Statistical Appendix

A. Statistical properties of the average cost of therapies

In this section we present further cross-sectional and time-series properties of the average monthly therapy costs to ground our econometric approach.

Figure A1 shows the distribution of costs over time for each patient in hemodialysis and Figure A2 for transplant. The need to control for cross-sectional heterogeneity is evident, because some patients have large swings on monthly costs and others have more constant costs and may also show different levels. In general, costs are more volatile for hemodialysis, also having more diversity in median values of costs over months. Therefore, it is important to control for patient-specific elements when estimating the evolution of costs.

Figure A1 – Distribution of costs over time, by patient in hemodialysis (Thousand USD)



Figure A2 – Distribution of costs over time, by patient in transplant (Thousand USD)



The following figures present the evolution of the average monthly cost in different renal therapies, by averaging the total cost in each month over patients. The values exhibited are only for months with at least 2 patients. Figure A3 shows that conventional and daily hemodialysis patients have similar patterns of cost evolution, with a slight downward trend. As noted before, average costs for DHD patients are larger, especially at the beginning of treatment and at the end.

Figure A3 – Evolution of average monthly cost of hemodialysis, by patient type (Thousand USD)



Figure A4 shows that the evolution of costs in transplant is quite different, starting at a higher level between USD 17.44 thousand and 21.8 thousand) and falling quickly, explaining the large variance visible in Table 2 of the main text. These results show that transplant and hemodialysis have very different dynamics, requiring separate estimation of the parameters for each therapy. In particular, the costs have strikingly different trends.

Figure A4 – Evolution of average monthly cost of transplant, by patient type (Thousand USD)



Figure A5 shows the Autocorrelation Function and Partial Autocorrelation Function for the average monthly costs of hemodialysis. The autocorrelation at lag 1 is the correlation coefficient between current cost and the previous month’s cost; at lag 2, it is the correlation between current cost and the cost from two months ago, and so on. The Autocorrelation Function presents the temporal dependency in the cost time series, suggesting that current costs may be a predictor for future costs. Similarly, the Partial Autocorrelation Function exhibits the correlation between current costs and lagged costs, but partials out any lags in between[[1]](#footnote-1).

Note that the cost series in hemodialysis apparently does not present a unit root, because the first-order autocorrelation does not exceed 0.7 and further lags tend to fall exponentially, losing significance after 3 months. The partial autocorrelation, despite the last lags presented, also indicates that the temporal dependence structure is limited to two lags.

For the average monthly cost of transplant, the time structure is even weaker, as Figure A6 shows. However, the great fall in costs in the first 5 months in therapy might generate a misleading autocorrelation. Indeed, limiting the estimation of the autocorrelations by dropping the first four months, all lags stay within the confidence intervals. Therefore, this series requires a time-series model that considers this peculiar evolution of transplant costs.

Figure A5 – Autocorrelation and Partial Autocorrelation functions for the average monthly cost of hemodialysis



Figure A6 - Autocorrelation and Partial Autocorrelation functions for the average monthly cost of transplant



B. Model choice

The goal is to obtain the best predictor for the desired time series. The previous section used monthly averages of patient costs to provide a picture of how therapy costs evolve over time. The problem with using monthly averages to estimate the threshold is that the temporal dependency of the costs would be ignored, undermining a more efficient inference and a more informed prediction. This happens because calculating the average cost over patients one month at a time effectively ignores the cost realization of previous periods, treating each month as if it were independent of the last. As Figures A5 and A6 showed, however, renal therapy costs may be positively and substantially correlated with their past realizations, so large costs today may be good predictors for large costs in the next month.

Suppose, for instance, that a complication arises during a hemodialysis session, such that the patient needs to receive additional medication or treatment in a particular month. If follow-up exams and further medication and treatment are necessary for a few more months to make sure that the complication does not arise again, or if it is discovered that the patient needs additional care to undergo future hemodialysis sessions, then a larger cost in one month has affected the total cost in the following months. This creates a correlation between the observations and months with above-average costs tend to predict future above-average costs.

