Gasoline and Diesel Consumption for Road Transport in Spain: a Dynamic Panel Data Approach

Rosa M. González-Marrero
Departamento de Análisis Económico, IUDR (ULL)

Rosa M. Lorenzo-Alegría
Departamento de Análisis Económico, IUDR (ULL)

Gustavo A. Marrero
Departamento de Análisis Económico (ULL)
Programa de Investigación Energía y Cambio Climático Focus-Abengoa-Fedea

Colección Estudios Económicos 04-2011
Programa de Investigación Energía y Cambio Climático Fedea – Focus Abengoa

ISSN 1988-785X
www.fedea.es
Gasoline and Diesel Consumption for Road Transport in Spain: a Dynamic Panel Data Approach

Rosa M. González-Marrero
Departamento de Análisis Económico, IUDR (ULL)

Rosa M. Lorenzo-Alegría
Departamento de Análisis Económico, IUDR (ULL)

Gustavo A. Marrero
Departamento de Análisis Económico (ULL)
Programa de Investigación Energía y Cambio Climático Focus-Abengoa-Fedea

ABSTRACT:

Using a panel data set for Spanish regions between 1998 and 2006, we study the factors explaining per capita fuel consumption for road transport at the macroeconomic level. The contributions of the article are the following. First, we specify a dynamic panel data (DPD) model for gasoline and diesel consumption. Second, we properly apply estimation techniques based on the system Generalized Methods of Moments (GMM) procedure of Arellano and Bover (1995). We find that more traditional estimation procedures (pooling-OLS, the Within-Group or the first difference GMM), which might generate bias estimate in a DPD framework, produce important differences that may even change policy recommendations. Finally, we find important differences between the results for the gasoline and the diesel model. While the estimated equation correctly fits the gasoline consumption behavior, results emphasizes the need to specify a different model for aggregate diesel consumption, which must include additional determinants than those traditionally used in fuel consumption models.

JEL: R41, O13, O56

KEY WORDS: Fuel consumption, road transport, Dynamic Panel Data model, GMM estimates.

---

1 Correspondence to: Rosa M. González-Marrero, Departamento Análisis Económico, Universidad de La Laguna, Campus de Guajara, La Laguna, 38071 Santa Cruz de Tenerife; Fax: 922 317204, Tfno: 922317113, rmglzmar@ull.es

2 Gustavo A. Marrero acknowledges financial support of FOCUS-ABENGOA through Programa de Investigación Energía y Cambio Climático Focus-Abengoa-Fedea.
1. Introduction

The transport sector in Spain represented almost 40% of final energy consumption in 2007, the highest between most important European countries, and the road transport accounts for almost 80% of the total energy consumed by the sector in 2006 (Spanish Government, 2009). Moreover, it generates about 25% of total CO₂ emissions, with road transport contributing the most to said emissions. Having such a large road transport sector poses a serious roadblock for Spain to reaching the goals set by the Kyoto Protocol and the recently proposed 20/20/20 plan. Hence, fuel consumption policies in road transport and environmental policies should be intimately related.


Using a panel data framework, this article studies the factors explaining per capita fuel consumption for road transport for Spanish regions between 1998 and 2006 at the macroeconomic level. It is important to notice that these results might differ from those

---

3 According to the EC(2009) report, only Luxemburg, Malta and Cyprus show a larger share of transport in final consumption.
factors explaining fuel consumption at the microeconomic level (Puller and Grrening, 1999; Kayser 2000; Nicol, 2003). Following the arguments of Shipper et al. (1993), when data come from household surveys (micro data), the collected information is about automobile fuel consumption (mostly of private-use), and it is not about trucks or busses (mostly of public or freight transport vehicles). However, macro data of gasoline and diesel consumption do not distinguish between private-use vehicles, trucks or busses. The problem is that while most of gasoline consumption is absorbed by private-use vehicles, diesel consumption is more distributed between private-use vehicles, the public transport and freight transport vehicles (Sterner, 2007). This fact may drive that results for the diesel consumption model are very different depending on whether we use micro or macro data.

Another relevant aspect of this article is the distinction made between gasoline and diesel consumption. Polemis (2006), Zervas (2006) and Labandeira and López-Nicolás (2002), among others, have already emphasized the importance of making this distinction. For the period analyzed, this distinction is especially relevant for the case of Spain, since it is probably, along with France, one of the countries in which the dieselization process has been the most significant during this period.4

A final contribution of this paper is to write the empirical fuel consumption model in a dynamics panel data (DPD) framework.5 With regards to the estimation procedure, we use the one-step system GMM estimator proposed by Arellano and Bover (1995) and developed by Blundell and Bond (1998), which handles the endogeneity and the weak instruments problems arised in these types of models, as we will discuss in Section 2. In the growth literature, Forbes (2000), Shioji (2001), Levine et al. (2000) and Bond et al. (2001), among

---

4 For instance, the diesel to gasoline consumption ratio in road transport in Spain, which was 1.7% in 1998, had risen to 3.8% by 2006.

5 The application of panel data estimation techniques to the study of energy demand was pioneered by Balestra and Nerlove (1966) and followed by many others authors. For a discussion on panel data estimation techniques and applications, see Baltagi (2001) and Baltagi and Baldev (1992), among others.
others, have used the one-step system GMM estimator that we consider in the paper. However, there are few exceptions in the energy and transport literature that seriously consider the weakness of traditional methods in estimating DPD models. For example, Halkos (2003), Gang (2004) and Metcalf (2008) address the endogeneity problem and use the first difference GMM estimator, but this method does not consider the weak instruments problem of this procedure when time series are persistent (Blundell and Bond, 1998), which is the case for aggregate fuel consumption. Huang et al. (2008), which revisits the causal relationship between energy consumption and GDP, and Marrero (2010), which studies the relationship between emissions and energy in Europe, are exceptions that properly address both the endogeneity and the weak instruments problems by considering the system GMM approach.\footnote{In order to show the importance of considering this estimation method, we follow Blundell et al. (2000), and compare the system GMM estimates with respect to alternative, more traditional methods – pooling-OLS, the within groups, the first difference GMM of Arellano and Bond (1991).}

As explanatory variable, the diesel and the gasoline model both include traditional factors in the transport literature: the real prices of gasoline and diesel, the real regional GDP per capita, the fleet of gasoline and diesel vehicles per capita (the motorization rate by fuel type) and the total fleet divided by the total kilometers of road as a proxy for the saturation of the road network. We find that most explanatory variables are significant – and with the appropriate signs according to economic theory - in explaining the evolution of gasoline consumption. However, we find that most of these factors are not significant at explaining the recent evolution of diesel consumption between Spanish regions, which is consistent with the idea, previously exposed, that diesel consumption is heavily used by the professional transport and, as a consequence, explanatory factors can differ from those of gasoline used in
private vehicles.\textsuperscript{7} With respect to the gasoline model, we find that the alternative estimation procedures produce important differences that may even change policy recommendations, which highlights the need to use appropriate econometric techniques when dealing with DPD models.

