Supplemental Appendix For: When Does Open Government Shut? Predicting Government Responses to Citizen Information Requests*

Benjamin E. Bagozzi,† Daniel Berliner,‡ and Zack W. Almquist§

August 30, 2019

Contents

*The authors declare no conflicts of interest. Bagozzi's research was supported in part the University of Delaware General University Research Fund. Almquist's research was supported in part by an ARO YIP Award #W911NF-14-1-0577.

†Department of Political Science & International Relations, University of Delaware. Email: bagozzib@udel.edu.

‡Department of Government, London School of Economics. Email: danberliner@gmail.com.

§Department of Sociology & eScience Institute, University of Washington. Email: zwa@uw.edu.

Overview

In this supplementary material, we first summarize the preprocessing steps used for our main access to information (ATI) request text corpus and then fully present the alternate sLDA (hyper)parameter selection routine that we briefly mentioned in main paper: simultaneous selection of k, α and η. We build upon this with a summary of our Spanish topwords (both stemmed and de-stemmed) corresponding to the top five "denied request" topics and top five "provided request" topics discussed in the main paper. This is followed by a series of in-sample χ^2 tests, time series plots, and "middle-leverage" topic interpretations (each referenced in the main paper) that together allow us to more extensively explore the uniqueness and representativeness of our key Denied and Provided topics.

We next present and discuss a set of alternative in-sample and out-of-sample sLDA classification comparisons, which apply sLDA to our request texts while withholding each request's *target agency* name as an additional feature. Finally, we fully compare our sLDA model's classification performance to that of LASSO, ridge, and logistic regression, followed by a comparison of our main in-sample sLDA topic output to that of a comparably specified structural topic model (STM).

Text Preprocessing

This summary describes the preprocessing steps that we applied to our merged ATI request texts. As discussed in the main paper, these merged request texts encompass respondents' main request text entries that appeared in INFOMEX's "DESCRIPCIÓN SOLICITUD" field, as well as any corresponding supplemental request text information where applicable. The latter supplemental request text material encompasses both the optional "OTROS DATOS" field within the INFOMEX system's publicly available metadata, and any optional attachment files. Linked attachment files were web-scraped separately and added to our corpus after being converted to plain text files. Conversion was done using optical character recognition software for relevant PDF and image files, and using a dedicated set of text extraction scripts for all remaining (e.g., Microsoft Word or Microsoft Excel) files. As noted in the main paper, we truncated exceptionally long combined documents at the 1000th string onwards. This step that only affected 0.02% of our documents.

During preprocessing, we next sought to remove any character or numeric entries that appeared frequently throughout the text, but without consistent meanings. This is in keeping with extant preprocessing steps for bag-of-words models such as those applied in our main paper (e.g., Puschmann and Scheffler, 2016; Bagozzi, Berliner and Almquist, 2016; Berliner, Bagozzi and Palmer-Rubin, 2018; Berliner et al., 2020; Zhang et al., 2018). To this end, we removed all numbers mentioned in our ATI request texts, including both Arabic and Roman numerals as well as Spanish words for relevant numbers. We similarly removed floating letters (e.g., 's', 't', 'x', etc.) and non-graphical characters. These can each occur within our merged request texts due to noise in the optical character recognition (OCR) of some requests' image attachment files. Likewise, we removed the Spanish names for months, and any websites that were mentioned the combined text.[1]

Requesters are required to supply their request within INFOMEX's main request-entry field, rather than supplying their ATI request solely as an attachment. However, many

[1]In the latter case, we specifically removed any character string beginning with "www" or "http".

requesters circumvented this requirement by including placeholder text within the request-entry field whilst providing their full ATI request within an accompanying attachment. As mentioned above, these attachment requests have been separately scraped, OCR'd (when applicable), and merged into our main request text corpus. However, for these cases of circumvention, we also sought to remove the most common forms of "placeholder text" that these individuals typically used within their main request text entries. This included the removal of entry text such as 'xxxxxxx' (of varying length), the removal of main request or OTROS DATOS text that simply re-stated "DESCRIPCIÓN SOLICITUD" or "OTROS DATOS," the removal of entry text that simply stated "ninguno" (i.e., "nothing"), and the removal of entry texts that corresponded to approximately 65 distinct Spanish phrases for statements such as "see the attached file," or "request attached." Importantly, in each of these cases, the actual request entry remained in our sample as a document, along with any remaining text that was included via an attachment (or via "OTROS DATOS").

Following the above steps, we next sought to address a number of misspellings that frequently arose within the main request text field of our ATI request documents.[2] First, we corrected grave accents to acute accents given the sole usage of the latter in Spanish. Some requesters also omitted accents entirely when typing their requests into INFOMEX. Hence, where appropriate, we standardized accents for all relevant variants of the following commonly used words: "información," "corrupción," "constitución," as well as for other relevant words ending in "-íon," "-pón," "-dón," "-tón," "-zón," "-ería," "-ísimo," and all Mexican state names that contain accents.[3] We next similarly standardized instances of "Mexico" to "México." We then converted all words to lowercase, removed punctuation and excess whitespace, and stemmed all Spanish words using the Porter stemming algorithm (Porter, 1980).[4] Finally, we omitted all sparse terms across our remaining corpus that did not occur in at least

[2]These misspellings appeared to arise because INFOMEX's request information was typically hand-typed into the INFOMEX system by each requester.

[3]Whilst simultaneously converting bigram and trigram Mexican state names to unigrams.

[4]We then also stemmed años to año separately, as this was not stemmed within the standard stemming implementation that we used.

0.1% of all retained request documents. These latter steps were implemented via the tm package in R (Feinerer, Hornik and Meyer, 2008) and have wide precedent as preprocessing steps within topic modeling analyses such as our own (e.g., Quinn et al., 2010; Roberts et al., 2014; Bagozzi, 2015; Puschmann and Scheffler, 2016; Bagozzi, Berliner and Almquist, 2016).

Alternate Selection of k, α, and η

Recall that our main paper's selection of k, α and η proceeded in two sequential steps. We first selected k based upon five fold cross-validation while holding our α and η hyper-parameters fixed at 1.0 and 0.1, respectively. After identifying $k = 250$ as ideal under this cross-validation routine, and based upon the out-of-sample AUCs and AUC-PRs obtained from our "denied request" sLDA models, we then proceeded to simultaneously select α and η with the use of a separate set of request-documents, and while holding k fixed at 250 (again while using five-fold cross-validation). Some readers may be concerned with the above approach, given that it did not simultaneously evaluate *all* possible combinations of α, η, and k. To address these concerns, this section performs and evaluates this more extensive cross-validation approach for robustness.

To implement the extended cross-validation parameter selection routine that is described immediately above, we return to the original subset of our request documents used for parameter selection of α and η in our main paper. Recall that this set of documents corresponded to roughly 25% of our total document sample, or to approximately 250,000 in total. We next randomly partitioned this set of documents into five folds of training and test data, and estimated "denied request" sLDA models using every three-pair combination of the following three vectors: $\alpha = \{0.01, 0.1, 0.5, 1, 5, 10\}$, $\eta = \{0.01, 0.1, 0.5, 1, 5, 10\}$, and $k = \{5, 20, 50, 100, 250, 500\}$ using five fold cross-validation, and stored the corresponding out-of-sample AUCs for these models. We then calculated *averaged* AUCs and AUC-PRs (across each set of five folds) for each α and η combination evaluated, separately for each k evaluated. We plot these averaged results, for each k of interest, in Figures A.1-A.2.

