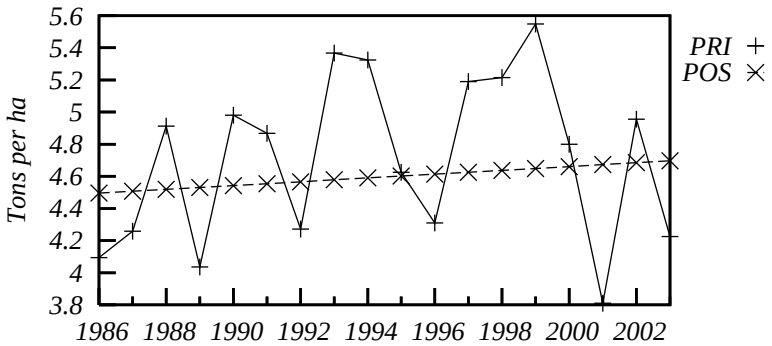
FR240000: PULS, Acreage



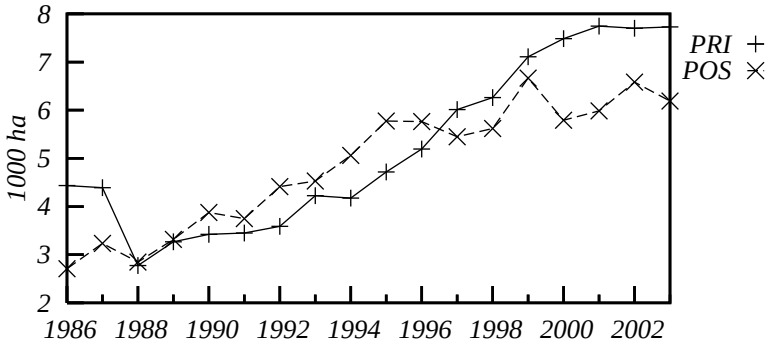
FR240000: PULS, Output Price



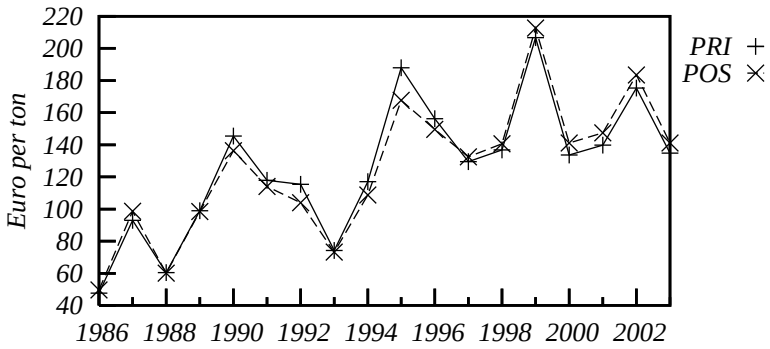
FR240000: PULS, Yield



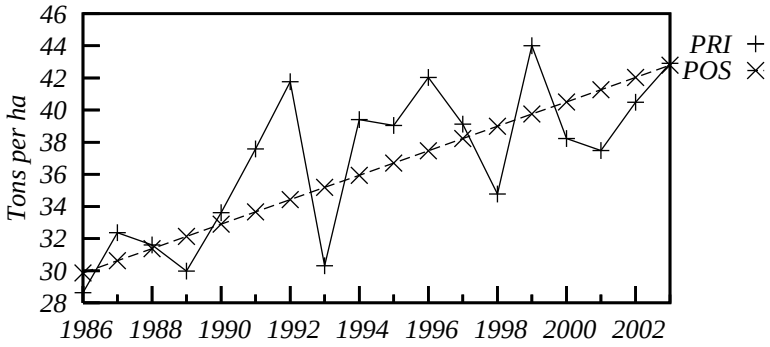
FR240000: POTA, Acreage



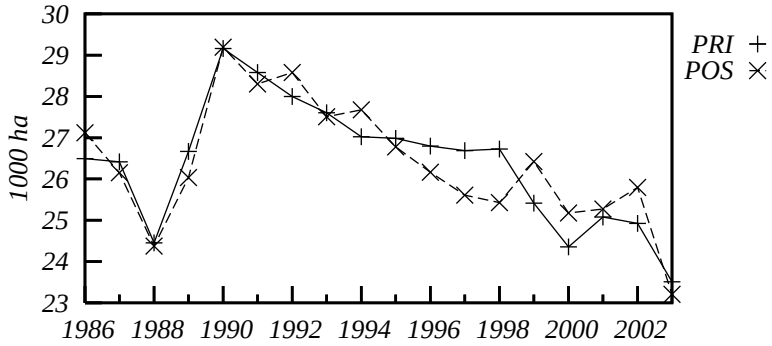
FR240000: POTA, Output Price



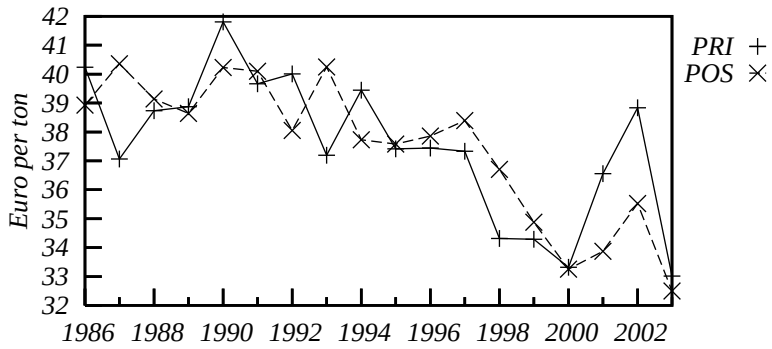
FR240000: POTA, Yield



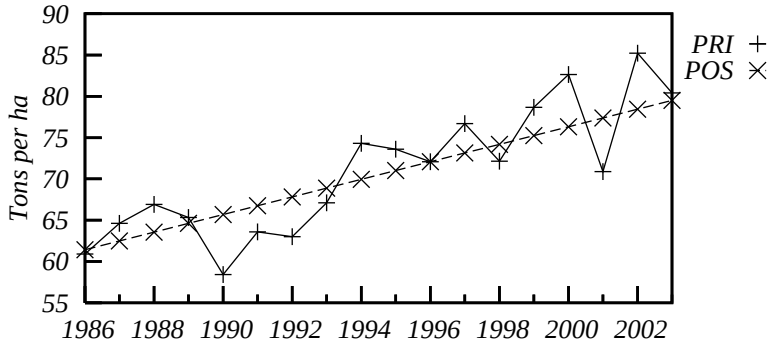