This paper is structured as follows. Section 2 presents the DPD fuel consumption model, describes the data used in the analysis and briefly comments on the system GMM estimation approach. Section 3 estimates the gasoline and diesel model and discusses the main results. Finally, Section 4 presents the main conclusions.

2. Fuel consumption for Spanish regions: dynamic model, data and econometric method

In this section we present a dynamic framework for fuel consumption, show a Dynamic Panel Data (DPD) model for gasoline and diesel consumption, describe the data set used and comment main aspects of the estimation procedure.

2.1. A dynamics framework for fuel consumption

A common starting point in the transport literature is to assume that aggregate fuel can be characterized like any aggregate products demand, that is, as a function of a measure of real income and fuel prices.\textsuperscript{8} Specifically, using a reduced-form specification, aggregate fuel (gasoline or diesel) desired demand at time $t$, $E^*_t$, can be assumed to be a lineal logarithm function of real gasoline price, $PG_t$, real diesel price, $PD_t$, real income, $Y_t$, and an error term, $u_t$, which is assumed to be identically and independently distributed with zero mean and constant variance:

\textsuperscript{7} For Spanish regions, diesel and gasoline consumption statistics do not distinguish between passenger and freight transport use. While that could be a problem for diesel consumption, it may not be for gasoline consumption, because gasoline is majority consumed by private vehicles.

\textsuperscript{8} See Dahl, 1986; Dahl and Sterner, 1991; Sterner, (2007); Basso y Oum (2007), among many others.
\[
\ln(E^*_t) = \delta_0 + \delta_1 \ln(PG_t) + \delta_2 \ln(PD_t) + \delta_3 \ln(Y_t) + \epsilon_t,
\]

(1)

Fuel consumption and income variables would be expressed in per capita terms. The relationship between the desired demand \((E^*_t)\) and current consumption \((E_t)\) is characterized by the following partial adjustment behavior,\(^9\)

\[
\ln(E_t) - \ln(E_{t-1}) = \theta(\ln(E^*_t) - \ln(E_{t-1}))
\]

(2)

where \(E_{t-1}\) is previous demand; \(\theta\) is parameter with range as \((0,1)\) and is referred to as the coefficient of adjustment. This model states that current fuel consumption growth is proportional to the difference between the current demand and previous consumption. The coefficient \(\theta\) indicates the speed of adjustment towards the desired level of demand. The time-lags have several explanations, including habit-persistence, slow turnover of car in the vehicle stock, and the time required for automobile manufacturers to change technologies in the cars they sell (Pollack and Wales, 1981).

Combining expressions (2) and (1), and rearranging terms, we obtain the following dynamic model for aggregate fuel consumption:

\[
\ln(E_t) = \alpha_0 + \beta \ln(E_{t-1}) + \alpha_1 \ln(PG_t) + \alpha_2 \ln(PD_t) + \alpha_3 \ln(Y_t) + \epsilon_t,
\]

(3)

where: \(\alpha_0 = \theta \delta_0\), \(\beta = 1 - \theta\), \(\alpha_1 = \theta \delta_1\), \(\alpha_2 = \theta \delta_2\), \(\alpha_3 = \theta \delta_3\) and \(\epsilon_t = \theta \epsilon_t\).

This specification (3) is commonly called a lagged endogenous model, where lagged endogenous variable can be seen as representing the inertia of the system.\(^{10}\)

2.2. A Dynamic Panel Data model for gasoline and diesel consumption

Taking (3) as the point of reference, we present and estimate a DPD model for gasoline consumption, GASO (the gasoline model) and another for diesel consumption, DISL (the

---

\(^9\) The partial adjustment model is a ad hoc model and, since the seminal paper of Houthaker et al. (1974), they have been used in many works about fuel demand (Baltagi et al, 2003, Pock, 2009).

\(^{10}\) For an interesting review of the approaches and methods that have been used in automobile fuel demand, see Basso and Oum (2007).
diesel model), extended with some additional explanatory variables, such as the road network and the stock of vehicles:

\[
\ln(GASO_{it}) = \alpha_i + \beta \ln(GASO_{i,t-1}) + \lambda_t \ln(Y_{it}) + \\
+ \lambda_2 \ln(PG_{it}) + \lambda_3 \ln(PD_{it}) + \lambda_4 \ln(FLEETG_{it}) + \lambda_5 \ln(FLEETD_{it}) + \lambda_6 \ln(SAT_{it}) + \varepsilon_{it}, \tag{4}
\]

\[
\ln(DISL_{it}) = \alpha_i + \beta \ln(DISL_{i,t-1}) + \lambda_t \ln(Y_{it}) + \\
+ \lambda_2 \ln(PG_{it}) + \lambda_3 \ln(PD_{it}) + \lambda_4 \ln(FLEETG_{it}) + \lambda_5 \ln(FLEETD_{it}) + \lambda_6 \ln(SAT_{it}) + \varepsilon_{it}, \tag{5}
\]

where FLEETG and FLEETD are the existing fleet of gasoline and gasoil vehicles per capita (motorization rate), respectively, and SAT is the total number of vehicles divided by total kilometers of road, which is a proxy of the saturation of the road network. Fixed factors \( \alpha_i \) are time-invariant and inherent to each region, and are not observed or not included in the model, such as geographical, social or local policy regional aspects or initial energy efficiency. Dynamic variables \((GASO_{i,t-1}\) and \(DISL_{i,t-1}\)) control for conditional convergence across regions in terms of fuel consumption, as it is standard in the dynamic literature. Finally, \( \varepsilon_{it} \) encompasses effects of a random nature which are not considered in the model, and it is assumed to have a standard error component structure:

---

11 Given the close relationship between energy and income, the dynamic specification is similar to that used in the convergence-growth literature (Barro and Sala-i-Martin, 1995, among many others). Álvarez et al. (2005) and Marrero (2010) have adapted this dynamic approach to a economic-pollution model.