Beginning with Figure A.1, we can first note that one's choice of α and η have relatively low influence on the sLDA models' abilities to accurately classify denied requests when one's choice of topics is fairly low ($5 \geq k \leq 50$). Furthermore, one can also observe in the top row of Figure A.1 that—no matter one's choice of α or η—sLDA models that employ fewer than 100 topics generally perform worse in classifying "denied requests" than do our three larger topic models (i.e., $k = 100, 250, 500$). This can be seen most clearly in the maximum AUCs achieved within each plot in Figure A.1: whereas the $k = 100, 250, 500$ models each at times yield AUCs above 0.70 depending on one's choice of hyperparamters, AUCs greater or equal to 0.7- are never achieved within the $k = 5, 25, 50$ sLDA models. Together this suggests that one should favor the $k = 100, 250, 500$ models over our smaller topic models no matter one's choice of α or η. This is confirmed by the AUC-PR results depicted in Figure A.2. Indeed, Figure A.2 demonstrates that one never receives an AUC-PR greater or equal to 0.20 when $k = 5, 25, 50$. At the same time, multiple combinations of α and η yield AUC-PRs that fall above 0.20, and in some cases 0.25, when k is assigned to a value in the 200-to-500 range.

As the bottom halves of Figures A.1-A.2 indicate, one's decisions with respect to α and η become more relevant in discriminating between the $k = 100$, $k = 250$, and $k = 500$ topic models. In each of these three sets of sLDA models, we can observe along the y-axes that $\alpha = 0.1$—which was also the α value identified in the main model selection routine reported in the main paper—consistently yields the highest AUC and AUC-PR values. In the case of η, we find in Figures A.1-A.2 that $\eta = 0.1$ consistently yields the highest achieved AUC when $\alpha = 0.1$ for the $k = 100$, $k = 250$, and $k = 500$ sLDA models; followed in most cases by $\eta = 0.5$. Given these results, we believe the extended cross-validation parameter selection routine evaluated here strongly suggests $\eta = 0.1$ to be the most optimal choice for this hyperparameter across our most optimal sLDA models. This is also the η hyperparameter value chosen within our main parameter selection routine, and for our primary analysis.

Finally, Figures A.1-A.2 suggest that—for these choices of α and η, the $k = 100$ sLDA model yields the lowest relative AUC and AUC-PR values, the $k = 250$ sLDA model yields the second highest AUC and AUC-PR values, and the $k = 500$ model yields the highest AUC and AUC-PR values. However, the $k = 500$ sLDA model's improvements over the $k = 250$ model are slight, especially in the case of AUC. In light of this, and in keeping with past evaluations of (cross-validated) out-of-sample model fit criteria for the purposes of topic number selection within LDA models (Barberá et al., 2014; Bagozzi, 2015; Berliner, Bagozzi and Palmer-Rubin, 2018; Berliner et al., 2020), there is strong justification for favoring a topic number that falls towards lower end of our optimal AUC and AUC-PR measure's range, i.e., $k = 250$. As was the case for the optimal α and η values above, this k was also the topic number chosen by our two-step parameter selection routine, and the topic number used in our main analysis.

Figure A.1: Simultaneous Model Selection Across k, α, and η Using Area under ROC

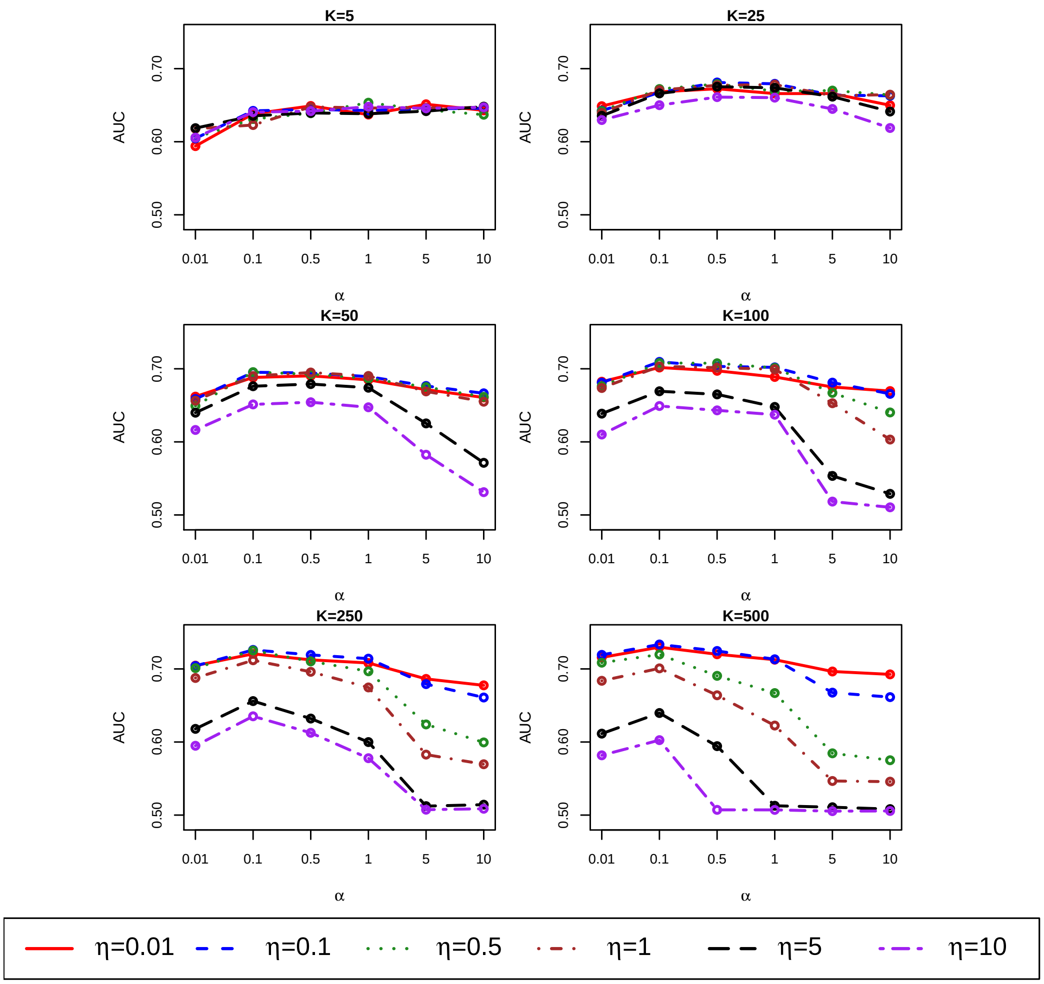
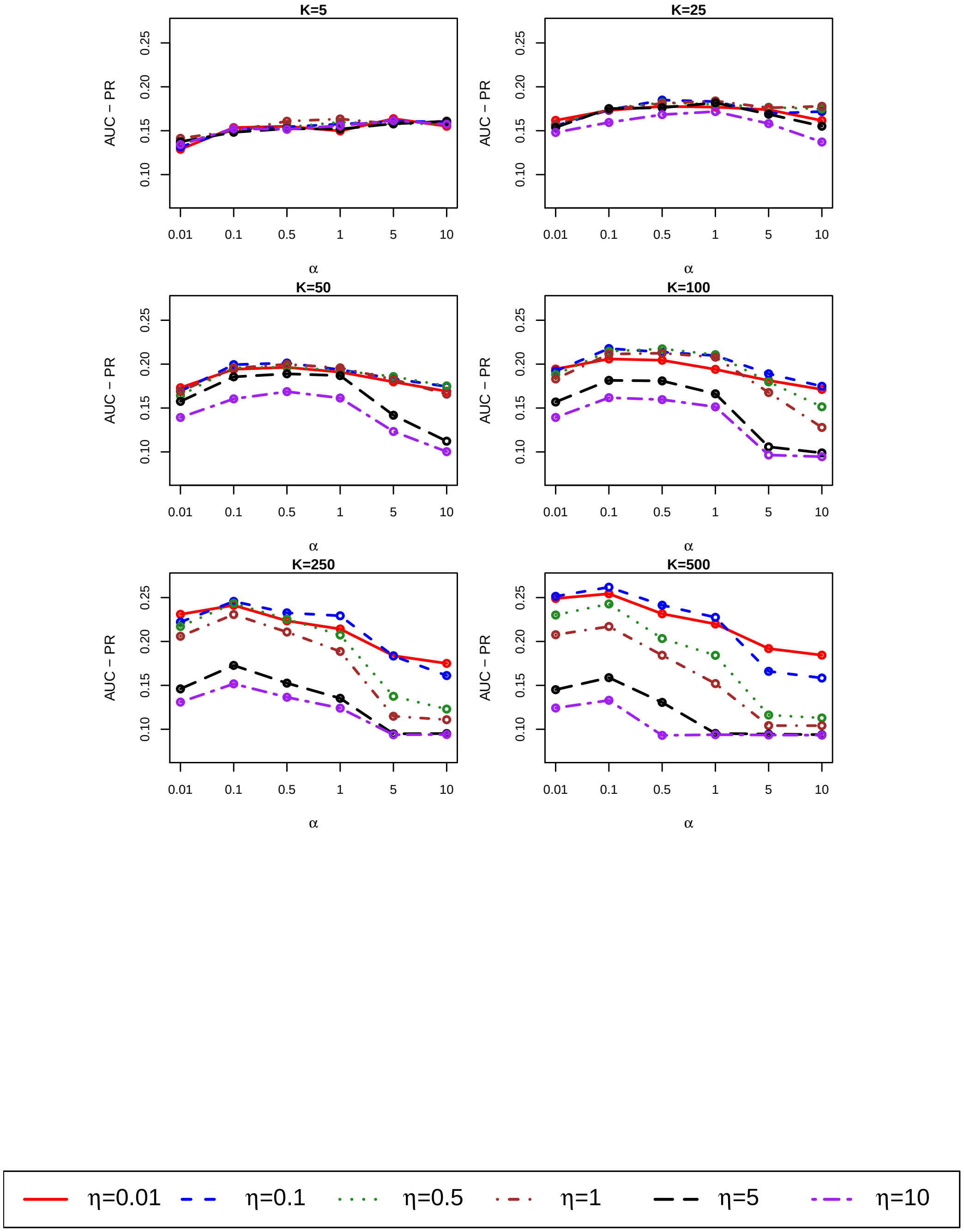


