

Indagine sulle variabili esogene che influiscono sui costi di distribuzione dell'energia elettrica per i piccoli distributori

Antonio Massarutto e Daniele Bortolotti

DIES, Università di Udine

Premessa

A seguito della pubblicazione da parte di ARERA del DCO 104/2018, ci è stato richiesto di effettuare alcune elaborazioni dirette a testare la significatività delle variabili esogene ipotizzate nel modello di calcolo della tariffa parametrica proposto dall'Autorità.

Le elaborazioni si sono basate sui dati forniti da un campione di imprese che hanno aderito volontariamente alla richiesta della Federazione, integrati da alcune indagini e verifiche dirette da parte del gruppo di ricerca.

Il campione originario, su cui si erano basate anche le elaborazioni già svolte nelle precedenti fasi del procedimento, è rappresentato da 55 imprese distributrici con meno di 100.000 punti di prelievo (POD).

Dato che nella presente fase si ipotizza di circoscrivere l'applicazione della tariffa parametrica alle sole imprese con meno di 25.000 POD, il campione è stato ridotto eliminando le imprese di dimensione superiore. Sono state inoltre espunte dal campione le imprese c.d. "integrate", in quanto mancanti del dato di output (misurato in funzione della tariffa di riferimento 2016). Ne risulta un campione che comprende 41 imprese distributrici.

La base di dati contiene informazioni relative alle variabili di scala (numero di utenti delle diverse categorie), alla consistenza tecnica delle reti (lunghezza delle linee AT, MT e BT, suddivise in base ad ulteriori caratteristiche (aeree e in cavo); numero e tipologia dei punti di trasformazione AT/MT e MT/BT; caratteristiche della rete (controalimentazione, accessibilità, presenza di vincoli idrogeologici o forestali, orografia del territorio servito). Solo per alcune aziende è stato rilevato il dato relativo all'energia distribuita, e pertanto questa variabile è stata esclusa.

Il modello ipotizzato nel DCO presenta formule differenziate per Opex e Capex, e si basa su una regressione condotta dall'Autorità e per suo conto dal Politecnico di Torino, utilizzando come variabile endogena i costi desunti dai conti separati di unbundling comunicati dalle aziende. La formula prevede, con diverse specificazioni per capex e opex, una funzione logaritmica dell'energia distribuita (EE), della densità di utenza, misurata come numero di utenti totali su totale delle linee MT e BT, di un "indice di vetustà". Le formule del capex presentano due diverse specificazioni a seconda che le imprese distributrici operino o meno in territorio montano.

Non disponendo dei dati di unbundling, né della quantità di energia distribuita, l'elaborazione è stata effettuata considerando come proxy dei costi totali i ricavi conseguiti moltiplicando le tariffe di riferimento provvisorie approvate per il 2016 per le rispettive variabili di scala. Ne deriva una parziale sottostima del costo totale riconosciuto, in quanto il valore non comprende né i ricavi per illuminazione pubblica, né quelli per la commercializzazione della distribuzione (COT). Tuttavia, ipotizzando che queste voci abbiano grosso modo la stessa proporzionalità rispetto agli altri utenti, l'approssimazione si può ritenere ragionevole.

Le finalità principali dell'elaborazione si possono così riassumere:

- Verifica della congruità dell'indicatore di densità rispetto ad altre possibili formulazioni che tengano conto della diversa tipologia delle linee
- Verifica della significatività di ulteriori variabili tecniche oltre alla mera lunghezza delle reti

- Verifica della significatività di variabili che descrivono le caratteristiche del territorio servito
- Non è stato invece possibile testare un altro fattore che alcuni operatori hanno segnalato, ossia la presenza sulla rete di inversioni di flusso legate alla presenza di punti di immissione, particolarmente centrali idroelettriche. Si è potuto testare solo l'impatto del numero di utenti attivi, che peraltro non è risultato significativo. Ci si riserva di effettuare ulteriori elaborazioni integrando la raccolta dati con l'informazione relativa ai kWh immessi nella rete nazionale.

