Working Paper Series

Please cite this paper as:

No. 184

Modelling Scale Effect in Cross-section Data:
The Case of Hedonic Price Regression

by

Duo QIN and Yimeng LIU

September, 2013

Department of Economics
School of Oriental and African Studies
London
WC1H 0XG
Phone: + 44 (0)20 7898 4730
Fax: 020 7898 4759
E-mail: economics@soas.ac.uk
http://www.soas.ac.uk/economics/

Department of Economics

SOAS Department of Economics
University of London

The **SOAS Department of Economics Working Paper Series** is published electronically by The School of Oriental and African Studies-University of London.

©Copyright is held by the author or authors of each working paper. SOAS DoEc Working Papers cannot be republished, reprinted or reproduced in any format without the permission of the paper's author or authors.

This and other papers can be downloaded without charge from:

SOAS Department of Economics Working Paper Series at
http://www.soas.ac.uk/economics/research/workingpapers/

Design and layout: O. González Dávila

Modelling Scale Effect in Cross-section Data: The Case of Hedonic Price Regression

Duo QIN[1]

Department of Economics, SOAS, University of London, UK

Yimeng LIU

School of Economics and Resource Management, Beijing Normal University, China

Abstract

An innovative and simple experiment with cross-section data ordering is carried out to exploit a basic and common feature between many economic variables – nonlinear scale dependence. The experiment is tried on hedonic price regression models using two data sets, one for automobiles and the other computers. The key findings are: (a) Hedonic price indices can be significantly biased if they are constructed using models which disregard possible nonlinear scale effects latent in random data samples; (b) Scale-based data ordering offers considerable potential to filter such scale-dependent information from cross-section samples; (c) The filtering can be easily carried out by systematic adoption of dynamic modelling methods.

---

[1] Contacting author; email: dq1@soas.ac.uk

1. Introduction

Many economic variables of interest are stochastically scale dependent such that adequate description of the scale effects constitutes a fundamental ingredient for the success of empirical modelling of relations between such variables. This is arguably best illustrated by the popularity of error-correction models and cointegration theory in dynamic models using time-series data, mainly because of the separate interpretation of the long-run and short-run effects pertinent to scale-dependent variables, e.g. consumption, investment, output and income.

However, such scale effects have not caught much attention in models using cross-section data. This appears to be primarily due to the lack of a data ordering scheme in cross-section samples, e.g. see Durlauf *et al* (2009, Section 24.7.5), unlike the single time-sequencing convention of time-series samples. But another important contributing factor is the staunch reliance of many applied modellers on *a priori* theories for providing the corrected formulated models as far as the parameters of interest are concerned. Such a textbook econometric attitude is particularly prevalent in microeconometrics, e.g. see Cameron and Trivedi (2005) and Qin (2013). An illuminating example can be found from Nerlove's classic study to measure the return-to-scale parameter using a cross-section sample of electric utility companies (1963), see also Berndt (1991, Chapter 3). The full-sample estimation of a cost equation based on the standard production function resulted in autocorrelated residuals, because the data observations were inadvertently ordered by output scale; sub-sample estimation was consequently carried out and different return-to-scale parameter estimates were obtained for different sub-groups of the companies. This approach has evolved into a formalised estimator-based procedure, e.g. a GMM-based estimation procedure for multi-stage samples by Bhattacharya (2005). Other popular methods following this approach include various semi-parametric or non-parametric estimators, e.g. see Henderson and Ullah (2005) and Su (2012).

The present study explores a novel alternative to the mainstream approach. We treat such scale effects as a combined issue of data ordering and model specification. Specifically, we take the diagnosed residual autocorrelation from cross-section models, when the data have been ordered by certain scale-dependent variables, as a strong indication of the presence of nonlinear scale effects, analogous to the long-run and short-run effects in the time-series context, whereas such effects are absent from the *a priori* postulated models. We therefore experiment with cross-section data ordering when key explanatory variables are clearly scale dependent and utilise 'dynamic', i.e. difference-equation, models to filter such scale-dependent information. The hedonic price regression model is chosen for the experiment here, not only because the model is relatively simple but also because the basic time dummy variable methods have been well-established and increasingly used in the public domain to generate various price indices, e.g. see Triplett (2004) and Hill (2012).[2] By comparing our results with the conventional hedonic price regression model results, we try to evaluate how much the extraction, through data ordering, of scale-dependent information latent in cross-section data can help increase the explanatory power of hedonic price models.

Our experiment is implemented on two well-known cases in hedonic price analyses – automobiles and computers. The data sets that we use are from two published studies. The first one is by Raff and Trajtenberg (1996) on the hedonic price indices for American automobile industry, see the next section. The second study is by Stengos and Zacharis (2006) on hedonic price analysis of the personal computer market, see section 3. Section 4 summarises the main findings of our experiment and discusses the methodological implications which may benefit future research.

---

[2] For the history and origins of the hedonic price regression model, see Berndt (1991, Chapter 4), Triplett (2004) and also Stapleford (2011).

2.   Scale Effects in Hedonic Price Indices for Automobiles

Modern hedonic price analysis stemmed from Griliches's pioneering work on the hedonic price indices of automobiles (1961). Raff and Trajtenberg (1996) adopted Griliches's approach to examine the history of the US automobile industry. The hedonic price indices constructed by Raff and Trajtenberg (1996) are based on the well-established time adjacent period dummy variable method, e.g. see Triplett (2004). Specifically, Raff and Trajtenberg use a semi-log model for a pooled cross-section sample of two adjacent periods:

(1) $$\ln(P_i) = \alpha_0 + \gamma D_i + \sum_{j=1}^{k} \beta_j X_{ij} + u_i$$

where P_i denotes the price of car i, X_{ij} denotes quality characteristic j of car i, and D_i is a time dummy variable with value 0 for the first period and 1 for the second period. Its parameter, γ, referred to as 'the hedonic coefficient' by Raff and Trajtenberg, constitutes the key parameter for constructing the hedonic price indices. Model (1) is estimated by the ordinary least squares (OLS) estimator T-1 times to get a series of the hedonic coefficients, $\{\hat{\gamma}\}$, when the full data set covers T periods. Their method forms the baseline comparison for our data-ordering experiment.