Figure A.2: Simultaneous Model Selection Across k, α, and η Using Area under PR-Curve



Topics: Spanish Topwords by Score

Table A.1: Topwords for Topics Associated with "Denied Request" (Destemmed Spanish Topwords, Based on Score)

Denied$_{\#1}$	policía, federal, parte, hecho, autorización, elementos, diario, pasado, seguridad, toma
Denied$_{\#2}$	bancaria, valor, deposito, banco, dicha, instituto, comisiones, ahorro, cuenta, acreditación
Denied$_{\#3}$	coordinación, administrión, general, republica, procuraduría, trabajo, precisa, federal, puesto, legislación
Denied$_{\#4}$	solicito, información, etc, naturaleza, refiere, mencionar, escrito, documentos, documental, contenga
Denied$_{\#5}$	aseguramiento, solicito, información, entrega, sur, decomisadas, fecha, requiero, ademas, especificar
Provided$_{\#1}$	educación, colegio, plantel, docente, horas, profesores, tecnología, bachillerato, dgeti, nombramiento
Provided$_{\#2}$	educación, escuela, sep, superior, primaria, secundaria, nivel, alumnos, escolar, docente
Provided$_{\#3}$	cuanto, cual, historia, existencias, pais, inah, arqueología, cada, monumentos, estan
Provided$_{\#4}$	presupuesto, asignado, destino, ejercicio, radio, anual, rubro, programa, egresos, televisiones
Provided$_{\#5}$	sueldo, salario, puesto, tabulador, mensual, nivel, percepciones, salarial, prestaciones, compensación

Table A.2: Topwords for Topics Associated with "Denied Request" (Stemmed Spanish Topwords, Based on Score)

Denied$_{\#1}$	polic, federal, part, hech, autor, element, dia, pas, segur, tom
Denied$_{\#2}$	bancari, valor, deposit, banc, dich, institu, comision, ahorr, cuent, acredit
Denied$_{\#3}$	coordin, administr, general, republ, procuradur, trabaj, precis, federal, puest, legisl
Denied$_{\#4}$	solicit, inform, etc, naturalez, refier, mencion, escrit, document, documental, conteng
Denied$_{\#5}$	asegur, solicit, inform, entreg, sur, decomis, fech, requier, ademas, especific
Provided$_{\#1}$	educ, colegi, plantel, docent, hor, profesor, tecnolog, bachiller, dgeti, nombramient
Provided$_{\#2}$	educ, escuel, sep, superior, primari, secundari, nivel, alumn, escol, docent
Provided$_{\#3}$	cuant, cual, histor, exist, pais, inah, arqueolog, cad, monument, estan
Provided$_{\#4}$	presupuest, asign, destin, ejerc, radi, anual, rubr, program, egres, television
Provided$_{\#5}$	sueld, salari, puest, tabul, mensual, nivel, percepcion, salarial, prestacion, compens

Topics: Spanish and English Topwords by Posterior Probability

Table A.3: Topwords for Topics Associated with “Denied Request”
(English Topwords, Based on Posterior Probability)

Denied$_{\#1}$	police, part, federal, fact, authorization, daily, elements, same, security, past
Denied$_{\#2}$	bank, value, deposit, said, bank, institute, account, comission, thus, saving
Denied$_{\#3}$	coordination, administration, general, republic, attorney general's office, work, accurate, federal, position, legislation
Denied$_{\#3}$	administration, coordination, general, information, federal, republic, attorney general's office, work, accurate, position
Denied$_{\#4}$	I request, information, etc., nature, mention, documents, refers, written, contain, documentary
Denied$_{\#5}$	I request, information, insurance, delivery, date, I require, south, confiscated, also, specify
Provided$_{\#1}$	education, school, staff, teacher, hours, professors, I request, medium, baccalaureate
Provided$_{\#2}$	education, school, SEP, higher, level, primary, secondary, students, school, teacher
Provided$_{\#3}$	how much, which, history, existence, country, each, INAH, archeology, they are, zone
Provided$_{\#4}$	budget, assigned, destination, exercise, radio, annual, item, program, expenditures, year
Provided$_{\#5}$	wage, salary, position, tabulator, monthly, level, perceptions, year, benefits, salary related

