Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Measuring Demand for Access to Regional Airports: An Application of Zero-Inflated Poisson Regression

Tony Diana

MEASURING DEMAND FOR ACCESS TO REGIONAL AIRPORTS: AN APPLICATION OF ZERO-INFLATED POISSON REGRESSION

Tony Diana

Abstract

The demand for access to regional airports is measured by the counts of area navigation (RNAV)/required navigation (RNP) procedures filed in the flight plan and collected by the Federal Aviation Administration's host computer system. In the present paper, the concept of access implies how easily and consistently equipped and authorized aircraft can utilize performance-based navigation at regional airports where there is no instrument landing systems (ILS). The demand for access to regional airports is likely to be strongly correlated to on-board surveillance equipment and to business jet type of aircraft whether privately or company-owned. The incidence of surveillance equipment is higher at airport such at Van Nuys (VNY) where the proportion of jets is the highest.

Introduction

Access is one of the eleven key performance areas (KPA) advocated by the International Civil Aviation Organization (ICAO) to measure the operation of the National Airspace System (NAS). *Access* comes from the Latin word *accedere* which means 'to approach'. Getting access to an airport is often associated with the concepts of equity and capacity: In the present paper, it implies how easily and consistently capable and authorized aircraft can utilize the resources of the NAS given regional airports' available capacity.

The sample includes three regional airports where an operator can use Localizer Performance with Vertical guidance (LPV) to land in the absence of an instrument landing system (ILS) and one that has both capabilities. Based on information provided by the Federal Aviation Administration (FAA)[1], LPV is similar to Lateral Navigation / Vertical Navigation (LNAV/VNAV) except:

1. It is much more precise (40-nautical-mile lateral limit),
2. It enables descent to 200-250 feet above the runway, and
3. It can only be flown with a Wide Area Augmentation System (WAAS) receiver.

LPV approaches are operationally equivalent to the legacy ILS but are more economical because no navigation infrastructure has to be installed at the runway. The use of LPV is important to the evolution of the air traffic system because it allows general aviation users to get access to

---

[1] The information can be retrieved at http://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/navservices/gnss/hill/media/WAAS-LPV-Q_A-020607.pdf

regional airports regardless of meteorological conditions, approach and terrain conditions.

This paper uses a Zero-Inflated Poisson (ZIP) regression model to evaluate whether the demand for performance-based navigation (PBN) at regional airports can be explained by selected variables such as type of aircraft, on-board surveillance equipment, aircraft size, and type of ownership. The demand for PBN is measured by the counts of RNAV/RNP procedures filed in the flight plan for a sample of four regional airports. Performance-based navigation refers to area navigation (RNAV) and required navigation performance (RNP). On-board performance monitoring and alerting requirements represent the main difference between RNAV and more demanding RNP procedures.

This paper intends to assess what factors are likely to influence the demand for PBN procedures and, by extension, measure the demand for access at regional airports. This is important for airport analysts who are considering investments to expand access at regional airports as well as for operators who try to assess the return on investment from their satellite-based equipage. After reviewing the characteristics of the Poisson distribution and the assumptions of the Zero-Inflated Poisson regression model, the paper will proceed with an explanation of the outcomes before concluding with some potential implications.

Methodology
Poisson and Zero-Inflated Poisson Regression Models

Readers interested in Poisson and negative binomial regression models are referred to Cameron and Trivedi (1998), Winkelmann (2008), and Hilbe (2011). A Poisson regression model is based on the following assumptions:

1. Each filed RNAV/RNP approach event is independent of one another. The fact that a precision approach was filed during the month of July 2011 has no bearing on how many events are counted in any other months.
2. The counts of filed RNAV/RNP approaches are random.

The Ordinary Least Squares model cannot be used in this study for the following reasons:

1. Counts are random, non-negative binary values,
2. The relationship between the dependent and independent variables is non-linear,
3. The counts are heteroskedastic, and

4. It is not possible to predict a negative value for filed RNAV/RNP procedures.

The Poisson regression models the log outcome rate as a linear function of a set of predictors. Given the vector of regressor X_i, the dependent variable Y_i is independently Poisson distributed with density:

$$f(Y_i \mid X_i) = e^{-\mu_i} \mu_i^{Y_i} / Y_i! \quad (1)$$

With $Y_i = 0, 1, 2..., n$ and mean parameter $\mu_i = \exp(x'\beta)$ where β is a $k * 1$ parameter vector.

