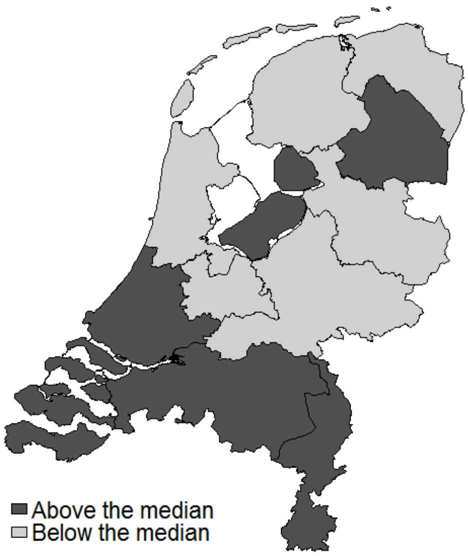
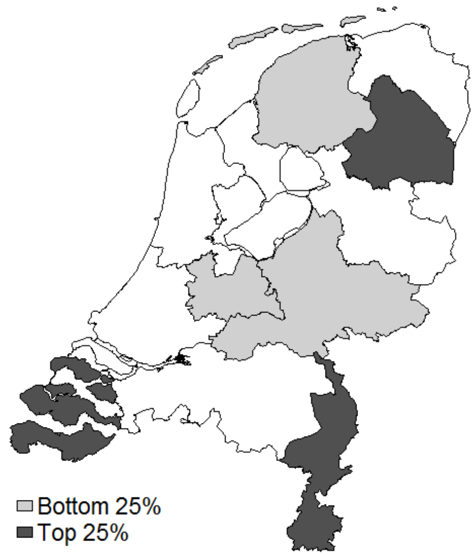
Causes of regional variation in Dutch healthcare expenditures: evidence from movers

Online Appendix

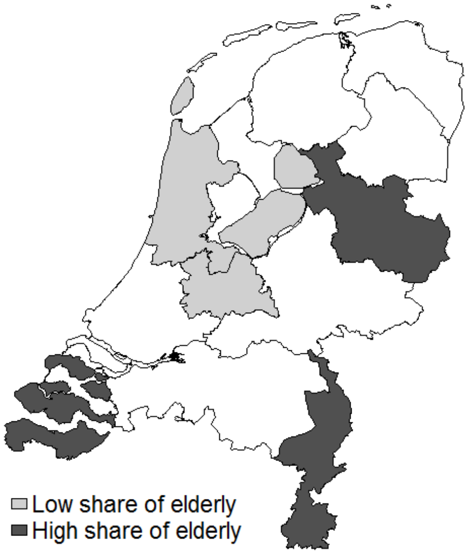
Appendix S1. Regions compared in the decomposition analysis



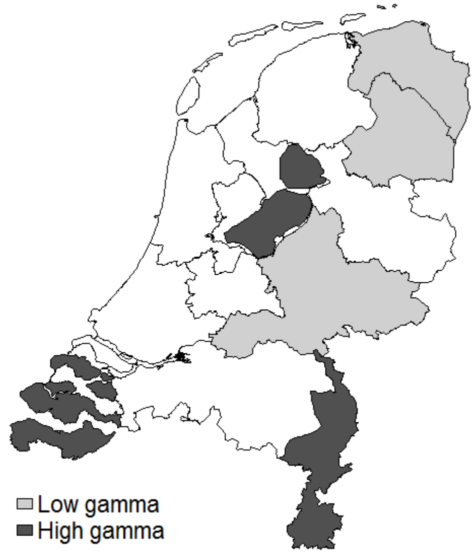
(a) Provinces above and below the median average log(healthcare expenditure+1)



(b) Provinces in the top and bottom 25% in terms of average log(healthcare expenditure+1)



(c) Provinces in top and bottom 25% in terms of share of elderly population



(d) Provinces in the top and bottom 25% of the (estimated) region fixed-effect

Appendix S2. Details on the dataset and data cleaning process

The original dataset covers 133,060,196 observations, with the number of individuals observed each period varying between 15.8 and 17.1 millions for the years 2006 and 2013, respectively. In the paragraphs below, the term observation refers to a given line in the dataset and the term individual refers to all observations associated with a given individual identifier.