FR240000: SUGB, Acreage



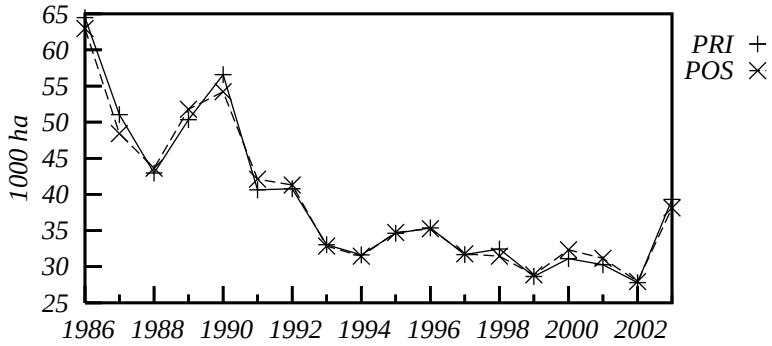
FR240000: SUGB, Output Price



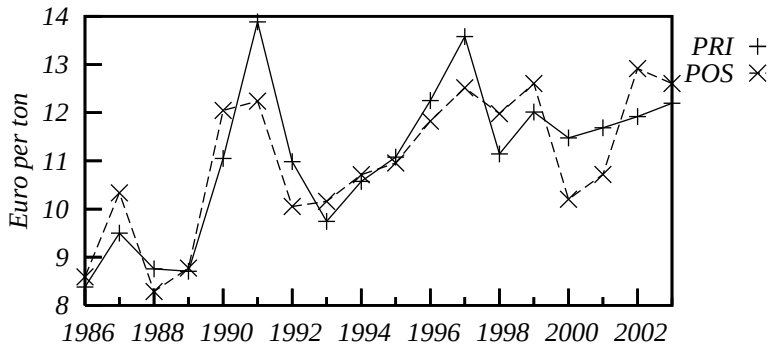
FR240000: SUGB, Yield



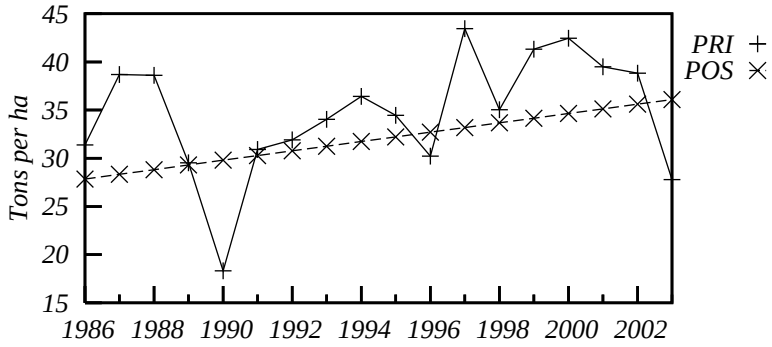
FR240000: MAIF, Acreage



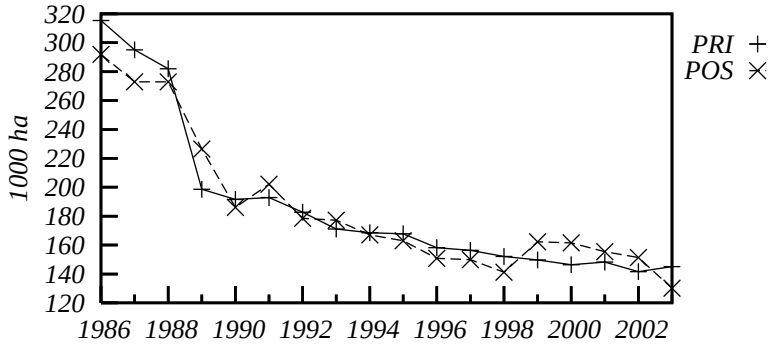
FR240000: MAIF, Output Price



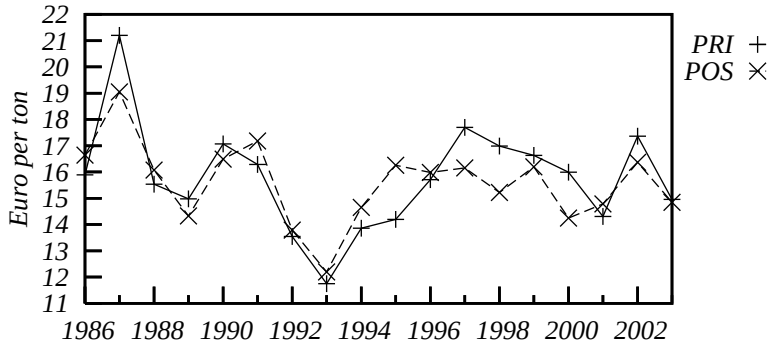
FR240000: MAIF, Yield



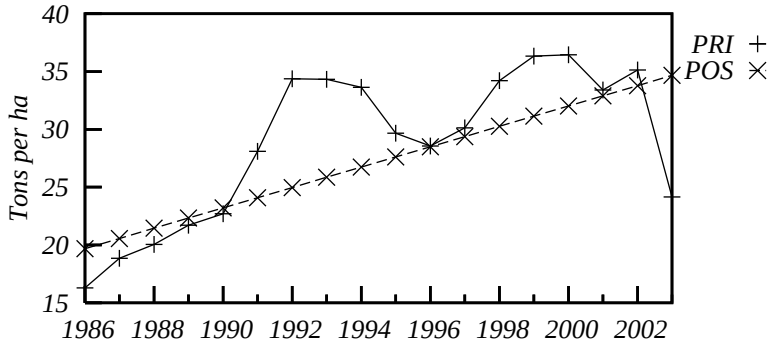
FR240000: OFAR, Acreage



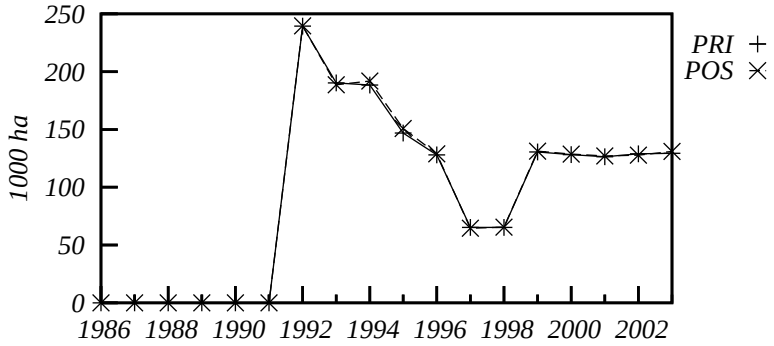
