

DISS. ETH NO. 21219

Essays in Applied Microeconomics

An Econometric Analysis of Swiss Gasoline Demand Using Panel Data

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

FABIAN MAX HEIMSCH

Master of Science in Management, Technology and Economics

ETH ZURICH

born 8th of September 1981

citizen of St. Gallen (SG)

accepted on the recommendation of

Prof. Dr. Massimo Filippini, examiner

Prof. Dr. Peter Egger, co-examiner

2013

Summary

This dissertation is composed of two essays on the econometric analysis of the demand for Swiss gasoline using panel data and a third essay on the econometric properties of the Arellano Bond and Blundell Bond estimators in the context of exogenous variables with a low within variation.

The goal of the first essay is to estimate a demand function for Swiss gasoline in Switzerland's border regions and simultaneously to quantify the amount of gasoline sold to foreigners. The share of gasoline sold to foreigners is often labelled as "gasoline tourism" and is driven by the existing price differentials across the border. Therefore, the price ratio is used as an explanatory variable in an econometric model and is also weighted with the distance from the border to determine the distance range from the border within which gasoline tourism is supposed to have an impact on the amount of gasoline sold. For this reason, gasoline sales data from the five largest gasoline retailers operating in Switzerland were collected. The sales from gasoline stations were averaged at the municipal level serving as the sales from a representative gasoline station for that municipality. A static panel data model accounting for fixed and random effects was estimated. We found a significant price elasticity of Swiss gasoline demand (in the border regions) with respect to the Swiss gasoline price of -0.211 when gasoline tourism was ruled out. The average price elasticity with gasoline tourism considered is -0.65. Accordingly, the demand for Swiss gasoline is considerably affected by foreigners purchasing gasoline in the Swiss border regions. The results indicate that gasoline tourism had an impact on municipalities up to a distance of some 30 kilometres from the border, but the main part was sold at stations located within 15 kilometres. The model shows that, compared to overall sales, which accounted for some 4.5 billion litres of gasoline, gasoline tourism reached values from some 250 million litres up to 450 million litres of gasoline on average. Further, it is shown that even a small increase in the Swiss gasoline price may lead to a substantial decrease in gasoline tourism.

The overall goal of Essay 2 is to estimate the demand for Swiss gasoline at the municipal level while taking spatial effects into account. The demand for gasoline can clearly be seen as a spatial story, since the consumption of gasoline not only depends on a municipality's car fleet or population, but also on exchange traffic. We use a panel data model with spatially lagged residuals and a spatially lagged dependent variable and account for random effects. We estimate a coefficient of the spatially lagged dependent variable of 0.34 and a coefficient of the spatially lagged residuals of 0.37. This implies that an increase in gasoline demand in one municipality

by 10.0% spreads over to other municipalities and leads to an increase of 3.4% in demand, given that the regions are neighbours. As a main result, we estimate an average price elasticity of Swiss gasoline demand of -0.655 (total effect). Spatial partitioning of this value leads to a direct effect of -0.58 on average. We estimate price elasticities ranging from -0.585 to -0.855, depending on the municipalities' locations. Due to the very different approach, we find a price elasticity of Swiss gasoline demand with respect to the foreign price of 0.32 (as the average total effect), which is significantly different from that reported in Baranzini et al. (2012) but not from the value obtained in Essay 1. As the demand for gasoline in one municipality affects demand in neighbouring municipalities, those results can impose important consequences on policy makers: First, there are regions which react more sensitively to change in the gasoline price than others (e.g., this change might result from the introduction of a CO₂ tax). From the spatial analysis, we conclude that the border regions and in general the urban areas of Switzerland respond more strongly to price changes than more rural or remote regions. One explanation for this is that public transport is more readily available in urban areas and therefore serves as a substitute.

