CONSOLIDATING BUREAUCRATIC SUPPLY? THE CASE OF ITALIAN HEALTH CARE (1982-2008)

SILVIA FEDELI, MICHELE SANTONI

Silvia Fedeli* and Michele Santoni**

Consolidating bureaucratic supply?
The case of Italian health care (1982-2008)

Abstract: The Italian publicly run National Health Service system is characterised by a rigid bureaucratic structure organised at the regional level and mostly financed by central government's transfers. In this paper, we first develop a two-stage stochastic model showing how the preferences of a regional government for differentiated health services can shape the regional health care organisation. The model shows that a regional government prefers the supply by one consolidated agency if the health services are perceived as complements, whereas it prefers the supply by two independent agencies if the health services are perceived as substitutes. Second, we use this prediction for an empirical analysis of the organisational structure of the Italian health care system. On the basis of a regional-level panel data set spanning from 1982 to 2008, we test whether and how the dimension of the basic health units which we take as a proxy measure for health care decentralisation is affected by complementary/substitutable health services demanded at the regional level. Our estimates confirm the predictions of the model.

*Università degli Studi di Roma 'La Sapienza',
Facoltà di Economia,
Dipartimento di Economia Pubblica,
Via del Castro Laurenziano 9,
00161 Roma,
ITALY.
Tel. +39 0649766399,
Fax +39 064461964
E-mail: silvia.fedeli@uniroma1.it

**Università degli Studi di Milano,
Dipartimento di Economia, Management e Metodi Quantitativi,
Via Conservatorio 7,
20122 Milano (MI),
ITALY.
Tel- +39 0250321511
E-mail: michele.santoni@unimi.it

1. Introduction

The aim of this paper is twofold. First, it aims at presenting a simple public choice theoretical model, which tries to explain, at least partially, the link between politicians' preferences, health care organisation and bureaucratic behaviour (see e.g. Sørensen, 2006, for a similar approach). Second, it aims at testing the model's predictions on data from the Italian public health care system.

The Italian National Health Service, NHS (SSN - Servizio Sanitario Nazionale) is a publicly run universal health care system that provides comprehensive health insurance coverage and uniform health benefits to the whole population, subject to user charges for certain services. The system, now organized on three levels of government (i.e. national, regional, and municipal government), was originally a highly centralized system. An increased degree of decentralization for hospital care had already begun in 1972, when the regions with ordinary bylaw were given the direct responsibility of managing hospital care assistance (see the Presidential decree D.P.R. 14/01/1972, n. 4). The national government was only left with residual competences (such as the definition of hospital care national standards which the regions should broadly refer to). However, it was only in 1978, with the law n. 833/78, that the supply of health care services took the current shape. On the expenditure side, each region provides primary care by contracting General Practitioners (GPs), who also refer patients to specialist care and hospitals. GPs are independent self-employed practitioners, who are paid on a capitation basis.[1] Decisions related to the supply of hospitals, ambulatory treatments, diagnostic and laboratory tests, specialist care, drugs, medical appliances and glasses are instead taken by the regional government in accordance with the national standards established by the *Istituto Superiore di Sanità*, an administrative body of the Health Minister. Regional governments can outsource the delivery of medical health services, including hospital care, to accredited private providers, with patients free to choose between public or private providers. Patients, at their own expenses, are also free to buy private health insurance and to receive treatment at non-contracted private hospitals and to consult private outpatient specialists.[2] On the financing side, since 1978 the amount of funds financing the Italian NHS was fixed by the central government, allocating resources for health care expenditures on a regional basis. Patients co-pay health care services at a price called *tickets*. These co-payments are determined by each region according to income, age, health conditions and other individual characteristics of patients. The separation between revenue raising responsibilities of the central government and expenditure responsibilities of the regional governments resulted in continuous

[1] There are two types of GPs, providing either paediatric care to people aged under fourteen or care for adults. Their number, role and pay are periodically determined by separate national contracts, which are broadly related to the regional population.

[2] Accordingly, individuals often opt to supplement their public health insurance with the purchase of private insurance and/or private services.

increases of the Italian public health care expenditures during the 1980s. As a result, the regions were regularly running budget deficits and accumulating large hidden regional debts, which were subsequently repaid by the state. Periodic attempts to stop this dynamics led to reforms of the Italian NHS's financing mechanisms.[3] However, state transfers from general taxation still remain the main source of funding, although at a decreasing rate.[4] Despite the reforms giving substantial financial autonomy to the regions since 1993, health care resources seem independent of the fiscal capacity of each region, being they based on historical spending and health care needs (see Francese and Romanelli, 2011, p. 8). Actually, since the '90s, the Italian regions held the expectation of the central government's intervention to bail out their past deficits (see Bordignon and Turati, 2009).

Given this institutional context, the basic idea of this paper is that the regional government's demand for specific health care services is a key determinant of the organisational design of the bureaucratic supply for these services. This is because the regions, by means of their ruling politicians, can exert the power of designing or modifying the bureaucratic organisation (see Moe, 1984, p. 761).[5] The regional governments can in principle choose whether to deal with either a decentralised bureaucracy (e.g. with health care services spread/diffused in the regional jurisdiction) or with a centralised one for the public provision of differentiated health services (such as recovery in hospital, visits to specialists and general practitioners, long-term care). This is accomplished through a network

[3] Initially, health care was fully financed by the state using general taxation revenues, while subsequently payrolls and compulsory social security contributions for health and a "health-tax" on self employed were introduced. The 1992/93 Italian NHS reform (see the Legislative Decrees n. 502/1992 and n. 517/1993) stated that the regions incurring budget deficits could rely on payroll contributions, previously collected by the central government, and earmarked for health care needs. The regions were also allowed to raise contribution rates (called *tickets*). These options, however, were never actually used by any of them. Therefore, the 1990 bail out of the regional health care deficits continued to affect regional funding. The 1997 fiscal reform aimed both at eliminating regional disparities in payroll tax contributions rates, and at introducing fiscal decentralisation (see the Legislative Decree n. 446/1997). The 2000 fiscal reform stated that the central government's health care funding would have been discontinued in 2001, with funding becoming a full regional responsibility (see the Legislative Decree n. 56/2000). The 2000 reform introduced the notion of LEAs, namely of essential and uniform assistance levels. LEAs include the essential benefit package covering all medical care defined as necessary, appropriate and cost-effective. The reform was based on the following principles: first, LEAs should be defined together with their financing, namely the capitation rate of public spending granted to each citizen. Second, the delivery of LEAs is given to the regional health authorities. Third, regions have to rely on regional taxes for funding both health care and other regionally-funded activities. Finally, the general objectives and fundamental principles of the Italian NHS are maintained at the national level.

[4] Financing from general taxation decreased from 41% in 1990 to 38% in 1998 and then increased to 46% in 1999. In 1994, social contributions (health dues) covered 52.2% of financing, while in 1986 they covered 62.5%. More specifically, with the 1997 fiscal reform the public financing of the Italian NHS was based on the following sources: a newly introduced income-type value added tax on regional productive activities (IRAP); user charges and tariffs paid directly from consumers to the local health management agencies (ASL and ASO); regional shares of general central taxation (personal income tax, VAT, excise on oil products), regional surcharges on personal income tax; an inter-regional equalization fund (transferring revenues collected from the most affluent regions to the poorest ones). In 2001, some regions increased both the IRAP tax rate and the surcharge personal income tax rate to finance past health deficits. Many regions used their autonomy for increasing tariffs and user charges.

[5] The standard public choice approach models the interaction between politicians and bureaucrats under the assumption of exogenous institutional structures. More specifically, most of the literature originating from Niskanen (1971) assumes one government dealing with one bureau, see for example Janssen et al. (2003). Fedeli (1999) considers instead one government dealing with two competing bureaus.

of independent local health management agencies: the ASLs - Aziende Sanitarie Locali, previously called USL, - and hospitals' trusts, the ASOs –Aziende Sanitarie Ospedaliere, introduced since 1995 Both the ASLs and the ASOs are independent bodies with full autonomy as regards their legal, organisational, administrative, financial, accounting, managerial and technical responsibilities. The regional governor appoints each ASL's and ASO's general director under private contracts, which are renewable every five years. The ASLs provide care either directly (through directly managed hospitals and territorial agencies offering primary care, ambulatory care, and other services) or indirectly by paying other public and accredited private health care providers (such as hospital, nursing homes and laboratory under contract to the NHS).[6] The ASOs are highly specialised hospital structures with a national profile, which also act as reference structures for emergency services within a local jurisdiction.

Given the bureaucratic organisation of the Italian health care system, public health managers actually bargain with the regional government for their budget, while the total amount of resources devoted to regional health care is basically determined at the national level on a per-capita basis. Therefore, a rational government, when choosing how to organise health care services at the regional level, will try to anticipate the effects of this decision on its negotiation with the public management over a total amount of given resources. Our stylised model, with one regional government and up to two bureaucratic health care agencies, predicts that a regional government demanding differentiated health care services that it perceives as being complementary (say, the demands for patients' recovery in hospital and for general practitioners assistance) is more likely to organise their supply into one single consolidated agency. On the contrary, a regional government demanding health care services perceived as being substitutes (say, the demands for specialists' visits and for general practitioners assistance) is more likely to organise the supply of these services into two independent agencies. The intuition for this result is that in both cases, and other things being equal, the government can maximise the agency outputs, while minimising bureaucratic rents (thus maximising political rents) at the same time.

These theoretical predictions are empirically tested for the Italian regions over the 1982-2008 years In particular, we test whether the regional governments' demand for differentiated health care activities influences the dimension of independent local health management agencies within a region (i.e. the number of ASLs plus ASOs, once controlling for their size in terms of population). By controlling both for the potential endogeneity of our proxy measures for differentiated health care activities, and for time-invariant unobservable region-specific components, and for common time-specific shocks our estimates show that the demand for substitutable (complementary) services lowers

[6] See Maio and Manzoli (2002) for further details.

(increases) the dimension of local health management agencies observed on average at the regional level.

This paper is related at least to two strands of the literature. First, as previously mentioned, the paper is related to the public choice literature studying the interaction between politicians and bureaucratic agencies delivering public services, when the politicians are able to influence the organizational form of bureaucratic supply (see e.g. Moe, 1984, Fedeli and Santoni, 2006, Sørensen, 2006; and Bates et al., 2011 for local health services).[7] Second, the paper is somehow related to the empirical literature seeking to understand a government's incentives for hospital merging in publicly run health care systems (see e.g. Bloom et al., 2010, and Gaynor et al., 2011, for the English NHS case). However, whereas this latter literature ultimately aims at assessing whether government's lead hospital merging is good or bad for health care outcomes (e.g. management performance, productivity, clinical quality and waiting times), this paper is only interested in understanding whether and how the differentiated nature of health care services demanded by the Italian regional governments may influence their choice on the number of independent management agencies.