One key characteristic of the Poisson distribution is equidispersion expressed mathematically as:

$$E[Y_i \mid X_i] = VAR[Y_i \mid X_i] = \exp(x'\beta) \quad (2)$$

When the equidispersion assumption is violated, a negative binomial regression usually represents an alternative to remedy the violation. That was not the case in this study since the variance of the response variable is relatively close to the mean of the response variable. For all the samples, the deviance and the scaled deviance have the same value, which suggests no overdispersion.

A Zero-Inflated Poisson regression model was selected because an excess of zeros (no filed RNAV/RNP approach) can bias the outcomes. In fact, excessive zeros can be a source of overdispersion. Therefore, ZIP regression models minimize the risk of biased estimates and underestimated standard errors. The ZIP regression model contains two components:

1. A Poisson counts model and
2. A logit model for predicting excess zeros.

The ZIP regression model was first introduced by Lambert (1992) to account for excessive zero counts in applications to defects in manufacturing. Zero counts are incorporated into the binary and count processes as detailed in the appendix.

The Airport Sample

The period of study is July 2011. The sample includes the following regional airports:
- Napa Valley County in California (APC,
- Naples in Florida (APF),
- Wicomico Regional Airport in Maryland (SBY),
- Van Nuys in California (VNY). Van Nuys offers both ILS and LPV capabilities.

Table 1 provides the total counts of arrivals and departures based on OPSNET data[2]. Scheduled air carriers are considered as operating under Part 121 of the Federal

---

[2] OPSNET data can be queried at https://aspm.faa.gov/opsnet/sys.

Aviation Regulations. Air taxi operations are governed by Part 135 of the Federal Aviation Regulations and involve aircraft operations with less than 30 passengers, a payload of 7,500 lbs. or less, for hire or compensation[3]. Commuter flights are included in the air taxi category. Local movements involve aircraft that depart from and return to the same airport.

---

[3] Title 14 Aeronautics and Space regulations can be retrieved at http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?&c=ecfr&tpl=/ecfrbrowse/Title14/14tab_02.tpl

Table 1

Average Daily Arrivals and Departures (Operations) in July 2011

| Facility | Itinerant | | | | | Local | | | Total |
	Air Carrier	Air Taxi	General Aviation	Military	Total	Civil	Military	Total	Operations
APC	0	24	78	1	103	116	0	116	220
APF	0	9	109	0	119	48	0	48	167
SBY	0	19	44	8	71	42	37	78	149
VNY	0	28	575	1	604	313	0	313	918
Total:	0	81	807	9	898	520	37	556	1,454

Source: OPSNET

The Traffic Flow Management System (TFMS) database[4] provides information about equipment and user type. Table 2 shows that jets are predominant at the APC, APF and VNY. Air taxi operations include commuters such as Piedmont (PDT) that flew 464 scheduled arrivals and departures[5] on behalf US Airways (USA) at SBY with Bombardier Q300 (DHC-8) turboprop aircraft. That explains the high number of turboprop aircraft at SBY compared with the other airports in the sample.

Table 2

Equipment Type by Airport in July 2011

	Jet	Piston	Turboprop
APC	64.3%	17.9%	17.8%
APF	50.9%	37.6%	11.5%
SBY	11.9%	18.8%	69.3%
VNY	68.4%	21.5%	10.1%

Source: Traffic Flow Management System

---

[4] The website to query TFMS data is https://aspm.faa.gov/tfms/sys.

[5] Based on the airline schedules published by Innovata (http://www.innovata-llc.com). Schedule for July 2011.

Variables and Data Sources

The data from the FAA's Advanced Technologies and Oceanic Procedures and Traffic Flow Management System (TFMS) were processed with the SAS® software. The GENMOD and COUNTREG procedures were used to derive the estimates and statistics of the Zero-Inflated Poisson regression models. There were 325 observations for APC, 350 for APF, 106 for SBY and 728 for VNY.

The dependent variable is the counts of filed Area Navigation (RNAV)/Required Navigation (RNP) approaches into selected regional airports. The variable is coded as '1' if the RNAV/RNP approaches were filed and '0' otherwise. There are several issues associated with tracking actually flown RNAV/RNP approaches. First, although pilots may file or request an RNAV/RNP procedure, they may actually not fly it because traffic mix makes it difficult (i.e. some aircraft in a landing queue are capable, whereas others are not). Second, air traffic controllers and pilots may not be trained to carry through PBN procedures. Third, some aircraft in an arrival stream are equipped to fly PBN procedures, while others are not. Finally, it is difficult to distinguish a PBN from an ILS procedure as their respective path can be overlaid.