To begin with, we exclude 1,619,384 (1.2%) observations with missing postal codes. These were due to the fact that some postal code areas are very small, thus raising privacy concerns. Then, we drop 1,008,072 (0.8%) observations whose registration time was above 1 year and 42,107 (0.0%) with no individual identifier.

Some individuals exhibit inconsistent age patterns over time and therefore were excluded from the dataset. In this case we delete all the observations associated with that individual, resulting in the exclusion of 8,647,602 (6.5%) observations.

Some individuals have more than one observation per year. This can be due to moving to a new postal code area, switching insurers, etc. If their demographic characteristics (age and gender) are not consistent, these individuals are excluded from the dataset. For individuals whose personal characteristics are consistent, we sum the registration times for all the lines corresponding to the same year. In case the total registration is above 1 year, the observation is dropped. In case the postal code varies between the multiple entries for an individual within a given year t, we check which of the postal codes corresponds to the destination region by looking and the next time period and attribute that postal code to year t. Because all information in the dataset refers to December 31 of each year, this is an harmless procedure. After fixing these issues, we sum the expenditures for each category of healthcare over the multiple lines for the same individual within a given year. Finally, we drop duplicated observations in terms of individual ID, year, gender, age, postal code, registration time and expenditure amounts, resulting from this procedure.

We exclude all observations whose total registration time is below 1 year (4,569,467, corresponding to 3.4% of the initial number of observations). Note that, since we had already excluded all observations with registration time above 1 year, this implies only observations with one year registration time are kept.

A few individuals exhibit negative expenditures in some categories of care. This is due to adjustments and reimbursements between the insurer and the insured that span across distinct years. We drop 35,223 observations corresponding to such situation.

We exclude 6,522,126 observations corresponding to individuals who exit and then re-enter the dataset, as such situations can also be related to moving (that is, we dropped non-consecutive observations within individuals). The fact that a given individual is not observed in a certain year can also be a result of some of the procedures described in the previous paragraphs which lead to the exclusion of some observations for a given individual.

We exclude individuals who move more than once during the time horizon under analysis (3,099,792 observations, corresponding to 2.3% of the initial number of observations). This ensures that we observe enough pre- and post-move years for each mover, which we need for the decomposition analysis. In addition, it allows to clearly identify the change in healthcare expenditures upon the move, which is crucial for our event-study analysis. Individuals moving in 2013, the last period of our sample, are not considered as movers because our empirical strategy requires that we observe healthcare expenditures both before and after the move.

Additional minor restrictions imposed on the data result in the exclusion of 187,446 observations.

There are a few observations with very high expenditures in certain time periods. Vektis specifically checks that those costs indeed were incurred. Thus, we keep them in our analysis.

Our final dataset consists on 107,364,200 observations, where all observations have 1 year registration time, each individual is observed only once a year, and individuals are observed continuously over time (though not necessarily over the same number of periods, ie. the panel is unbalanced)

Appendix S3. Additional analysis using the event-study framework

A. Different functional forms of the outcome variable

In this section, we assess the robustness of our baseline results with respect to the functional form of the outcome variable. Specifically, instead of $\log(\text{totexp}_{it} + 1)$, we estimate equation (1) using $\log(\text{totexp}_{it} + 0.1)$, $\log(\text{totexp}_{it} + 10)$, $\log_{10}(\text{totexp}_{it} + 1)$, the inverse hyperbolic sine transformation[1], expenditures in levels, and binary variables for being above certain expenditure percentiles as dependent variable.

The results are shown in Table Appendix S3A. Changing the functional form of the outcome variable to other logarithmic forms and to the inverse hyperbolic sine leaves the baseline estimates unchanged. Using a levels specification results in a somewhat lower estimate for supply-side share (16%). This is because the log specifications and the inverse hyperbolic sine transformation place more weight on differences at the lower end of the expenditure distribution, so the fact that we obtain a lower supply-side share when giving equal weight to all differences suggests that there is less room for supply-side variations at the top of the expenditure distribution. This is in line with our estimates of the supply-side share when defining the outcome variable as an indicator for being above the 50th, 75th, and 90th expenditure percentiles.