Table A.4: Topwords for Topics Associated with “Denied Request”
(Destemmed Spanish Topwords, Based on Posterior Probability)

Denied$_{\#1}$	policía, parte, federal, hecho, autorización, diario, elementos, mismo, seguridad, pasado
Denied$_{\#2}$	bancaria, valor, deposito, dicha, banco, instituto, cuenta, comisiones, asi, ahorro
Denied$_{\#3}$	administración, coordinación, general, información, federal, republica, procuraduría, trabajo, precisa, puesto
Denied$_{\#4}$	solicito, información, etc, naturaleza, mencionar, documentos, refiere, escrito, contenga, documental
Denied$_{\#5}$	solicito, información, aseguramiento, entrega, fecha, requiero, sur, decomisadas, ademas, especificar
Provided$_{\#1}$	educación, colegio, plantel, docente, horas, profesores, tecnología, solicito, medio, bachillerato
Provided$_{\#2}$	educación, escuela, sep, superior, nivel, primaria, secundaria, alumnos, escolar, docente
Provided$_{\#3}$	cuanto, cual, historia, existencias, pais, cada, inah, arqueología, estan, zona
Provided$_{\#4}$	presupuesto, asignado, destino, ejercicio, radio, anual, rubro, programa, egresos, año
Provided$_{\#5}$	sueldo, salario, puesto, tabulador, mensual, nivel, percepciones, año, prestaciones, salarial

Table A.5: Topwords for Topics Associated with "Denied Request" (Stemmed Spanish Topwords, Based on Posterior Probability)

$\text{Denied}_{\#1}$	polic, part, federal, hech, autor, dia, element, mism, segur, pas
$\text{Denied}_{\#2}$	bancari, valor, deposit, dich, banc, institu, cuent, comision, asi, ahorr
$\text{Denied}_{\#3}$	administr, coordin, general, inform, federal, republ, procuradur, trabaj, precis, puest
$\text{Denied}_{\#4}$	solicit, inform, etc, naturalez, mencion, document, refier, escrit, conteng, documental
$\text{Denied}_{\#5}$	solicit, inform, asegur, entreg, fech, requier, sur, decomis, ademas, especific
$\text{Provided}_{\#1}$	educ, colegi, plantel, docent, hor, profesor, tecnolog, solicit, medi, bachiller
$\text{Provided}_{\#2}$	educ, escuel, sep, superior, nivel, primari, secundari, alumn, escol, docent
$\text{Provided}_{\#3}$	cuant, cual, histor, exist, pais, cad, inah, arqueolog, estan, zon
$\text{Provided}_{\#4}$	presupuest, asign, destin, ejerc, radi, anual, rubr, program, egres, año
$\text{Provided}_{\#5}$	sueld, salari, puest, tabul, mensual, nivel, percepcion, año, prestacion, salarial

Examples from Highly Associated Requests

For each of our primary Denied and Supplied topics, this section reports two representative ATI request text examples. Herein, we selected two ATI requests from the 50 most highly associated requests for each relevant topic. In instances where the selected text was extremely long, we truncated the reported request text at a reasonable length, while still endeavoring to capture the context of that request. For circumstances where the randomly selected request contained an individual's name, we have removed that individual's name from the reported text below. In cases where the selected request was made in all capitalized letters, we have re-typed the request using standard capitalization conventions for readability purposes. We have left any and all remaining typographical errors within each original Spanish request as they appeared within the original text. Below, we first present our selected requests for topics Denied$_{\#1}$-Denied$_{\#5}$, followed by our example requests for Provided$_{\#1}$-Provided$_{\#5}$. In addition to our reporting of the original Spanish version of each example request text, we also provide a brief English-language translation of that request.

Denied$_{\#1}$ Examples
August 27, 2005: POLICÍA FEDERAL ANTES POLICÍA FEDERAL PREVENTIVA

- Cantidad de policías y efectivos de todos los rangos de la Policía Federal Preventiva que han sido despedidos de la dependencia o que han salido por voluntad propia en los últimos cinco años. Detallar las razones de la salida de cada uno de los miembros y cuántos lo han hecho por voluntad propia. Más detalles en archivo [...].

- (Number of Federal police and troops of all ranks who have been dismissed from their own unit or have left voluntary during the last five years. Detail the reasons for the departure of each of these members and how many departed voluntarily. More details in the attachment [...].)

October 25, 2011: SECRETARÍA DE SEGURIDAD PÚBLICA

- Tomo como ejemplo el caso de los cazadores desaparecidos en Zacatecas por policías municipales como lo mencionan en sus notasmedios como CNN en su portal del día lunes 13 de diciembre de 2010. Aunque este ejemplo la responsabilidad de la información que solicito es de carácter estatal insisto en conocer información sobre autoridades FEDERALES detenidas por estar implicadas en desaparición forzada en el rango de tiempo antes mencionado

- (I take as example the case of disappeared hunters in Zacatecas by the municipal police, as mentioned on Monday the 13th of December, 2010. Although this is only an example, the state has a responsibility to provide the information that I request. I insist on knowing information about Federal authorities that have been detained for their involvement in forced disappearances during the aforementioned time frame.)

Denied$_{\#2}$ Examples
March 27, 2015: COMISIÓN NACIONAL BANCARIA Y DE VALORES

- Solicito los documentos ya sean actas minutas acuerdos resoluciones o documentos de cualquier otra denominación en los que se haya registrado y se sustente la intervención Gerencial de FICREA por parte de la CNBV de conformidad con el comunicado que la CNBV entregó a los ahorradores de FICREA en la reunión que se llevó a cabo en las instalaciones de CODUSEF con funcionario de la CNVB CONDUSEF y el interventor gerente de FICREA el día 19 de noviembre de 2014 y que a su vez la CNBV publicó en su página de internet el mismo día. Adjunto la minuta de acuerdos de la reunión mencionada así como el comunicado que publicó FICREA en su página de internet. [...]

- (I request documents—be they minutes, resolutions, or documents of any other form—in which the FICREA intervention was registered and sustained by the CNBV in accordance with the communique that the CNBV delivered to FICREA depositors in the meeting held at CODUSEF with the CNVB CONDUSEF official and FICREA's manager on November 19, 2014; which was in turn published by CNBV on its website that same day. I enclose the agreement minutes of the aforementioned meeting as well as the statement published by FICREA on its website [...])

June 20, 2011: INSTITUTO PARA LA PROTECCIÓN AL AHORRO BANCARIO

- Solicito copia certificada de la autorizacion concedida por el instituto para la proteccion del ahorro bancario a Banco Nactional de Mexico S. A. Para la venta y/o cesion de derechos de creditors en favor de Basilisk Seis S. De R. L. De C. V. Incluyendo el listado de creditos cedidos.

- (I request a certified copy of the authorization granted by the Institute for the Protection of Bank Savings to the National Bank of Mexico, SA for the sale and/or reassignment of creditors rights in favor of Basilisk Six, including the list of credits reassigned.)