FR240000: OFAR, Output Price



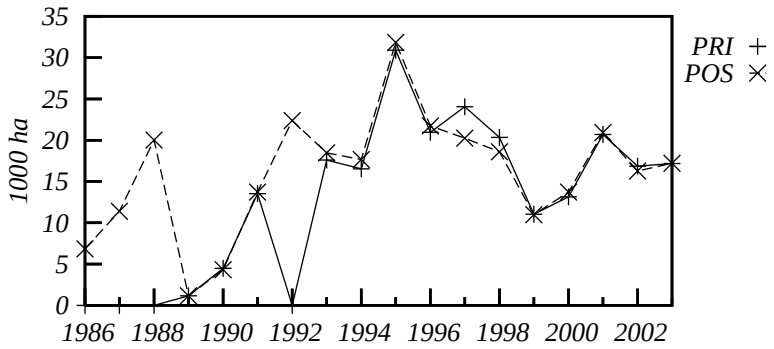
FR240000: OFAR, Yield



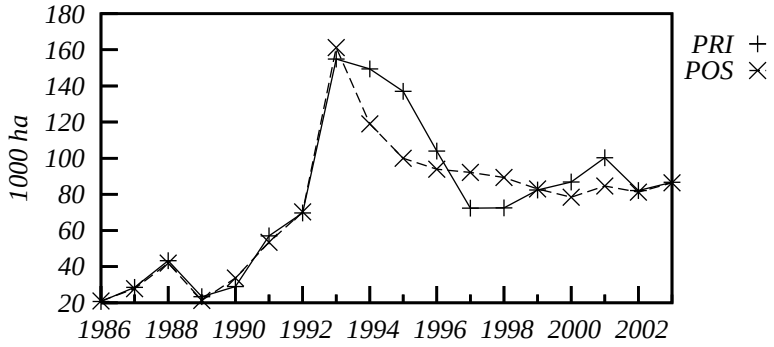
FR240000: OSET, Acreage



FR240000: VSET, Acreage



FR240000: FALL, Acreage



Chapter 6 Discussion

1. Conclusions

This thesis attempted to provide a general approach to econometric specification of constrained optimization models. Special attention was given to issues that arise when (1) inequality constraints are involved in the model that is to be estimated, and (ii) the estimation problem is ill-posed, i.e. the parameters are not identified. It can be argued that both of those problems are common in empirical work, and thus the treatment provided in this thesis should be of general interest. The solution approaches contained elements from bilevel programming, errors-in-variables modelling and Bayesian estimation, and was developed in two theoretical and two empirical chapters.