Given that past costs may be important predictors for current costs, it is standard to model such a time-series dependency by an autoregressive process. Calling total monthly cost in month as , the linear regression model

|  |  |
| --- | --- |
|  | (1) |

is an improvement over simply estimating the average of every month, because it explicitly allows for past costs to affect current costs and could be estimated by ordinary least squares, following standard assumptions over the residual term .

In equation (1), given an estimate for parameters and , say and  , is the mathematical formula do predict for each month and the uncertainty can be calculated as by how far the predictor  will miss the real value , on average, over different months.

Although the autoregressive model in (1) may suffice for many applications, in practice it is advisable to inspect further dynamics, by adding more lags and a polynomial on time trends, and so on. Including time trends allows the model to capture long-term tendencies in the data. For instance, Figure A3 shows that hemodialysis monthly costs tend to go down in time, whether in conventional or daily hemodialysis, even if there are ups and downs in between.

An additional feature of the dataset is that there are many patients, so one may model not only the time-series part of the data, but also the cross-sectional part of it. In other words, we exploit the fact that we observe the correlation between and not only once in time, but once for each patient. Thus, the full econometric model is a rich expansion of (1), yielding:

|  |  |
| --- | --- |
|  | (2) |

where indexes whether the patient is in hemodialysis or transplant and whether she is of CHD or DHD type. Index refers to the patient and , the time. Variable is the total average monthly cost that patient had in the -eth month in treatment.

Model (2) includes the lagged costs , and to control for short-run dynamics and also time trend variables and to account for the general tendency of cost data. It also features an overall constant so that each patient-therapy combination may have a different mean and a patient-specific parameter . The patient-specific parameter, also known as fixed effects or individual heterogeneity, controls for the fact that different patients may have a different expected cost of treatment. This is particularly true to renal therapies, because the performance of the treatment may vary depending on child-specific characteristics, such as weight before transplant, age and other biological predispositions.

By combining time-series and cross-section observations, a panel dataset is formed, which allows for the estimation of the econometric model (2) for the evolution of costs. It is possible to control for patient heterogeneity and generate predictions as function of past values, while also quantifying the uncertainty with confidence intervals. The dataset is a balanced panel where 30 patients are observed over an average of 20 months in hemodialysis and 27 for transplant.[[2]](#footnote-2)

As is common in the dynamic panels literature, the inclusion of lagged dependent variables in the model generates a bias when time is fixed and the within or first-differences estimator is used. Therefore, the estimator employed is the System GMM that instruments the transformed endogenous variables and the level endogenous variables by lagged levels and first-differences of the dependent variable.

C. Regression results and confidence intervals

The following tables present the different estimated specifications. Models (1), (2) and (3) differ only on the number of lags added. The model choice was based on the resulting cost dynamics, i.e., the model with the most adherent projections was chosen, given the statistical significance of the variables.

Tables C1 and C2 present the estimation results for the six specifications for each patient type – CHD and DHD – for hemodialysis. In conventional hemodialysis, it is evident that including two lags is necessary, but the third one isn’t significant. Comparing the predictions generated by models (2) and (5), we opted for model (2), illustrating the results in Figure C1 against the observed cumulative costs. For daily hemodialysis, model (4) was chosen over model (5).

Table C1 – Results for conventional hemodialysis



Table C2 – Results for daily hemodialysis



Figure C1 – Predictions and observed data for the cumulative average costs in hemodialysis (Thousand USD)



Using the same criteria for transplant, Table C3 shows that model (5) is the best model. In Table C4, no model had a satisfactory comparison between predicted and observed costs. This happens due to the steep growth in three months, from zero to USD 43.59 thousand, according to Figure C2. Therefore, the models were estimated once again including a dummy variable that equals 1 for the first three months and zero otherwise. In the end, model (1) was chosen, with the added dummy.