The plan of the paper is as follows. Section 2 presents the theoretical model of the regional governments' choices related to the health care system. Section 3 empirically tests the model with Italian data. Section 4 concludes with final remarks.

2. The model

The model presented in this section is a stochastic version of Fedeli and Santoni (2006) applied to health care issues. Consider a regional government willing to offer two differentiated health care services to its local population. These services can be produced either separately by two independent bureaus, each specialising in the supply of one service, or jointly by one single consolidated bureau. The bureaus are, for example, two local public health authorities (called ASL and ASO: *Aziende sanitarie locali e ospedaliere*) supplying, say, primary care (e.g., visits to general practitioners) and secondary care (e.g., visits to specialists working in hospitals and recovery and medical assistance in hospital).[8] The government, on the basis of the people's preferences, can perceive these services as being either substitutable or complementary sources of medical care (see Atella and Deb, 2008, Fortney et al., 2006, and Scott, 1996, for a discussion). The bureaucratic structure, the level of health care services supplied, and – possibly – the political and bureaucratic rents observed in equilibrium are assumed to be the outcome of a two-stage sequential game. Given the amount of resources

[7] Bates et al. (2011) analyse the decision of two or more independent local communities in Connecticut to consolidate local public health services, with no explicit role given to bureaucracies in the process.

[8] Alternatively, the two bureaus can be interpreted as one local government (municipalities, called *Comuni* in Italy) and one public health authority supplying, say, long-term care services to the ageing population or mental healthcare services by offering house care/nursing homes and hospital care.

devoted to the regional health care budget by the state, at stage one the regional government chooses whether the services should be offered jointly by one integrated independent public health authority or separately by two independent authorities. In making this choice, the government is concerned about both the level of health care services and the political rents it can extract from the budgetary process. In particular, the regional government evaluates health care services on the basis of the people's/its own electoral constituency's preferences. It is assumed that a random shock to the demand for the publicly-provided health care services occurs after stage one. The government has knowledge of the distribution of the shock as of stage one, while the realised value of the shock becomes public information at the beginning of stage two. At stage two, there is a compliance game à la Miller (1977) between the government and the health authority(ies). The government chooses the bureaucratic budget, namely the fraction of the resources devoted to the production of each health care service, while the bureau(s) simultaneously chooses the fraction of this budget that is actually allocated to producing the service. Following Niskanen (1971, p. 29), the bureau(s) evaluates the health care services it supplies on the basis of the government's demands, while aiming at gaining rents from this activity at the same time. The government and the bureau(s) are assumed to be the residual claimants of the resources allocated to health care services (see Shleifer and Vishny, 1994, for a discussion). Thus, in equilibrium, their compliance levels determine both the actual amounts of health care services provided and the level of political and bureaucratic rents.

2.1 Government and bureau(s) preferences with shocks to the demand for health care services

Before introducing the shock, this section presents the government's and the bureau's preferences in the deterministic case. The government's preferences are:

$$MG = \underbrace{\alpha[Q_1 + Q_2] - \left[\frac{\beta\left(Q_1^2 + Q_2^2\right)}{2}\right] - 2\gamma Q_1 Q_2}_{\text{government's evaluation function of goods 1 and 2}} + \underbrace{\sum_{i=1}^{2}(R - B_i)}_{\text{political rents}} \qquad (1)$$

In equation (1), the first term represents the government quadratic evaluation function of the bureaucratic production of the two differentiated health care services Q_i , i=1,2, where $\alpha > 0$, $\beta > 0$, $\beta^2 > 4\gamma^2$ and $\beta > 2|\gamma|$ (see Singh and Vives, 1984, for a discussion). We assume that this term reflects the preferences for health care services of the patients (i.e. the government electoral constituency, the people) that can be inferred, say, from surveys. The second term represents the political rents that the government can obtain from the production of each service. These rents are equal to the difference between the exogenous and symmetric amount of resources officially allocated to the production of each service, R, and the budget the regional government assigns to the bureau(s) for this purpose, B_i.

R is considered as an exogenous variable (coming from the central government as a transfer);[9] B is endogenously determined as the outcome of a compliance game with the bureau(s) (see below). More specifically, $B_i = Rg_i$, where $g_i \in [0,1]$ is the share of resources the government allocates to the bureau(s) for producing the health care service i=1, 2. Government's maximisation of (1) with respect to Q_i yields

$$V_i^M = \alpha - \beta Q_i - 2\gamma Q_j \qquad (2)$$

In equation (2), $V^M{}_i$, i={1,2}, i≠j, represents the government willingness to pay for the health care service Q_i when it also demands Q_j. For $\gamma<0$, the two services are complements. This is the case, for example, if the willingness to pay for general practitioners' assistance increases when the demand for patients' recovery in hospital rises. For $\gamma>0$, they are substitutes. This is the case, for example, if the willingness to pay for visits to specialists in hospital falls when the demand for visits to the primary care physicians rises (see Atella and Deb, 2008, for Italian evidence supporting this claim). For $\gamma=0$, the two services are independent.

Turning to the bureau(s) preferences, with two independent public health authorities they write

$$MH^{12}{}_i = \underbrace{V_i^M Q_i}_{\substack{\text{bureau -i's evaluation}\\ \text{of its own services}}} + \underbrace{B_i(1-h_i)}_{\text{bureau's rents}} \qquad i=1,2 \quad \text{and } i \neq j \qquad (3)$$

In equation (3) the superscript "12" indicates 1 government dealing with 2 bureaus, whereas the subscript i indicates the bureau producing good i. In the case of one single consolidated bureau, its payoff corresponds to the sum of the two individual bureaus' payoff, namely

$$MH^{11} = \sum_{i=1}^{2} MH^{12}{}_i = \underbrace{\sum_{i=1}^{2} V_i^M Q_i +}_{\substack{\text{bureau's evaluation}\\ \text{of its own services}}} + \underbrace{\sum_{i=1}^{2} B_i(1-h_i)}_{\text{bureau's rents}} \qquad (4)$$

In equation (4) the superscript "11" denotes 1 government dealing with 1 bureau supplying both types of health care services. As previously mentioned, following Niskanen (1971) we assume that the bureau(s) evaluates its own activity on the basis of the government's preferences. This implies that, in equations (3) and (4), V_i^M reflects the government's willingness to pay for service i=1, 2. (This is given by equation 2 for the deterministic case, see below for the stochastic case). The bureau(s) produces the health care service i=1, 2 by devoting a fraction $h_i \in [0,1]$ of the budget B_i it obtains from the government, while keeping the residual budget as its own rents. Production of bureaucratic activities occurs under a constant-return-to-labour technology. This implies the total cost function $TC_i = cQ_i$, where $c>0$, with $c<\alpha$, represents the minimum (symmetric) cost of producing good i. This

[9] In Italy, as mentioned in the introduction, health care expenditure responsibilities are decentralised at the regional level, whereas revenue raising responsibilities are still in practice highly centralised.

cost function can be seen as a good proxy of the short-run cost function for health care services that are heavily labour intensive. We will exploit the cost function in solving the compliance game below.[10]

Let us now consider the demand shocks. Following Klemperer and Meyer (1986), the random shock to the demand curve is modelled in three different ways. First, it is an intercept shock to the residual demand curve for each service (namely, a shock affecting linearly the reservation price given the demand curve for the other service). Second, it is a shock affecting the demand slope. This latter random disturbance is modelled such that the rotation of the demand curve about the vertical intercept occurs with both the government's reservation price and the degree of differentiation between services remaining unchanged. Such intercept and slope shocks might reflect, for example, the shocks to the demand for primary and secondary care that are associated with undocumented and documented migrant workers inflows and outflows in the region. Third, it is a disturbance affecting the degree of differentiation between health care services. For example, shocks to the electorate's tastes may induce the government to revise its trade-off between primary care and secondary care. The next section determines the sub-game perfect Nash equilibrium of the sequential game, when the model is solved by backward induction under alternative assumptions as regards the demand shock.

2.2 The choice of bureaucratic organisation with additive uncertainty

Assuming a linear random shock to the vertical intercept, the demand for health care services is

$$V_i^M = \alpha + \varepsilon - \beta Q_i - 2\gamma Q_j \qquad (2')$$

where i, j={1, 2} and i≠j; ε is a random variable with $E(\varepsilon) = 0$ and $E(\varepsilon^2) = s^2$. Following Klemperer and Meyer (1986), we assume that the support of the shock is small enough that prices and health care service levels (thus compliance levels) are always positive. The shock occurs after the government has already chosen the bureaucratic structure at stage one; its realised value becomes publicly observable henceforth.

Solving by backward induction, consider now stage two of the game, when the realisation of the demand shock is common knowledge and the bureaucratic structure is given. Assuming that the government allocates the share $g_i \in [0,1]$ of resources R to the bureau(s) as health care budget and that the bureau(s) actually devotes a fraction $h_i \in [0,1]$ of this budget to producing the services, given that total production costs are TC_i=cQ_i, we can write $h_i g_i R = cQ_i$, implying $Q_i = h_i g_i R / c$, for i=1, 2. Using this fact, substituting equation (2'), into equations (3) and (4), the bureau(s) preferences can be

[10] We abstract from the presence of fixed costs that would generate economies of scale. Had we assumed that fixed costs are determined by a unilateral long-run decision of the government, their presence would not alter Proposition 1 and 2. However, the conditions for consolidation there derived would only become necessary.