We also estimate equation (2) for each of the alternative functional forms of the outcome variable. In Figure Appendix S3A we present the corresponding plots and assess the existence of trends before and after the move. When testing whether the coefficients for the periods before the year of move are jointly statistically different from zero, we find evidence of a small pre-trend in most specifications. This may result from people already using care in the destination region before they officially move, for example if they move only some time after starting a job or a new relationship in the destination region. When looking at the periods post-move, the corresponding coefficients are never statistically different from each other,

[1]The inverse hyperbolic sine transformation is defined as $\log(\text{total exp} + \sqrt{\text{total exp}^2 + 1}$.

Table Appendix S3A: Event-study Analysis: estimates of θ for alternative outcome variables

Model	Estimate	St. Error
Distinct functional form of outcome variables:		
θ, Log(total expenditures + 10)	0.267***	0.023
θ, Log(total expenditures + 0.1)	0.272***	0.029
θ, Log_{10}(total expenditures + 1)	0.274***	0.026
θ, Log(total exp + $\sqrt{\text{total exp}^2 + 1}$)	0.273***	0.027
θ, Total expenditure (levels)	0.160***	0.028
θ, Expenditure > median	0.283***	0.024
θ, Expenditure > percentile 75	0.096***	0.027
θ, Expenditure > percentile 90	0.155***	0.043

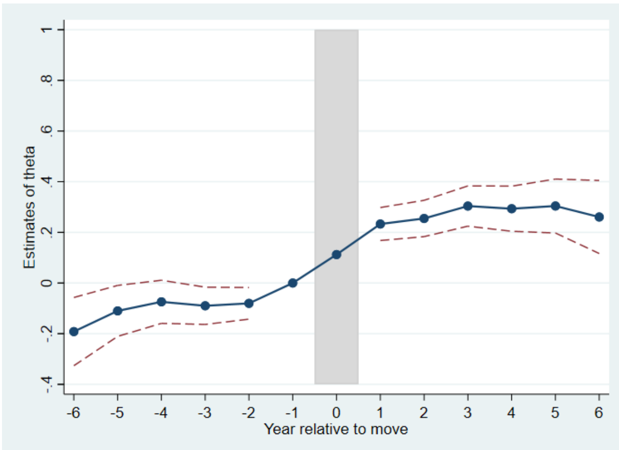
NOTES: Estimates are based on equation (1). In the levels specification we winsorized the top 5% of the expenditure distribution, in order to avoid having extreme outliers affecting our estimates. In the specifications using the distribution percentiles, the dependent variable is an indicator for whether the individual is above a given percentile of the expenditure distribution of all observations in a certain year; The number of observations is 4,146,945 , corresponding to 549,500 movers. Standard errors are robust standard errors, clustered at individual level. * significant at 10%; ** significant at 5%; *** significant at 1%.

suggesting no gradual adjustments post-move.

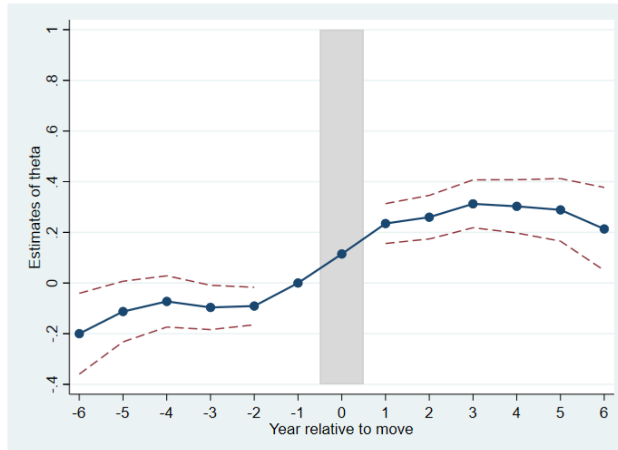
B. Robustness checks to the definition of regions

We redefine the relevant regions to the 25 GHOR regions instead of 12 provinces. GHOR stands for Geneeskundige Hulpverleningsorganisatie in de Regio (Regional Medical Emergency Preparedness and Planning) and it is responsible for the coordination and management of medical aid in case of disasters and crises. The GHOR also provides advice to governments and other organizations. There are 25 GHOR offices in the Netherlands and each of them is responsible for a specific territorial area. In this robustness check, we thus consider as movers individuals who move to a new GHOR region once and only once during the time period under analysis. Using GHOR regions results in 724,952 movers, corresponding to 5,458,133 observations. Among GHOR regions, differences in terms of healthcare expenditures are even higher than those among provinces, reaching 30 percentage points over our sample period.