Denied$_{\#3}$ Examples
March 16, 2009: PROCURADURÍA GENERAL DE LA REPÚBLICA

- Sustentando mi Derecho a la Información Publica Gubernamental y en especifico a la que obra en poder de la ProcuradurÃŋa General de la República y en virtud de la propia y especial naturaleza de esta Solicitud de Información la cual esta cimentada y apoyada en base y términos de los Preceptos Legales que se tutelan en la Ley Federal de Transparencia y Acceso a la Información Publica Gubernamental misma Legislación que se identifica por sus siglas como LFTAIPG me permito solicitar con valido y legitimo Fundamento en el Apartado B del ArtÃŋculo 123 de la ConstituciÃşn PolÃŋtica Federal de los Estados Unidos Mexicanos [...] solicito se me informe a detalle y de modo claro preciso y pormenorizado tomando como referencia el periodo cronológico Fiscal y/o Presupuestal de 2000 2001 2002 2003 2004 2005 2006 2007 2008 y 2009 [...] y siendo inherente a la naturaleza de que en los Catálogos Generales de Puestos del Gobierno Federal y en los Tabuladores de Percepciones Mensuales en menciÃşn se contiene y en ellos se plasman los PUESTOS y/o PLAZAS autorizados a la Procuraduría General de la República por la Secretaría de Hacienda y Crédito Público para su ejercicio Presupuestal y Nominal relativo a los Trabajadores al Servicio del Estado que laboran en esa Institución identificable por sus siglas como P.G.R. al pago de Salarios y Prestaciones de la Plantilla de Personal y/o Trabajadores al Servicio del Estado que laboran en esa Institución aunado a que en el precisado Tabulador se exponen y presentan entre otros elementos formales los Códigos Presupuestales Denominaciones del Puesto o Plaza Niveles Salariales y Percepciones o Emolumentos entendiéndose por dichos Conceptos el Sueldo y Compensación Garantizada solicito se me informe a detalle y de modo preciso y pormenorizado ¿Cuáles son las funciones actividades atribuciones y responsabilidades asignadas y/o conferidas que de manera específica y concreta le corresponden y debe desempeÃśar conforme a la legal Normatividad que aplique en el ámbito de la Procuraduría General de la República la Categoría Laboral y/o Puesto Plaza que se identifica con Código Presupuestal: CF53096 y un Nivel Salarial que se precisa con la Clave o Código: MA1 cuya Denominación del Puesto-Plaza es: COORDINADOR ADMINISTRATIVO DE LA OFICINA DE S.P.S. 36 (H)? [...]

- (Consistent with my right to public government information and specifically to the Attorney General's public information, which is guaranteed and supported on the basis of LFTAIPG, I would like to request with valid and legitimate justification in Section B of Article 123 of the Federal Political Constitution of the United Mexican States in the Federal Law of Workers [...] I request to be provided in a clear and precise manner in reference to the fiscal and/or budgetary period of 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009 [...] the monthly tabulations and posts/places authorized for the Attorney General's office by the Secretariat of Finance and Public Credit for budgetary and nominal purposes relative to state employees in that Institution, identified by pavements of salary and benefits for state works at that Institution [...])

March 17, 2009: PROCURADURÍA GENERAL DE LA REPÚBLICA

- ProcuradurÃŋa General de la República y/o Unidad de Enlace para el Acceso a la Información Pública de la Procuraduría General de la República.- Se hace la pertinente aclaración que la presente Solicitud de Información versa y es relativa a la naturaleza Homologa Análoga Equivalente Homogénea y/o Equiparable que guarda y/o tiene o puede tener y/o guardar un Puesto-Plaza mismo que se encuentra debidamente indicado en relación con aquellos que están enlistados y precisados en el cuerpo de esta Solicitud por tanto y dado el numero de caracteres que la integran y componen la Solicitud de Información en cuestión se encuentra en Archivo Adjunto a la presente y se da aquí por reproducida para todos los efectos legales jurídicos y/o administrativos que se originen o haya lugar en base al ARCHIVO DE LA DESCRIPCIÓN que se consigna y que otorga la Autenticidad de la Información del Archivo Adjunto y del Acuse de esta Solicitud [...]

- (I would like to make a pertinent clarification to the requested information related to nature of the Homologa Análoga Equivalente Homogénea and/or Equiparable that maintains and/or has or may have a position that is duly indicated in relation to those that are listed and specified int he body of this application, and in recognition of content of the relevant request for information in the corresponding attachment, and is hereby reproduced for all legal and/or administrative purposes that originate or are located in the attached description that is co-signed and that guarantees the authenticity of the attached file and of the acknowledgment of the request. [...])

Denied$_{\#4}$ Examples October 28, 2009: FONATUR CONSTRUCTORA S.A. DE C.V.

- Solicito el documento de la naturaleza documento que consigne el monto total que se ejericó en el mes de Julio de 2009 mediante adjudicaciones directas el monto total que se ejercio en el mismo periodo mediante invitaciones a cuando menos tres personas y el monto total que se ejercio en el mismo periodo mediante licitaciones publicas.

- (I request the document or document source material that shows the total amount that was allocated in July 2009 through direct awards, the total amount that was allocated in that same period through invitations to at least three people, and the total amount that was allocated in the same period through public tenders.)

February 23, 2012: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- Esta Contraloría Social le solicita a la Dirección de Administración y Evaluación de Delegaciones del IMSS a nivel central LOS RESULTADOS O LAS CONCLUSIONES de los estudios comparativos que se elaboraron para analizar las diferencias que se han suscitado entre las distintas bases de licitación PARA LA ADQUISICIÓN DE ALIMENTOS GRUPO DE SUMINISTRO 480 VÍVERES. En la resolución 1480/11 el IFAI ordenó que la información que estoy solicitando se hiciera pública una vez que el Instituto tomara la decisión final al respecto y una vez que fueran publicados en el COMPRANET los proyectos de convocatoria para la adquisición de alimentos. [...]

- (This social comptroller requests the Directorate of Administration & Evaluation of the IMSS to provide the results or conclusions of the comparative studies that were prepared to analyze differences arising between the different bidding bases for the purchase of food supply group #480. In resolution 1480/11, the IFAI ordered that the information I am requesting be made public once the Institute made its final decision on this topic, and once the awards for the procurement of food were published in COMPRANET.)

Denied$_{\#5}$ Examples
February 17, 2013: SERVICIO DE ADMINISTRACIÓN TRIBUTARIA

- Solicito información sobre la cantidad de bienes inmuebles decomisados enajenados o asegurados ministerialmente a [XXXX] por año y estado. Especificando dirección si fue enajenación decomiso aseguramiento o expropiación el monto estimado del valor de la propiedad y la situación jurídica del inmueble es decir si fueron vendidos o arrendados. Además requiero especifiquen cual fue el monto decomisado o asegurado a [XXXX] al momento de su detención en qué lugar se dio y a dónde se destinó dicho dinero.

- (I request information on the amount of property confiscated, seized, or ministerially insured to [XXXX] by year and state. Specify the address, estimated amount, and current legal status of the seized, confiscated or insured property, as well as whether it was sold or leased. I also require you to specify the amount confiscated or insured to [XXXX] at the time of his/her arrest, the location where this ocurred, and where that money went.)