formulated in terms of both compliance strategies g_i and h_i (see Miller, 1977) and the value of the shock ε:

$$MH^{12}{}_i = (\alpha+\varepsilon)\left(\frac{g_i h_i R}{c}\right) - \beta\left(\frac{g_i h_i R}{c}\right)^2 - 2\gamma\left(\frac{g_i h_i R}{c}\right)\left(\frac{g_j h_j R}{c}\right) + Rg_i(1-h_j) \qquad (3')$$

$$MH^{11} = (\alpha+\varepsilon)\left[\left(\frac{g_1 h_2 R}{c}\right) + \left(\frac{g_2 h_2 R}{c}\right)\right] - \beta\left[\left(\frac{g_1 h_2 R}{c}\right)^2 + \left(\frac{g_2 h_2 R}{c}\right)^2\right] - 4\gamma\left(\frac{g_1 h_2 R}{c}\right)\left(\frac{g_2 h_2 R}{c}\right) + R\left[g_1(1-h_1) + g_2(1-h_2)\right] \qquad (4')$$

i, j={1, 2} i≠j

Equations (3') represent the preferences of the two independent public health authorities, while equation (4') gives the preferences of one single independent and consolidated authority . Similarly, using equation (1), after appropriate substitutions the government's payoff in terms of compliance is

$$Mg = \alpha\sum_{i=1}^{2}\left(\frac{g_i h_i R}{c}\right) - \frac{\beta}{2}\sum_{i=1}^{2}\left(\frac{g_i h_i R}{c}\right)^2 - 2\gamma\prod_{i=1}^{2}\left(\frac{g_i h_i R}{c}\right) + \sum_{i=1}^{2}\left[R(1-g_i)\right] \qquad (5)$$

Note that the realised value of the random shock has no effect on the government's payoff. However, by affecting the equilibrium compliance levels, it will influence the value function of the government's payoffs in the compliance game, which we consider in turn. The government chooses g_i and g_j by maximising equation (5), given h_i and h_j. When facing one single independent health authority, the government simultaneously chooses h_i and h_j so as to maximise equation (4'), given g_i and g_j. The symmetric Nash equilibrium solution for this case is

$$\hat{g}_1^{11} = \hat{g}_2^{11} = \frac{[\alpha^2 - (\varepsilon - c)^2]}{4R(\beta + 2\gamma)}$$
$$\hat{h}_1^{11} = \hat{h}_2^{11} = \frac{2c}{\alpha + c - \varepsilon} \qquad (6)$$

When there are two independent health authorities, each chooses simultaneously its own compliance variable h_i by maximising equation (3'), given the compliance levels of both the other bureau h_j and the government g_i and g_j. The symmetric Nash equilibrium solution for this case is

$$\hat{g}_1^{12} = \hat{g}_2^{12} = \frac{(\alpha - c + \varepsilon)(\alpha\beta + (\beta + 2\gamma)(c - \varepsilon))}{4R(\beta + \gamma)^2}$$
$$\hat{h}_1^{12} = \hat{h}_2^{12} = \frac{2c(\beta + \gamma)}{(\alpha\beta + (\beta + 2\gamma)(c - \varepsilon))} \qquad (7)$$

At stage one the government knows the distribution of the random shock affecting the equilibrium compliance and its own payoffs at stage two. Thus, when choosing the bureaucratic organisation, it must compute the expectations of its own payoff function in both of these cases. By substituting equation (6) back into (5) and taking expectations, the government's expected payoff under bureaucratic consolidation is

$$E(\hat{M}G^{11}) = 2R + \frac{(\alpha - c)^2}{4(\beta + 2\gamma)} + \frac{s^2}{4(\beta + 2\gamma)} \qquad (8)$$

since $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{E}(\varepsilon^2) = s^2$. Equation (8) shows that additive uncertainty about demand increases the government's expected payoff over and above its deterministic payoff by a term that is proportional to the variance of the shock. Moreover, uncertainty distorts the government's trade-off between utility from the political rents and utility from the outputs in favour of the former. Appendix 1 discusses further this result, by providing explicit comparisons with the deterministic case.

Similarly, by substituting (6) back into (5) and taking expectations, the government's expected payoff when negotiating with two independent bureaus is

$$E(\hat{M}G^{12}) = 2R + \frac{(\alpha - c)^2(\beta + 2\gamma)}{4(\beta + \gamma)^2} + \frac{s^2(\beta + 2\gamma)}{4(\beta + \gamma)^2} \qquad (9)$$

Uncertainty (through the effect of the variance of the shock) raises the government's expected payoff over and above its deterministic level. By comparing (8) and (9), it follows that at stage one of the game the government chooses bureaucratic consolidation if and only if

$$E(\hat{M}G^{11}) - E(\hat{M}G^{12}) = -\gamma[(\alpha - c)^2 + s^2]\left[\frac{(2\beta + 3\gamma)}{4(\beta + \gamma)^2(\beta + 2\gamma)}\right] \qquad (10)$$

yielding

Proposition 1. *Suppose that the residual demand for health care services is subject to a continuous random shock to the vertical intercept, whose realised value, but not its distribution, is unknown to the government when choosing the bureaucratic organization. Then, the government has the incentive to choose one consolidated bureau when the goods are complements and two independent bureaus when they are substitutes. This optimal incentive is increasing in the variance of the shock.*

Uncertainty on the demand side increases the government's optimal incentive to bureaucratic consolidation or independence in the same direction as pointed out by Fedeli and Santoni (2006) for the deterministic case. The same qualitative effect is obtained when the random disturbance affects the slope of the inverse demand curve for each of the two health care services, without affecting the degree of substitutability between services. Appendix 2 reports the detailed derivation of this solution.

2.3 The choice of bureaucratic organisation with uncertainty on the degree of substitutability

This section considers uncertainty as regards the degree of substitutability between the two services. This implies the following demand function for health care services:

$$V_i = \alpha - \beta Q_i - 2(\gamma + \varepsilon) Q_j \qquad (2''')$$

The random shock is now assumed to be uniformly distributed $\varepsilon \sim U(\varepsilon_L, \varepsilon_H)$, with ε_L <0 and ε_H >0.[11] In this case, the demand for health care services is positive provided that $\beta^2 > 4(\gamma+\varepsilon)^2$, which we assume. In turn, this restriction implies $-(4\beta+\gamma) < \varepsilon < 4\beta - \gamma$, as long as β>2|γ| for the second order condition to be satisfied. Solving as usual the model by backward induction under current-stage full information (see section 2.2 above for details), the symmetric Nash equilibrium solution of the compliance game with one independent health authority is

$$\begin{aligned} \tilde{g}_1^{11} = \tilde{g}_2^{11} &= \frac{(\alpha - c)\left[(\alpha+c)(\beta+2\gamma)+4\alpha\varepsilon\right]}{4R\left[\beta+2\gamma+2\varepsilon\right]^2} \\ \tilde{h}_1^{11} = \tilde{h}_2^{11} &= \frac{2c(\beta+2\gamma+2\varepsilon)}{(\alpha+c)(\beta+2\gamma)+4\alpha\varepsilon} \end{aligned} \qquad (11)$$

Substituting (11) back into (5), yields the value function for the government

$$\tilde{M}G^{11} = 2R + \frac{(\alpha - \mathrm{c})^2(\beta+2\gamma)}{4(\beta+2\gamma+2\varepsilon)^2} \qquad (12)$$

Similarly, the symmetric Nash equilibrium solution with two independent health authorities is

$$\begin{aligned} \tilde{g}_1^{12} = \tilde{g}_2^{12} &= \frac{(\alpha - c)\left[\alpha\beta+c(\beta+2\gamma)+2\alpha\varepsilon\right]}{4R\left[\beta+\gamma+\varepsilon\right]^2} \\ \tilde{h}_1^{12} = \tilde{h}_2^{12} &= \frac{2c(\beta+\gamma+\varepsilon)}{\alpha\beta+c(\beta+2\gamma)+2\alpha\varepsilon} \end{aligned} \qquad (13)$$

Substituting (13) back into (5), the government's value function at stage two is

$$\tilde{M}G^{12} = 2R + \frac{(\alpha - \mathrm{c})^2(\beta+2\gamma)}{4(\beta+\gamma+\varepsilon)^2} \qquad (14)$$

[11] The special case in which the support of the shock is as follows: $\varepsilon_L = -\varepsilon_H > 0$ is analysed below.

At stage 1 the government takes the expectation of equations (12) and (14). Thus, it chooses bureaucratic consolidation if and only if (see Appendix A.3 for a derivation)

$$E(\tilde{M}G^{11}) - E(\tilde{M}G^{12}) = -\frac{(\alpha - c)^2(\beta+2\gamma)}{4}\left[\frac{\beta(\gamma+\varepsilon_H+\varepsilon_L)+3\gamma(\gamma+\varepsilon_H+\varepsilon_L)+3\varepsilon_H\varepsilon_L}{(\beta+\gamma+\varepsilon_H)(\beta+\gamma+\varepsilon_L)(\beta+2\gamma+2\varepsilon_H)(\beta+2\gamma+2\varepsilon_L)}\right]>0 \qquad (15)$$

The sign of equation (15) depends on the sign of: $-[\beta(\gamma+\varepsilon_H+\varepsilon_L)+3\gamma(\gamma+\varepsilon_H+\varepsilon_L)+3\varepsilon_H\varepsilon_L]$. If we further assume that $\varepsilon_L = -\varepsilon_H > 0$, it follows that $E(\varepsilon)=0$, $E\left(\frac{\gamma+\varepsilon}{\beta}\right)=\frac{\gamma}{\beta}$ with $\varepsilon \sim U(-\varepsilon_H, \varepsilon_H)$. In this special case, the government expects that the degree of product differentiation is on average equal to γ. Thus, it turns out that: Sign of $E(MG^{11}) - E(MG^{12})$ = Sign of $\left[\varepsilon_H{}^2 - \gamma\left(\gamma+\frac{2}{3}\beta\right)\right]$, yielding

Proposition 2: *Suppose that the demand for health care services is subject to a continuous random shock affecting the degree of substitutability between the goods with uniform distribution and zero mean, whose realised value, but not its distribution, is unknown to the government when choosing the bureaucratic organization. Then, the government will* unambiguously *choose bureaucratic consolidation, if and only if:* $4\beta-\gamma > \varepsilon_H \equiv \varepsilon_H{}^* > \sqrt{\gamma\left(\gamma+\frac{2}{3}\beta\right)}$ *for* $\gamma>0$ *(if* $4\beta-\gamma > \varepsilon_H \equiv \varepsilon_H{}^* > 0$ *otherwise).*

In Proposition 2 the first inequality follows from the restrictions imposed on the demand parameter, while the second inequality follows from equation (15). Proposition 3 gives the cut-off values of the upper limit of the uniform distribution such that the government prefers consolidation of health care services in a unique public health authority even when health care services are substitutes. Some numerical example can help us interpreting this condition. Let β=1, implying -1/2<γ<1/2 from the parametric restrictions on government's demand. For γ=1/3, the government chooses bureaucratic consolidation if the inequality $3.67>\varepsilon_H>0.57$ holds. For γ=1/4, the relevant inequality is $3.75>\varepsilon_H>0.48$. Thus, the higher is the expected degree of substitutability γ, the more restrictive becomes the condition reported in Proposition 2.