The data set used by Raff and Trajtenberg (1996) is available from Inter-University consortium for Political and Social Research (ICPSR) and covers 1906-1941 annually.[3] Our experiment uses only the first six year (1906-11) data[4] and two key attributes for X_{ij}: (a) 'horsepower', X_{i1}, as indication of engine power and (b) 'wheelbase', X_{i2}, as indication of vehicle size (see the original paper for a more detailed data description). In the original paper, Raff and Trajtenberg only reported the regression results for two years apart as the adjacent periods, e.g. 1906-08, 1908-10 etc.. We carry out the analysis annually here for five times.

---

[3] The website of ICPSR is: www.icpsr.umich.edu and the ICPSR study number for this set is: 31762.
[4] One key reason we end the experiment at 1911 is that the price series is missing in the 1912 data file.

Now, before pooling two adjacent year data together, we order the observations of each year first. Since both X_{i1} and X_{i2} exhibit the scale-dependent feature, we experiment with two data-sorting schemes, the first using X_{i1} as the primary ordering base and X_{i2} as the secondary base, and the second scheme vice versa. For example, the observations in the first scheme are sorted by $X_{11} \leq X_{21} \leq \cdots \leq X_{n_t 1}$ and then by $X_{12} \leq X_{22} \leq \cdots \leq X_{n_t 2}$. We then run (1) repeatedly using the five pooled adjacent-period sets of the ordered data. Nine out of the ten regressions of the two schemes result in significant residual first-order autocorrelation and the correlation is much more pronounced in scheme one than scheme two. We therefore abandon scheme two.

The detected autocorrelated residuals indicate that the shadow prices, β_j, are inadequately specified in (1), in that $\ln(P_i)$ reacts to X_{ij} not only in levels, as represented by β_j, but also incrementally, i.e. there is certain relation between $\Delta \ln(P_i)$ and ΔX_{ij},[5] similar to the short-run effect in the time-series context. Since the hedonic model approach emphasises the market demand side more than the production supply side, e.g. see Stapleford (2011), such additional price effects can be interpreted as illustrating consumers' willingness to pay an extra margin for an enhancement in the quality of X_{ij}.

One existing and prevailing approach to deal with spatial autocorrelation is the spatial autoregressive model (SAR), e.g. see Anselin and Lozano-Gracia (2009). However, unlike the SAR model where data information on spatial dependence has to be obtained somewhere else external to the data sample and represented in a spatial weight matrix, the present data ordering method enables us to filter the cross-section autocorrelation information directly with 'dynamic' models. Notice that the SAR approach under the present situation would amount to extending (1) simply by an autoregression of the residual term, say, a first-order autoregression:

---

[5]The symbol Δ denotes differencing, e.g. $\Delta X_{i1} = X_{i1} - X_{i-1,1}$.

(2) $\qquad u_i = \rho u_{i-1} + \varepsilon_i .$

But we should note that the addition of (2) effectively revise (1) into a common factor (COMFAC) model, which imposes a nonlinear parameter restriction, via ρ, to equate the short-run effect (i.e. the incremental effect) with the long-run effect (i.e. the level effect), e.g. see Hendry and Mizon (1978) and Hendry (1995, Section 7.7). This COMFAC restriction is testable from the following (first-order) autoregressive distributed lag (ADL) model:

$$(3) \qquad \ln(P_i) = \alpha_0 + \alpha_1 \ln(P_{i-1}) + \gamma D_i + \sum_{j=1}^{k}\left(\beta_{0j} X_{ij} + \beta_{1j} X_{i-1,j}\right) + \varepsilon_i .$$

Since the COMFAC restriction, e.g. $\alpha_1 \beta_{0j} + \beta_{1j} = 0$ on (3), has been frequently rejected in time-series models, we adopt (3) and test the restrictions before deciding if we should follow the SAR approach.

From the cross-section perspective, we can regard model (3) as a scale-wise difference-equation model. Such a viewpoint is particularly useful here since one of the main unsettled issues in hedonic regression analyses is what the appropriate functional form should be. Difference-equation models are well-known expedients to approximate unknown functions. In our modelling experiment, we follow the convention of using Ramsey's RESET test (1969) to check on the functional form. The following log model is also tried as an alternative to (3):

$$(4) \qquad \ln(P_i) = \alpha_0 + \alpha_1 \ln(P_{i-1}) + \gamma D_i + \sum_{j=1}^{k}\left(\beta_{0j} \ln(X_{ij}) + \beta_{1j} \ln(X_{i-1,j})\right) + \varepsilon_i .$$

Another issue particularly important for the robustness of hedonic price indices is parameter constancy, especially the constancy of $\hat{\gamma}$, although it has been, unfortunately, largely ignored in the literature. During the experiment, we run recursive estimations to check the constancy and also assess it by Hansen's parameter stability test (1992).

Tables 1, 2 and 3 report the main results of models (1), (3) and (4) respectively. It is clear from Table 1 that model (1) suffers from both residual autocorrelation and inadequately

specified functional form.[6] The first problem is fixed by both models (3) and (4) whereas the second problem is largely resolved by model (4), as shown from Tables 2 and 3. The presence of incremental scale effects is unequivocal, as shown by the strong significance of the estimated $\hat{\alpha}_1$ in all the cases. Moreover, almost all the parameter estimates in model (4) have passed the Hansen parameter instability test (see Table 3), especially for the 1910-11 case where the sample has the largest size and parameter non-constancy is the most pronounced in models (1) and (3). Hence, model (4) outperforms (3) overall. Noticeably, the COMFAC restriction is rejected in most of the cases, even though the data-ordering scheme implies that the interval of the incremental (short-run) variable ΔX_{i1} is defined by the 'distance' between X_{i1} and $X_{i-1,1}$. The SAR approach is thus inadequate for remedying spatial correlation here. In addition to fixing model specification problems, the explicit adoption of the ADL model and methods from the LSE (London School of Economics) dynamic modelling procedure (e.g. see Hendry, 1995 and Qin, 2013) is shown to have enhanced the model explanatory power significantly, up to 20% if judged by the reduction of the residual standard errors, or above 30% if judged by the increase of the adjusted R^2.