12 This is one of the alternatives that can be used for analyzing the demand of fuel. It consists of using an equation of fuel demand where some measure of the stock of vehicles is explicitly include (ie, Banfi et. al., 2005; Kayser, 2000; and Puller and Greening, 1999). An alternative procedure is to employ a system of simultaneous equations in which the demand for vehicles and the demand for fuel are analyzed jointly (ie. Belhaj, 2002 and Chandarisi, 2006).

13 In addition to the dynamic term, the other explanatory variables assumed in (4) and (5) are among those traditionally considered as indicators for characterizing the behavior of the road transportation sector [Eltony (1993), Bentzen (1994), Kirby et al. (2000), Alves and Bueno (2003), Polemis (2006)]. Moreover, we also distinguish between the total diesel and gasoline fleet, as proposed by Pock (2009).

14 Fixed effects, such as differences in the initial energy efficiency, would be omitted in a standard OLS pool regression, resulting in bias estimates (i.e., the \( \beta \) estimates is upward bias.) See Anderson and Hsiao (1982) and Hsiao (1986) for more details about this point.

15 Indeed, the interpretation of equation (1) depends on the level of \( \beta \). A \( \beta \) smaller than one is consistent with the conditional convergence hypothesis, which means that regions relatively close to their steady-state per capita fuel consumption levels will experience a slowdown in their consumption growth. In this case, \( \alpha_i \) and all explanatory variables affect to the steady-state the fuel consumption of region \( i \) is converging to. On the other hand, if \( \beta \) is greater than one, there is no convergence effect and \( \alpha_i \) and all regressors would measure differences in steady-state energy consumption growth rates. Estimated \( \beta \) will be lower than one in all cases, hence we will focus on the conditional convergence interpretation.
We also consider a common assumption in DPD models [Arellano and Bond (1991)], which is that $GASO_{it}$ and $DISL_{it}$ are predetermined,

\[ E[GASO_i | e_{it}] = E[DISL_i | e_{it}] = 0, \text{ for } i = 1,...,N \text{ and } t = 2,...,T \]

2.3. Data

Table 1 shows annual growth rates between 1998 and 2006 for each variable used in models (4) and (5) at a regional level.

**TABLE 1 ABOUT HERE**

Per capita gasoline consumption fell between 1998 and 2006 in all Spanish regions at an average rate of 4.1%, while per capita diesel consumption increased for each region at a higher average rate of 5.7%. During this period, gasoline and diesel real prices increased by 2.8% and 4.8% in Spain, respectively. Although significant differences were noted in their growth rates in the time dimension, differences between regions are very small. The large time-volatility of fuel prices resulted from Spain’s enormous (nearly 100%) dependence on foreign oil, on the important fluctuations in the euro/dollar exchange rate, and on changes in fuel taxes and their repercussions on the final price.\(^{16}\) The data on per capita GDP growth rate showed an annual increase of 2.6% nationally, varying between 1.6% in Valencia and 3.8% in Extremadura. Overall, per-capita GDP growth showed a notable regularity among the different regions, as evidenced by the generalized slowdown between 2001 and 2003-2004 and the subsequent recovery until 2006 for most of the regions.\(^ {17}\)

\(^{16}\) For example, in 2001 the price in dollars of a barrel of Brent crude fell 14%, while in 2004 and 2005 it rose by 33% and 42%, respectively.

\(^{17}\) Spanish regional GDP is built on the basis of regional physical, economic indicators, such as retail sales, industrial production index, car sales, overnight stay of tourists, consumption of cement, etc, which are highly related to the domestic level of activity in each Region. De la Fuente (2002), among others, has used the Spanish regional GDP dataset to study the source of convergence in Spain. Petrol prices are retail prices and deflated by each regional CPI. For more details about how petrol prices are determined in Spain, see Perdiguero (2006). GASO and DISL are measured in Kilo-tones per inhabitant.
The motorization rates by type of fuel (gasoline fleet pc and diesel fleet pc) show a trend similar to those of their associated fuel consumption series. For most regions, it decreased for gasoline vehicles (2.0% in Spain), while it increased for diesel (9.2% in Spain). As is the case with fuel consumption, the dieselization process has changed the vehicle fleet composition in Spain: the ratio of diesel to gasoline vehicles, which was 43% in 1998, had risen to 105% by 2006. However, the intensity of this substitution process varied greatly depending on the Region. Thus, Madrid is the region that experienced the most significant substitution process from gasoline to diesel vehicles, with a decline in the motorization rate for gasoline vehicles of 4.2% and an increase for diesel of 10.6%. Other regions show a different pattern. For example, in Extremadura, the motorization rate for gasoline vehicles remained relatively stable, while that of diesel experienced the largest increase, within all Spanish regions, of 12% per year. As for the number of vehicles versus road network kilometers (the saturation level of road), it increased in all regions, showing the highest average annual growth rate in Castilla La Mancha (5.4%) and the lowest in Aragon (2.1).

2.4. Estimation procedure

Traditional procedures for estimating a DPD models like (4) and (5) (i.e., fixed or random effects methods or pooling-OLS) are known to be unsuitable [Anderson and Hsiao (1982); Hsiao (1986)]. Holtz-Eakin et al. (1988) and Arellano and Bond (1991) propose an alternative approach, where first differences in the regression equation are taken to remove unobserved time-invariant country specific effects and then particular moment conditions for lagged variables are exploited to find a set of instruments and construct a GMM-based estimator. Their GMM approach (GMM-DIF) allows us to handle endogeneity problems. However, the GMM-DIF approach shows important bias problems in small sample when variables are highly persistent, which is the case of aggregate fuel consumption and other
macroeconomic variables. Under these circumstances, the instruments used in the GMM-DIF estimator have proven to be weak and the first difference estimator is poorly behaved, and Arellano and Bover (1995) and Blundell and Bond (1998) proposed an alternative GMM procedure which might overcome the weak instruments problem. This procedure estimates a system of equations in both first-differences and levels, where the instruments in the level equations are lagged first differences of the variables. In this paper we use the one-step system GMM estimator. In contrast to the two-step version, the one-step GMM estimator has standard errors that are asymptotically robust to heteroskedasticity and have been found to be more reliable for finite sample inference.18

3. Gasoline and diesel model results

Our goal of this section is twofold. First, we want to emphasize the importance of considering an appropriate quantitative approach when estimating a dynamic fuel consumption model; second, we point out the differences between the gasoline and the diesel model results.