Generalising the implications of the theoretical model, it is possible to derive two testable predictions for the Italian public health care system. First, the government has an incentive to consolidate the bureaucratic health care organization in a smaller number of large independent agencies (i.e. one in our model), when it perceives that the services are complementary, other things being equal. If these services are perceived as substitutes, the government prefers instead bureaucratic supply by a larger number of small independent agencies (i.e. two in our model). In practice, this should imply a positive (negative) correlation between the supply of substitutable (complementary) health care services and the dimension of local public health authorities. Note that this argument for consolidation depends only on the nature of the health care services demanded by the government, not on the government's willingness to obtain cost-savings by exploiting scale economies, as it is often pointed out in the literature.[12] To be sure, at least until 2008, it is unlikely that scale economies were driving the Italian regional governments' decision to consolidate local health authorities, as long as cost-savings were not a major concern for regional governments, due to the presence of a soft budget constraint related to the NHS financing system(see Bordignon and Turati, 2009).

Second, if the demand side shock affects either the vertical intercept or the slope, higher uncertainty raises the incentives to organise complementary (substitutable) health care services by reducing (increasing) the dimension of independent local public health management agencies, other things being equal. In practice, this shock should strengthen the positive (negative) correlation between the supply of substitute (complementary) health care services and the dimension of local health management agencies. If the demand shock affects the degree of differentiation between health care services, however, positive realisations of the demand side shock, under some conditions, may induce consolidation of health care supply, other things being equal. The next section aims at testing these predictions by considering health care data for the Italian regions over the 1982-2008 years.

3. Testing the model for the Italian health care services (1982-2008)

In this section we present an empirical test of the theoretical model presented in section 2 above. The test is based on a yearly panel dataset containing health care data and economic, political and socio-demographic variables for all the Italian regions from 1982 to 2008. The main theoretical prediction of the paper is that a regional government has an incentive to consolidate (demerge) public health agencies, when it perceives their services as complementary (substitutable). Moreover, under some conditions, these incentives may be reinforced by shocks to the demand for health care services. In order to empirically test this prediction, dealing with longitudinal panel data, we refer to the following model:

$$y_{it} = \gamma Y_{it} + \beta X_{it} + \nu_{it} \quad (16)$$

In equation (16), the dependent variable y_{it} - *asl_aso_pop* - is the average regional dimension, measured as a share of regional population (i.e., per 1,000 inhabitants) of the local health care management agencies (i.e. the ASLs *plus* the hospital trusts, the ASOs), observed in region i=1,…, 20 at time t=1982,…, 2008. Recall that the local health agencies (the ASLs) form the basic elements of the Italian NHS for providing health care services to the population. Each ASL is financed from its region under a global budget with a weighted capitation system. In addition, since 1995, a number of

[12] Bates et al. (2011), for example, point out the trade-offs between scale economies and the loss of decision-making authority faced by a local government considering consolidation of local health care services with other local governments. The original argument that consolidation may depend on the complementarity/substitutability of goods is due to Horn and Wolinsky (1988).

public hospitals have been qualified as "hospital trusts" (ASO). Hospital trusts work as independent providers of health services and have the same level of administrative responsibility as ASL.

Y_{it} is a vector of variables capturing differentiated health care services. The nature of these variables is a source of potential endogeneity bias. Therefore, we assume that Y_{it} is a vector of endogenous variables included as covariates, and these variables are allowed to be correlated with the errors ν_{it}, i.e., $E(Y_{it}, \nu_{it}) \neq 0$. γ is a vector of coefficients related to Y_{it}. Among the likely endogenous variable that might influence y_{it}, we consider three empirical proxies for the complementary and the substitutability of the health care services.

The first variable is **Ratio_med,** defined as the ratio between the number of general practitioners per 1,000 inhabitants and the number of physicians working for NHS hospitals per 1,000 inhabitants. In this respect, we already noticed that Atella and Deb (2008) find that in Italy the general practitioners and physicians/specialists working for NHS hospitals are substitute sources of medical care from the point of view of the patients. As a consequence, we would expect that an increasing value of this ratio (e.g. more general practitioners given the specialists) should be associated with an increasing demand for substitute health care services and thus, according to our theoretical model, should be positively correlated with our dependent variable.

On the basis of the institutional characteristics of the Italian health care service system, we identify a second empirical proxy capturing complementary health care services. In Italy, as well as in other EU countries, a patient's recovery in hospitals belonging to the publicly provided health care system is usually decided and has to pass through the general practitioners. In this respect, general practitioners act as "gatekeepers" of recovery services in hospitals (see Maio and Manzoli, 2002, for Italian evidence, and Brekke et al., 2007, for a theoretical model).[13] Therefore, we argue that recovery in hospital and general practitioners can be considered as complementary health care services. We take as a proxy for hospital recovery the number of beds in hospitals. Thus, the ratio between the number of general practitioners per 1,000 inhabitants and the number of beds in hospitals per 1,000 inhabitants, which is denoted by **Me_gen_letti,** is our measure of complementary sources of medical care. As a consequence, we should expect that an increase in **Me_gen_letti** is negatively correlated with the number of ASLs and ASOs per 1,000 inhabitants.

The third empirical proxy capturing substitutability is given by the ratio between physicians working for public hospitals and hospitals' beds, where the former can be taken as a proxy for the demand for services by fee paying patients and the latter is taken as a proxy for health care services associated with recovery in hospital. As already noticed, in Italy there is the possibility of private provision of

[13] As pointed out in the literature, politicians may be willing to introduce general practioners' gatekeeping both to control the expenditures and to improve the efficiency (i.e. patient/hospital matching) of secondary care services, see Brekke et al. (2007) for a discussion.

health care services within the public health care system, the so called *intramoenia* (in hospital) regime.[14] Under the *intramoenia* regime, physicians working for public hospitals, other than for medical care to recovered patients, may act as specialists for privately paid visits to outsiders and insider patients. Physicians are paid a fixed salary for their standard activities. However, outside the contracted hours of service, they can privately visit paying patients at a fee that is autonomously established by each region. It is widely believed that private visits often reduce the need for recovery by providing diagnostic by specialists and diagnostic tests that would be otherwise part of the recovery service package, thus improving the cost-effectiveness of health care services (see Turchetti, 2009, pp. 116-117). Thus, the ratio between the number of specialists working in hospitals per 1,000 inhabitants and the number of beds in hospitals per 1,000 inhabitants, which is denoted by **Me_osp_letti,** can be considered as a further measure of substitute sources of medical care. This implies that, according to the predictions of our theoretical model, we would expect that an increase in the ratio **Me_osp_letti** is positively correlated with the dependent variable.

In equation (16), X_{it} is a set of regressors assumed to be exogenous, i.e., $E(X_{it}, v_{it}) = 0$, and β is the corresponding vector of coefficients. The set of exogenous variables that we test as instruments for our endogenous regressors is composed by those variables - capturing political, economic, socio-demographic and institutional features of the Italian health care system - that may affect the regional governments' preferences and, therefore, their organisational choices. In particular, we take into consideration the standard structural, socio-economic and demographic characteristics of a region that may affect the demand for regional health care services (see for example Francese and Romanelli, 2011) such as: the logarithm of per capita GDP at constant prices (and its square), capturing possibly non linear effects of regional income on the demand for health care services; the total unemployment rate (and its square), measuring differently the same kind of effects; the economic activity rate (i.e., the percentage of the population, both employed and unemployed, who constitutes the supply of the labor market), capturing the demand for formal health care services (assuming that informal care services can be continuously provided within the household only by those not actively seeking for a job); the birth rate in the resident population and the proportion of people aged 65 or more, both capturing the effect of the age structure of the population on demand; the net yearly regional migration rate, measuring the expected change in the regional demand for health care services. We moreover consider the abortion ratio and the logarithm of the average age to abortion, the mortality rate and the infant mortality rate, which we take as proxies for the quality of health care services.

---

[14] The *intramoenia* regime was formally introduced in 1996 (see Law n. 662/1996). However, since 1978, doctors working in public hospitals had the right to exercise private professional activities. Between 1991 and 1996, doctors were able to exercise privately as long as their activity was carried out outside contracted hours and not in private health care agencies unrelated with the NHS, see Turchetti (2009), pp. 112-113.

Finally, we a control and test as instruments a set of political and institutional dummies defined as follows: **(i)** the type of electoral system under which the regional government was elected. This was considered by means of two dummy variables. The first regional dummy variable *(electoral_law)* takes on a value equal to 0 under the proportional system (up to 1994) and a value equal to 1 under the plurality system. The second regional dummy variable (*president_elec)* captures the direct election of the president of the region, who becomes formally and substantially responsible of regional administration. In 1999, direct presidential elections were introduced for the Italian regions with ordinary bylaw and in 2001 for the Italian regions with straordinary bylaw (i.e. Valle d'Aosta, Trentino Alto-Adige, Friuli, Sardinia and Sicily). **(ii)** The test has been also carried out considering the political colours of the regional ruling coalitions, by constructing a dummy variable (*sinis_gov)* equal to 1 in the years and regions ruled by centre-left wing parties' coalitions and equal to zero otherwise.[15] **(iii)** Political and local elections are captured by the variable *elez* that takes on value 1 for the years of general political election and/or local elections and 0 otherwise. As regards the set of political and institutional variables considered, only one resulted significant as an instrument: the variable *sinis_gov* (see below).

In addition, X_{1it} includes temporal dummies. time-invariant fixed regional effects and a constructed variable that we identify as a measure of the time-varying idiosyncratic regional shock to the demand for health care services. This variable, denoted as **migraBF**, is obtained by the product of the net yearly migration rate in each region and the number of new regional residence permits given to immigrants since 2003 and until 2008, thus it is set equal to zero up to 2002. Our purpose is to consider the impact on the regional health care services of the biggest immigration amnesty, regularising in 2002 about 702,156 undocumented workers out of the 1,486,392 undocumented workers regularised in Italy over the full sample period (see Fasani, 2009).[16] The law n. 189 of 30th July 2002, known as the "Bossi-Fini law" after the names of the politicians who proposed it, amended the previous immigration law and introduced new clauses regulating undocumented immigration in Italy. It came into force in August 2002 and was followed by a decree-law on the procedures for regularizing the situation of undocumented immigrants already in the country in September 2002. In particular, the decree-law, issued on 6th September 2002, provided for the regularization of the position of two types of undocumented immigrant workers: those employed as domestic workers and

[15] A further set of variables combining the ruling party and the electoral system (proportional or plurality electoral system) over time is built, capturing the following features. In the period of proportional electoral system, we distinguish the regional governments ruled by either the Christian Democratic (DC) party, the Communist party alone and with the Socialist party, centre-left coalitions (with the DC party), or Local Autonomies (i.e., Union Valdotain, Lega Lombarda and SVP in Trentino-Alto Adige). During the period of plurality electoral system the regional government coalitions are simply separated into two main groups: the centre-right and centre-left ruling coalitions.