September 2, 2013: CONSEJERÍA JURÍDICA DEL EJECUTIVO FEDERAL

- Solicito información sobre la cantidad de bienes inmuebles decomisados o asegurados ministerialmente a [YYYY] por año y estado. Especificando dirección si fue decomiso aseguramiento o expropiación el monto estimado del valor de la propiedad y la situación jurídica del inmueble es decir si fueron vendidos o arrendados. Además requiero especifiquen cual fue el monto decomisado o asegurado a [YYYY] al momento de su detención en qué lugar se dio y a dónde se destinó dicho dinero.

- (I request information on the amount of property confiscated, seized, or ministerially insured to [YYYY] by year and state. Specify the address, estimated amount, and current legal status of the seized, confiscated or insured property, as well as whether it was sold or leased. I also require you to specify the amount confiscated or insured to [YYYY] at the time of his/her arrest, the location where this ocurred, and where that money went.)

Provided$_{\#1}$ Examples

December 10, 2012: SECRETARÍA DE EDUCACIÓN PÚBLICA

- Buenas tardes por este medio les pido por favor tengan a bien poderme apoyar con la siguiente información que necesito del INSTITUTO TECNOLÓGICO DE GUSTAVO A. MADERO: 1. Plantilla de docentes con el número de horas que tienen cada uno y su categoría dictaminada. Medios tiempos tres cuartos de tiempo y tiempos completos. 2. Edad sexo y formación académica de los profesores de tiempo completo. 3. Los profesores de tiempo completo que categoría tienen RITULAR A B C o profesor investigador. GRACIAS

- (Good afternoon. I politely ask you to provide me with the following information that I need from the Technological Institute of Gustavo A. Madero: 1. Teaching staff, with the number of hours each staff member is assigned and their assigned title. Include half time, three quarters time, and full time. 2. Age, sex, and academic training of full time teachers. 3. The full-time professors of category RITULAR A, B, C, or research professor. Thank you.)

February 20, 2012: INSTITUTO POLITÉCNICO NACIONAL

- estimados señores: les agradeceré que me proporcionen el conjunto de información publica correspondiente a: la fecha de ingreso a laborar al instituto politécnico nacional el tipo y la cantidad en horas de nombramientos de base o interinato (por semestre y por plantel en que haya laborado) según sea el caso que haya tenido asignadas desde su ingreso a laborar como docente del ipn; la antigüedad como docente del instituto politécnico nacional la cantidad de horas en propiedad actuales la ultima categoría académica el ultimo grado de estudios los diversos planteles del ipn en que haya laborado desde su ingreso y la totalidad de planteles del ipn en los que actualmente se encuentra laborando así como la totalidad de las asignaturas (incluyendo plantel semestres grupos horarios y la totalidad de horas en calidad de base e interinato que se hayan asignado en cada caso para impartir dichas asignaturas) que haya impartido desde su ingreso al ipn de la PROFESORA [ZZZZ] y la totalidad de las asignaturas (incluyendo grupos horarios planteles del ipn y horas asignadas en calidad de base e interinato según sea el caso) que en semestres proximos pasados y que actualmente se le hayan dado a dicha profesora [ZZZZ] de la escuela superior de medicina. Gracias.

- (Dear sirs: I would be grateful if you would provide me with information corresponding to: the work-start date at National Polytechnic Institute, the type and number of hours of basic or interim appointment (per semester and per campus where worked) since appointment as a teacher of IPN; the level of seniority at the National Polytechnic Institute; the number of hours of ownership, the previous academic title; the last grade completed; the various schools of the ipn in which he/she has worked since appointment and the totality of ipn campuses in which he/she is currently working as well as the total number of subjects (including the six semester hour groups and the total number of hours as a base and intern that have been assigned in each case for instruction of these subjects) that he/she has taught since appointed as Professor [ZZZZ] and the totality of subjects (including ipn group timetables and assigned hours as a base and intern as the case may be) in the past semesters that have been given to this teacher [ZZZZ] of the medical school. Thank you.)

Provided$_{\#2}$ Examples
November 7, 2012: SECRETARÍA DE EDUCACIÓN PÚBLICA

- Del Total de escuelas ubicadas en la Delegación Gustavo A. Madero del Distrito Federal en el ciclo escolar 2010-2011 solicito se me proporcione: 1. El número ALUMNOS inscritos en las escuelas Públicas de nivel Preescolar 2. El número ALUMNOS inscritos en las escuelas Públicas de nivel Primaria 3. El número ALUMNOS inscritos en las escuelas Públicas de nivel Secundaria 4. El número ALUMNOS inscritos en las escuelas Públicas de nivel Profesional Técnico 5. El número ALUMNOS inscritos en las escuelas Públicas de nivel Bachillerato

- (Out of the total number of schools located in the Gustavo A. Madero delegation of Mexico City in the 2010-2011 school year, I request that you provide me with: 1. The number of students enrolled in public schools at the pre-school level. 2. The number of students enrolled in public schools of primary level 3. The number of students enrolled in public schools of secondary level 4. The number of students enrolled in public schools of professional technical level 5. The number of students enrolled in public schools at the Bachelor level.)

November 26, 2009: ADMINISTRACIÓN FEDERAL DE SERVICIOS EDUCATIVOS EN EL DISTRITO FEDERAL (AFSEDF)

- Las calificaciones bimestrales de todos los estudiantes de primaria oficiales y particulares incorporadas a la SEP del DF para los ciclos escolares 2000-2009.

- (The bimonthly grades of all the official and private elementary students incorporated to the SEP of the D.F. for the 2000-2009 school cycles.)

Provided$_{\#3}$ Examples
February 19, 2015: INSTITUTO NACIONAL DE ANTROPOLOGÍA E HISTORIA

- Registro de zonas arqueológicas sitios históricos y/o monumentos históricos inmuebles ubicados en Zapotlanejo Jalisco. (Proporcionar los datos históricos de cada lugar)
- (Register of archaeological sties, historical sites, and/or historical landmarks located in Zapotlanejo, Jalisco. (Provide historical data for each place))

September 24, 2005: INSTITUTO NACIONAL DE ANTROPOLOGÍA E HISTORIA

- Solicito informacÃŋón sobre el presupuesto asignado por el Instituto Nacional de Antropología e Historia a cada uno de los lugares sitios o zonas de monumentos históricos considerados patrimonio cultural de la humanidad en México. Solicito esta información durante los años 2000 2001 2002 2003 y 2004.
- (I request information on the budget assigned by the National Institute of Anthropology and History to each of the places, sites, or areas, of historical monuments considered a cultural heritage site in Mexico. I request this information for the years 2000, 2001, 2002, 2003, 2004.)

Provided$_{\#4}$ Examples
October 19, 2009: SECRETARÍA DE HACIENDA Y CRÉDITO PÚBLICO

- Presupuesto ejercido con cargo a la partida 3701 durante los años 2006 2007 2008 el aprobado para 2009 y ejercido hasta septiembre de dicho año.
- (Budget exercised under heading 3701 during the years 2006, 2007, 2008 approved for 2009 and exercised until September of that year.)