Densità, linee di bassa e media tensione

La nostra indagine prende le mosse da una perplessità, segnalata da molti operatori, riguardo all'indicatore di densità proposto dall'Autorità. Questo infatti considera indifferentemente linee di bassa e media tensione, quando l'esperienza degli operatori suggerisce che le linee MT rappresentino una voce di costo assolutamente dominante, sia come investimento che come costo di gestione. Le linee MT infatti sono utilizzate per il trasporto dell'energia sul territorio, connettendo i diversi centri abitati. Esse attraversano spesso territori difficilmente raggiungibili soprattutto nei mesi invernali, sono maggiormente soggette a guasti di origine meteorologica (neve, fulmini) e idrogeologica (frane, valanghe).

Altri operatori hanno segnalato la possibile incidenza sui costi di altri fattori, come ad esempio:

- Il fatto che le linee siano aeree o interrate (con maggiore vulnerabilità ai guasti delle prime, ma più alti costi di investimento per le seconde)
- Il fatto che siano presenti rilevanti punti di immissione, che determinano la necessità di manovre per gestire le inversioni di flusso, richiedendo alternativamente investimenti speciali (smart grid) ovvero maggiore dotazione di personale tecnico specializzato adibito alle manovre
- Il fatto che le linee siano per ragioni obiettive esposte a danneggiamenti (neve, caduta di alberi etc) e/o richiedano comunque un presidio manutentivo.

Abbiamo pertanto effettuato una prima serie di stime, volte a confrontare una specificazione del modello simile a quello proposto dall'Autorità, con altre specificazioni che tengano conto separatamente delle diverse tipologie di linee.

Ciò è stato fatto dapprima regredendo il costo totale riconosciuto (proxy: ricavo totale da TR 2016) sul volume di attività (numero di utenti) e sulla lunghezza totale delle linee BT e MT. Si è scelto di utilizzare una specificazione logaritmica, sia perché questa forma è adottata anche nella proposta di ARERA, sia perché ciò consente di considerare una per una ciascuna variabile invece del valore medio. Questo modello è illustrato nella prima colonna della tabella 1. Pur non replicando la formula parametrica del DCO, esso si basa sulla medesima logica.

Sono state successivamente affiancate a questo modello base alcune versioni alternative, che considerano separatamente la lunghezza delle linee BT e MT. È stata inoltre introdotta una variabile dummy (cluster_MT), che vale 1 per le aziende che presentano un rapporto tra LMT e LBT particolarmente elevato.

Come si vede dalla tabella 1, il modello base – che riprende la struttura del modello di ARERA – conferma la significatività della variabile densità. Tuttavia, analizzando separatamente l'impatto di LBT e LMT, risulta chiaramente che soltanto la seconda è significativa, considerando la variabile continua (colonna 6).

Anche considerando LTOT e Cluster_MT (colonna 4), si nota come la presenza della variabile Cluster_MT, pur non aggiungendo molto al coefficiente R^2 , consente di migliorare la qualità della stima. Infatti, in questo modo è possibile tener conto della presenza di aziende con un rapporto MT/BT particolarmente elevato.

Tabella 1 – Modello base

| | (1) $y = \log \left(\frac{RICAVI}{UTENTI} \right)$ | (2) $y = \log (RICAVI)$ | (3) $y = \log (RICAVI)$ | (4) $y = \log (RICAVI)$ | (5) $y = \log (RICAVI)$ | (6) $y = \log (RICAVI)$ | (6) $y = \log (RICAVI)$ |
|---|--|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| (Intercetta) | 0.26 (0.64) | 0.26 (0.64) | 0.26 (0.64) | 0.37 (0.49) | 1.11* (0.06) | 1.10** (0.05) | 1.16** (0.04) |
| log (UTENTI) | 0.11** (0.04) | 1.11*** (0.00) | 0.85*** (0.00) | 0.83*** (0.00) | 0.76*** (0.00) | 0.77*** (0.00) | 0.75*** (0.00) |
| $\log \left(\frac{UTENTI}{LTOT} \right)$ | -0.26** (0.02) | -0.26** (0.02) | | | | | |
| log (LTOT) | | | 0.26** (0.02) | 0.26** (0.02) | | | |
| CLUSTER_MT | | | | 0.38** (0.05) | 0.21 (0.27) | 0.21 (0.25) | |
| log (LBT + 1) | | | | | 0.01 (0.94) | | |
| log (LMT + 1) | | | | | 0.29*** (0.00) | 0.29*** (0.00) | 0.31*** (0.00) |
| R ² | 0.20 | 0.93 | 0.93 | 0.93 | 0.94 | 0.94 | 0.94 |
| Num. osservaz. | 41 | 41 | 41 | 41 | 41 | 41 | 41 |
| Media di Δ% | 24 | 24 | 24 | 22 | 20 | 19 | 20 |
| Massimo di Δ% | 104 | 104 | 104 | 101 | 121 | 121 | 127 |

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. p -value tra ().