[16] Five immigration amnesties were introduced in Italy over the sample period (in 1986, 1990, 1995, 1998, 2002). The 2002 amnesty mainly interested workers living in the North and Center of Italy (about 81% of the total).

home-helpers and the dependent workers involved in other kinds of sub-ordinate employment. Immigrants whose residence permits had expired could also regularize their situation, provided that they have not received a deportation order. All regularized immigrants workers received a residence permit with duration equal to the duration of their employment contract. The residence permit gave them the right to directly access the public health care system (see Devillanova, 2008, on undocumented immigrants access to health care in Italy shortly before the 2002 amnesty). Notice that after their regularization these workers could apply for "re-conjunction", i.e. they were able to ask for their wives/husbands and children living outside Italy to joining them. As long as re-conjunction took place in the years following the immigration amnesty, the regional demand shock occurred not only in 2003, but also in the following years, which is what we assume.

Table 1. Summary statistics of the considered variables. Italian regions 1982-2008.

Variable	Description	Obs	Mean	Std.Dev.	Min	Max
asl_aso_pop	Regional yearly number of local health firms (ASLs) and hospital trust (ASOs) per 1,000 inhabitants	540	0.009058	0.004497	0.001283	0.021323
me_osp_letti	Regional yearly ratio between the number of specialists working in hospitals per 1,000 inhabitants and the number of beds in hospitals per 1,000 inhabitants	540	0.33662	0.146413	0.107864	0.704976
me_gen_letti	Regional yearly ratio between the number of general practitioners per 1,000 inhabitants and the number of beds in hospitals per 1,000 inhabitants	540	0.149403	0.054172	0.036557	0.289394
ratio_med	Regional yearly ratio between the number of general practitioners per 1,000 inhabitants and the number of public specialists working in hospitals per 1,000 inhabitants.	540	0.478516	0.14476	0.116597	0.987473
Unempl. Rate	Regional yearly total unemployment rate	560	10.00714	5.385751	2.5	26.8
Unempl. Rate Sq.	Square of the regional yearly total unemployment rate	560	129.0974	136.1186	6.25	718.2399
Activity rate	Regional yearly economic activity rate	560	49.24946	3.790323	38.3	57.8
Ln Average Abortion age	Logarithm of the regional yearly average age to abortion	560	3.394863	0.017252	3.299534	3.430756
Birth rate	Regional yearly birth rate population	560	9.387036	1.830801	5.9	15.9
Pop. Share over65	Regional yearly proportion of people aged 65 or more	560	17.39457	3.677671	9.37	26.8
Infant Mort.rate	Regional yearly infant mortality rate	540	64.96544	31.19822	2.06	183.387
sinis_gov	Dummy taking on value 1 for centre-left wing parties' coalitions governments	560	0.789286	0.408181	0	1
Migration rate	Regional yearly net migration rate	560	0.785294	2.596218	-16.1578	11.57902
migraBF	Product of the regional yearly net migration rate and the number of residence permits since 2003.	560	42878.98	187422.7	-444336	1508787
t_i , i=1,2….27	Time dummies. t_i =1 if year =X_i, X_i =1982, ….2009 for i=1,2….27. t_i =0 otherwise	560	0.035714	0.185743	0	1
Year	Time variable	560	1995.5	8.084969	1982	2009
Region	Panel ID variable	560	10.5	5.771437	1	20

Data Sources. Political variables: Ministero degli Interni; Socio-demographic and economic variables: ISTAT; Health care data: Ministero della Salute and ISTAT (Health for all dataset).

Table 1 above reports the summary statistics of the variables resulting significant for the analysis. As previously mentioned, the dataset is a yearly panel data for the 20 Italian regions over the 1982-2009 period. In the following tables 2 and 3 we shall report two-step estimates that yield theoretically robust results (see below for the battery of tests used) for both the first and the second stage estimates. In either table the output is divided into two parts: the first part, labelled A. FIRST

STAGE FIXED EFFECTS ESTIMATION, shows the OLS estimates of the three endogenous regressors (i.e., Me_osp_letti, Me_gen_letti and Ratio_med); the second part , labelled B. IV (2SLS) estimation, shows the impact of the endogenous regressors on the dependent variable. Table 2 presents the estimates excluding from the regressors the regional shock migraBF, whereas table 3 includes it.

As for the first stage estimates, notice here that, among the instruments, the political/institutional variables are never significant with the exception of *sinis_gov*, whereas the traditional structural, socio-economic and demographic characteristics of a region are correctly signed and significant for at least one of the endogenous variables. Among the latter, the significant instruments are: the total unemployment rate and the square of the total unemployment rate, entering with positive and negative signs, respectively, thus suggesting a U-shaped effect of regional unemployment on the demand for health care;[17] the economic activity rate, taking on a positive sign, suggesting that higher labour market participation rises the demand for formal health care services; the proportion of people aged 65 or more and the birth rate, when statistically significant, enter with positive and negative sign, respectively, implying that the demand for health care is higher for older people and is lower in regions with higher natality: this latter result may reflect the fact that, in Italy and over the sample, the regional birth rate is negatively correlated with regional income; the infant mortality rate, when statistically significant at conventional levels, enters with a positive sign (see ratio_med), whereas the logarithm of the average age to abortion, the net migration rate, and the interaction between net migration and the number of regularised immigrant workers under the Bossi Fini law (migraBF), when statistically significant, all take on a negative sign. Regarding the effects of net migration and migraBF on the demand for health care services, table 3, a possible explanation is as follows. On the one hand, higher net migration is likely to increase the regional demand for formal health care services. On the other hand, migrants tend to cluster in the richest areas of the country (see also footnote 16 above) and in unskilled occupations (see Pellizzari, 2011, for 2007 Italian evidence). As long as the Italians are relatively more educated in richer regions, this implies a potential higher demand for personal services, including old age care, in these regions. As a consequence, an increase in the supply of unskilled migrant jobs may induce natives to substitute formal health care (through the NHS) with informal health care (i.e. within the households). If the latter indirect effect of higher net migration dominates the former direct effect, one would observe a negative correlation between the demand for health care services, which we observe here.

---

[17] A possible rationale for this U-shaped effect is that an increase in unemployment, being associated with a reduction in income, initially gives rise to a reduction in the demand for health care; however, above a certain threshold, it may be associated with an increasing deterioration of the population health, leading to higher demand for health care.

Table 2. Fixed-effects IV regression

A. FIRST STAGE FIXED EFFECTS ESTIMATION

Number of groups = 20 Obs per group: min = 27; avg = 27.0; max = 27

First-stage regression of me osp letti (OLS estimation)

		Number of obs	540
		F(35, 485)	300.86
		Prob > F	0.0000
Total (centered) SS	10.40895583	Centered R2	0.9373
Total (uncentered) SS	10.40895583	Uncentered R2	0.9373
Residual SS	.6527355166	Root MSE	.03669

me_osp_letti	Coef.	Robust Std.Err.	t	P>|t|	[95% Conf.	Interval]
Time dummies	YES	YES	YES	YES	YES	YES
Unempl. Rate	-.0112125	.002914	-3.85	0.000	-.0169382	-.0054868
Une. Rate Sq.	.0002446	.000089	2.75	0.006	.0000698	.0004194
Activity rate	.0034562	.0012749	2.71	0.007	.0009511	.0059613
Ln Average Abortion age	.003204	.2258189	-4.44	0.000	-1.446908	-.5594996
Birth rate	.0022653	.0028657	0.79	0.430	-.0033655	.0078961
Pop Share over 65	.0138351	.0045567	3.04	0.003	.0048818	.0227884
Infa.Mort.ra	-.0002732	.0001729	-1.58	0.115	-.0006129	.0000665
Sinis_gov	.0134916	.0059882	2.25	0.025	.0017256	.0252575
Migration rat	-.0028358	.0014999	-1.89	0.059	-.005783	.0001114

First-stage regression of me gen letti (OLS estimation)

		Number of obs	540
		F(35, 485)	40.75
		Prob > F	0.0000
Total (centered) SS	1.191061776	Centered R2	0.7139
Total (uncentered) SS	1.191061776	Uncentered R2	0.7139
Residual SS	.340818027	Root MSE	.02651

me_gen_letti	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Time dummies	YES	YES	YES	YES	YES	YES
Unempl. Rate	-.01201	.0023374	-5.14	0.000	-.0166026	-.0074173
Une.Rate Squ	.0003181	.0000718	4.43	0.000	.000177	.0004593
Activity rate	.0027963	.0009741	2.87	0.004	.0008824	.0047102
Ln Average Abortion age	-.0439131	.1673726	-0.26	0.793	-.372778	.2849518
Birth rate	-.0102306	.0023195	-4.41	0.000	-.0147882	-.005673
Pop. share over 65 €	.0121398	.0034949	3.47	0.001	.0052728	.0190069
Infa Mort ra	.0001418	.0001192	1.19	0.235	-.0000925	.0003761
Sinis_gov	.0039707	.0037318	1.06	0.288	-.0033617	.0113031
Migration rat	-.0021669	.0009512	-2.28	0.023	-.0040358	-.000298

First-stage regression of ratio med (OLS estimation)