Table C3 – Results for transplant in CHD patients



Table C4 – Results for transplant in DHD patients



Figure C2 – Predictions and observed data for the cumulative average costs in transplant (Thousand USD)



Given the models above, the expected cost may be predicted for each patient and its confidence interval is constructed. The predictions are made without including the fixed effects, so that the average patient is represented. Then, the predictions for each patient are aggregated into monthly averages. Letting be predicted value for patient *i* in month *t* of therapy type *j*, and its standard deviation, we calculate:

Finally, the cumulative predicted cost is:

and its standard deviation is estimated by:

Thus, the uncertainty increases over time as it accumulates the variance of each period’s predictions. We use the standard deviation estimate to form the 95% confidence intervals around the predictions, yielding figures 2 and 3 in the main text.

D. Analysis of hemodialysis and transplant procedures

This section analyzes the evolution of costs for hemodialysis sessions and transplant hospitalizations. Figure 4 in the main text has shown that the temporal evolution of average monthly costs of the procedures are similar in nature to the evolution observed for total costs depicted in Figures A3 and A4.

The cost dynamics of hemodialysis sessions is pictured in Figure D1 by the autocorrelation and partial autocorrelation functions, confirming that this subgroup’s dynamics is similar to the aggregate case. However, for transplant, it doesn’t make sense to analyze a time series that is a constant starting on the third month. The temporal dependency of transplant therapy is generated by the costs from other service types, such as complications and consults.

Figure D1 – Autocorrelation and Partial Autocorrelation Functions for the average monthly costs in hemodialysis sessions



The cumulative average costs over months in treatment are presented in Figure D2. The strong rising trend for hemodialysis, specially for the daily type, contrasts with the step-function-like growth of transplant. Already in the first month, the transplant patient generates a USD 19.62 thousand – USD 21.79 thousand cost, which is only surpassed by hemodialysis sessions on the 7th or 10th month. However, hospitalization costs in transplant therapy do not repeat over the following months, so the cumulative cost remains constant, while hemodialysis session costs keep growing.

Figure D2 – Evolution of the cumulative average costs for hemodialysis sessions (left) and transplant hospitalization (right), by patient (Thousand USD)



The following tables show the estimated equations for hemodialysis and Figure D3 compares the predictions with observed data. The model chosen for CHD patients is column (5) on Table D1, because it a parsimonious representation of the dynamics and generated a good mean prediction. For DHD type patients, the only model to return statistically significant coefficients was model (5) on Table D2, which also generates data-compatible predictions.

Table D1 – Results for conventional hemodialysis sessions



Table D2 – Results for daily hemodialysis sessions



Figure D3 – Predictions and cumulative average cost data for hemodialysis sessions (Thousand USD)



Since no time series predictions can be made for transplant hospitalization costs, the cumulative average cost was estimated by calculating the running sum of average (over patients) monthly costs. The confidence interval was constructed by estimating with the mean’s standard deviation. The results for conventional hemodialysis patients are in Figure D4, while Figure D5 has those for daily hemodialysis patients.

Similarly to the previous case, there is threshold month such that hemodialysis’ cumulative costs overtake transplant’s costs. For conventional hemodialysis patients, that threshold is around 15 months, when the lower bound of the cumulative predicted costs surpass the upper bound for transplant’s upper bound. For DHD patients, this threshold lies between 8 and 9 months, due to the higher frequency of sessions. The threshold month is associated with a USD 30.51 thousand cumulative cost in therapy.

Figure D4 - Comparison of cumulative costs: patients from conventional hemodialysis (Thousand USD)



Figure D5 – Comparison of cumulative costs: patients from daily hemodialysis (Thousand USD)



1. For further details, refer to Enders (2014). [↑](#footnote-ref-1)
2. Thus the framework chosen is that of a dynamic panel with asymptotics in the cross-sectional dimension and fixed time length. The econometric errors $ε\_{it}^{j}$ are independently distributed over patients, but robust standard errors are employed for heteroscedasticity and arbitrary serial correlation. [↑](#footnote-ref-2)