Table 4 lists the hedonic price indices calculated from the three models. It is striking to see there that the inadequately specified scale effects in model (1) can result in an over-estimated hedonic price reduction by roughly 40% in a five-year span, i.e. the difference between 45% and 67% of 1911.

3.   Scale Effects in Hedonic Price Indices for Personal Computers

Hedonic price analysis for computers attained an official status as early as 1986, when the US Bureau of Economic Analysis first released official price indices for computers based in part on the results of hedonic regression models, e.g. see Berndt (1991, Chapter 4). The data

---

[6] We choose not to consider the diagnostic tests on heteroscedasticity here since this problem is relatively unrelated to the scale effects.

set that we use has been provided by Stengos and Zacharis (2006) and deposited at the Journal of Applied Econometrics Data Archive.[7] The data set was collected from *PC Magazine* covering 1993(1)-1995(11) monthly. Nonlinearity in pricing constitutes a central issue tackled by Stengos and Zacharis and their remedy is to estimate a conventional hedonic price model by the semi-parametric method using data of the entire sample. However, they have not constructed a price index from their estimation. In fact, there still lacks well-established ways of constructing price indices based on semi-parametric or non-parametric methods, e.g. see Hill (2012).

Here, we carry on with the adjacent period time dummy variable methods but follow the basic linear model specification by Stengos and Zacharis (2006):

$$\ln(P_i) = \alpha_0 + \gamma D_i + \sum_{j=1}^{k} \beta_j \ln(X_{ij}) + \sum_{j=1}^{h} \delta_j Z_{ij} + u_i \tag{5}$$

where X_{ij} denotes characteristic variables within the continuous value domain and Z_{ij} are those characteristic variables of the binary type. Specifically, there are four X_{ij}: 'speed'$= X_{i1}$, 'hard-drive size'$= X_{i2}$, 'RAM size'$= X_{i3}$, 'screen size'$= X_{i4}$, and Z_{ij} consists of three variables: 'CD'$= Z_{i1}$ (whether the PC has a CD-ROM), 'MULTI'$= Z_{i2}$ (whether the PC has a multimedia kit) and 'FIRM'$= Z_{i3}$ (whether the PC manufacturer is IBM or COMPAC). Model (5) is used as the baseline model here. In view of the value ranges of X_{ij} (see Table 1 in Stengos and Zacharis (2006)), we choose to order the observations of each month using the following sorting scheme: first by 'hard-drive size'$= X_{i2}$ and then by 'RAM size'$= X_{i3}$. We choose to use the subperiod of 1993(3)-1995(3) only, mainly for the reason that the monthly sample sizes during this period exceed 100.[8]

---

[7] The archive website: http://econ.queensu.ca/jae/.
[8] Data for the binary variable, 'MULTI', starts only from 1993(6).

It is clearly shown from Table 5 that model (5) suffers from both residual autocorrelation and inadequately specified functional form, the same problems as found in the previous case. However, the order of residual autocorrelation is found to be generally three instead of one. Since there are seven characteristic variables in addition to the time dummy variable, a third-order ADL model similar to (3) or (4) would result in a clumsily parameterised model. Nevertheless, we run such ADL models in order to conduct the COMFAC restriction test as well as the RESET test. We subsequently try to reduce the model by the automatic model selection procedure designed by Hendry and Krolzig (2001), see also Hendry and Krolzig (2005). Table 6 reports the key results of these modelling exercises. Specifically, the key results are: (a) More than half of the cases have passed the COMFAC restriction test; (b) The constancy of the estimated parameters for $\hat{\gamma}$ is much improved if compared to Table 5 (this is true for many of the parameters of other regressors as well, but the results are not reported to keep the tables compact); (c) The goodness of fit increases by 6% on average as shown from the adjusted R^2 and the model information gain is over 15% if judged by the reduction of the residual standard errors; (d) Both the residual serial-correlation problem and the functional form problem are alleviated. On the whole, our ADL model results outperform substantially the small increase of the adjusted R^2 from 0.791 to 0.811 reported by Stengos and Zacharis (2006) via the semi-parametric estimation of the entire sample (see Table II of their paper).

Since the automatically reduced ADL models are still fairly large, the COMFAC specification becomes an attractive alternative for its compact model size. We therefore run the 3rd-order autoregressive least squares (RALS) on model (5) and report the main results in Table 7. It is noticeable from the table that the overall model performance of this alternative matches that of the ADL models (Table 6) if judged by both the adjusted R^2 and the residual standard errors. Somewhat unexpectedly, however, the resulting hedonic price series turns out to be closer to that based on model (5) than the series based on the automatically reduced ADL

model, as shown from Table 8. A close reflection on this points the explanation to the COMFAC restrictions again. Each addition of u_{i-l} in the RALS estimation implies not only the addition of the corresponding lagged X_{ij} and Z_{ij} in the specifically parametrically restricted way, but also the addition of the lagged time-dummy variable as well. In contrast, we only have one time-dummy variable in the automatically reduced ADL model specification.[9] Considering the highly collinear nature of having the current and the lagged time-dummy variables in a regression and also the role of the time-dummy variable method in hedonic regressions, we believe that the hedonic price index calculated from the automatically reduced ADL model is more reliable than those from the COMFAC extension of (5) even though its overall performance appears as good as the ADL model. If we compare the annual hedonic price reductions for the period of April 1994 – March 1995, the average reduction estimated by model (5) is roughly 25% whereas it is only 12% by the ADL model. This result reinforces what we have obtained from the previous automobile case, namely that the conventional hedonic regression model generates substantially over-estimated hedonic price changes due to inadequately filtered scale effects associated to those characteristic variables which exhibit the scale-dependent feature.

4.   Concluding Remarks

Our experiment with the simple model of hedonic price regression has demonstrated clearly the potential of cross-section data-ordering schemes in exploiting a basic and common feature between many economic variables – nonlinear scale dependence. The approach is simple, economically intuitive, and its precision gains appear quite substantial, over 15% to 20% of reduction in the residual standard errors from the two cases discussed in the previous sections. Moreover, the resulting hedonic price indices demonstrate a significantly slower changing trend than those constructed from conventional hedonic regression models, indicating

---

[9] We impose the dummy variable as unreducible during the automatic model reduction.

that those old indices are systematically biased due to model mis-specification of the nonlinear scale effects latent in cross-section data.