July 7, 2011: SECRETARÍA DE HACIENDA Y CRÉDITO PÚBLICO

- Si existe algun presupuesto asignado a la Zona Metropolitana de Villahermosa-Nacajuca en Tabasco según el Presupuesto de Egresos de la Federación para los Ejercicio Fiscal 2009 2010 y 2011. Si existe presupuesto asignado delimitar por año y Planes estudios programas proyectos obras o acciones seleccionadas para asignarlo
- (If available, any budget allocated to the Metropolitan Area of Villahermosa-Nacajuca in Tabasco according to the Budget of Expenditures of the Federation for Fiscal Years 2009, 2010, and 2011. If available, budgets broken down by year and studied plans, programs, projects, works, or actions selected for assignment.)

Provided$_{\#5}$ Examples
January 26, 2012: SECRETARÍA DE HACIENDA Y CRÉDITO PÚBLICO

- Por medio del presente escrito el C VICTOR MANUEL LABASTIDA HERNANDEZ solicito tabulador regional de los años 2010 2011 y 2012 del puesto TECNICO MEDIO que ocupaba como trabajador en activo en la SECRETARIA DE COMUNICACIONES Y TRANSPORTES. Es de señalar que desconozco si el puesto señalado en líneas anteriores siga existiendo o haya sufriendo algún cambio de puesto nivel o código. Solicitando de la misma forma me sea señalado en la misma contestación para poder saber en estos tabuladores cuales el puesto que hoy ocupa mi plaza como activo Así también es de especificar que lo que solicito es un tabulador regional de forma desglosada es decir que sea señalado el sueldo sobre sueldo y compensaciones manifestando que no solicito tabulador con sueldo base. Es necesario que se especificado el sueldo sobre sueldo y compensaciones. Por su atención gracias

- (By means of the current document, [WWWW] requests a regional tabulation of the years 2010, 2011, and 2012 for the Medium Tecnico position that he occupied as an active worker in the Secretariat of Communication and Transportation. It is worth noting that I do not know if the position indicated above still exists or has undergone a change of position level of title. Along these lines, I request that you indicate to me in your response the current position that now occupies my place. Also, I am requesting a regional tabulation in disaggregated form for salary and computations, rather than basic salary. It is necessary that the salary be specified in terms of salary and compensations. Thanks for your attention.

October 6, 2014: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- Solicito cotizacion de salario en el IMSS de la categoria de medico familiar anual de los años 2011 al 2014 asi como porcentaje de incrementos salariales por quinquenio y conceptos adicionales al salario. De un trabajador con fecha de ingreso en el 2003 con una antiguedad de 8 años 4 quincenas. Gracias.

- (I request an annual IMSS salary quote for the category of family doctor for the years 2011 to 2014, as well as the percentage of salary increases for five years and additional concepts to the salary. This is for a worker with a start date of 2003 with 8 years experience. Thank you.)

Variation in (In-Sample) Request Volume Over Time

Table A.6: In-Sample Associations between Top Topics and Presidential Election Indicator

Topic	χ^2 value	p-value
$\text{Denied}_{\#1}$	14.18	0.0002
$\text{Denied}_{\#2}$	22.37	2.243e-06
$\text{Denied}_{\#3}$	5.98	0.0145
$\text{Denied}_{\#4}$	0.28	0.5966
$\text{Denied}_{\#5}$	5.50	0.019
$\text{Provided}_{\#1}$	0.27	0.6070
$\text{Provided}_{\#2}$	0.22	0.6358
$\text{Provided}_{\#3}$	0.0002	0.9901
$\text{Provided}_{\#4}$	5.67	0.0017
$\text{Provided}_{\#5}$	0.0006	0.9802

Note: $N = 101,494$, $DF = 1$

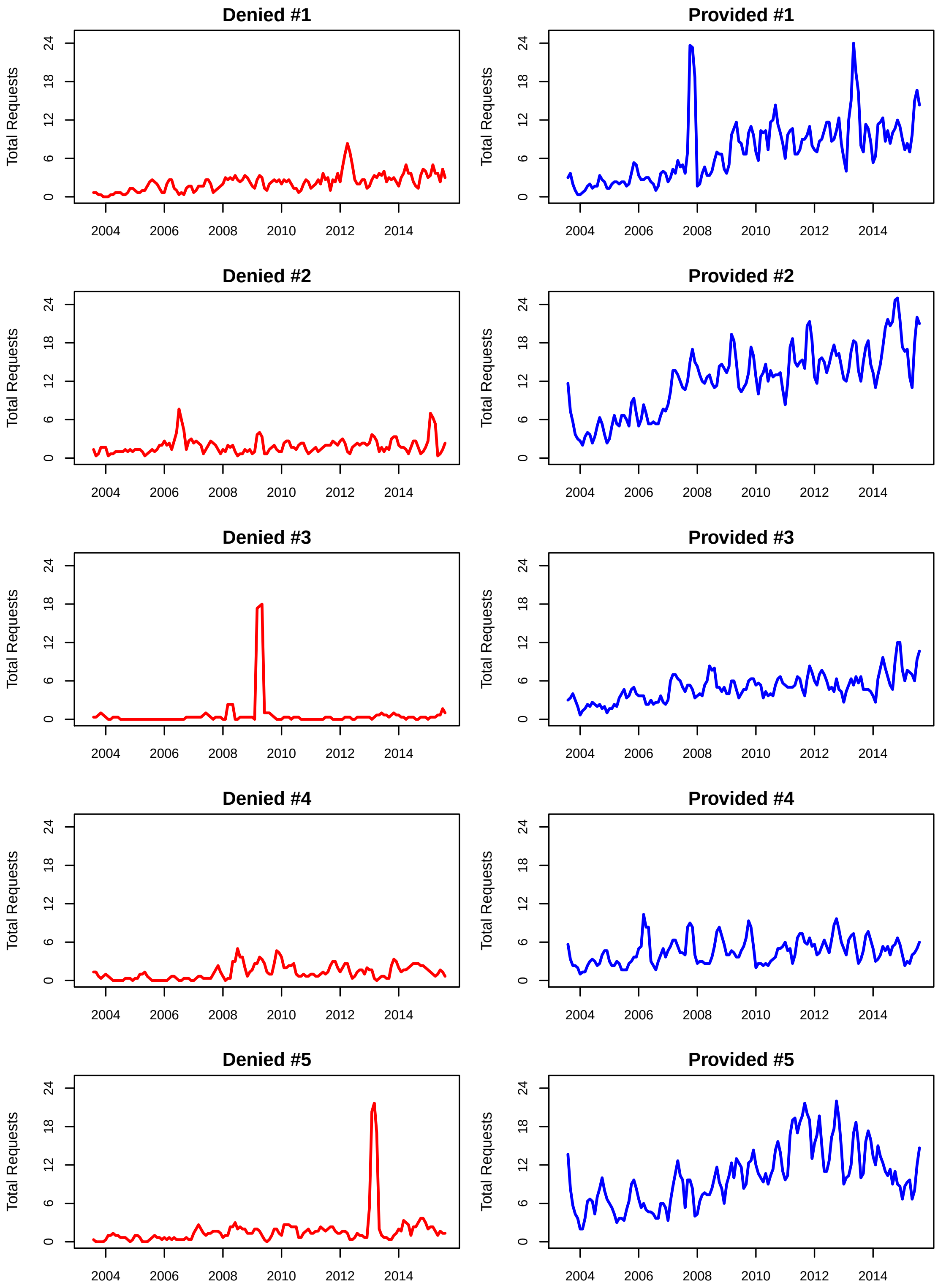


Figure A.3: Three-Month Moving Averages of Request Volume ("Denied Request" Model)