3.1. Alternative estimation procedures

We first find evidence supporting the good properties of the system GMM estimates. We follow the practical rule proposed by Blundell et al. (2000), and compare the results of alternative estimation methods for the gasoline and the diesel model: the OLS pooling estimates (OLS-POOL), the Within Group estimates (WG), the first-difference GMM approach (GMM-DIF) of Arellano and Bond (1991) and the system GMM method (GMM-SYS). For the GMM-based methods, we employ the one-step GMM estimator, with

---

18 See Blundell and Bond (1998), Blundell et al. (2000) or Baltagui (2001), among others. See the technical appendix for more details about this point.
heteroskedasticity-consistent asymptotic standard errors.19 Tables 2 and 3 show the results for the gasoline and diesel model, respectively, for all these alternative methods. The p-value of the t-significance test associated with each parameter is shown. We also show standard specification tests for each model. Notice that the Haussman test rejects the null hypothesis of random effects at any standard level of significance. For any GMM-based estimates, we show the $m_1$ and the $m_2$ tests and conclude that moment conditions underlying GMM estimates seem to be robustly supported.20

**INSERT TABLE 2 AND 3 ABOUT HERE**

Based on the results shown in Tables 2 and 3 and following the practical rule proposed by Blundell et al. (2000), OLS-POOL seems to give an upward-biased estimate of the $\beta$ coefficient (0.853 for the gasoline and 0.998 for the diesel model), while WG appears to give a downward-biased estimate of this coefficient (0.387 for the gasoline and 0.495 for the diesel model). Using GMM-DIF, the $\beta$ coefficient is barely lower than the WG estimates, suggesting the possibility of important finite sample bias due to the weak instruments problem [Blundell and Bond (1998)]. This comparison also highlights how the estimated coefficients of the remaining regressors, which are our main interest, differ among the alternative procedures. Hence, using a method resulting in bias estimates (the OLS-POOL, WG or the GMM-DIF) might lead to misleading conclusions. For example, the coefficients associated with the network saturation variable in the gasoline and diesel model are not

---

19 For a given cross-sectional sample size, the use of too many instruments in models with endogenous regressors may result in seriously biased estimates (Álvarez and Arellano, 2003; Arellano and Bond, 1998). Hence, even when computing speed is not an issue, these authors recommend not using the entire series history as instruments. We use instruments up to t-3. Including more lags does not change results significantly.

20 The most frequently used tests to validate the assumptions underlying GMM methods are the $m_1$, $m_2$ and Sargan tests. The $m_1$ and $m_2$ tests are based on the standardized average residuals autocovariance, which are asymptotically $\text{N}(0,1)$ distributed under the null hypothesis of no autocorrelation. The Sargan test, in contrast, is distributed chi-squared with degrees of freedom equal to the number of moment restrictions minus the number of parameters, estimated under the null hypothesis that moment conditions are valid. However, the Sargan test is less meaningful since it requires that the error terms be independently and identically distributed, which is not expected in our case. Hence, we will consider primarily the $m_1$ and $m_2$ tests.
significant under the WG and GMM-DIF estimates, while it is significant under the GMM-SYS procedure; for the diesel model, the per capita GDP variable is significant under the WG, while it is not under the GMM-SYS and OLS-POOL; the magnitude of the estimated price-elasticities are smaller under the GMM-SYS than under the WG estimates in the diesel model.

In summary, this comparison suggests that the WG estimates are severely biased, that there exists a problem with weak instruments and hence that the GMM-DIF is biased similarly to WG, and that the GMM-SYS approach is a convenient way to overcome the weak instruments problem. This conclusion is an important contribution of the paper, and not always properly considered in the related literature.

3.2. System-GMM estimation results

Focusing on the second goal of the paper, we compare the results of the gasoline and diesel models, and we find important differences in the magnitude and significance of the variables. We will mainly focus our attention on the one-step GMM-SYS estimates from now on.

In general, we find that the coefficients associated with the explanatory variables are less significant in the diesel model than in the gasoline model. This result could be due to different reasons. Following Sterner (2007), the diesel is used heavily in professional transport (buses, heavy trucks, etc.) with different explanatory factors than those used for private use. Moreover, the dieselization process that has taken place in Spain since 1994, aproximatly, has resulted in diesel consumption being exposed to important regulatory factors, which might have modified expected results (Burguillo et al., 2009). Likewise, since fuel consumption data include both passenger and freight transport, and heavy trucks mainly

---

21 For the diesel model, we try with alternative specifications, considering variables such as the percentage of heavy trucks in the total fleet, or the percentage of buses. But these alternative specifications also fail in explaining the recent evolution of diesel consumption in Spain, being in line with the results obtained by Baltagi and Griffin (1983).
uses diesel, the diesel model must distinguish between consumption in the passenger and freight transport sector. But, diesel consumption statistics for Spanish regions do not distinguish between consumption in the passenger and the freight transport sectors. Moreover, there are also specific technical characteristics of diesel vehicles, such as a different engineering development, which should also lead to introduce additional diesel-specific variables in the diesel model. Therefore, it would be necessary to further study the characteristics of diesel consumption, which goes beyond the objectives of this work, but they represent a promising extension for future research.

For the diesel and gasoline model, we next comment the estimations for each variable. The parameters estimated for the $DISL_{t-1}$ variable and the $GASO_{t-1}$ are positive and lower than one at the 1% level of significance. The estimate is 0.867 for the diesel model and 0.558 for the gasoline model. Hence, the evidence for conditional convergence is significant in both cases, though it is greater for the gasoline case. The estimates indicate that the rate of conditional convergence for the per capita fuel consumption ratio is about 13% for diesel consumption and about 44% for gasoline.

A common result in the gasoline and diesel model is that per capita GDP is not significant in explaining per capita fuel consumption. From Section 2.3, we showed how per capita fuel consumption and GDP experienced important increases from 1998 to 2006. However, it seems that the former evolved independently of per capita GDP when, for a given period, we compare different regions. For example, we find that regions with different per capita GDP levels shared similar fuel consumption patterns. This is the case of the Basque Country, with

---

22 A similar result is found in Pock (2009), which estimates a demand function for aggregate gasoline, and finds that, in some models, all coefficients are significant with the exception of the income term. He points out that this result may be due to multicollinearity problems. For the case of diesel demand, Burguillo et al. (2009) find the same result: the coefficient associated to real income was negative, but non-significant. Baltagi and Griffin (1997) obtain that the income elasticities are frequently insignificant in the long run. There exits other articles in the literature which estimates a model for aggregate energy consumption or CO2 emissions whose income-elasticity estimation is very low, and even negative in some cases [Schmalensee et al. (1998), Judson et al. (1999), Holtz-Eakin and Selden (1995) or Marrero (2010)]. Baltagi and Griffin (1997) obtain the income elasticities in long run ar
a large per capita GDP, Castilla and Leon, with intermediate per capita GDP, and Extremadura, with one of the smallest per capita real GDP in Spain. However, they have experienced a similar increase in per capita fuel consumption for the analyzed period (between 3.6% and 3.9%).