L'osservazione appena fatta permette di suggerire un approccio che, anziché stimare nuovamente le variabili del modello utilizzando le sole linee MT come variabile esplicativa, possa continuare ad utilizzare la specificazione proposta dall'Autorità, inserendo tuttavia uno specifico correttivo per le aziende che possiedono un rapporto LMT/LBT particolarmente elevato e disomogeneo rispetto al campione.

Una simile soluzione permetterebbe tra l'altro di evitare di effettuare nuove stime, potendo continuare ad utilizzare il modello proposto dall'Autorità, correggendone gli esiti solo per un numero assai limitato di imprese.

Per identificare le suddette imprese, è ovviamente necessario formulare un'ipotesi circa il valore-soglia che permette di selezionare le imprese da includere.

Al fine di scegliere la formulazione di Cluster_MT più adatta, si è effettuata un'analisi del campione volta ad evidenziare la presenza di outlier. Nella fig. 1 si evidenzia, attraverso un diagramma "box-and-whisker", la struttura del campione analizzato. Risulta evidente la presenza di alcuni casi palesemente eccentrici rispetto alla distribuzione.

È interessante notare come il valore medio del campione risulta pari a 0,4496, molto vicino al valore di riferimento nazionale di e-distribuzione, per la quale il parametro vale 0,445.

Sono state effettuate tre prove distinte, utilizzando tre diversi valori, pari a 1,5, 1,7 e 2 volte il valore medio del campione (o il valore medio nazionale, essendo i due valori praticamente identici).

La tabella 2 contiene il risultato dell'elaborazione. Come si può vedere, la variabile mantiene la significatività in tutti e tre i casi, ma risulta più significativa nella versione "più inclusiva". Il valore più elevato di R² e il valore più basso dello scostamento medio si ottengono adottando la versione intermedia (LMT/LBT > 1,75), mentre lo scostamento massimo è più limitato adottando la versione più ampia (LMT/LBT > 1,5).

Nella tabella 3 si illustrano i valori degli scostamenti dei valori modellati rispetto al dato effettivo. Come si può osservare, l'introduzione della variabile cluster_MT riduce sensibilmente il valore dello scostamento (aziende evidenziate). È altrettanto interessante osservare che questa riduzione opera in una doppia

direzione, sia riducendo lo scostamento negativo per le aziende con molte linee MT, sia, al contrario, riducendo quello positivo che verrebbe altrimenti attribuito alle aziende con prevalenza di linee BT.

Figura 1 – Diagramma box-and-whisker della variabile LMT/LBT e relativa distribuzione

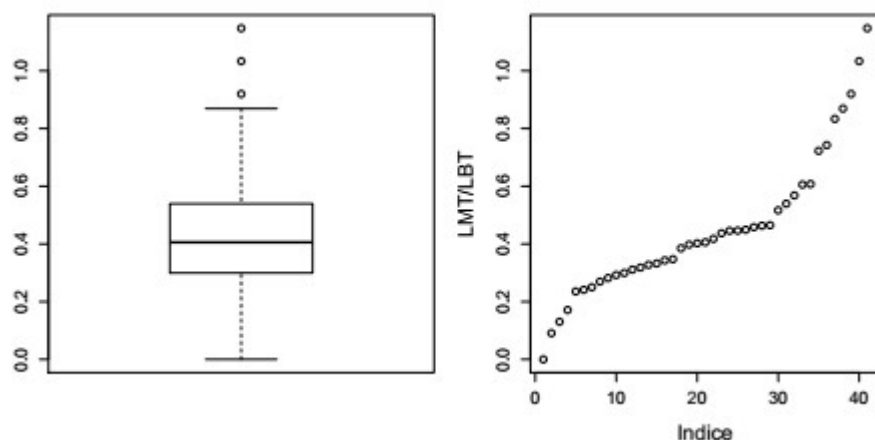


Tabella 2 – Significatività della variabile cluster_MT in tre diverse specificazioni