On the methodological front, the experiment has taught us mainly the following.

First, spatial correlation or functional form problems detected from empirical modelling using cross-section data are strongly indicative of inadequate model design. As such, remedies following the estimation-based strategy cannot be fundamentally effective. Our experiment results show that those detected problems generally disappear once the scale effects are explicitly taken into consideration via data ordering and model specification. The results also reveal to us that mis-specified scale effects could be a major cause of the functional form problem. Our approach thus offers a useful alternative to the practice of correcting the problem by adding quadratic terms into the explanatory variable set.

Secondly, improvement in model design entails an intelligent combination of substantive knowledge, data understanding and econometric tools. Here, applied modellers should be particularly on the alert against pitfalls of estimator-induced implicit model revisions. Our results show how the RALS estimation of the 'static' hedonic regression could generate dubious results due to the implicit COMFAC restrictions. Even when the restrictions are shown to be data permissible, they could still be indefensible from the substantive ground. This finding cautions us particularly against models which rely on using regression error terms as explanatory variables.

Thirdly, the data-ordering approach enables us to take advantage of various available time-series econometric techniques. In fact, many of the 'statistical complications' of economic time series are the result of scale effects simply because 'Rome was not built in a day'. Notice from our ADL results that these models could be readily reparameterised into error-correction models to facilitate separate interpretation of the level (long-run) scale effects and incremental (short-run) scale effects. We can easily proceed along this line if we want to further study the

shadow prices of individual characteristic variables in the hedonic price cases. Notice also that the reparameterisation offers a more efficient way to tackle the multicollinearity problem than those ad hoc methods commonly used in cross-section data modelling practice.

Fourthly, active data ordering schemes do not violate the fundamental assumption of random sampling in statistics. The data-ordering experiment is based on data sets which have been collected following the random sampling rule. All the data ordering does is to try and better organise random data samples using economic knowledge so as to get more useful information out of the unorganised random samples. If we were to re-order time-series data by a random sorting sequence, the nonlinear scale effects captured by dynamic models would largely disappear. It is thus a superficial diagnosis to blame mechanically the OLS for not being statistically optimal when it is applied to static models using time-sequenced time-series data.

Finally, the data-ordering process can increase the probability of relocating those relatively anomalous observations towards the two ends of the cross section, thus enabling modellers to decide whether these observations are representative of the population in consideration and also to examine how much these observations aggravate the heterogeneity problem which prevails in micro data. We have noticed this from graphs of certain cross-section samples, but we have not pursued further of this data selection issue here.

References

Anselin, L. and N. Lozano-Gracia (2009) Spatial hedonic models, in T.C. Mills and K. Patterson eds. *Palgrave Handbook of Econometrics, vol 2: Applied Econometrics*, Palgrave MacMillan, pp. 1213-50.

Berndt, E. R. (1991) *The Practice of Econometrics: Classic and Contemporary*. Addison-Wesley Publishing Company.

Bhattacharya, D. (2005) Asymptotic inference from multi-stage samples. *Journal of Econometrics*, **126**, 145–71.

Cameron, A.C. and P.K. Trivedi (2005) Microeconometrics: Methods and Applications, Cambridge University Press.

Durlauf, S.N., P.A. Johnson and R.W Jonathan (2009) the methods of growth econometrics, in T.C. Mills and K. Patterson eds. *Palgrave Handbook of Econometrics, vol 2: Applied Econometrics*, Palgrave MacMillan, pp. 1119-79.

Griliches, Z. (1961) Hedonic price indexes for automobiles: An econometric analysis of quality change, in Price Statistics Review Committee ed., *The Price Statistics of the Federal Government,* National Bureau of Economic Research, 173–96; reprinted in D. Qin ed., (2013) *The Rise of Econometrics*, vol.3, London: Routledge, pp. 96-124.

Hansen, B.E. 1992. Parameter instability in linear models. *Journal of Policy Modeling*, **14**, 517-33.

Henderson, D.J. and Ullah, A. (2005) A nonparametric random effects estimator. *Economics Letters*, **88**, 403-7.

Hendry, D.F. (1995) *Dynamic Econometrics*, Oxford University Press.

Hendry, D.F. and Mizon, G.E. (1978) Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money by the Bank of England. *Economic Journal*, **88**, 549-63.

Hendry, D.F. and Krolzig, H-M. (2001) *Automatic Econometric Model Selection Using PcGets 1.0*. Timberlake Consultants Ltd.

Hendry, D.F. and Krolzig, H-M. (2005) The properties of automatic Gets modelling. *Economic Journal*, **115**, C32-C61.

Hill, R.J. (2012) Hedonic price indexes for residential housing: A survey, evaluation and taxonomy. *Journal of Economic Survey*, doi: 10.1111/j.1467-6419.2012.00731.x.

Nerlove, M. (1963) Returns to scale in electric supply, in C.F. Christ ed., *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*, Stanford University Press, pp. 167-98.

Qin, D. (2013) *A History of Econometrics: The Reformation from the 1970s*, Oxford University Press.

Ramsey, J.B. (1969) Tests for specification errors in classical least-squares regression analysis, *Journal of the Royal Statistical Society B*, **31**, 350-71.

Raff, D. and M. Trajtenberg (1996) Quality-adjusted prices for the American automobile industry: 1906-1940, in T.F. Bresnahan & R.J. Gordon eds. *The Economics of New Goods*, National Bureau of Economic Research, pp. 71-108.

Stapleford, T. A. (2011) Reconceiving quality: Political economy and the rise of hedonic price indexes. *History of Political Economy* (Annual Supplement), **43**, 309-28.

Stengos, T. and E. Zacharias (2006) Intertemporal pricing and price discrimination: A semiparametric hedonic analysis of the personal computer market. *Journal of Applied Econometrics*, **21**, 371–86.

Su, L.-J. (2012) Semiparametric GMM estimation of spatial autoregressive models. *Journal of Econometrics*, **167**, 543-60.

Triplett, J. (2004) *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*. Directorate for Science, Technology and Industry Working Paper 2004/9, Paris: OECD.