Regarding other papers studying the relationship between gasoline demand and income, Dahl and Sterner (1991) showed that short-term income elasticity on gasoline demand varied between 0.30 and 0.52 in the different studies they considered. However, notice that our GDP elasticity under the WG estimate was significant and about 0.44 for diesel, which is indeed consistent with the Dahl-Sterner range; this, however, is the result of a bias estimation procedure, as discussed above. With this example, we are not claiming that Dahl-Sterner estimates are wrong. In fact, differences between their estimates and ours may only be due to differences in the data sample used. We are just stressing the importance of considering an appropriate estimation approach for handling fuel consumption models, because, otherwise, results could lead to misleading conclusions.

Regarding the real price of fuel, the GMM-SYS procedure estimate for its elasticity is negative and significant for the case of gasoline, though its magnitude is well below one (-0.29). This result confirms the evidence that the elasticity of the demand price for gasoline is low in the short term, as verified by, among others, Kayser (2000) with data for the United States and Baltagi and Griffin (1997) with data for OECD countries. Results at the international level place the price-elasticity in the -0.2 and -0.3 range [Dahl and Sterner (1991)], which is consistent with our results. This result indicates that fuel demand is highly

---

23 We are aware that the short time dimension of the data can influence this result. Also, the transport network is very different across regions as it is the availability of public transport services. The response to income of consumers in predominantly rural areas could be different than in predominantly urban areas, at least because of the substitution possibilities. Once again, we do not have data that distinguish between consumption in rural and urban areas. However, the importance of this result deserves further research, which goes beyond the purpose of this paper.

24 More recently, studies such as that by Koshal et al. (2007) gave values of 0.29. For a detailed review, see Graham and Glaister (2002) and Goodwin et al. (2004).
inelastic, at least at current price levels. Moreover, it supports a result commonly discussed in the literature: fuel taxes are convenient for increasing fiscal revenues, but they are not effective enough to reduce fuel consumption [Kirby et al. (2002)].

In addition, the real price of diesel is significant in explaining short-term changes in per capita gasoline consumption, although its parameter is small (0.21). This last result is also consistent with the transport literature due to the strictness that exists in substituting types of vehicles in the short term [Polemis (2006)]. We should emphasize that the recent intensive switch from gasoline to diesel vehicles has been basically due to regulatory reasons (dieselization) rather than to a change in the price of the alternative fuel. For the case of the diesel consumption model, neither its own price nor the price of gasoline is significant. As commented above, differences between the markets of the gasoline (most for passenger transport use) and the diesel (passenger and freight transport are important) could explain this result for the diesel model. Once again, this finding also highlights the need to consider different models for gasoline and diesel consumption, and to go further in the research of the determinants of diesel consumption in Spain.

The remaining variables are specific to the road transport sector and include relevant aspects that can affect fuel consumption. The per capita diesel and gasoline fleet variables show the motorization rate for each type of fuel vehicle. For the gasoline model, the coefficient of the per capita gasoline fleet variable is highly positive and significant (0.64), while the coefficient of the per capita diesel fleet is negative but much smaller in magnitude (-0.08). Regarding these variables, results for the diesel model are controversial, as with other variables.25

The coefficients of the measure of the degree of saturation of the road network (the ratio between total fleet and road network) are negative and significant in both models. Moreover,

25 Further investigation on this important topic would constitute a prominent extension of this paper.
their estimates are similar: -0.048 for diesel and -0.059 for gasoline. The fact that estimated coefficients are similar in both models is a clear indication that road saturation affects both diesel and gasoline vehicles in a similar way. This result suggests that a reduction in road congestion promotes mobility, which may induce an increment in per capita fuel consumption.26

5. Final Remarks

This paper has analyzed the factors explaining the fuel consumption for road transport in Spain in a dynamic panel data framework in a macroeconomic perspective. Two features on this study are the use of a balanced panel using regional data and the distinction between gasoline and diesel. As explanatory variables, we considered real GDP and fuel prices, which are the most commonly used in the related literature, as well as other relevant and different variables, such as the motorization rate and the congestion of the road network.

A novelty in this paper is the use of a the one-step system GMM estimation method of Arellano and Bover (1995) and Blundell and Bond (1998), which has been shown to solve many of the problems that arise in traditional panel data procedures. When compared with the system GMM results, we found that traditional panel data estimation procedures (the within-group estimates, OLS-pooling or the first difference GMM approach of Arellano and Bond, 1991) might exhibit significantly biased estimates, which might even change policy recommendations.

For the sample used, we found that most explanatory variables are significant in explaining the evolution of gasoline consumption in Spain, while diesel consumption was found to have

26 As noted by Goodwin (1996), improving the infrastructure has an induced effect on the demand for transport. Moreover, Cervero and Hansen (2002) provided empirical evidence of the existence of a direct relationship between investing in roads and the demand for transport, namely that an expansion of infrastructure generates demand for transport, which in turn induces the creation of infrastructure.
less dependence on most of these factors. These conspicuous differences between the results for the gasoline and diesel models imply that the conclusions derived from an analysis of overall fuel consumption could be misleading.

Our estimates confirm that the price elasticity of demand for fuel consumption is low – even negligible for diesel - in the short term, which supports the view that the policy of taxing fuel has little effect on reducing fuel consumption. Our results are also consistent with the evidence of small cross price elasticities for gasoline and diesel consumption. Another important finding of this work is the negative and significant relationship between the degree of saturation of the road network and both types of per capita fuel consumption. This result shows that reducing road network saturation – i.e., by increasing the road network -, could promote mobility and a higher transport demand ("induced travel demand"), which can favour higher levels of fuel consumption.

Several are the implications of our research. On the one hand, we emphasize the need to revisit DPD fuel consumption results obtained with traditional procedures, and show the relevance of considering a suitable estimation method; ant the other, it necessary to investigate the determinants of diesel consumption using a different model than that used for gasoline consumption, which is a promising extension of this paper.
REFERENCES


• Burguillo-Cuesta M., Jorge-Garia I. and Romero D. Dieselization of passenger cars in the EU-15 in the nineties: environmental implications for transport policy. 8th International Conference of the European society for Ecological Economics, June 2009, Slovenia.