The goal of Essay 3 is to assess the estimation accuracy in terms of bias, variance and root mean squared error (RMSE) of the FD-GMM estimator and of the SYS-GMM estimator when an exogenous regressor exhibits a low within variation. For this reason, a Monte Carlo experiment is carried out. We vary the number of cross-sectional units N , the number of observations per unit T , the coefficient of the lagged dependent variable and the within variation of the exogenous regressor over a parameter range which is usually of interest or which can be met in applied empirical studies. Each experiment is replicated 1,000 times. For several parameter combinations, the bias of estimated coefficients is lower for FD-GMM. However, if the variance is preferred as the decision parameter to discriminate between the two estimators, SYS-GMM should be preferred in almost all situations. As a reasonable compromise, we use the root mean squared error, which combines variance and bias of the estimates in one measure. Using the RMSE of the estimates, the SYS-GMM estimator should generally be preferred over the FD-GMM estimator when the within variation of the exogenous regressor is low. For instance, for the situation where a small panel is used (e.g. $N = 50$ and $T = 5$), the RMSE of the short-run effects are on average almost 20% lower for the SYS-GMM estimator. For situations with a relatively low within variation compared to the between variation e.g. of only 10%, the RMSE of the SYS-GMM estimates are as much as 67% below the FD-GMM estimates. Interestingly, for a small panel with low within variation of the exogenous regressor, results in terms of bias of the short-run effect are significantly in favour of the FD-GMM estimator. However, the efficiency gain in terms of a lower variance is strongly in favour of the SYS-GMM estimator, which compensates its relatively higher bias. Similar findings can be reported for the long-run effects and the coefficient of the lagged dependent variable itself. All experiments were carried out with either only a subset of instruments or the full set of instruments used. Further, a decision matrix is created with which the researcher can decide either to use FD-GMM or SYS-GMM, dependent on how bias is weighted against variance and dependent on the within variation of the exogenous regressor, on the number of observations N and T and on the supposed persistence of the dependent variable, γ .

Zusammenfassung

Diese Doktorarbeit besteht aus vier Teilen. Der erste Teil entspricht den Vorgaben für Doktorarbeiten des Departements für Management, Technologie und Ökonomie (D-MTEC) und stellt das Einführungskapitel dar. Die restlichen Teile sind für die drei Aufsätze reserviert, welche die eigentliche Dissertation bilden.

Das Ziel des ersten Aufsatzes ist die Schätzung einer Nachfragefunktion für Benzin in den Schweizer Grenzregionen und gleichzeitig die Quantifizierung der Benzinmenge, welche an Ausländer verkauft wurde. Der Anteil an Benzin, welcher an ausländische Fahrzeughalter verkauft wird, wird oft als "Tanktourismus" bezeichnet und wird durch die Preisdifferenzen über die Grenze getrieben. Daher wird das Preisverhältnis als erklärende Variable in einem ökonometrischen Modell benutzt und wird zudem gewichtet mit der Distanz zur Grenze, um aus dem Modell dann die Reichweite zu bestimmen, in welcher Tanktourismus eine Rolle spielt. Zu diesem Zweck wurden Verkaufsdaten von Tankstellen von den fünf grössten Benzinverteilern, welche in der Schweiz operieren, erhoben. Aus diesen Verkaufsdaten wurde dann ein Durchschnittswert der Absätze pro Tankstelle in einer Gemeinde gebildet, welche dann als Absätze einer "Referenztankstelle" der jeweiligen Gemeinde dient. Ein statisches Paneldaten Modell für fixe und zufällige Effekte wurde geschätzt. Wir haben eine Elastizität der Schweizerischen Benzinnachfrage in Bezug auf den Schweizer Preis von -0.211 gefunden, solange der Tanktourismus nicht berücksichtigt wird. Die durchschnittliche Preiselastizität unter Berücksichtigung des Tanktourismus beträgt allerdings -0.65 . Somit ist die Nachfrage nach Schweizer Benzin in den Grenzregionen beträchtlich durch die Nachfrage der ausländischen Fahrzeughalter beeinflusst. Die Resultate zeigen ferner, dass der Tanktourismus ungefähr innerhalb einer Reichweite von 30 Kilometern der Grenze eine Rolle spielt, der Hauptanteil dieser Verkäufe an ausländische Fahrzeughalter stellen jedoch Tankstellen, die maximal 10 bis 15 Kilometer zur Grenze entfernt sind. Das Modell zeigt ausserdem, dass je nach Jahr 250 bis 450 Millionen Liter der totalen Benzinabsätze von 4.5 Milliarden Litern in der Schweiz dem Tanktourismus zugeschrieben werden müssen. Ausserdem wird gezeigt, dass auch nur eine kleine Erhöhung des Schweizer Benzinpreises den Tanktourismus empfindlich einbrechen lassen könnte.