Table 1. Hedonic Price Regression of (1) for Automobiles

Sample	1906-07	1907-08	1908-09	1909-10	1910-11
$\hat{\gamma}$	0.0002	-0.291	0.0318	-0.075	-0.308
(t-statistic)	(0.005)	(-6.78)**	(0.784)	(-1.65)	(-8.43)**
(Hansen statistic)	(0.1548)	(0.5141)*	(0.2262)	(0.2116)	(5.742)**
$\hat{\beta}_1$	0.0041	0.0056	0.0079	0.0135	0.0248
(t-statistic)	(1.95)*	(2.34)*	(3.29)**	(3.6)**	(15.1)**
(Hansen statistic)	(0.2423)	(0.1548)	(0.2985)	(0.4307)	(7.562)**
$\hat{\beta}_2$	0.0347	0.0304	0.0302	0.0108	0.0105
(t-statistic)	(15.8)**	(12.3)**	(12.0)**	(2.7)**	(6.71)**
(Hansen statistic)	(0.5399)*	(0.3395)	(0.2365)	(0.5394)*	(5.061)**
Number of observations	286	233	271	224	821
$\overline{R}^2$	0.7117	0.769	0.769	0.5	0.572
$\hat{\sigma}_u$	0.353	0.311	0.291	0.338	0.344
AR(1)F-test	4.3023[0.039]*	33.7 [0.000]**	46.23 [0.000]**	78.83 [0.000]**	540.1 [0.000]**
RESET	4.6033 [0.011]*	6.007 [0.003]**	17.93 [0.000]**	33.75 [0.000]**	47.51 [0.000]**

Note: $\overline{R}^2$ denotes the adjusted R^2, $\hat{\sigma}_u$ denotes the residual standard error; AR(1) is residual first-order autocorrelation test, RESET is the Ramsey test and *p*-values are given in the squared brackets. * and ** indicate 5% and 1% significance levels respectively.

Table 2. Hedonic Price Regression of (3) for Automobiles

Sample	1906-07	1907-08	1908-09	1909-10	1910-11
$\hat{\gamma}$	-0.0117	-0.1686	0.0269	-0.055	-0.1134
(t-statistic)	(-0.26)	(-3.89)**	(0.739)	(-1.48)	(-3.83)**
(Hansen statistic)	(0.1134)	(0.1756)	(0.0962)	(0.0656)	(1.179)**
$\hat{\alpha}_1$	0.1561	0.3643	0.4243	0.5165	0.632
(t-statistic)	(2.68)**	(5.89)**	(7.71)**	(9.53)**	(23.6)**
(Hansen statistic)	(0.2924)	(0.1042)	(0.1059)	(0.1528)	(1.037)**
$\hat{\beta}_{0,1}$	-0.0057	0.0205	0.0215	0.0064	0.0276
(t-statistic)	(-1.86)	(4.11)**	(6.25)**	(1.38)	(7.02)**
(Hansen statistic)	(0.1644)	(0.1087)	(0.1053)	(0.1338)	(1.637)**
$\hat{\beta}_{1,1}$	0.0078	-0.0188	-0.0182	-0.0014	-0.021
(t-statistic)	(2.84)**	(-3.66)**	(-5.42)**	(-0.296)	(5.3)**
(Hansen statistic)	(0.1777)	(0.1172)	(0.1141)	(0.1427)	(1.699)**
$\hat{\beta}_{0,2}$	0.029	0.0257	0.0196	0.01	0.0079
(t-statistic)	(10.8)**	(8.68)**	(6.19)**	(2.87)**	(6.1)**
(Hansen statistic)	(0.307)	(0.1016)	(0.1034)	(0.1445)	(0.959)**
$\hat{\beta}_{1,2}$	0.0032	-0.0046	-0.0025	-0.0039	-0.001
(t-statistic)	(0.909)	(-1.36)	(-0.763)	(-1.48)	(-0.767)
(Hansen statistic)	(0.264)	(0.1076)	(0.1056)	(0.1558)	(1.094)**
$\overline{R}^2$	0.735	0.808	0.811	0.645	0.749
$\hat{\sigma}_u$	0.339	0.283	0.255	0.271	0.263
AR(2) F-test	0.0233[0.977]	0.899[0.408]	0.312[0.732]	0.4244[0.655]	2.858[0.058]
RESET	1.444[0.238]	6.943[0.001]**	9.661[0.000]**	1.2532[0.288]	4.377[0.013]*
COMFAC χ^2 test	20.41 [0.000]**	9.748 [0.008]**	13.82 [0.001]**	2.182 [0.336]	10.35 [0.006]**

Note: see the note in Table 1; the COMFAC test involves testing: $\alpha_1 \beta_{01} + \beta_{11} = 0$; $\alpha_1 \beta_{02} + \beta_{12} = 0$.

Table 3. Hedonic Price Regression of (4) for Automobiles

Sample	1906-07	1907-08	1908-09	1909-10	1910-11
$\hat{\gamma}$	0.0084	-0.158	0.053	-0.0815	-0.0645
(t-statistic)	(0.176)	(-3.64)**	(1.5)	(-2.16)*	(-2.14)*
(Hansen statistic)	(0.1017)	(0.1843)	(0.0856)	(0.103)	(0.3449)
$\hat{\alpha}_1$	0.1905	0.3816	0.3694	0.4954	0.594
(t-statistic)	(3.34)**	(6.47)**	(6.44)**	(9.27)**	(20.9)**
(Hansen statistic)	(0.5399)*	(0.1191)	(0.0621)	(0.2235)	(0.335)
$\hat{\beta}_{01}$	-0.177	0.4412	0.7155	-0.0207	0.944
(t-statistic)	(-2.22)*	(4.0)*	(7.8)**	(-0.242)	(8.93)**
(Hansen statistic)	(0.2423)	(0.1948)	(0.061)	(0.2309)	(0.3655)
$\hat{\beta}_{11}$	0.1476	-0.42	-0.3359	-0.2812	-0.485
(t-statistic)	(2.0)*	(-3.97)**	(-3.42)**	(-3.01)**	(-4.4)**
(Hansen statistic)	(0.5399)*	(0.2015)	(0.063)	(0.2492)	(0.368)
$\hat{\beta}_{02}$	2.792	2.559	1.686	2.164	0.323
(t-statistic)	(10.1)**	(9.7)**	(5.06)**	(5.32)**	(3.43)**
(Hansen statistic)	(0.5399)*	(0.1272)	(0.0618)	(0.2184)	(0.287)
$\hat{\beta}_{12}$	0.3764	-0.365	-0.502	0.7117	0.129
(t-statistic)	(1.11)	(-1.1)	(-1.49)	(1.68)	(1.36)
(Hansen statistic)	(0.5399)*	(0.1283)	(0.0624)	(0.2223)	(0.302)
$\overline{R}^2$	0.722	0.811	0.823	0.658	0.759
$\hat{\sigma}_u$	0.347	0.281	0.247	0.266	0.256
AR(2) F-test	0.1744 [0.84]	1.621 [0.20]	0.012 [0.988]	0.927[0.397]	0.367[0.693]
RESET	4.6033 [0.011]*	1.478[0.23]	2.862[0.059]	1.506[0.224]	2.934[0.054]
COMFAC χ^2 test	21.58 [0.000]**	11.65 [0.008]**	0.934 [0.627]	22.77 [0.000]**	18.63 [0.000]**