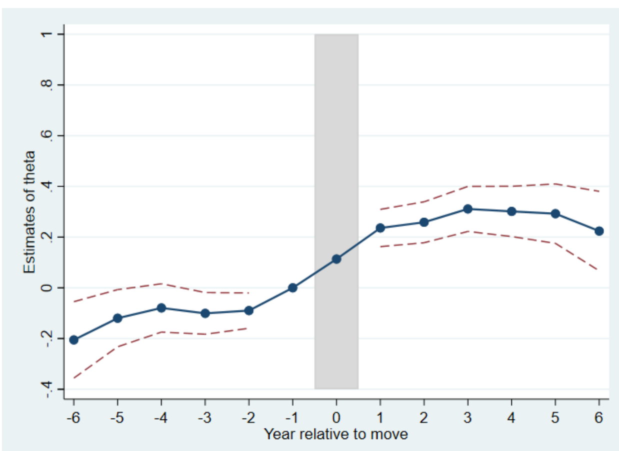
Figure Appendix S3A: Assessment of pre- and post-move trends for alternative outcome variables



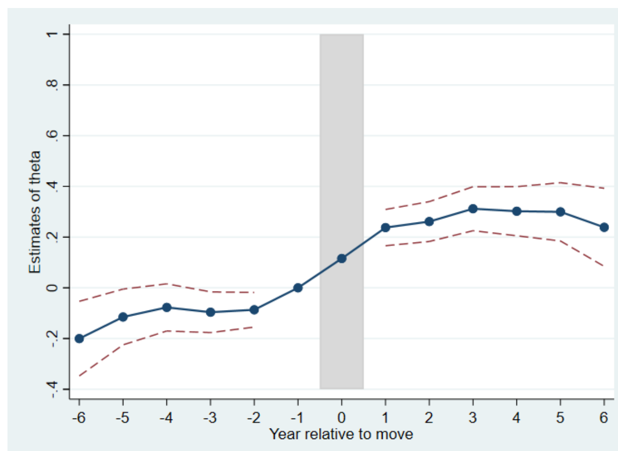
(a) Log(total expenditures + 10)



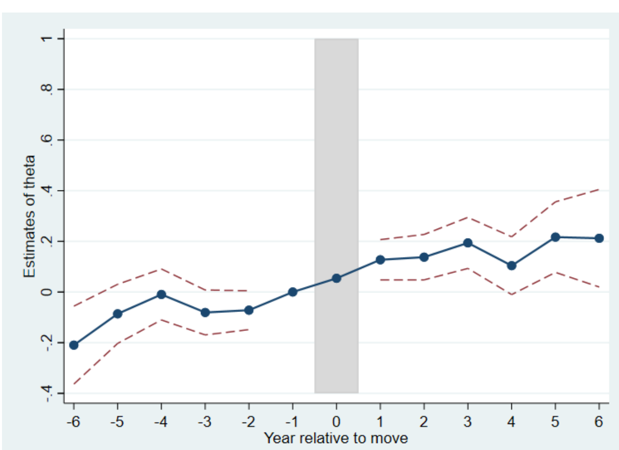
(b) Log(total expenditures + 0.1)



(c) Inverse hyperbolic sine transformation



(d) Log base 10



(e) Levels, winsorized at the top 5%

NOTES: The figures plot the estimated coefficients θ_r based on equation (2), for different functional forms of the outcome variable. The coefficients for the year just before the move, $r = -1$, were normalized to zero. The solid lines connect all estimated coefficients and the dashed lines connect the upper and lower bounds of their 95% confidence intervals. The sample consists of all 549,500 individuals who are movers.

The event-study results using GHOR regions are shown in Table Appendix S3B. The estimated supply-side shares of regional variation are slightly lower than the ones obtained using provinces as the relevant region. With GHOR regions, about 20% of the variations in total healthcare expenditures can be attributed to the supply-side. As with provinces, the supply-share is similar across distinct categories of care.