• Goodwin P. Empirical evidence on induced traffic. Transportation 1996, 23(1), 35-54.


• Mazzarino M. The economics of the greenhouse effect: evaluating the climate change impact due to the transport sector in Italy. Energy Policy 2000, 28, 957-966.
• Perdiguero G. J. Dinámica de precios en el mercado español de gasolina: un equilibrio de colusión tácita, Documento de Trabajo de Funcas 2006, nº 253.
TABLE 1: Lists the average annual variation rates for all variables for the period 1998-2006

<table>
<thead>
<tr>
<th>REGION</th>
<th>consumption gasoline pc</th>
<th>consumption diesel pc</th>
<th>gasoline real price</th>
<th>diesel real price</th>
<th>per capita GDP</th>
<th>gasoline fleet pc</th>
<th>diesel fleet pc</th>
<th>total fleet/road network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andalucia</td>
<td>-3.44</td>
<td>6.19</td>
<td>2.78</td>
<td>5.09</td>
<td>2.97</td>
<td>-1.48</td>
<td>11.47</td>
<td>4.84</td>
</tr>
<tr>
<td>Aragón</td>
<td>-3.90</td>
<td>6.25</td>
<td>2.63</td>
<td>4.78</td>
<td>2.79</td>
<td>-1.77</td>
<td>9.64</td>
<td>2.14</td>
</tr>
<tr>
<td>Asturias</td>
<td>-3.44</td>
<td>5.51</td>
<td>3.01</td>
<td>5.18</td>
<td>2.99</td>
<td>-1.36</td>
<td>8.23</td>
<td>2.39</td>
</tr>
<tr>
<td>Cantabria</td>
<td>-3.78</td>
<td>6.67</td>
<td>2.93</td>
<td>5.03</td>
<td>3.07</td>
<td>-1.30</td>
<td>9.46</td>
<td>4.28</td>
</tr>
<tr>
<td>Castilla y León</td>
<td>-3.33</td>
<td>5.91</td>
<td>2.72</td>
<td>4.90</td>
<td>3.26</td>
<td>-1.11</td>
<td>9.78</td>
<td>2.94</td>
</tr>
<tr>
<td>Castilla La Mancha</td>
<td>-4.76</td>
<td>5.87</td>
<td>3.03</td>
<td>5.23</td>
<td>2.14</td>
<td>-1.71</td>
<td>10.43</td>
<td>5.40</td>
</tr>
<tr>
<td>Catalonia</td>
<td>-5.16</td>
<td>4.11</td>
<td>2.71</td>
<td>4.85</td>
<td>2.03</td>
<td>-2.53</td>
<td>8.33</td>
<td>2.85</td>
</tr>
<tr>
<td>Valencia</td>
<td>-4.33</td>
<td>5.13</td>
<td>3.01</td>
<td>5.05</td>
<td>1.64</td>
<td>-2.33</td>
<td>7.85</td>
<td>3.72</td>
</tr>
<tr>
<td>Extremadura</td>
<td>-2.89</td>
<td>6.96</td>
<td>3.03</td>
<td>5.06</td>
<td>3.75</td>
<td>-0.82</td>
<td>12.01</td>
<td>3.94</td>
</tr>
<tr>
<td>Galicia</td>
<td>-3.24</td>
<td>3.74</td>
<td>3.00</td>
<td>4.99</td>
<td>2.99</td>
<td>-1.32</td>
<td>7.94</td>
<td>3.18</td>
</tr>
<tr>
<td>Madrid</td>
<td>-6.72</td>
<td>6.42</td>
<td>2.91</td>
<td>3.86</td>
<td>2.13</td>
<td>-4.15</td>
<td>10.63</td>
<td>3.19</td>
</tr>
<tr>
<td>Murcia</td>
<td>-4.62</td>
<td>6.58</td>
<td>2.29</td>
<td>4.54</td>
<td>1.98</td>
<td>-2.42</td>
<td>8.94</td>
<td>4.82</td>
</tr>
<tr>
<td>Navarre</td>
<td>-3.33</td>
<td>6.04</td>
<td>2.64</td>
<td>4.60</td>
<td>2.65</td>
<td>-2.62</td>
<td>7.38</td>
<td>2.74</td>
</tr>
<tr>
<td>Basque Country</td>
<td>-3.99</td>
<td>6.28</td>
<td>2.49</td>
<td>4.59</td>
<td>3.26</td>
<td>-2.01</td>
<td>7.62</td>
<td>2.86</td>
</tr>
<tr>
<td>La Rioja</td>
<td>-4.18</td>
<td>4.07</td>
<td>2.36</td>
<td>4.43</td>
<td>1.72</td>
<td>-2.66</td>
<td>8.08</td>
<td>3.26</td>
</tr>
<tr>
<td>SPAIN</td>
<td>-4.07</td>
<td>5.72</td>
<td>2.77</td>
<td>4.01</td>
<td>2.63</td>
<td>-1.97</td>
<td>9.19</td>
<td>3.50</td>
</tr>
</tbody>
</table>