Note: see the note in Table 2.

Table 4. Hedonic Price Indices for Automobiles

	Raff & Trajtenberg (1996)	$\hat{\gamma}$ from Table 1	$\hat{\gamma}$ from Table 2	$\hat{\gamma}$ from Table 3
1906	100	100	100	100
1907		95.72	94.59	96.51
1908	70.0	73.13	81.67	84.22
1909		73.13	81.27	86.02
1910	54.0	61.26	69.46	71.60
1911		45.02	62.02	67.13

Note: All the indices are deflated by the CPI-based US inflation rates. The indices constructed by Raff & Trajtenberg (1996) are derived from two-year apart adjacent-period time-dummy regression model.

Table 5. Key results of Model (5) for PCs

Sample	3-4(93)	4-5(93)	5-6(93)	6-7(93)	7-8(93)	8-9(93)
Sample size	208	247	315	425	547	544
$\hat{\gamma}$	0.007	-0.064	-0.037	-0.003	-0.018	-0.008
(t-statistic)	(0.459)	(-5.12)**	(-3.71)**	(-0.242)	(-1.60)	(-0.727)
(Hansen statistic)	(0.2451)	(0.1256)	(0.0512)	(0.5377)*	(0.5422)*	(0.4002)
$\overline{R}^2$	0.8	0.847	0.881	0.8	0.774	0.808
$\hat{\sigma}_u$	0.101	0.089	0.085	0.114	0.127	0.121
AR(3) F-test	15.95[0.00]**	12.52[0.00]**	2.88[0.036]*	47.29[0.00]**	95.78[0.00]**	83.98[0.00]**
RESET	2.93[0.089]	0.44[0.5095]	0.03[0.873]	4.97[0.026]*	23.98[0.00]**	25.93[0.00]**
Sample	9-10(93)	10-11(93)	11-12(93)	12-1(93/4)	1-2(94)	2-3(94)
Sample size	529	542	534	491	508	559
$\hat{\gamma}$	-0.033	-0.003	-0.035	0.01	-0.062	-0.023
(t-statistic)	(-3.23)**	(-0.282)	(-3.24)**	(0.875)	(-5.92)**	(-2.73)*
(Hansen statistic)	(0.4999)*	(0.6659)*	(0.5794)*	(0.2531)	(0.3069)	(0.2913)
$\overline{R}^2$	0.813	0.779	0.755	0.751	0.809	0.845
$\hat{\sigma}_u$	0.114	0.118	0.124	0.126	0.114	0.100
AR(3) F-test	72.19[0.00]**	79.42[0.00]**	76.5[0.00]**	69.54[0.00]**	63.73[0.00]*	42.50[0.00]**
RESET	20.51[0.00]**	29.16[0.00]**	41.93[0.00]**	32.04[0.00]**	12.07[0.00]**	17.31[0.00]**
Sample	3-4(94)	4-5(94)	5-6(94)	6-7(94)	7-8(94)	8-9(94)
Sample size	574	646	612	498	473	453
$\hat{\gamma}$	-0.034	-0.019	-0.021	-0.028	-0.013	0.004
(t-statistic)	(-4.06)**	(-2.40)*	(-2.26)*	(-2.99)**	(-1.23)	(0.36)
(Hansen statistic)	(0.2577)	(0.4975)*	(0.5894)*	(0.5556)*	(0.6771)*	(0.3615)
$\overline{R}^2$	0.848	0.83	0.799	0.84	0.842	0.827
$\hat{\sigma}_u$	0.097	0.102	0.109	0.102	0.110	0.118
AR(3) F-test	32.67 [0.00]**	70.22[0.00]**	84.72[0.00]**	37.93[0.00]**	47.4[0.00]**	53.95[0.00]**
RESET	8.14[0.00]**	10.49[0.00]**	25.96[0.00]**	6.00 [0.015]*	3.57[0.06]	9.45 [0.002]**
Sample	9-10(94)	10-11(94)	11-12(94)	12-1(94/5)	1-2(95)	2-3(95)
Sample size	367	353	373	340	310	315
$\hat{\gamma}$	-0.035	-0.037	-0.026	-0.02	-0.017	-0.054
(t-statistic)	(-2.72)**	(-3.24)**	(-2.87)**	(-2.17)*	(-1.49)	(-5.88)**
(Hansen statistic)	(0.1282)	(0.2507)	(0.2204)	(0.1948)	(0.2543)	(0.2674)
$\overline{R}^2$	0.817	0.834	0.87	0.868	0.843	0.869
$\hat{\sigma}_u$	0.121	0.106	0.088	0.086	0.092	0.081
AR(3) F-test	35.88[0.00]**	31.51[0.00]**	30.08[0.00]**	17.38[0.00]**	14.04[0.00]**	16.72[0.00]**
RESET	13.59[0.00]**	18.71[0.00]**	11.95[0.00]**	7.04[0.008]**	10.81[0.00]**	16.53[0.00]**

Note: Data for Z_{i2} is not available for the first three adjacent-period regressions. The parameter estimates of other regressors are omitted to keep the table short. Almost all the regressors are significant. As an example, the following is the estimated result for 9-10 (1993):

$$\ln\left(\hat{P}_i\right) = \underset{\substack{(13.8)** \\ (0.537)*}}{3.58} - \underset{\substack{(-3.23)** \\ (0.499)*}}{0.03} D_i + \underset{\substack{(16.1)** \\ (0.545)*}}{0.24} \ln(X_{i1}) + \underset{\substack{(15)** \\ (0.56)*}}{0.24} \ln(X_{i2}) + \underset{\substack{(5.44)** \\ (0.819)**}}{0.08} \ln(X_{i3})$$

$$+ \underset{\substack{(7.24)** \\ (0.537)*}}{0.72} \ln(X_{i4}) + \underset{\substack{(6.74)** \\ (0.248)}}{0.09} Z_{i1} - \underset{\substack{(-1.72) \\ (0.115)}}{0.045} Z_{i2} - \underset{\substack{(-11.7)** \\ (0.557)*}}{0.18} Z_{i3}$$

See also the note in Table 1.