| | (1) $y = \log\left(\frac{RICAVI}{UTENTI}\right)$ Stima OLS | (2) $y = \log(RICAVI)$ Stima OLS | (3) $y = \log(RICAVI)$ Stima OLS | (4) $y = \log(RICAVI)$ Stima OLS | (5) $y = \log(RICAVI)$ Stima OLS | (6) $y = \log(RICAVI)$ Stima OLS |
|---|--|--|--|--|--|--|
| (Intercetta) | 0.26 (0.64) | 0.26 (0.64) | 0.26 (0.64) | 0.37 (0.49) | 0.37 (0.42) | 0.11 (0.83) |
| $\log(UTENTI)$ | 0.11** (0.04) | 1.11*** (0.00) | 0.85*** (0.00) | 0.83*** (0.00) | 0.82*** (0.00) | 0.85*** (0.00) |
| $\log\left(\frac{UTENTI}{LTOT}\right)$ | -0.26** (0.02) | -0.26** (0.02) | | | | |
| $\log(LTOT)$ | | | 0.26** (0.02) | 0.26** (0.02) | 0.27*** (0.01) | 0.26** (0.01) |
| $\frac{LMT}{LBT} > 2 \times \mu_{\frac{LMT}{LBT}}$ | | | | 0.38** (0.05) | | |
| $\frac{LMT}{LBT} > 1.75 \times \mu_{\frac{LMT}{LBT}}$ | | | | | 0.54*** (0.00) | |
| $\frac{LMT}{LBT} > 1.5 \times \mu_{\frac{LMT}{LBT}}$ | | | | | | 0.35*** (0.01) |
| R ² | 0.20 | 0.93 | 0.93 | 0.93 | 0.95 | 0.94 |
| Num. osservaz. | 41 | 41 | 41 | 41 | 41 | 41 |
| Media di $ \Delta\% $ | 24 | 24 | 24 | 22 | 20 | 22 |
| Massimo di $ \Delta\% $ | 104 | 104 | 104 | 101 | 95 | 92 |

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. p -value tra ().

Di quanto sopra discusso si potrebbe tenere conto in modo relativamente semplice, mantenendo l'impianto del modello proposto nel DCO.

Un modo per farlo potrebbe essere ad esempio quello di introdurre un coefficiente $g > 1$ per quelle aziende che dovessero presentare un rapporto superiore alla soglia individuata.

Tabella 3 – Scostamento del valore modellato dal valore effettivo per diverse specificazioni del modello