Das Ziel des zweiten Aufsatzes ist die Schätzung einer Nachfragefunktion nach Schweizer Benzin auf Gemeindeebene unter Berücksichtigung von räumlicher Korrelation. Die Nachfrage nach Benzin kann auf jeden Fall als ein räumliches Phänomen verstanden werden, da die Nachfrage nach Benzin in einer Gemeinde nicht nur durch die Bevölkerung oder durch den Fahrzeugbestand in jener Gemeinde beeinflusst wird, sondern auch durch den Verkehr zwischen

den Gemeinden. Wir benutzen somit ein Paneldaten Modell mit räumlich korrelierten Residuen und einer räumlich korrelierten abhängigen Variablen. Wir schätzen einen Koeffizienten für die räumlich korrelierte abhängige Variable von 0.34 und einen Koeffizienten für die räumlich korrelierten Residuen von 0.37. Das bedeutet, dass eine Erhöhung der Benzinnachfrage um 10.0% in einer Gemeinde zu einer 3.4%igen Erhöhung der Benzinnachfrage in benachbarten Gemeinden führt. Wir schätzen zudem eine durchschnittliche Preiselastizität der Benzinnachfrage von -0.655 (als totalen Effekt). Die räumliche Aufteilung dieses Wertes führt zu einem durchschnittlichen direkten Effekt von -0.58. Je nach geographischer Lage einer Gemeinde, können sich die berechneten Elastizitäten von -0.585 und -0.855 bewegen. Aufgrund der sehr unterschiedlichen Schätzmethoden finden wir eine Preiselastizität im Bezug auf den ausländischen Benzinpreis von 0.32 (als den durchschnittlichen totalen Effekt), welcher sich signifikant von jenem unterscheidet, der durch Baranzini et al. (2012) geschätzt wurde, aber nicht von jenem, welchen wir im ersten Aufsatz berechnet haben. Da die Nachfrage nach Benzin in einer bestimmten Gemeinde auch die Nachfrage in den benachbarten Gemeinden beeinflusst, können die erhaltenen Resultate auch energiepolitisch wichtige Konsequenzen haben: Zum Beispiel reagieren nicht alle Regionen gleich sensitiv auf Änderungen im Benzinpreis (zum Beispiel über die Einführung einer CO₂-Steuer). Aus der räumlichen Analyse können wir schliessen, dass Grenzregionen und generell städtische Gebiete sensitiver auf den Benzinpreis reagieren als Gebiete in eher abgelegenen oder alpinen Regionen. Eine mögliche Erklärung für diese Beobachtung ist, dass in städtischen Regionen unter anderem eine höhere Verfügbarkeit von öffentlichen Verkehrsmitteln und somit eine Substitutionsmöglichkeit zum privaten Verkehr besteht, was in abgelegenen Regionen häufig nicht der Fall ist.