		Number of obs	540
		F(35, 485)	33.81
		Prob > F	0.0000
Total (centered) SS	8.770071403	Centered R2	0.6158
Total (uncentered) SS	8.770071403	Uncentered R2	0.6158
Residual SS	3.369231373	Root MSE	.08335

ratio_med	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Time dummies	YES	YES	YES	YES	YES	YES
Unempl rate	-.0275105	.0076579	-3.59	0.000	-.0425572	-.0124637
Une.Rate Squ	.0008034	.0002438	3.30	0.001	.0003244	.0012825
Activity rate	.0041563	.0032156	1.29	0.197	-.002162	.0104746
Ln Average abortion age	1.667612	.5446741	3.06	0.002	.5973992	2.737824
Birth rate	-.0319119	.0082033	-3.89	0.000	-.0480303	-.0157936
Pop. Share Over 65	.0069713	.0098634	0.71	0.480	-.0124091	.0263516
Infa Mort ra	.0011188	.0003815	2.93	0.004	.0003693	.0018683
Sinis_gov	-.0130259	.0102579	-1.27	0.205	-.0331812	.0071295
Migration rat	-.0052851	.0030636	-1.73	0.085	-.0113046	.0007344

```
Summary results for first-stage regressions
-------------------------------------------
Variable     | Shea Partial R2 |   Partial R2    |  F(  9,   485)    P-value
me_osp_letti|      0.0661      |     0.1293      |        6.46       0.0000
me_gen_letti|      0.0735      |     0.1069      |        5.58       0.0000
ratio_med   |      0.0736      |     0.1031      |        3.62       0.0002
NB: first-stage F-stat heteroskedasticity-robust
Underidentification tests
Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)
Ha: matrix has rank=K1 (identified)
Kleibergen-Paap rk LM statistic           Chi-sq(7)=26.78    P-val=0.0004
Kleibergen-Paap rk Wald statistic         Chi-sq(7)=32.72    P-val=0.0000
Weak-instrument-robust inference
Tests of joint significance of endogenous regressors B1 in main equation
Ho: B1=0 and overidentifying restrictions are valid
Anderson-Rubin Wald test     F(9,485)= 6.32      P-val=0.0000
Anderson-Rubin Wald test     Chi-sq(9)=61.03     P-val=0.0000
Stock-Wright LM S statistic  Chi-sq(9)=49.14     P-val=0.0000
NB: Underidentification, weak identification and weak-identification-robust test statistics heteroskedasticity-robust
Number of observations             N  =        540
Number of regressors               K  =         29
Number of instruments              L  =         35
Number of excluded instruments     L1 =          9
```

B.IV (2SLS) estimation

```
Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity
                                                      Number of obs =      540
                                                      F( 29,   491) =    34.31
                                                      Prob > F      =   0.0000
Total (centered) SS     =  .0074323093                Centered R2   =   0.4853
Total (uncentered) SS   =  .0074323093                Uncentered R2 =   0.4853
Residual SS             =  .0038254168                Root MSE      =  .002712
------------------------------------------------------------------------------
             |               Robust
 asl_aso_pop |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
me_osp_letti |   .0753945    .022547     3.34   0.001     .0312031    .1195859
me_gen_letti |  -.0999362   .0449321    -2.22   0.026    -.1880016   -.0118709
   ratio_med |   .0243438   .0129517     1.88   0.060    -.0010411    .0497288
Time dummies        YES        YES        YES     YES        YES          YES
------------------------------------------------------------------------------
Underidentification test (Kleibergen-Paap rk LM statistic): 26.782 Chi-sq(7) P-val = 0.0004
LM test of redundancy of specified instruments: 80.836 Chi-sq(15) P-val = 0.0000
Instruments tested: unemployment rate, activity rate, Log average abortion age, Birth rate, activity rate, population share over 65, infant mortality rate
------------------------------------------------------------------------------
Hansen J statistic (overidentification test of all instruments): 8.669 Chi-sq(6) P-val = 0.1931
Hansen J statistic (eqn. excluding suspect orthog. conditions): 7.579  Chi-sq(5) P-val = 0.1810
C statistic (exogeneity/orthogonality of suspect instruments): 8.669   Chi-sq(6) P-val =    0.1931
Instruments tested:    unemployment rate,activity rate, log average abortion age, natality rate, population share over 65, infant mortality rate. Endogeneity test of endogenous regressors: 41.471 Chi-sq(3) P-val = 0.0000
Regressors tested:    me_osp_letti me_gen_letti ratio_med
------------------------------------------------------------------------------
Instrumented:         me_osp_letti me_gen_letti ratio_med
Included instruments: t1 t2 t3 t4 t5 t6 t7 t8 t9 t10 t11 t12 t13 t14 t15 t16 t17 t18 t19 t20 t21 t22 t23 t24 t25 t26
Excluded instruments: unemployment rate, unemployment rate squared, activity rate, log average abortion age, Birth rate, population share over 65, infant mortality rate, sinis_gov, migration rate
------------------------------------------------------------------------------
```

Table 3. Fixed-effects IV regression

A. FIRST STAGE FIXED EFFECTS ESTIMATION

Number of groups = 20 | Obs per group: min = 27; avg = 27.0; max = 27

First-stage regression of me osp letti (OLS estimation)

Total (centered) SS	= 10.40895583	Number of obs =	540
Total (uncentered) SS	= 10.40895583	F(36, 484) =	310.47
Residual SS	= .6443552401	Prob > F =	0.0000
		Centered R2 =	0.9381
		Uncentered R2 =	0.9381
		Root MSE =	.03649

me_osp_letti	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Time dummies	YES	YES	YES	YES	YES	YES
Unemplo. rate	-.0110537	.0028782	-3.84	0.000	-.016709	-.0053984
Une.rate squa	.0002497	.0000876	2.85	0.005	.0000775	.0004219
Activity rate	.0034107	.0012687	2.69	0.007	.0009179	.0059034
Ln average Abortion age	-1.004271	.2303065	-4.36	0.000	-1.456795	-.5517468
Natality rate	.0046763	.0029533	1.58	0.114	-.0011266	.0104792
Population Share over 65	.0129069	.0044048	2.93	0.004	.004252	.0215618
Infan.mort.ra	-.0002785	.0001728	-1.61	0.108	-.0006179	.0000609
Sinis_gov	.0114896	.0061131	1.88	0.061	-.0005219	.0235011
Migration rat	-.0025013	.0014961	-1.67	0.095	-.005441	.0004385
MigraBF	-3.23e-08	1.00e-08	-3.22	0.001	-5.21e-08	-1.26e-08

First-stage regression of me gen letti (OLS estimation)

Total (centered) SS	= 1.191061776	Number of obs =	540
Total (uncentered) SS	= 1.191061776	F(36, 484) =	40.01
Residual SS	= .3289293494	Prob > F =	0.0000
		Centered R2 =	0.7238
		Uncentered R2 =	0.7238
		Root MSE =	.02607

me_gen_letti	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Time dummies	YES	YES	YES	YES	YES	YES
Unemplo. rate	-.0118209	.0022517	-5.25	0.000	-.0162451	-.0073966
Une.rate squa	.0003242	.0000698	4.65	0.000	.0001871	.0004614
Activity rate	.002742	.0009606	2.85	0.004	.0008545	.0046296
Ln average Abortion age	-.0451841	.1739996	-0.26	0.795	-.3870719	.2967038
Birth rate	-.0073589	.0022741	-3.24	0.001	-.0118272	-.0028907
Population Share over 65	.0110343	.0032052	3.44	0.001	.0047366	.0173321
Infan.mort.ra	.0001355	.0001203	1.13	0.261	-.0001009	.0003719
Sinis_gov	.0015862	.0037752	0.42	0.675	-.0058316	.009004
Migration rat	-.0017684	.0009868	-1.79	0.074	-.0037074	.0001705
MigraBF	-3.85e-08	8.93e-09	-4.32	0.000	-5.61e-08	-2.10e-08

First-stage regression of ratio med (OLS estimation)

Total (centered) SS	= 8.770071403	Number of obs =	540
Total (uncentered) SS	= 8.770071403	F(36, 484) =	33.58
Residual SS	= 3.367043354	Prob > F =	0.0000
		Centered R2 =	0.6161
		Uncentered R2 =	0.6161
		Root MSE =	.08341

ratio_med	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Time dummies	YES	YES	YES	YES	YES	YES
Unemplo. rate	-.0274293	.0076745	-3.57	0.000	-.0425088	-.0123498
Une.rate squa	.000806	.0002435	3.31	0.001	.0003276	.0012845
Activity rate	.004133	.0032219	1.28	0.200	-.0021976	.0104637
Ln average Abortion age	1.667066	.5484536	3.04	0.002	.5894222	2.74471
Birth rate	-.03068	.0087136	-3.52	0.000	-.0478011	-.0135588
Population Share over 65	.006497	.0098754	0.66	0.511	-.012907	.025901
Infan.mort.ra	.0011161	.0003827	2.92	0.004	.0003641	.0018681
Sinis_gov	-.0140488	.0102541	-1.37	0.171	-.0341969	.0060993
Migration rat	-.0051141	.0031347	-1.63	0.103	-.0112735	.0010452
MigraBF	-1.65e-08	1.99e-08	-0.83	0.406	-5.55e-08	2.25e-08

```
Summary results for first-stage regressions
-------------------------------------------
Variable     | Shea Partial R2 |   Partial R2    |  F( 10,   484)    P-value
me_osp_letti|     0.0781      |     0.1405      |        6.41       0.0000
me_gen_letti|     0.1085      |     0.1381      |        7.63       0.0000
ratio_med   |     0.0938      |     0.1036      |        3.98       0.0000
NB: first-stage F-stat heteroskedasticity-robust
Underidentification tests
Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)
Ha: matrix has rank=K1 (identified)
Kleibergen-Paap rk LM statistic            Chi-sq(8)=22.64    P-val=0.0039
Kleibergen-Paap rk Wald statistic          Chi-sq(8)=25.32    P-val=0.0014
Weak-instrument-robust inference
Tests of joint significance of endogenous regressors B1 in main equation
Ho: B1=0 and overidentifying restrictions are valid
Anderson-Rubin Wald test     F(10,484)=6.70      P-val=0.0000
Anderson-Rubin Wald test     Chi-sq(10)=72.02    P-val=0.0000
Stock-Wright LM S statistic  Chi-sq(10)=50.61    P-val=0.0000
NB:  Underidentification,  weak  identification  and  weak-identification-robust  test  statistics
heteroskedasticity-robust
Number of observations              N  =       540
Number of regressors                K  =        29
Number of instruments               L  =        36
Number of excluded instruments      L1 =        10
```