There are four independent variables or predictors:

- The *type of surveillance equipment* (mode) refers to the type of aircraft transponder on board an aircraft. If an aircraft is equipped with an 'S' transponder (including both pressure-altitude and identification transmission), the record flag is '1' and '0' otherwise. The other transponder types are 'C' transponder (Mode A and C) and 'P' transponder (Mode S, including pressure-altitude, but no aircraft identification transmission).

- The *wake category* (wakecat) is a variable coded as '1' for larger aircraft and '0' otherwise. Wake category denotes the impact of wake turbulence on other aircraft. It is more significant as the size of the aircraft increases.

- The variable *type of owner* (owner) is coded as '1' if the aircraft is identified by an alpha-numeric combination including the 'N' country registration code (ex. N2345). In this case, the aircraft is assumed to be flown privately. It is '0' if the flight is identified by a three-letter airline code plus numbers (ex. OPT123, PDT123). In the study sample, the records with airline codes information are either fractional flights[6] or commuter flights.

- The *type of aircraft* (bizjet) refers to business jet (coded as '1' for business jets) versus turbo-prop aircraft (coded as '0'). A business jet is defined in this study as a turbo-fan jet aircraft with seating capacity from 3 (i.e. Eclipse 500 jet) to about 15 passengers (i.e. Gulfstream V) used for private use (aircraft identified by an 'N' number) or revenue service (in the case of fractional jets, for instance).

Outcomes

Table 3 shows the estimates for the Poisson model and those for the logit model (inflated estimates in the shaded area).

---

[6] Factional flights are operated by companies such Flight Options flying mostly under Title 14 of the Code of Federal Regulations (CFR) Part 91, Subpart K instead of Part 135 (under commercial aviation regulations).

Table 3

Zero-Inflated Regression Model Outputs (July 2011)

Parameter	APC Estimate	APC Approx. Pr > |t|	APF Estimate	APF Approx. Pr > |t|	SBY Estimate	SBY Approx. Pr > |t|	VNY Estimate	VNY Approx. Pr > |t|
Intercept	-2.0759	0.0003	-2.4424	0.0009	-0.4296	0.7415	-1.6013	0.0132
mode	1.1692	0.0009	1.9559	0.0012	-0.2216	0.8063	0.7393	0.2407
wakecat	0.2073	0.7197	-0.9742	0.026	-2.2865	0.086	-2.0984	<.0001
owner	-0.3550	0.0551	0.0464	0.8569	-0.2159	0.5957	-0.1992	0.0719
bizjet	0.7052	0.1691	-0.1513	0.7714	0.0026	0.9979	0.6148	0.0917
Inf_Intercept	-34.7012	<.0001	-0.4844	0.8978	28.7162	0.8146	15.6714	<.0001
Inf_mode	0.4332	0.7984	2.4698	0.5356	-41.8278	0.8927	-54.6324	.
Inf_wakecat	36.2383	<.0001	-1.5937	0.357	-71.9167	0.0026	-33.5272	<.0001
Inf_owner	-16.5073	<.0001	-15.1805	.	28.0319	0.9196	-30.6492	<.0001
Inf_bizjet	15.7086	<.0001	-19.7458	.	-29.5424	0.8094	15.8893	<.0001

☐	Poisson model for the RNAV/RNP approaches that are not certain zero cases.
▨	Logit model for the RNAV/RNP approaches that are certain zero cases.

Inf_ Inflated variable.

As mentioned earlier, the ZIP regression model has two components. On the one hand, the logistic model predicts whether or not an RNAV/RNP was filed in the flight plan: It is used to predict whether a filed RNAV/RNP approach is a certain zero. On the other hand, the Poisson model is generated to predict the counts for those filed RNAV/RNP approaches that are not certain zeros.