Das Ziel des dritten Aufsatzes ist es, den Einfluss von exogenen Variablen mit kleinen Varianzen zwischen den Gruppen auf die Schätzgenauigkeit des Arellano-Bond Schätzers (in der Folge FD-GMM genannt) und des Blundell-Bond Schätzers (SYS-GMM) zu erklären. Ein oft vorkommendes Problem in empirischen Studien ist die Tatsache, dass Koeffizienten von Variablen mit sehr kleinen Varianzen innerhalb der Gruppen nur sehr ungenau geschätzt werden können. Um jedoch Vorteile wie das Berücksichtigen von individueller Heterogenität mit Paneldaten Schätzern voll auszunutzen, müssen die Variablen im Modell entsprechend transformiert werden. Dies wird je nach Schätzer unterschiedlich erreicht: Bei statischen Modellen mit fixen Effekten werden die Daten transformiert, indem gruppenspezifische Mittelwerte subtrahiert werden. Bei dynamischen Paneldaten Modellen werden üblicherweise die ersten Differenzen gebildet (Differenzen von der aktuellen Periode zur vorherigen Periode). Solche Transformationen können einen beträchtlichen Teil der Varianzen zwischen den Gruppen eliminieren. Daher weisen Variablen, welche eine kleine Varianz innerhalb der Gruppen haben, nach der Transformation auch eine kleine Varianz auf. In der angewandten empirischen Forschung tritt dieser Effekt besonders häufig auf, wenn sozio-ökonomische Daten in den Modellen verwendet werden, was oft der Fall ist. Typische Variablen mit einer kleinen Varianz innerhalb der Gruppen (in Bezug auf die Varianz zwischen den Gruppen) sind Einkommensdaten, Bevölkerungsdaten oder die Anzahl der Fahrzeuge in einer Gemeinde, um einige zu nennen. Clark und Linzer (2012) haben den Einfluss von kleinen Varianzen innerhalb

der Gruppen der erklärenden Variablen auf statische Paneldaten Modelle mit fixen und zufälligen Effekten analysiert.

Dementsprechend führen wir ein Monte Carlo Experiment durch, um die Fragestellung für den Fall von dynamischen Paneldaten Schätzern zu erklären. Wir variieren die Anzahl der Gruppen N , die Anzahl Beobachtungen innerhalb der Gruppen T , den Koeffizienten der zeitlich verzögerten abhängigen Variablen und die Varianz innerhalb der Gruppen eines exogenen Regressors über einen Parameterbereich, der typischerweise in empirischen Studien anzutreffen ist. Jedes Experiment wurde 1'000 mal wiederholt.

Für einige der genannten Parameter Kombinationen ist die Abweichung der geschätzten Koeffizienten vom wahren Wert signifikant geringer für FD-GMM als für SYS-GMM. Wird jedoch die Varianz der Schätzer als Qualitätskriterium hinzugezogen, so sollte der SYS-GMM Schätzer in (fast) jedem Fall vorgezogen werden. Als guter Kompromiss gilt die Betrachtung der Wurzel des mittleren quadrierten Fehlers (RMSE = root mean squared error), da dieser beide Sichtweisen, jene der Abweichung zum wahren Wert und jene der Varianz der Schätzer, einbezieht. Benützt man den RMSE als Qualitätskriterium, so ist SYS-GMM generell dann vorzuziehen, wenn die Varianz des exogenen Regressors innerhalb der Gruppen klein ist. Für eine Situation mit einem kleinen Paneldatensatz ($N = 50$ und $T = 5$), der RMSE der kurzfristigen Effekte (geschätzte Koeffizienten des exogenen Regressors) ist im Mittel 20% kleiner für den SYS-GMM Schätzer. Wenn zusätzlich nun die Varianz innerhalb der Gruppen noch klein ist, so beträgt der Unterschied bis zu 67%. Interessanterweise ist die mittlere Abweichung der Koeffizienten zum wahren Wert für die genannten Situationen im Mittel kleiner für den FD-GMM Schätzer als für den SYS-GMM Schätzer. Die Varianz des SYS-GMM Schätzers ist aber viel geringer, so dass der SYS-GMM Schätzer auch mit dem RMSE Gütekriterium zu bevorzugen ist. Ähnliche Schlussfolgerungen können auch für die langfristigen Effekte des exogenen Regressors und den Koeffizienten der zeitlich verzögerten Variablen gemacht werden.