| ID | (1) $y = \log \left(\frac{\text{RICAVI}}{\text{UTENTI}} \right)$ Stima OLS | (2) $y = \log (\text{RICAVI})$ Stima OLS | (3) $y = \log (\text{RICAVI})$ Stima OLS | (4) $y = \log (\text{RICAVI})$ Stima OLS | (5) $y = \log (\text{RICAVI})$ Stima OLS | (6) $y = \log (\text{RICAVI})$ Stima OLS |
|-------------------------|---|--|--|--|--|--|
| 9 | 45 % | 45 % | 45 % | 37 % | 31 % | 40 % |
| 10 | -6 % | -6 % | -6 % | -11 % | -14 % | -9 % |
| 11 | -23 % | -23 % | -23 % | -27 % | -30 % | -27 % |
| 12 | 26 % | 26 % | 26 % | 19 % | 15 % | 20 % |
| 13 | 24 % | 24 % | 24 % | 18 % | 13 % | 18 % |
| 14 | 15 % | 15 % | 15 % | 9 % | 5 % | 10 % |
| 15 | 47 % | 47 % | 47 % | 40 % | 34 % | 40 % |
| 16 | 10 % | 10 % | 10 % | 5 % | 2 % | 5 % |
| 17 | -13 % | -13 % | -13 % | 23 % | 37 % | 18 % |
| 19 | -12 % | -12 % | -12 % | -15 % | -18 % | -16 % |
| 20 | 15 % | 15 % | 15 % | 11 % | 7 % | 9 % |
| 21 | 18 % | 18 % | 18 % | 14 % | 10 % | 12 % |
| 22 | 19 % | 19 % | 19 % | 15 % | 11 % | 13 % |
| 23 | -18 % | -18 % | -18 % | -20 % | -23 % | -22 % |
| 24 | 20 % | 20 % | 20 % | 17 % | 12 % | 14 % |
| 25 | -43 % | -43 % | -43 % | -20 % | -9 % | -24 % |
| 26 | -14 % | -14 % | -14 % | -16 % | -19 % | -19 % |
| 27 | 16 % | 16 % | 16 % | 12 % | 8 % | 9 % |
| 30 | -21 % | -21 % | -21 % | -23 % | 27 % | 5 % |
| 31 | 104 % | 104 % | 104 % | 101 % | 95 % | 92 % |
| 32 | -19 % | -19 % | -19 % | -21 % | -24 % | -25 % |
| 33 | 16 % | 16 % | 16 % | 14 % | 10 % | 8 % |
| 34 | 1 % | 1 % | 1 % | 0 % | -5 % | -6 % |
| 35 | -4 % | -4 % | -4 % | -5 % | -8 % | -10 % |
| 36 | -26 % | -26 % | -26 % | -27 % | -29 % | -31 % |
| 37 | 6 % | 6 % | 6 % | 6 % | 2 % | 40 % |
| 38 | 30 % | 30 % | 30 % | 30 % | 26 % | 71 % |
| 39 | -28 % | -28 % | -28 % | -28 % | -30 % | -33 % |
| 41 | 81 % | 81 % | 81 % | 83 % | 78 % | 66 % |
| 43 | 19 % | 19 % | 19 % | 16 % | 11 % | 11 % |
| 44 | 4 % | 4 % | 4 % | 2 % | -3 % | -3 % |
| 45 | 6 % | 6 % | 6 % | 6 % | 1 % | -3 % |
| 46 | 22 % | 22 % | 22 % | 18 % | 14 % | 15 % |
| 47 | 23 % | 23 % | 23 % | 22 % | 17 % | 15 % |
| 48 | -24 % | -24 % | -24 % | -25 % | -27 % | -29 % |
| 50 | -16 % | -16 % | -16 % | -18 % | -21 % | -21 % |
| 51 | -28 % | -28 % | -28 % | -30 % | -32 % | -32 % |
| 52 | 10 % | 10 % | 10 % | 5 % | 2 % | 6 % |
| 53 | -28 % | -28 % | -28 % | 2 % | 16 % | -2 % |
| 54 | -66 % | -66 % | -66 % | -67 % | -45 % | -55 % |
| 55 | -2 % | -2 % | -2 % | -5 % | -9 % | -8 % |
| Media di $ \Delta\% $ | 24 % | 24 % | 24 % | 22 % | 20 % | 22 % |
| Mediana di $ \Delta\% $ | 19 % | 19 % | 19 % | 18 % | 15 % | 16 % |
| Massimo di $ \Delta\% $ | 104 % | 104 % | 104 % | 101 % | 95 % | 92 % |

Impatto di variabili tecniche e geografiche

Come ulteriore contributo, si è cercato di vedere se altre variabili potessero aggiungere elementi informativi al modello. Data l'esiguità del campione, non è stato possibile effettuare un test simultaneo delle possibili variabili esplicative. Si è adottata la scelta di inserirle ad una ad una, a partire dal modello base sopra esaminato.

Sono state esaminate in particolare:

- Presenza di linee interrate
- Numero di punti di trasformazione
- Collocazione in ambito montano (dummy che attribuisce valore 1 ai comuni situati in "alta montagna" in base alla classificazione Istat)
- Un indicatore di accessibilità (% di linee situate in luoghi distanti dalla viabilità ordinaria)
- Un indicatore di complessità del territorio servito (% di linee non controalimentate)
- Il dislivello (misurato come differenza tra la quota altimetrica massima e minima del territorio servito)
- La superficie del comune in rapporto alla popolazione
- Il numero di utenti attivi

In generale, tra le variabili esaminate nessuna sembra mostrare livelli di significatività elevati. Solo la superficie e la popolazione (con i segni attesi) e l'indice di accessibilità mostrano una qualche influenza, peraltro debole. È possibile che quest'ultimo indicatore sia "sporco" per via del fatto che si basa su un'autovalutazione qualitativa fornita dalle aziende e quindi potrebbe scontare margini di indeterminatezza.