Chapter two showed how efficiency can be gained if the estimation is treated as a bilevel programming problem, and also suggested a numerical method from the field of mathematical programming with equilibrium constraints that can be used to find the estimates in the case when the optimization model to estimate contains inequality constraints.

Chapter three applied the results obtained in chapter 2 to a spatial price equilibrium model of crop production in Benin. Trade costs, prices and regional excess demand were estimated for one year, given observations of prices, excess demand and distances between markets. The estimates compare rather well with results of empirical studies, albeit they were found not to be robust.

In chapter four, the focus was on the inclusion of information from diverse data sources using a Bayesian approach. It was shown how a highest posterior density estimator can be regarded a more general alternative to some currently used methods for inclusion of prior information, most notably maximum entropy and generalized cross entropy.

In chapter five, the estimator proposed in chapter four was applied to the supply model of the large scale modelling system CAPRI. The estimation was formulated as a bilevel programming problem, which was elaborated to an additive measurement error model, and a highest posterior density estimator designed to include prior information on several parameters. In contrast to the application in chapter three, no inequality constraints were included. All in all, supply parameters of supply models with up to 23 crops were estimated in 165 regions in the EU. The results were compared to aggregated results of other studies for France, Denmark and the Netherlands.

It is tempting to try to condense the entire body of this thesis into a concise "recipe" that in a few points addresses the title subject. With risk of oversimplifying, the general approach to the estimation of parameters of constrained optimization models where there are observations correlated with model outcomes can thus be condensed into the following steps:

1. Formulate the necessary and sufficient optimality conditions for the model to estimate.

2. Formulate a plausible and workable error model that explains how observations relate to model parameters or variables.

3. Include any available prior information on the parameters of the error model.

4. Choose an estimation criterion that fits with the selected error model. If prior information is to be included (and also otherwise), the highest posterior density estimator is an important option. If the problem is ill-posed even after prior information has been included, the posterior mean may still be defined.

5. Find the estimates by solving the bilevel programming problem defined by optimizing the estimation criterion subject to the optimality conditions, error model equations and prior distributions. If the inner problem contains inequality constraints, apply bilevel programming techniques for a numerical solution.

2. Outlook

Providing a fully general treatment of the selected topic is a tremendous task that in fact is too vaguely defined to be suitable for inquiry. Some areas that could be considered as belonging to this topic were hardly touched upon at all in this thesis. Perhaps the most obvious such deficit is the issue of hypothesis testing and regions of confidence. Only chapter two is concerned with the small sample properties of the parameters, and none of the preceding chapters provided a way of testing the significance of the estimating equations or provide regions of confidence for the estimated parameters.

If the estimation of all parameters is performed within the general framework provided here, the possibility is opened up for statistical tests for parts or whole models. If the model at hand is a *normative* model, such testing is not really interesting, since the model is not intended to answer questions of how reality *is* behaving, but of how it *should* behave. But precisely the application of empirical estimation procedures like the ones advocated here shows that the borderline be-

tween *normative* and *positive* models is fuzzy, since the estimation aims at adjusting the model to better reflect *actual* behaviour.

Theoretically, the work presented here falls within the area of extremum estimators, a class of problems for which several potentially useful results are available, at least in cases where the inner problem has no inequality constraints. Analytical deduction of estimator properties is, however, likely to be difficult in most cases, depending on the complexity of the model at hand, and especially on the existence of inequality constraints. A workable alternative could be numerical simulation methods like bootstrapping. It was considered outside the scope of this thesis to also treat this subject at any depth.

For the author in his role as an applied economist, the work presented in this thesis has provided a general, readily applicable framework for estimating parameters in a wide array of models, also facing sparse data or identification problems. When this is written (spring 2007), the framework has already proved useful in several applications not reported here (not published work). It is the modest wish of the author that it may facilitate consistent and transparent model parameter specification also for a wider community of modellers.

The End.