B. IV (2SLS) estimation

```
Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity
                                                     Number of obs =      540
                                                     F( 29,   491) =    33.02
                                                     Prob > F      =   0.0000
Total (centered) SS     =  .0074323093               Centered R2   =   0.4735
Total (uncentered) SS   =  .0074323093               Uncentered R2 =   0.4735
Residual SS             =  .0039133279               Root MSE      =  .002743
------------------------------------------------------------------------------
             |               Robust
 asl_aso_pop |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
me_osp_letti |   .0774866   .0194566     3.98   0.000     .0393523    .1156208
me_gen_letti |  -.1055601   .0318819    -3.31   0.001    -.1680475   -.0430726
   ratio_med |   .0257037   .0104419     2.46   0.014      .005238    .0461695
Time dummies        YES        YES        YES     YES       YES         YES
------------------------------------------------------------------------------
Underidentification test (Kleibergen-Paap rk LM statistic): 22.638  Chi-sq(8) P-val =    0.0039
LM test of redundancy of specified instruments): 75.694  Chi-sq(15) P-val =   0.0000
Instruments tested: Log average abortion age, Birth rate, activity rate, population share over 65,
infant mortality rate
------------------------------------------------------------------------------
Hansen J statistic (overidentification test of all instruments): 8.434  Chi-sq(7) P-val = 0.2959
Hansen J statistic (eqn. excluding suspect orthog. conditions): 7.355  Chi-sq(6) P-val = 0.2893
C statistic (exogeneity/orthogonality of suspect instruments): 8.073   Chi-sq(6) P-val =    0.2328
Instruments  tested:  unemployment  rate,  activity  rate,  Log  average  abortion  age,  Birth  rate,
activity rate, population share over 65, infant mortality rate
Endogeneity test of endogenous regressors:  43.739   Chi-sq(3) P-val = 0.0000
Regressors tested:    me_osp_letti me_gen_letti ratio_med
------------------------------------------------------------------------------
Instrumented:         me_osp_letti me_gen_letti ratio_med
Included instruments: t1 t2 t3 t4 t5 t6 t7 t8 t9 t10 t11 t12 t13 t14 t15 t16 t17 t18 t19 t20 t21
t22 t23 t24 t25 t26
Excluded  instruments:  unemployment  rate,  unemployment  rate  squared,  activity  rate,  log  average
abortion  age,  Birth  rate,  population  share  over  65,  infant  mortality  rate,  sinis_gov,  migration
rate
------------------------------------------------------------------------------
```

As for the second stage estimates reported in tables 2 and 3 above, the parts labelled B. IV (2SLS) estimation show the impact of the regressors on the dependent variable. Recall that, in both tables, we have considered three endogenous regressors capturing, respectively, substitutability (i.e. Ratio_med and Me_osp_letti) and complementary (i.e. Me_gen_letti) between health care services. In both cases, the estimated coefficients are all statistically significant at the 5% level or less and are correctly signed according to the predictions of our theoretical model, with the exception of ratio_med, which is significant at the 6% level in Table 2. On the one hand, ratio_med (the ratio between the number of general practitioners and the number of specialist physicians working in hospitals, who, following Atella and Deb (2008), we consider as substitute sources of health care services), as well as Me_osp_letti (which captures substitutability between the visits by specialist doctors working in hospitals and recovery in hospital), take on a positive sign: an increase in these ratios (representing an increasing demand for substitute health care service) raises the average dimension of regional local health management agencies. On the other hand, Me_gen_letti capturing the complementarity between general pratictioners and recovery in hospital takes on negative sign.

Table 3 also considers the shock on the demand for health care as represented by the variable migraBF (i.e., the product of the regional yearly net migration rate and the number of residence permits since 2003)**,** capturing the pressure of regularised immigrants, who, since 2003 were fully entitled to the services of the Italian national health care system. At the first stage, as already discussed, this variable is significant at the 5% level or less and negatively signed for both me_gen_letti and me_osp_letti. At the second stage, the impact of this variable shows up in a slightly bigger absolute value of all the estimated parameters of the endogenous regressors, according to the predictions of our theoretical model. Actually, the estimated coefficients of me_osp_letti passes from about 0.075 to about 0.077; that of me_gen_letti passes from about -0.0999 to about -0.1056; ratio_med from about 0.024 to about 0.026.

As previously mentioned, for both cases reported in tables 2 and 3, the two-step estimates yield theoretically robust results. More specifically, we refer to the following battery of tests. We apply the two-step estimator in order to obtain the robust Sargan test, i.e., the (robust) Hansen J-test, which is not available in one-step estimation. The Hansen J-statistic tests the null hypothesis of correct model specification and valid overidentifying restrictions, i.e. the validity of instruments (see Baum, 2006). The rejection of the null hypothesis means that either or both assumptions are questionable. The Hansen test of overidentifying restrictions does not reject the null at conventional level of significance (p-value=0.1931 in table 2 and p-value=0.2959 in table 3). Hence, we take it as an indication that both models have valid instruments. While the Hansen J-test evaluates the entire set of overidentifying restrictions/instruments, it is also important to test the validity of subsets of instruments. For this purpose, we use the C-test (see Baum, 2006; Roodman, 2006). This consists in estimating the models with and without a subset of suspect instruments, enabling investigation of the validity, i.e. exogeneity of any subset of instruments, as well as their contribution to the increase in J-test (see Roodman, 2007, p. 11). The null hypothesis of the C-test is that the specified variables are proper instruments, i.e. that the set of examined instruments is exogenous. Here, we use the C-test for the following instruments: unemployment rate, activity rate, log of the average abortion age, natality rate, share of population over 65, infant mortality rate. As we may see from Tables 2 and 3, we do not have enough evidence to reject the null hypothesis of exogeneity of any of the instruments used, as well as the validity of standard IV instruments. Endogeneity tests of all the endogenous regressors have also been implemented. Under the null hypothesis that the specified endogenous regressors can actually be treated as exogenous, the test statistic is distributed as chi-squared with degrees of freedom equal to the number of regressors tested. The underidentification test is an LM test of whether the equation is identified, i.e., that the excluded instruments are relevant, meaning correlated with the endogenous regressors. The test is essentially the test of the rank of a matrix: under the null hypothesis that the equation is underidentified, a rejection of the null indicates that the matrix is full column rank; i.e., the model is identified. We also test whether a subset of excluded instruments is redundant. Excluded instruments are redundant if the asymptotic efficiency of the estimation is not improved by using them. Breusch et al. (1999) show that the condition for the redundancy of a set of instruments can be stated in several equivalent ways: e.g., in the reduced-form regressions of the endogenous regressors on the full set of instruments, the redundant instruments have statistically insignificant coefficients, or the partial correlations between the endogenous regressors and the instruments in question are zero. Here, we test the rank of the matrix cross-product between the endogenous regressors and the possibly redundant instruments, the ranktest is used to test whether the matrix has zero rank. The test statistic is an LM test and is numerically equivalent to a regression-based LM test. Under the null hypothesis that the specified instruments are redundant, the rejection of the null indicates that the instruments are not redundant.

5. Summary of the results and preliminary conclusions

This paper has considered a stylised model of the determination of health care organisation by a regional government, bargaining over the budget allocation with health care management agencies possibly supplying a mix of health services. In our model, the central government can choose whether to deal with one consolidated bureau or with two competing bureaus for the public provision of differentiated outputs. Our model shows that, if the government sees the bureaucratic outputs as complements (respectively, substitutes), it will prefer dealing with one consolidated public health agency (respectively, two competing agencies), and that these incentives may be increased when there is more uncertainty as regards the demand for health care services. Generalising the implications of the theoretical model, we derived two testable predictions for the Italian public health care system. First, the government has an incentive to consolidate the bureaucratic health care organization into a smaller number of larger independent agencies when it perceives health care services as complementary, whereas it prefers a larger number of smaller independent agencies when these services are substitutes. In practice, this should imply a positive (negative) correlation between the supply of substitutable (complementary) health care services and the dimension of local public health authorities. These predictions have been tested by using a panel data set for the 20 Italian regions over the period 1982-2008. As proxies for substitutable health care services, we have used the ratio between the number of general pratictioners and the number of physicians working for NHS hospitals and the ratio of public specialists and the number of beds (a proxy for recovery service) per 1,000 inhabitants. As a proxy for complementary services, we have used the ratio between physicians working for public hospitals and hospitals' beds per 1,000 inhabitants. The model has been estimated with a fixed effects IV regression – that took into account as instruments also the political economic and socio, economic and demographic covariates. The results show, first, that when the demand for substitute (complementary) activities rises, a regional government has on average an incentive to increase (lower) the dimension of local health management agencies, as measured by the sum of the ASLs and ASOs in a region per 1,000 inhabitants. Second, it turns out that the impact of the 2002 immigration amnesty, which we take a proxy for higher uncertainty in the demand shock, shows up into a bigger effect of substitutable health care services on the average regional incentives to consolidate.

References

Atella, V. and Deb. P. (2008). 'Are primary care physicians, public and private sector specialists substitutes or complements? Evidence from a simultaneous equations model for count data', *Journal of Health Economics*, 27, 770-785.

Bates, L.J., Lafrancois B.A., Santerre R.E., (2011) 'An empirical study of the consolidation of local public health services in Connecticut', *Public Choice* 147: 107–121.

Baum, F. C. (2006) *An Introduction to Modern Econometrics Using Stata*, Texas: Stata Press.

Bloom, N., Propper, C., Seiler, S. and van Reenen, J. (2010). 'The impact of competition on management quality: Evidence from public hospitals'. NBER WP n. 16032.

Bordignon, M. and Turati, G. (2009). 'Bailing out expectations and public health expenditure', *Journal of Health Economics*, 28: 305-321.

Brekke, K.R., Nuscheler, R. and Straume, O.R. (2007). "Gatekeeping in health care", *Journal of Health Economics.*

Breusch, T., H. Qian, P. Schmidt, and D. Wyhowski. 1999. Redundancy of moment conditions. Journal of Econometrics 9: 89-111.

Devillanova, C. (2008). 'Social networks, information and health care utilization: Evidence from undocumented immigrants in Milan, *Journal of Health Economics*, 27: 265-286.

Fasani, F. (2009) 'Undocumented migration in Italy: A country report'. *CLANDESTINO project* - Counting the Uncountable. Data and Trends Across Europe – 6th FP – European Commission.

Fedeli, S. (1999). 'Competing bureaus and politicians: A compliance approach to the diversion of public funds'. *Public Choice*, 100: 253-270.

Fedeli, S. and Santoni, M. (2006). 'The government choice of bureaucratic organisation: an application to Italian state museums', *Journal of Cultural Economics*, 2006, vol. 30, pp. 41-72.

Fortney, J.C., Steffick, D.E., Burgess, J.F. Jr, Maciejewski, M.L. and Petersen, L. A. (2005). 'Are primary care services a substitute or complement for specialty and inpatient services?', *Health Research and Educational Trust*, 40: 1423-1442.

Francese, M. and Romanelli, M. (2011). 'Health care in Italy: expenditure determinants and regional differentials', *Bank of Italy Working Paper* n. 828.