Table Appendix S3B: Event-study Analysis: estimates of θ using GHOR regions

Model	Estimate	St. Error
GHOR regions:		
θ, Total expenditures	0.190***	0.019
θ, GP expenditures	0.189***	0.018
θ, Hospital expenditures	0.223***	0.016
θ, Pharmaceutical expenditures	0.228***	0.012

NOTES: Estimates are based on equation (1). The dependent variables are $\log(\text{tot exp}_{it} + 1)$ for total expenditures, $\log(\text{GP}_{it} + 1)$ for GP care, $\log(\text{Hospital}_{it} + 1)$ for hospital care, and $\log(\text{Pharma}_{it} + 1)$ for pharmacy care. The number of observations is 5,458,133, corresponding to 724,952 individuals who are classified as movers when using GHPR as the relevant regions. Standard errors are robust standard errors, clustered at individual level. * significant at 10%; ** significant at 5%; *** significant at 1%.

C. Robustness checks to the definition of the sample

We assess the robustness of the baseline event-study results to the sample used. Specifically, we restrict the estimation sample to a balanced sample, we exclude the most common move paths between Noord-Holland and Zuid-Holland, and we redefine movers to include only individuals moving distances longer than 75km an 100km so as to minimize the chance that after the move they still seek care in the same health care facilities and are treated by the same physicians which they visited before the move. This means they would not be exposed to different supply conditions.

The corresponding results are shown in Table Appendix S3C, and are of similar magnitude to the baseline results.

Table Appendix S3C: Event-study Analysis: estimates of θ, additional robustness checks

Model	Estimate	St. Error	N. Obs.
Sample restrictions:			
θ, balanced sample	0.299***	0.027	3,611,944
θ, excluding moves between NH and ZH	0.278***	0.026	3,823,868
θ, moves > 75km	0.349***	0.031	2,166,909
θ, moves > 100km	0.285***	0.035	1,544,721

NOTES: Estimates are based on equation (1), using $\log(\text{tot exp}_{it}+1)$ as dependent variable. The number of mover-year observations varies according to the restriction imposed, as shown in the last column of the table. Standard errors are robust standard errors, clustered at individual level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix S4. Assessment of threats to the exogeneity assumption