La non significatività della variabile "montagna" può dipendere dal fatto che le imprese del campione sono collocate pressoché interamente in territorio montano. Essa diventa non a caso significativa estendendo l'analisi all'intero campione delle imprese sotto i 100.000 utenti, laddove nella fascia 25 – 100 kPOD sono comprese soprattutto aziende non montane.

Come si è detto, non è stato possibile testare la significatività della variabile "inversione di flusso", mancando per la maggior parte delle aziende il dato relativo all'energia immessa in rete nazionale. Questa è una verifica che, in base all'esperienza suggerita dagli operatori, potrebbe avere senso provare.

Si può comunque sostenere che l'analisi svolta conferma la poca utilità di affannarsi nella ricerca di parametri e grandezze esogene troppo specifiche, anche perché in ogni caso queste rifletteranno solo indirettamente le vere cause determinanti dei costi, che si possono sempre, in ultima analisi, riassumere nell'esigenza di mantenere un presidio operativo sul territorio e/o in costi di investimento più elevati.

Conclusioni

Dall'analisi svolta, certamente parziale e limitata anche dalla indisponibilità di alcune delle informazioni necessarie, si possono comunque evincere alcune interessanti conclusioni che possono essere utili a calibrare il modello parametrico proposto dall'Autorità.

Si conferma, in primo luogo, la difficoltà di catturare variabili esogene misurabili con effetti significativi al di là di quelle più semplici e ovvie, già contemplate nel modello di ARERA. Ciò suggerisce l'inopportunità di insistere in ulteriori ricerche in questa direzione.

Ciononostante, gli scostamenti dei dati modellati rispetto a quelli effettivi permangono piuttosto elevati, determinando variazioni potenzialmente anche molto grandi tra i ricavi ammessi fino al 2016 e quelli dovuti alla tariffa parametrica, in entrambe le direzioni.

Il fatto che i modelli econometrici non riescano più di tanto a rendere ragione di questi scostamenti non autorizza a concludere che si tratti di operatori inefficienti (o viceversa assai efficienti). È semmai più probabile che i piccoli numeri enfatizzino la rilevanza delle situazioni site-specific, difficili da modellare in quanto non si manifestano ovunque con la stessa intensità. Simili effetti verrebbero mascherati in gestioni di

maggiori dimensioni, in quanto con l'aumentare delle dimensioni le diverse fonti di rumore statistico più facilmente si compensano. Non si può tuttavia confondere questo effetto con un'economia di scala.

Il risultato sicuramente più robusto riguarda la necessità di tenere conto in qualche modo della diversa struttura delle reti, con particolare riferimento all'incidenza relativa delle linee BT e MT. È accertato in modo chiaro che i costi sostenuti dagli operatori risentono molto più direttamente della presenza delle seconde. Ciò sarà presumibilmente più vero in quei casi in cui la dispersione delle località sul territorio obbliga la rete a disporre di rami relativamente lunghi di "adduzione principale" in MT.

Un modo per tenere conto di questo effetto senza stravolgere le formule già calibrate potrebbe essere quello di inserire un correttivo (coefficiente g) per le aziende che presentano un rapporto tra LMT e LBT superiore alla media nazionale; il nostro studio suggerisce che gli effetti siano visibili quando tale rapporto risulti almeno a 1,5 volte la media nazionale, mentre scegliendo il parametro più elevato (2 volte la media) il numero di aziende identificate nel cluster sarebbe troppo piccolo (due soltanto).

Concretamente, quanto suggerito potrebbe essere attuato prevedendo un coefficiente moltiplicativo da applicare alla capex (o al costo totale) in funzione dell'appartenenza o meno al cluster_MT (o, con valore inferiore all'unità, a quelle aziende che presentino un'incidenza delle linee MT molto più bassa).

Rimane da verificare la possibilità di un'incidenza della presenza di utenti attivi – misurata dalla quantità di energia immessa nella rete nazionale con inversione di flusso. Qualora ciò fosse dimostrato, se ne potrebbe tenere conto nella formula inserendo ad esempio un addendo (di segno opposto a quello che esprime l'impatto dell'energia distribuita) che sia funzione dell'energia restituita in cabina primaria.