Gaynor, M., Laudicella, M. and Propper, C. (2011). 'Can governments do it better? Merger mania and hospital outcomes in the English NHS'. NBER WP n. 17608.

Horn, H. and Wolinsky, A. (1988). "Bilateral monopolies and incentives to merge". *Rand Journal of Economics*, 19, 408-19.

Janssen, R.T.J.M., Leers, T., Meijdam, L.C. and Verbo H.A.A. (2003). 'Bureaucracy versus market in hospital care'. *Public Choice*, 114: 477-489.

Klemperer, P. and Meyer, M. (1986). 'Price competition vs. quantity competition: the role of uncertainty'. *RAND Journal of Economics*, 17: 618-639.

Maio, V. and Manzoli, L. (2002). *The Italian Health care System: W.H.O. Ranking Versus Public Perception*, Vol. 27 No. 6, June P&T 301.
Available at: http://gepp.it/docs/pubblicazioni/ManzoliLPharmacyandTherapeutics2002.pdf

Miller, G. J. (1977). 'Bureaucratic compliance as a game on the unit square'. *Public Choice*, 29: 37-51.

Moe, T.M. (1984). 'The new economics of organization'. *American Political Science Review*, 88: 739-77.

Niskanen, W.A. (1971). *Bureaucracy and representative government*. Chicago: Aldine-Atherton.

Pellizzari, M. (2011). "The use of welfare by migrants in Italy", IZA DP n, 5613.

Roodman, D. (2006). "How To Do xtabond2: An Introduction to "Difference" and "System" GMM in Stata", *Center for Global Development Working Paper No. 103.*

Roodman, D. (2007) A Short Note on the Theme of Too Many Instruments, *Center for Global Development Working Paper No. 125.*

Scott, A. (1996). 'Primary or secondary care? What can economics contribute to evaluation at interface?', *Journal of Public Health Medicine*, 18: 19-26.

Shleifer, A. and Vishny, R.W. (1994). ‘Politicians and firms’. *Quarterly Journal of Economics,* 109: 995-1025.

Singh, N. and Vives, X. (1984). ‘Price and quantity competition in a differentiated duopoly’. *RAND Journal of Economics*, 15: 546-554.

Sørensen , R. J. (2006). ‘Local government consolidations: The impact of political transaction costs’. *Public Choice*, 127: 75–95.

Turchetti, G. (2009). “The interaction of public and private systems in healthcare provision: the Italian two-faced Janus”. *European Papers On The New Welfare* , 11 : 110-122.

APPENDICES

APPENDIX 1 Comparing the stochastic and deterministic solution with one consolidated bureau

Section 2.2 of the main text has shown that, when the demand shock affects linearly the reservation price for health care services, the government expected payoff under bureaucratic consolidation is

$$E(\hat{M}G^{11}) = 2R + \frac{(\alpha - c)^2}{4(\beta + 2\gamma)} + \frac{s^2}{4(\beta + 2\gamma)}$$

Thus, the government expected payoff is increasing in the variance of the shock, other things being equal. It is useful to decompose the stochastic component as follows

$$\frac{s^2}{4(\beta + 2\gamma)} = \underbrace{\frac{2s^2}{4(\beta + 2\gamma)}}_{\text{stochastic gains from rents}} - \underbrace{\frac{s^2}{4(\beta + 2\gamma)}}_{\text{stochastic loss from production}} . \qquad \text{(A.1)}$$

(A.1) shows that the stochastic shock alters the government trade-off between utility from political rents and utility from the outputs. The economic intuition for this result is as follows. The government expected compliance level in the stochastic case, $E\hat{g}^{11} = \frac{\alpha^2 - c^2 - s^2}{4R(\beta + 2\gamma)}$ (using equation 6 in the main text), is less than the level it would choose in the deterministic case, $\hat{g}^{11} = \frac{\alpha^2 - c^2}{4R(\beta + 2\gamma)}$ (see Fedeli and Santoni, 2006: 65). This implies that the government expected political rents are higher in the stochastic than in the deterministic case. However, although it can be shown that the equilibrium level of health care services is the same in both cases, $E\left(\frac{\hat{g}^{11}\hat{h}^{11}R}{c}\right) = E\left(\frac{\alpha + \varepsilon - c}{2(\beta + 2\gamma)}\right) = \frac{\alpha - c}{2(\beta + 2\gamma)}$, the government expected indirect utility from the outputs is lower in the stochastic than in the deterministic case, as long as $E\left[\left(\frac{\hat{g}^{11}\hat{h}^{11}R}{c}\right)^2\right] = \left[\frac{(\alpha - c)^2 + s^2}{4(\beta + 2\gamma)^2}\right] > \left[E\left(\frac{\hat{g}^{11}\hat{h}^{11}R}{c}\right)\right]^2 = \left(\frac{\alpha - c}{2(\beta + 2\gamma)}\right)^2$ by Jensen's inequality. Considering both effects together, and given that the government payoff is concave in outputs, in the stochastic case, relatively to the deterministic case, the increase in the expected political rents more than compensate the reduction in the expected gains from the outputs. As a result, the government expected payoff becomes an increasing function in the variance of the shock.

APPENDIX 2 Bureaucratic consolidation with random shocks to the demand slopes

Following Klemperer and Meyer (1986), assume that the random shock influences the slopes of the demands for health care services as follows:

$$V_i^M = \alpha - \frac{\beta}{\varepsilon} Q_i - \frac{2\gamma}{\varepsilon} Q_j \qquad \text{(A2.1)}$$

with=1, 2 and i≠j and $E(\varepsilon) = 1, E(1/\varepsilon) > 1, E(\varepsilon^2) = s^2 > 1$. Note that the random shock leaves the degree of substitutability between products unaffected. Solving the model in the usual way, it turns out that the players' compliance levels at a symmetric Nash equilibrium when the government deals with one consolidated bureau are

$$\breve{g}_1^{11} = \breve{g}_2^{11} = \frac{(\alpha - c)[2\alpha - (\alpha - c)\varepsilon]\bar{\varepsilon}}{4R(\beta + 2\gamma)}$$
$$\breve{h}_1^{11} = \breve{h}_2^{11} = \frac{2c}{2\alpha - (\alpha - c)\varepsilon} \qquad \text{(A2.2)}$$

Thus, at stage one, the government's expected payoff is

$$E(MG^{11}) = \underbrace{2R}_{\text{deterministic payoff}} + \underbrace{\frac{2s^2(\alpha - c)^2}{4(\beta + 2\gamma)}}_{\text{stochastic gain in terms of rents}} - \underbrace{\frac{s^2(\alpha - c)^2}{4(\beta + 2\gamma)}}_{\text{stochastic loss in terms of output evaluation}} \qquad \text{(A2.3)}$$

Similarly, the symmetric Nash equilibrium compliance levels when the government deals with two independent bureaus are

$$g_1^{12} = g_2^{12} = \frac{(\alpha - c)[2\alpha(\beta + \gamma) - (\alpha - c)(\beta + 2\gamma)\varepsilon]\bar{\varepsilon}}{4R(\beta + \gamma)^2}$$
$$h_1^{11} = h_2^{11} = \frac{2c(\beta + \gamma)}{2\alpha(\beta + \gamma) - (\alpha - c)(\beta + 2\gamma)\varepsilon} \qquad \text{(A2.4)}$$

from which follows the government's expected payoff

$$E(MG^{12}) = \underbrace{2R}_{\text{deterministic payoff}} + \underbrace{\frac{2s^2(\alpha - c)^2(\beta + 2\gamma)}{4(\beta + \gamma)^2}}_{\text{stochastic gain in terms of rents}} - \underbrace{\frac{s^2(\alpha - c)^2(\beta + 2\gamma)}{4(\beta + \gamma)^2}}_{\text{stochastic loss in terms of output evaluation}} \qquad \text{(A2.5)}$$

Equations (A2.3) and (A2.5) show that the government's expected payoff in either case increases with the variance of the shock. Using (A2.3) and (A2.5), yields

$$E(MG^{11}) - E(MG^{12}) = -\gamma[(\alpha - c)^2 s^2 \left[\frac{2\beta + 3\gamma}{4(\beta + 2\gamma)(\beta + \gamma)^2}\right] \geq 0 \qquad \text{(A2.6)}$$

Uncertainty on the slope of demand does not change the government's incentives to choose bureaucratic consolidation (separation) with complements (substitutes), or γ<0 (γ>0). However, uncertainty reinforces such an incentive, other things being equal. This is the same qualitative result as for the case of an additive random shock explicitly considered in section 2.2 of the main text.

APPENDIX 3 Bureaucratic consolidation with random shocks to the degree of substitutability

With a uniformly distributed random shock $\varepsilon \sim U(\varepsilon_L,\ \varepsilon_H)$, $\varepsilon_L < 0$ and $\varepsilon_H > 0$, the government's expected payoff from bureaucratic consolidation is from equation (12)

$$E(\tilde{M}G^{11}) = 2R + \frac{(\alpha - c)^2(\beta + 2\gamma)}{4} E\left[\frac{1}{(\beta + 2\gamma + 2\varepsilon)^2}\right] =$$

$$= 2R + \frac{(\alpha - c)^2(\beta + 2\gamma)}{4} \int_{\varepsilon_L}^{\varepsilon_H} \frac{1}{(\beta + 2\gamma + 2\varepsilon)^2} \frac{d\varepsilon}{\varepsilon_H - \varepsilon_L} = \tag{A3.1}$$

$$= 2R + \frac{(\alpha - c)^2(\beta + 2\gamma)}{4(\beta + 2\gamma + 2\varepsilon_H)(\beta + 2\gamma + 2\varepsilon_L)}$$

Similarly, the government's expected payoff under bureaucratic separation is from equation (14)

$$E(MG^{12}) = 2R + \frac{(\alpha - c)^2(\beta + 2\gamma)}{4} E\left[\frac{1}{(\beta + \gamma + \varepsilon)^2}\right] =$$

$$= 2R + \frac{(\alpha - c)^2(\beta + 2\gamma)}{4} \int_{\varepsilon_L}^{\varepsilon_H} \frac{1}{(\beta + \gamma + \varepsilon)^2} \frac{d\varepsilon}{\varepsilon_H - \varepsilon_L} = \tag{A3.2}$$

$$= 2R + \frac{(\alpha - c)^2(\beta + 2\gamma)}{4(\beta + \gamma + \varepsilon_H)(\beta + \gamma + \varepsilon_L)}$$

The government chooses bureaucratic merging if its expected payoff (A3.1) is larger than (A3.2), which is the condition (15) given in the main text.