Table 6. Key results from 3rd-orderADL Models for PCs

Sample	3-4(93)	4-5(93)	5-6(93)	6-7(93)	7-8(93)	8-9(93)
COMFAC χ^2 test	16.71[0.729]	10.31[0.975]	35.43[0.062]	43.34[0.01]*	33.81[0.088]	44.79[0.006]**
No. of parameters after model reduction	13	20	16	19	21	22
$\hat{\gamma}$	-0.001	-0.039	-0.026	0.008	-0.005	-0.001
(t-statistic)	(-0.059)	(-2.96)**	(-2.64)*	(0.786)	(-0.572)	(-0.091)
(Hansen statistic)	(0.0936)	(0.0671)	(0.0395)	(0.1512)	(0.0853)	(0.0771)
$\overline{R}^2$	0.836	0.863	0.889	0.849	0.858	0.875
$\hat{\sigma}_u$	0.09	0.083	0.082	0.098	0.099	0.097
AR(2) F-test	0.192 [0.825]	0.443 [0.693]	0.408 [0.665]	0.126 [0.881]	2.65 [0.072]	0.21 [0.811]
RESET	0.33 [0.719]	0.765 [0.467]	0.228 [0.796]	1.108 [0.331]	2.593 [0.076]	2.929 [0.054]
Sample	9-10(93)	10-11(93)	11-12(93)	12-1(93/4)	1-2(94)	2-3(94)
COMFAC χ^2 test	77.28[0.000]**	48.88[0.002]**	47.62[0.003]**	39.24[0.026]*	40.57[0.019]*	29.04[0.219]
No. of parameters after model reduction	25	21	22	21	18	19
$\hat{\gamma}$	-0.011	0.005	-0.012	0.001	-0.023	-0.014
(t-statistic)	(-1.34)	(0.58)	(-1.42)	(0.138)	(-2.56)*	(-1.86)
(Hansen statistic)	(0.0729)	(0.1721)	(0.0742)	(0.0401)	(0.0796)	(0.0515)
$\overline{R}^2$	0.888	0. 872	0. 861	0.859	0.875	0.877
$\hat{\sigma}_u$	0.087	0.089	0.093	0.095	0.091	0.089
AR(2) F-test	0.108 [0.898]	1.895 [0.151]	2.243 [0.107]	2.614 [0.074]	2.098 [0.124]	4.45 [0.012]*
RESET	2.929 [0.054]	1.198 [0.274]	0.903[0.406]	0.049 [0.952]	0.17 [0.844]	0.421 [0.656]
Sample	3-4(94)	4-5(94)	5-6(94)	6-7(94)	7-8(1994)	8-9(94)
COMFAC χ^2 test	59.01[0.000]**	48.36[0.002]**	33.55[0.093]	35.07[0.067]	38.94[0.028]*	61.57[0.000]**
No. of parameters after model reduction	26	22	21	22	24	25
$\hat{\gamma}$	-0.021	-0.006	-0.006	-0.013	-0.004	-0.003
(t-statistic)	(-2.69)*	(-0.97)	(-0.726)	(-1.57)	(-0.49)	(-0.308)
(Hansen statistic)	(0.0712)	(0.0892)	(0.0983)	(0.1723)	(0.1658)	(0.1055)
$\overline{R}^2$	0.881	0.887	0.87	0.882	0.895	0.897
$\hat{\sigma}_u$	0.085	0.082	0.087	0.089	0.092	0.094
AR(2) F-test	4.25 [0.015]*	1.53 [0.218]	1.41 [0.245]	0.675 [0.51]	2.959 [0.053]	2.297 [0.102]
RESET	0.032 [0.968]	0.23 [0.795]	1.07 [0.344]	0.866 [0.421]	0.115 [0.89]	3.89 [0.021]*
Sample	9-10(94)	10-11(94)	11-12(94)	12-1(94/5)	1-2(95)	2-3(95)
COMFAC χ^2 test	33.95[0.086]	19.69[0.714]	30.95[0.155]	35.18[0.053]	41.42[0.015]*	30.93[0.132]
No. of parameters after model reduction	20	18	19	17	18	16
$\hat{\gamma}$	-0.019	-0.016	-0.015	-0.01	0.007	-0.038
(t-statistic)	(-1.71)	(-1.51)	(-1.78)	(-1.16)	(0.587)	(-4.32)**
(Hansen statistic)	(0.0288)	(0.081)	(0.0872)	(0.1206)	(0.0954)	(0.1206)
$\overline{R}^2$	0.868	0.867	0.899	0.889	0.873	0.89
$\hat{\sigma}_u$	0.102	0.094	0.076	0.078	0.082	0.074
AR(2) F-test	0.258 [0.773]	0.761 [0.468]	1.557 [0.212]	0.17 [0.844]	0.165 [0.85]	3.78[0.024]*
RESET	0.209 [0.812]	1.023 [0.321]	0.831 [0.437]	0.89 [0.412]	2.185 [0.114]	2.974[0.053]

Note: There are 24 COMFAC restrictions for most of the regressions, since there are three restrictions for each regressor. Take X_{i1} for example, its COMFAC restrictions are (α_i is parameter of the lagged explained variable): $\alpha_1\beta_{01} + \beta_{11} = 0$; $\alpha_2\beta_{01} + \beta_{21} = 0$; $\alpha_3\beta_{01} + \beta_{31} = 0$.

Other than the COMFAC test, all the other results are from the automatically reduced ADL models.