Table 2: Estimates of the gasoline DPD model

<table>
<thead>
<tr>
<th></th>
<th>Traditional methods</th>
<th></th>
<th></th>
<th></th>
<th>GMM methods</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS-POOL</td>
<td>WG-Fixed effects</td>
<td>GMM-DIF</td>
<td>GMM-SYS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>estimates</td>
<td>p-value</td>
<td>estimates</td>
<td>p-value</td>
<td>estimates</td>
<td>p-value</td>
<td></td>
</tr>
<tr>
<td>Lag of gasoline</td>
<td>0.853</td>
<td>0.000</td>
<td>0.387</td>
<td>0.000</td>
<td>0.343</td>
<td>0.010</td>
<td>0.558</td>
</tr>
<tr>
<td>consumption pc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gasoline real price</td>
<td>-0.417</td>
<td>0.000</td>
<td>-0.375</td>
<td>0.000</td>
<td>-0.377</td>
<td>0.000</td>
<td>-0.292</td>
</tr>
<tr>
<td>Diesel real price</td>
<td>0.175</td>
<td>0.021</td>
<td>0.181</td>
<td>0.007</td>
<td>0.186</td>
<td>0.000</td>
<td>0.212</td>
</tr>
<tr>
<td>Real GDP pc</td>
<td>0.009</td>
<td>0.622</td>
<td>0.241</td>
<td>0.244</td>
<td>0.293</td>
<td>0.254</td>
<td>-0.011</td>
</tr>
<tr>
<td>Gasoline fleet pc</td>
<td>0.163</td>
<td>0.013</td>
<td>0.640</td>
<td>0.000</td>
<td>0.707</td>
<td>0.000</td>
<td>0.639</td>
</tr>
<tr>
<td>Diesel fleet pc</td>
<td>-0.065</td>
<td>0.002</td>
<td>-0.264</td>
<td>0.006</td>
<td>-0.284</td>
<td>0.009</td>
<td>-0.083</td>
</tr>
<tr>
<td>Total fleet / Road</td>
<td>-0.023</td>
<td>0.002</td>
<td>0.006</td>
<td>0.957</td>
<td>0.005</td>
<td>0.964</td>
<td>-0.059</td>
</tr>
<tr>
<td>Network</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0.959</td>
<td></td>
<td>0.946</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hausman, random</td>
<td></td>
<td></td>
<td>53.68</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>effect test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m1-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2.564</td>
<td>0.010</td>
<td>-3.368</td>
</tr>
<tr>
<td>m2-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.022</td>
<td>0.307</td>
<td>-0.510</td>
</tr>
</tbody>
</table>

Note: ‘WG’ is Within Groups estimation, OLS-POOL is OLS applied to the entire pool of data. For GMM estimates, we take as instruments the lagged levels of y and the endogenous regressors dated t-2 and earlier and the pre-determined regressors dated t-1 and earlier. We use the lagged difference of y and all regressors dated t-1 as additional instruments in the system GMM estimation. For the DIF-GMM and SYS-GMM, we report their one-step estimations. The null of the Hausman test is the existence of random effects. The null of the m1 and m2 test is the absence of first- and second-order serial correlation between...
regressors and residuals, respectively. Number of regressors: 8; number of cross sections: 15 (all Spanish regions except Ceuta and Melilla, Balears and Canary islands); number of time periods: 9 (1998-2006); number of time periods adjusted for GMM-DIF and GMM-SYS: 6 (2001-2006).

Table 3: Estimates of the Diesel DPD model

<table>
<thead>
<tr>
<th></th>
<th>OLS-POOL</th>
<th>WG-Fixed effects</th>
<th>GMM-DIF</th>
<th>GMM-SYS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>estimates</td>
<td>p-value</td>
<td>estimates</td>
<td>p-value</td>
</tr>
<tr>
<td>Lag of gasoil consumption pc</td>
<td>0.998</td>
<td>0.000</td>
<td>0.495</td>
<td>0.000</td>
</tr>
<tr>
<td>Gasoline real price</td>
<td>-0.113</td>
<td>0.285</td>
<td>-0.101 0.265</td>
<td>-0.116</td>
</tr>
<tr>
<td>Diesel real price</td>
<td>-0.049</td>
<td>0.502</td>
<td>-0.083 0.205</td>
<td>-0.075</td>
</tr>
<tr>
<td>Real GDP pc</td>
<td>-0.004</td>
<td>0.852</td>
<td>0.445</td>
<td>0.030</td>
</tr>
<tr>
<td>Gasoline fleet pc</td>
<td>0.012</td>
<td>0.732</td>
<td>-0.010 0.920</td>
<td>0.052</td>
</tr>
<tr>
<td>Diesel fleet pc</td>
<td>-0.047</td>
<td>0.088</td>
<td>0.199</td>
<td>0.059</td>
</tr>
<tr>
<td>Total fleet / Road Network</td>
<td>-0.001</td>
<td>0.947</td>
<td>-0.019</td>
<td>0.852</td>
</tr>
<tr>
<td>R2</td>
<td>0.986</td>
<td>--</td>
<td>0.954</td>
<td>--</td>
</tr>
<tr>
<td>Hausman, random effect test</td>
<td>--</td>
<td>--</td>
<td>37.206</td>
<td>0.000</td>
</tr>
<tr>
<td>m1-test</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>m2-test</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Note: See Note on Table 2.
Appendix: System GMM Estimation of DPD Models

The fixed effect treatment leads to the well known within group estimator (WG) [Hsiao (1986)], which has been applied to multiple frameworks. However, the within transformation in a panel dynamic model implies a correlation of order $1/T$ between the lagged dependent term $y_{it-1}$ and the error $\varepsilon_{it}$, which leads to biased estimates [Anderson and Hsiao (1981); Hsiao (1986)]. In addition, a fuel consumption equation such as (4) or (5) suffer from endogeneity and, maybe, from measurement errors problems. For instance, gasoline and diesel prices are jointly determined with gasoline and diesel consumption. The WG method neither properly handle these problems.

Holtz-Eakin et al. (1988) and Arellano and Bond (1991), among others, point out these problems and propose a GMM-based estimation approach. The current response of these authors is to first difference the model equation, remove the fixed effect term and then use the following orthogonally conditions, which, under assumptions A1 and A2 (see Section 2 of the paper), are valid for the first difference model:

$$E[y_{it-1}, \Delta \varepsilon_t] = 0, \ t = 3, \ldots, T \ and \ 2 \leq s \leq t - 1, \ for \ i = 1, \ldots, N, \ (3)$$

Regressors in the gasoline and diesel models are either endogenous (prices, GDP and registrations) or pre-determined (the road network and the vehicle fleets ratios). Assuming a similar condition to A2 but for the regressors in $X$,

$$A3: \ E[x_{it}, \varepsilon_t] = 0, \ for \ i = 1, \ldots, N \ and \ t = 2, \ldots, T,$$

we have additional $0.5(T-1)(T-2)$ moment conditions,

27 In the case of exogenous regressors, additional moment conditions are available. See Arellano and Bond (1991) for more detail about this point.
\[ E[x_{it-2}\Delta e_i] = 0, \ \text{for each endogenous regressor}, \text{and another (T-1)(T-2) moment conditions}, \]

\[ E[x_{it-2}\Delta e_i] = 0, \ \text{for each pre-determined regressor}. \]

For the case of K=1 and endogenous regressor, we have a total of Nd=(T-1)(T-2) moment conditions. Conditions in (3) and (4.a) can be written more compactly as

\[ E[Z_i\Delta e_i] = 0, i = 1, \ldots, N, \]

where \( Z_i \) is a \((T-2)\times Nd\) matrix, given by

\[
Z_{iDIF} = \\
\begin{pmatrix}
y_{i1} & x_{i1} & 0 & \ldots & 0 \\
0 & y_{i1} & x_{i1} & x_{i2} & \ldots & \ldots & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \ldots & \ldots & \ldots & y_{i1} & \ldots y_{iT-3} x_{i1} \ldots x_{iT-3} \\
0 & 0 & \ldots & \ldots & \ldots & 0 & \ldots y_{i1} \ldots y_{iT-2} x_{i1} \ldots x_{iT-2}
\end{pmatrix}
\]

These are the moment conditions exploited by the standard first-difference GMM estimator (GMM-DIF).