The Poisson regression models the log of the expected counts as a function of the independent variables. Not all the predictors in the count and inflated model are statistically significant at a 95% confidence level. Taking the example of APC, the coefficient for mode (1.1692) represents the proportional change in the conditional mean of the number of filed RNAV/RNP procedures if the counts of aircraft with 'S' surveillance equipment changes by one unit. The estimates can also be interpreted as follows: The expected number of filed PBN procedure changes by 3.22

[exp(1.1692)] for each unit increase in the counts of aircraft equipped with the 'S' surveillance mode, holding other predictors constant. The expected number of aircraft with 'S' transponder is 3.22 times [exp(1.1692)] the expected counts of aircraft with 'C' or 'P' transponders. For any additional filed RNA/RNP approach, the odds that an aircraft would have a 'C' or 'P' transponder is 1.54 [exp(0.4332)], that is, about 60% [1.54/(1+1.54)]. At SBY, the expected number of filed PBN procedure changes by 0.80 respectively for each unit increase in the number of aircraft equipped with the 'S' surveillance mode or for each unit increase in the number of private owned aircraft flying into SBY, holding other predictors constant. Moreover, the expected number of filed PBN procedure changes by 1.00 respectively for each unit increase in the number of larger aircraft, holding other predictors constant. The difference between SBY and the other airports can be explained by the incidence rate ratios explained later.

While the logit model handles the 'certain zero cases' with the inflated model, the PBN procedures that were actually filed are handled by the Poisson model. Table 3 shows the type of surveillance equipment was not significant at a 95% confidence level when no RNAV/RNP procedure was filed, holding other variables constant. The counts of aircraft with mode 'S' equipment is significant when the counts of filed RNAV/ RNP are not certain zero counts.

In the cases of APC and VNY, the counts of aircraft privately owned is significant whether or not a filed RNAV/RNP is a certain zero. When no procedure was filed, then the wake category (larger aircraft) was significant in all the cases. The outcomes point out that when the counts of filed RNAV/RNP are certain zeros, then the wake category, the type of owner and the type of aircraft are all significant estimates at a 95% confidence level at APC and VNY.

Another important element in counts models is the incidence rate that Hilbe (2011) defined as "the rate at which counts enter a time period or area" and the incidence rate ratio as "a variety of risk ratio" (p. 111). Since β as the regression coefficient is the difference between $\log(\mu_{x+1})$ and $\log(\mu_x)$, the difference of two logs can be expressed as the log of their quotient ($\log(\mu_{x+1}/\mu_x)$). Therefore, incidence rate ratios refer to the estimate as the log of the ratio of expected counts. So, if the number of aircraft with 'S' transponder was to increase by one unit, the rate ratio for the number of filed RNAV/RNP would be expected to increase by a factor of 3.10 at APC, holding all other factors in the model constant. The incidence ratios were 6.68 and 4.92 respectively at VNY and APF were the proportion of jet aircraft represented respectively 68.4% and 50.9% of the operations. The incidence ratio was the lowest at SBY where 69.3% of the observed equipment was turboprop (Table 2).

Table 4

The Incidence Rate Ratios

Parameter	APC	APF	SBY	VNY
mode	3.1000	4.9247	1.2698	6.6813
wakecat	0.3313	0.4488	0.5826	0.1679
owner (0)*	1.2484	0.8288	1.4771	1.2064
bizjet	1.2549	2.3382	4.5797	1.7294

*When the owner is not private.

Another important statistic is the average marginal effects: It is the product of the mean value of the number of filed RNAV/RNP approaches at each airport multiplied by the coefficient in for each factor in Table 3. For instance, the average marginal effect of aircraft equipped with "S" surveillance equipment on the counts of filed RNAV/RNP approaches is computed as the mean times the coefficient, that is 4.71 multiplied by 1.1692. Therefore, for each additional aircraft with "S" equipment, there is an average of 5.5 additional filed RNAV/RNP approaches. The exception in this sample is SBY where for each additional aircraft with "S" surveillance equipment, there is practically no change in filed RNAV/RNP approaches. The coefficients are listed in Table 5:

Measuring Demand for Access to Regional Airports

Table 5

The Average Marginal Effects

Parameter	APC	APF	SBY	VNY
mode	5.5066	6.3093	-0.2140	8.1801
wakecat	0.9761	-3.1425	-2.2076	-23.2175
owner	-1.6718	0.1498	-0.2084	-2.2037
bizjet	3.3213	-0.4882	0.0025	6.8023

For non-inflated coefficients.

At VNY where 68.4% of the aircraft are jets and 21.5% are piston, there is an average of 23.21 fewer filed RNAV/RNP approaches for each additional large aircraft that intend to land there. This implies that the traffic mix is likely to have an impact on the counts of filed RNAV/RNP approaches. There is a positive, linear relationship between the counts of filed RNAV/RNP and the counts of large aircraft at VNY (the Pearson correlation coefficient is ρ = 0.55 and significant at a 95% confidence level).