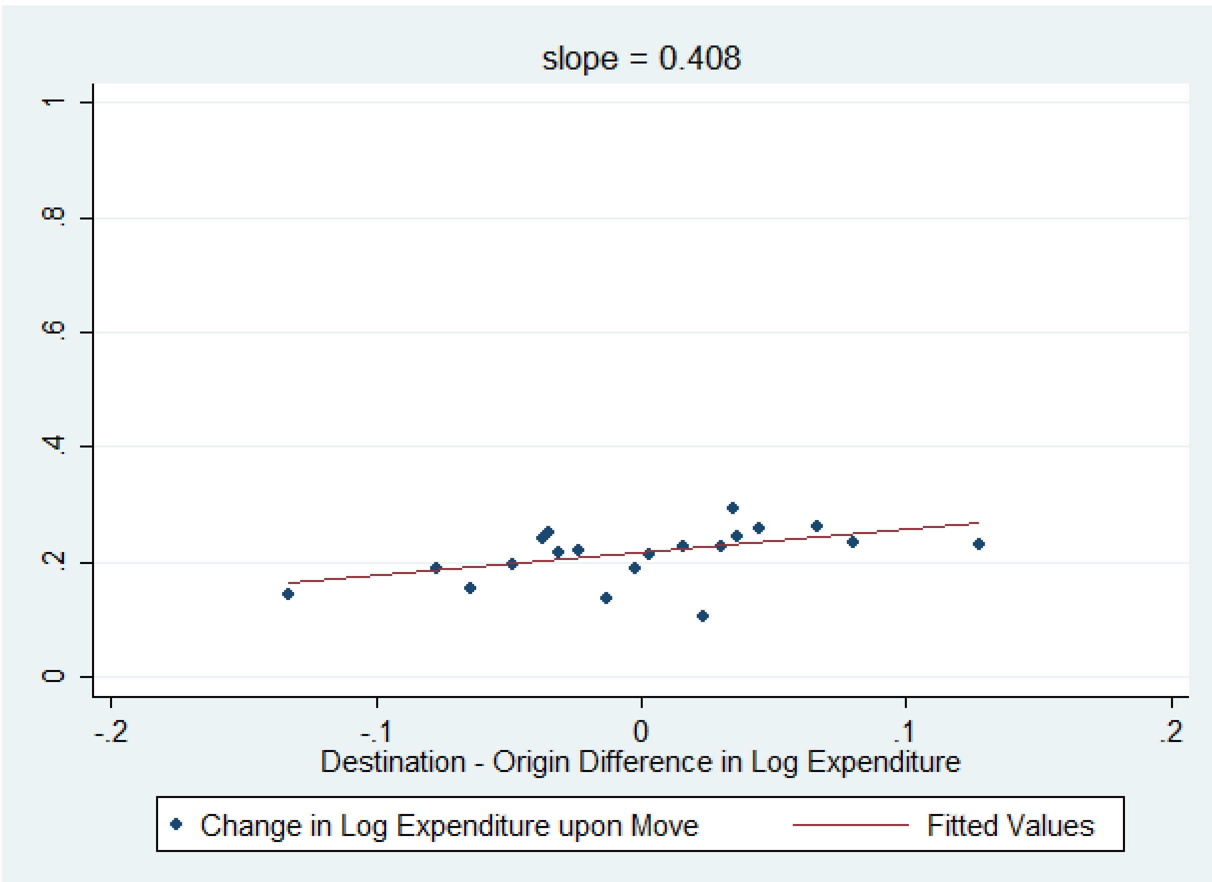
In the main text, we discuss a possible violation of the exogeneity assumption in our model, which relates to the existence of unobserved individual times trends that are systematically related to δ_i. We test for it by estimating equation (2) and examining trends pre- and post-move. An alternative way to assess whether there is evidence of unobserved individual times trends related to δ_i is to restrict the estimation sample around the year of move. We thus restrict the estimation sample to 1, 2, and 3 years around the year of move. The results are reported in the top panel of Table Appendix S4A and are similar to those at baseline.

There are additional threats to the exogeneity assumption required by our baseline model. Below we discuss the plausibility of our assumptions that the effect of δ_i on healthcare expenditures is linear and that θ is time-invariant.

Our baseline model specification assumes that the change in healthcare expenditures at the time of the move is linear in δ_i. This assumption would be violated for example if individuals respond differently to positive and negative δ_i's, i.e. the effect of moving to a region with 10% higher average expenditures can be different in magnitude from that of moving to a region with 10% lower average expenditures. In order to test whether assuming a linear effect of δ_i on healthcare expenditures is reasonable, we divide the sample of movers into 20 equally sized bins according to δ_i, and for each of the bins we compute the average change in healthcare expenditures upon the move (i.e. the average of the difference between average annual expenditure after the move and before the move). We then assess the plausibility of the assumption that the effect of δ_i is linear using a visual approach. Figure Appendix S4A plots the 20 equally sized bins of movers according to δ_i on the horizontal axis and the associated average change in individual healthcare expenditure around the move on the vertical axis. These are the 20 points in the plot. The line is a regression line connecting all 20 points. One can see that the points lie close to the regression line and the figure does not suggest the existence of non-linear effects of δ_i. Note that the slope of the regression line connecting all 20 points is 0.4, which is above our estimated θ in the baseline regression. This is due to the fact that this plot is only assessing a correlation and no covariates are being accounted for.

Another key assumption of our baseline model specification is that θ is time-invariant. This implies that the event-study equation (1) assumes the relative importance of demand and supply-side factors as drivers of regional variations in healthcare expenditures to be constant over time. As previously mentioned, on January 1st 2006, the Dutch healthcare system was subject to a systematic reform, which introduced managed competition and a single compulsory health insurance scheme for all individuals. Simultaneously, 2006 is the first year in our dataset. Thus, it may well be that in the years after the reform many adjustments were taking place as patients, physicians, hospital managers, and other agents in the healthcare sector learned the new rules of the game. This could result in θ varying over time. In order to assess the stability of θ over time, we estimate the event-study regression separately in two distinct time periods (the early period between 2006-2009, and the late period between 2010-2013). Additionally, we estimate the event-study equation distinguishing individuals who moved in the first and the second half of our study period. In each of the cases, testing the equality of the θ coefficients allows assessing whether θ is time-invariant. The corresponding results are shown in Table Appendix S4A and suggest that our assumption of a time-invariant θ is reasonable. When distinguishing between the early and late sample periods, or estimating the