Table 7. Key results of RALS Estimation of Model (5) for PCs

Sample	3-4(93)	4-5(93)	5-6(93)	6-7(93)	7-8(93)	8-9(93)
$\hat{\gamma}$ (t-statistic)	0.021 (0.577)	-0.07 (-3.15)**	-0.036 (-2.79)**	-0.017 (-0.601)	-0.019 (-0.584)	-0.013 (-0.42)
$\hat{u}_{i-1}$ (t-statistic)	0.314 (4.35)**	0.304 (4.60)**	0.17 (2.93)**	0.346 (7.07)**	0.292 (6.85)**	0.247 (5.77)**
$\hat{u}_{i-2}$ (t-statistic)	0.254 (3.38)**	0.152 (2.22)*	0.019 (0.33)	0.151 (2.94)**	0.271 (6.29)**	0.299 (7.08)**
$\hat{u}_{i-3}$ (t-statistic)	0.09 (1.21)	0.045 (0.652)	0.053 (0.922)	0.166 (3.39)**	0.178 (4.18)**	0.19 (4.47)**
Pseudo $\overline{R}^2$	0.837	0.866	0.883	0.85	0.855	0.872
$\hat{\sigma}_u$	0.091	0.083	0.084	0.098	0.101	0.098
Sample	9-10(93)	10-11(93)	11-12(93)	12-1(93/4)	1-2(94)	2-3(94)
$\hat{\gamma}$ (t-statistic)	-0.036 (-1.21)	-0.001 (-0.033)	-0.014 (-0.354)	0.042 (0.895)	-0.058 (-1.69)	-0.016 (-0.734)
$\hat{u}_{i-1}$ (t-statistic)	0.324 (7.53)**	0.443 (10.3)**	0.495 (11.4)**	0.489 (10.8)**	0.395 (8.79)**	0.294 (6.89)**
$\hat{u}_{i-2}$ (t-statistic)	0.24 (5.42)**	0.194 (4.18)**	0.125 (2.60)*	0.142 (2.83)**	0.211 (4.46)**	0.22 (5.03)**
$\hat{u}_{i-3}$ (t-statistic)	0.183 (4.29)**	0.144 (3.40)**	0.181 (4.16)**	0.191 (4.23)**	0.165 (3.71)**	0.137 (3.21)**
Pseudo $\overline{R}^2$	0.879	0.867	0.857	0.856	0.875	0.877
$\hat{\sigma}_u$	0.091	0.091	0.095	0.096	0.092	0.089
Sample	3-4(94)	4-5(94)	5-6(94)	6-7(94)	7-8(1994)	8-9(94)
$\hat{\gamma}$ (t-statistic)	-0.028 (-1.47)	-0.035 (-1.38)	-0.052 (-1.78)	-0.037 (-1.82)	-0.031 (-1.21)	-0.02 (-0.603)
$\hat{u}_{i-1}$ (t-statistic)	0.273 (6.45)**	0.386 (9.79)**	0.452 (11.4)**	0.362 (8.06)**	0.319 (6.92)**	0.276 (5.82)**
$\hat{u}_{i-2}$ (t-statistic)	0.225 (5.26)**	0.187 (4.55)**	0.079 (1.80)	0.034 (0.724)	0.16 (3.36)*	0.269 (5.68)**
$\hat{u}_{i-3}$ (t-statistic)	0.123 (2.92)**	0.176 (4.52)**	0.241 (6.10) **	0.21 (4.74)**	0.194 (4.23)**	0.186 (3.96)**
Pseudo $\overline{R}^2$	0.876	0.884	0.869	0.874	0.885	0.882
$\hat{\sigma}_u$	0.087	0.084	0.087	0.09	0.094	0.098
Sample	9-10(94)	10-11(94)	11-12(94)	12-1(94/5)	1-2(95)	2-3(95)
$\hat{\gamma}$ (t-statistic)	-0.06 (-1.92)	-0.04 (-1.46)	-0.035 (-1.48)	-0.033 (-1.69)	-0.037 (-1.48)	-0.053 (-2.60)*
$\hat{u}_{i-1}$ (t-statistic)	0.237 (4.49)**	0.223 (4.13)**	0.24 (4.54)**	0.336 (5.91)**	0.324 (5.41)**	0.312 (5.32)**
$\hat{u}_{i-2}$ (t-statistic)	0.273 (5.21)**	0.26 (4.88)**	0.269 (5.14)**	0.144 (2.49)*	0.194 (3.23)**	0.202 (3.42)**
$\hat{u}_{i-3}$ (t-statistic)	0.157 (2.98)**	0.167 (3.10)**	0.167 (3.15)**	0.09 (1.61)	0.099 (1.68)	0.084 (1.45)
Pseudo $\overline{R}^2$	0.866	0.871	0.898	0.887	0.866	0.891
$\hat{\sigma}_u$	0.103	0.093	0.077	0.079	0.084	0.074

Note: Hansen statistics are not available for RALS estimation.

Table 8. Hedonic Price Indices for PC

	$\hat{\gamma}$ from Table 5	$\hat{\gamma}$ from Table 6	$\hat{\gamma}$ from Table 7
March (1993)	100	100	100
April (1993)	100.70	99.90	102.12
May (1993)	94.46	96.08	95.22
June (1993)	91.03	93.61	91.85
July (1993)	90.76	94.36	90.30
August (1993)	89.14	93.89	88.60
September (1993)	88.43	93.80	87.46
October (1993)	85.56	92.77	84.37
November (1993)	85.30	93.24	84.28
December (1993)	82.37	92.13	83.11
January (1994)	83.19	92.22	86.68
February (1994)	78.19	90.12	81.79
March (1994)	76.41	88.87	80.49
April (1994)	73.86	87.02	78.27
May (1994)	72.47	86.50	75.58
June (1994)	70.96	85.98	71.75
July (1994)	69.00	84.87	69.14
August (1994)	68.11	84.54	67.03
September (1994)	68.39	84.28	65.70
October (1994)	66.03	82.70	61.88
November (1994)	63.64	81.38	59.45
December (1994)	62.00	80.17	57.41
January (1995)	60.77	79.37	55.54
February (1995)	59.75	79.93	53.53
March (1995)	56.61	76.95	50.76