Table 5 also shows that there is an average of 3.32 and 6.80 additional filed RNAV/RNP approaches for each additional business jets flying respectively into APC and VNY. These are most likely high-performance general aviation aircraft with sophisticated equipment on board. The Pearson correlation coefficient between the counts of filed RNAV/RNP procedures and business jets is respectively .71 and .90 at APC and VNY compared with .72 at APF and .58 at SBY.

Final Remarks

Facilitating access is a priority for air traffic control agencies since aviation is a key enabler of human and economic exchanges. Access to regional airports is all the more important as the Next Generation of Transportation System (NextGen) aims to transform the present air traffic controlled into the future air traffic managed system relying on satellite navigation instead of radars. In this paper, access is as the ability of equipped aircraft operators to utilize satellite navigation technologies at regional airports where there is no ILS. VNY served as a control airport in the sample since it offered both capabilities at the time of the survey.

The demand for access was measured by a Zero-Inflated Poisson regression models that handles the counts of filed RNAV/RNP approaches when they are certain zeros (logit model) and when the counts for those filed RNAV/RNP approaches are not certain zeros (Poisson model). Although procedures are filed, it is not possible, at the time of this writing, to determine with accuracy whether a performance-based navigation procedure has been actually flown, especially when RNAV/RNP procedures are laid over traditional ILS approaches.

The demand for RNAV/RNP procedures at regional airports is expected to depend on the technology onboard turbo-fan aircraft with 3 to 15 seats, traditionally called 'business jets'. The outcomes of the models show that the use of mode 'S' transponder was not significant at a 95% confidence level when there was a demand for RNAV/RNP procedures. When there are no filed PBN procedures, it appears that large aircraft, private owners and business jets are significant only at VNY and APC where the percentage of jet aircraft is the highest and the average marginal effects of private ownership may be significant. If the counts of aircraft equipped with a mode 'S' transponder was to increase by one unit at VNY, the rate ratio for the counts of filed RNAV/RNP approaches would be expected to increase by a factor of 6.68 at VNY where jet aircraft account for 68% of the traffic, holding all other factors in the model constant. VNY is also the only airport in the sample to offer both LPV and ILS capabilities. Business jets are more likely to request RNAV/RNP approaches at SBY where most of the traffic is turboprop.✈

Tony Diana is the division manager, NextGen Performance at the U.S. Federal Aviation Administration. He received his Ph.D. and Master's Degree in policy analysis from the University of Maryland Baltimore County (UMBC). He is involved in the measurement of operational outcomes resulting from the implementation of NextGen initiatives at U.S. airports, metroplexes and airspaces. Prior to that position, he was deputy division manager, forecasting and performance analysis in the Office of Aviation Policy and Plans of the FAA where he managed the aviation system performance metrics data warehouse. At the Maryland Aviation Administration, Dr. Diana was involved in performance measurement and route development. His main interests are performance evaluation and benchmarking, as well as the study of delay.

38

References

Cameron, C. and Trivedi, P.K. (1998). *Regression analysis of count data (Econometrics Society Monographs)*. Cambridge: Cambridge University Press.

Hilbe, J. (2011). *Negative binomial regression*. Cambridge: Cambridge University Press.

Lambert, D. (1992). *Zero-inflated Poisson regression with an application to defects in manufacturing. Technometrics*, 31, 1-14.

Winkelmann, R. (2008). *Econometric analysis of count data* (Fifth Edition). Berlin: Springer Verlag.

Appendix A

The Zero-Inflated Regression Model[7]

A ZIP regression model has a Poisson counts component. Given

$$g(Y_i) = e^{-\mu} \mu_i^{Y_i}/Y_i!$$

where $\mu_i = \exp(x'\beta)$, the Zero-Inflated Poisson model can be defined as follows:

$$Pr(Y_i = 0 \mid X_i, Z_i) = F_i + (1 - F_i) \exp(-\mu_i) \text{ and } Pr(Y_i \mid X_i, Z_i) = (1 - F_i) e^{-\mu} \mu_i^{Y_i}/Y_i, Y_i > 0$$

It also has a logit component for predicting excess zeros. For a ZIP regression model with logistic link function, the probability Φ_i is computed as

$$\Phi_i = \exp(Z_i'\gamma)/(1 + \exp(Z_i'\gamma))$$

---

[7] SAS/Stat (R) 9.2 (2nd Ed.). (2013). User's Guide. Retrieved from
http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm.

40