However, the GMM-DIF estimator has been found to have large finite sample bias and poor precision when the set of instruments is weak [Blundell and Bond (1998).], which is the case of our fuel consumption model. To deal with this problem, Arellano and Bover (1995) and Blundell and Bond (1998) assume additional conditions to \( A1, A2 \) and \( A3 \).

\[ ^{28} \text{For ease of exposition, we restrict notation to the case of only one endogenous regressor (i.e., K=1). The extension of the general case is straightforward.} \]
A4: \( E[\eta_i \Delta y_{i2}] = 0, \ i = 1, ..., N \),

A5: \( E[\eta_i \Delta x_{i2}] = 0, \ i = 1, ..., N \)

which allows the use of other \( 2 \cdot (T-2) \) moment conditions for a model in levels,

\[
E[u_{it} \Delta y_{i,t-1}] = 0, \ t = 3, ..., T, \quad (7)
\]

\[
E[u_{it} \Delta x_{i,t-1}] = 0, \ t = 3, ..., T \quad (8)
\]

which stay informative even for high persistent time series. Their proposal consists in a stacked system of all \( (T-2) \) equations in first differences and all \( (T-2) \) equations in levels for \( t=3,4, ..., T \), and combine restrictions (3), (4), (7) and (8) to form a linear system GMM estimator (GMM-SYS) based on the following instrument matrices:

\[
Z_i = \begin{pmatrix} Z_{i\text{DIFF}} & 0 \\ 0 & Z_{i\text{SYS}} \end{pmatrix}, \quad (9)
\]

with \( Z_{i\text{DIFF}} \) given by (6) and \( Z_{i\text{SYS}} \) by

\[
Z_{i\text{SYS}} = \begin{pmatrix}
\Delta y_{i2} & \Delta x_{i2} & 0 & ... & 0 \\
0 & \Delta y_{i3} & \Delta x_{i3} & ... & ... \\
. & ... & ... & ... & ... \\
. & ... & ... & \Delta y_{iT-2} & \Delta x_{iT-2} \\
0 & 0 & ... & 0 & \Delta y_{iT-1} \Delta x_{iT-1}
\end{pmatrix};
\]
Monte Carlo analysis has shown that using GMM-SYS greatly reduces the finite sample bias and improves the precision of the estimator in presence of weak instruments. The linear GMM estimator is given by $\left( \bar{X}' Z \bar{H}_N Z' \bar{X} \right)^{-1} \left( \bar{X}' Z \bar{H}_N Z' \bar{Y} \right)$, where, for the GMM-DIF,

$$\bar{X} = \begin{bmatrix} \Delta X_1 \\ \vdots \\ \Delta X_N \end{bmatrix}; \bar{Y} = \begin{bmatrix} \Delta Y_1 \\ \vdots \\ \Delta Y_N \end{bmatrix}; Z = \begin{bmatrix} Z_{1,DIF} \\ \vdots \\ Z_{N,DIF} \end{bmatrix},$$

while for the GMM-SYS case,

$$\bar{X} = \begin{bmatrix} \Delta X_1 \\ \vdots \\ \Delta X_N \\ X_1 \\ \vdots \\ X_N \end{bmatrix}; \bar{Y} = \begin{bmatrix} \Delta Y_1 \\ \vdots \\ \Delta Y_N \\ Y_1 \\ \vdots \\ Y_N \end{bmatrix}; Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_N \end{bmatrix}.$$

For each GMM-FIF and GMM-SYS case, two different choices of $H_N$ result in two different GMM estimators. The one-step estimator sets

$$H_{N,GMM 1} = \left( \frac{1}{N} \sum_{i=1}^{N} Z_i' H Z_i \right)^{-1},$$

where the $H$ matrix is a (T-2) square matrix with 2’s on the main diagonal, -1 on the first off-diagonals and zeros elsewhere. The two-step GMM estimator uses

$$H_{N,GMM 2} = \left( \frac{1}{N} \sum_{i=1}^{N} Z_i' \Delta \tilde{a}_i \Delta \tilde{a}_i' Z_i \right)^{-1},$$

Indeed, Blundell and Bond (1998) and Bond et al. (2001) shows that an optimal combination of differenced and level equations allow us to calculate a GMM estimator using the full set of linear moment conditions implied by assumptions A1-A5.
where estimated residuals are from a consistent one-step estimator (i.e., the one-step), which is an asymptotically efficient GMM estimator.

Under spherical disturbances, GMM1 and GMM2 are equivalent in the first-difference model. Otherwise, GMM2 is more efficient. However, Monte Carlo studies have shown that the efficiency gains of the two-step estimator are generally small. It also has the problem of converging to its asymptotic distribution relatively slowly. Hence, in finite samples, its variance-covariance matrix can be seriously biased. Moreover, for the case where the total number of instruments is large relative to the cross-section dimension of the panel, there may be computational problems in calculating the two-step estimates and serious estimation errors may arise [Arellano and Bond (1998); Doran and Schmidt (2006)]. With this in mind, most empirical works with a relatively small cross-section dimension report results of the one-step GMM estimator, which has standard errors that are asymptotically robust to heteroskedasticity and have been found to be more reliable for finite sample inference [Blundell and Bond (1998), Blundell et al. (2000); Windmeijer (2005); Bond (2002)]. This is the strategy considered in this paper.

There exist some tests to validate the assumptions underlying GMM methods. The standard approach for testing the validity of the moment conditions in GMM estimation is the Sargan test of overidentifying restrictions and the m2 second-order serial correlation test [Arellano and Bond (1991)]. Under the null hypothesis that moment conditions are valid, the Sargan test is distributed chi-squared with degrees of freedom equal to the number of moment restrictions minus the number of parameters estimated. The m2 test is normally distributed under the null hypothesis of the absence of second-order serial correlation between regressors and residuals.