Figure Appendix S4A: Testing the linearity assumption on δ_i



NOTES: This figure was constructed by grouping movers into 20 equally sized bins according to their δ_i. The horizontal axis measures the average change in log expenditure for movers in each bin upon the move. The trend line was estimated by OLS using the 20 data points shown. The sample consists of all 549,500 individuals who are movers.

baseline model among a sample of either early movers or late movers, we find no statistically significant differences in the estimated coefficients at the 5% significance level.

Table Appendix S4A: Event-study Analysis: estimates of θ, assessing model assumptions

Model	Estimate	St. Error	N. Obs.
Years around the move:			
θ, up to 3 years around move	0.266***	0.026	3,166,863
θ, up to 2 years around move	0.257***	0.027	2,499,843
θ, up to 1 years around move	0.221***	0.028	1,633,074
Early vs. Late sample:			
θ, sample from 2006 to 2009	0.209***	0.050	2,051,943
θ, sample from 2010 to 2013	0.226***	0.042	2,095,002
Early vs. Late movers:			
θ, individuals who moved between 2006 and 2009	0.235***	0.035	4,146,945
θ, individuals who moved between 2010 and 2013	0.329***	0.039	4,146,945

NOTES: Estimates are based on equation (1), using $\log(\text{tot exp}_{it} + 1)$ as dependent variable. The number of observations varies according to each restriction imposed, as shown in the last column of the table. Standard errors are robust standard errors, clustered at individual level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix S5. Decomposition Analysis per type of care

Table Appendix S5: Additive decomposition by type of care, GP and Hospital

	Above/below median	Top/bottom 25%	Old/ Young	High/ Low $\hat{\gamma}$
Difference in overall log GP expenditure				
Overall ($\Delta\bar{y}$)	0.048	0.079	0.069	0.054
Due to place ($\Delta\hat{\bar{\gamma}}$)	0.015	0.021	0.008	0.031
Due to patients	0.033	0.059	0.061	0.023
Share of difference due to				
Place	0.312	0.260	0.112	0.573
Patients	0.688	0.740	0.888	0.427
Difference in overall log Hospital expenditure				
Overall ($\Delta\bar{y}$)	0.182	0.216	0.136	0.178
Due to place ($\Delta\hat{\bar{\gamma}}$)	0.073	0.082	-0.023	0.134
Due to patients	0.109	0.134	0.160	0.044
Share of difference due to				
Place	0.400	0.379	-0.167	0.753
Patients	0.600	0.621	1.167	0.247
Difference in overall log Pharma expenditure				
Overall ($\Delta\bar{y}$)	0.205	0.353	0.353	0.297
Due to place ($\Delta\hat{\bar{\gamma}}$)	0.056	0.112	0.112	0.139
Due to patients	0.149	0.242	0.242	0.158
Share of difference due to				
Place	0.273	0.316	0.316	0.468
Patients	0.727	0.684	0.684	0.532

NOTES: The provinces belonging to the groups of expenditure percentiles being compared in columns 1, 2, and 4 are not constant across types of care. This is because the rank of provinces in terms of GP expenditure can be very different from that in terms of hospital expenditure and the same applies to the estimated region fixed-effects. Thus, we use the specific expenditure in the type of care under analysis in order to determine which provinces belong to the groups of percentiles being compared. This analysis uses all movers and non-movers and excludes the year of move, amounting to 106,800,653 observations.

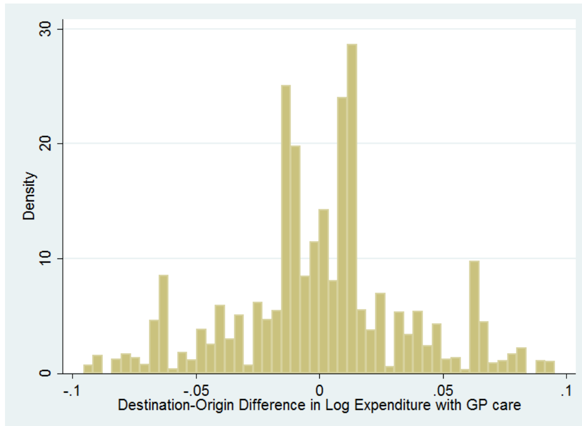
Appendix S6. Decomposition Analysis using GHOR regions

Table Appendix S6: Additive decomposition of log total healthcare expenditures using GHOR regions

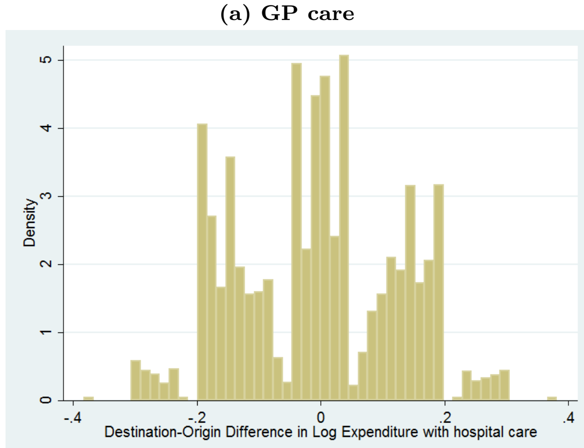
	Above/below median expenditure	Top/bottom 25% expenditure	Top/bottom 10% expenditure	High/Low share elderly	High/ Low $\hat{\gamma}$
Difference in overall log total expenditures					
Overall ($\Delta\bar{y}$)	0.047	0.093	0.098	0.085	0.095
Due to place ($\Delta\bar{\hat{\gamma}}$)	0.018	0.031	0.025	0.010	0.062
Due to patients	0.029	0.062	0.073	0.075	0.033
Share of difference due to					
Place	0.376	0.330	0.253	0.115	0.650
Patients	0.624	0.670	0.747	0.885	0.350
	(0.061)	(0.110)	(0.077)	(0.182)	(0.036)
95% CI for Patient share	[0.505, 0.743]	[0.394, 0.826]	[0.597, 0.897]	[0.529, 1.241]	[0.280, 0.420]

NOTES: Results based on equation (3) with y_{ijt} = log(total expenditures+1). The columns indicate the groups of GHOR regions being compared. The first row shows the difference in average log expenditures between the two groups of GHOR regions; the second and third rows report the difference in average log expenditures due to place and patients, respectively; rows 4 and 5 report the estimated shares attributable to supply (place) and demand (patients), respectively; finally, rows 6 and 7 show the standard errors for the patient share and the corresponding 95% confidence interval. The standard errors for the patient share are obtained by bootstrapping with 50 repetitions drawn at the individual level. The sample consists of movers and non-movers and excludes the year of move, amounting to 106,814,700 observations.

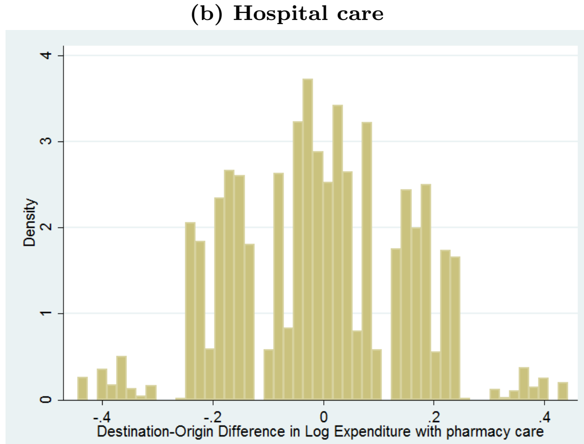
Appendix S7. Histograms of δ_i by type of care



(a) GP care



(b) Hospital care



(c) Pharmaceutical care

Figure Appendix S7: Distribution of destination-origin difference in expenditures by type of care (δ_i) *NOTES:* The figures show the histograms of δ_i, the destination-origin difference in the average log individual expenditures, by type of care. Regions are defined as provinces. The histograms were built using 50 bins and the sample of all 549,500 individuals who are movers, corresponding to 4,146,945 observations.