Topics: Spanish and English Middle Leverage Topics

As mentioned in the main paper, we further evaluate the uniqueness of our five identified Provided topics, and our five identified Denied topics, through an examination of the five most middle leverage topics. To identify our five most middle leverage topics, we returned to the in-sample "denied request" coefficient estimates from our primary sLDA model, which were discussed in our main paper. Using the same sorted rankings of these 250 coefficient estimates that was used to identify our top five Provided and top five Denied topics in the paper, we then instead identified the five topics that were ranked 123rd, 124th, 125th, 126th, and 127th in their estimated effects on the probability of a "denied request." Given the overall distribution of our in-sample sLDA coefficient estimates—as reported in Figure 3 of the main paper—these five topics are each associated with a decrease the probability of a "denied request." However, the estimated effects in each regard are much smaller in magnitude than was the case for Provided$_{\#1}$-Provided$_{\#5}$. Accordingly, it can be concluded that these five "middle leverage" topics exhibit less discriminant power than either our Provided$_{\#1}$-Provided$_{\#5}$ or Denied$_{\#1}$-Denied$_{\#5}$ request topics, and they explored in light of this below.

We report the Score-based topwords associated with topics Middle$_{\#1}$-Middle$_{\#5}$ in Tables A.7-A.9, and the corresponding topwords based instead upon highest posterior probability in Tables A.10-A.12. As was the case for our primary topics in the main paper, we base our interpretations of Middle$_{\#1}$-Middle$_{\#5}$ on their top words, as well as upon a close reading of the 50 documents most highly associated with each topic. Overall, we find that topics Middle$_{\#1}$-Middle$_{\#5}$ do not appear to exhibit significant potential for political scrutiny or political sensitivity. Rather, these topics pertain to fairly straightforward requests in areas such as (i) aggregate government statistics, (ii) procurement information from actual service providers or suppliers, (iii) tourism-oriented market research, and (iv) higher education. We now turn to a more detailed interpretation of each Middle topic in these regards.

We begin our assessment with Middle$_{\#1}$, which is perhaps the topic that has the most political potential among topics Middle$_{\#1}$-Middle$_{\#5}$. Based upon Middle$_{\#1}$'s topwords in Table

A.7, this topic clearly encompasses requests for summary statistics on aggregate government budgets, expenses, and fiscal spending totals. While such information can be used to identify government corruption, excesses, or inefficiencies, the 50 most associated requests for this topic suggest that this is usually note the intention. For example, the most associated request for this topic begins with the statement "For an academic project [...]" before requesting aggregate information on the market share of various port terminals. On the other hand, the second most associated request with this topic *is* potentially investigative in nature, in that it pertains to a requester's demands for evidence (from the ATI liaison unit of the Secretary of Health) to justify the Secretary's previous delays in replying to the requester's earlier ATI requests. Nevertheless, a majority of the remaining associated requests for this topic largely mirror the former request format, in that they pertain to requests for aggregate information on items such as quarterly tourism investment, budget totals, or numbers of hospital beds with little to suggest that the information is being sought for reasons other than academic, market, or government supplier research.

The topics $\text{Middle}_{\#2}$ and $\text{Middle}_{\#3}$ each encompass procurement-focused requests from the perspective of government contractors (or potential government contracts) rather than from those seeking to scrutinize procurement contracts. For instance, each of these topics includes topwords related to medical devices (dental, brushes, container, or catheter) which in the case of $\text{Middle}_{\#2}$ often corresponds to desired information on the handling, number, or status of various[5] medical devices that do not (or may not) satisfy technical or quality control standards—seemingly from the provider or from a competing provider. $\text{Middle}_{\#2}$'s topwords further suggest that this topic may encompass similar (potential) provider-initiated requests relating to government procurement in construction; though the top 50 associated documents for $\text{Middle}_{\#2}$ are all requests related to medical and/or dental devices. As noted, and based upon its topwords and associated documents, $\text{Middle}_{\#3}$ similarly captures government service provider or supplier requests for procurement information. Oftentimes these requests

[5]But always very well specified, down to product names and product reference numbers.

are again in relation to medical devices, which in this case are especially focused on information related to the total current inventory of specific medical products, information on whether specific products require sanitary registration, or requests for specific procurement (compliance) documents.

The requests underlying Middle$_{\#4}$ relate to tourism (topwords: tourists, tourism, development, project, hotel, promotions) in Mexico (topwords: Baja California Sur, La Paz, Mexican). A majority of this topic's associated requests are apolitical in nature. For example, one such request specifies that the requested information is desired for market research as part of a school project, before proceeding to ask for various aggregate Mexican tourism load statistics. Other representative requests instead ask for specific master plans, business plans, current load capacities, and hotel(-room) numbers for hospitality projects or for entire tourist-focused regions.[6] In each of these cases, the requests appear to be seeking information for the purposes of market research and/or for developers' preparation of future proposals or permit requests, rather than for politically sensitive reasons. On the other hand, at least three highly associated requests do seek environmental impact statements for specific tourism (construction) projects,[7] which may have political potential given the longstanding tensions between the environment and development in many parts of Mexico. While such scrutiny—if that is the intention of these requests—would likely be aimed at the developers and investors involved in the identified projects—as opposed to government agencies—these requests nevertheless suggest that a minority of Middle$_{\#4}$'s requests exhibit at least some political potential.

The topwords associated with Middle$_{\#5}$ are less straightforward than the earlier Middle topics. Nevertheless, they together suggest that this topic may encompass requests pertaining to various activities or sectors ('sport,' 'weather,' 'engineering') within universities and related centers of higher education ('national', 'institute,' 'polytechnic') in Mexico. This

[6]This suggests that this topic exhibits a degree of thematic overlap with the zoning and land use requests that were associated with Provided$_{\#3}$.

[7]Either for single projects/locations or more frequently for very long lists of specific projects.

topic thus exhibits at least some thematic overlap with Provided$_{\#1}$ and Provided$_{\#1}$. Examination of the top 50 requests associated with this topic reinforces this interpretation. To this end, we note that several of the most associated requests for this topic seek extensive information (by name) on specific university employees' labor histories and records (including salary, responsibilities, promotions, student evaluations, and curriculum vitae). Other associated requests seek (i) individual signed documents related to higher education, (ii) information on university accreditation or budgets, (iii) aggregate salary information for university employees, (iv) credentials or curriculum vitae for specific university employees, or (v) meeting minutes. While several of these items may be related to ongoing labor disputes—either at the individual or labor union levels—the majority appear to be fairly benign, and to be made in a manner that is sufficiently clear and direct so as to preclude any possibility of obfuscation or delay in agency response.

As was the case for our top-5 Denied or top-5 Provided topics, we also compare (i) binary indicators of whether or not each in-sample request was associated with one of our Middle five topics to (ii) a binary indicator denoting whether or not an in-sample request was made during a Mexican presidential election cycle. If our Middle topics are indeed less politicized than our top-5 Denied topics (but perhaps marginally more politicized than our top-5 Provided topics), we would expect the associations between our Middle topic indicators and our presidential election indicator to fall somewhere in between the election indicator associations identified for our Denied and Provided indicators in the main paper (and in Table A.6 above). In assessing the association between (i) each of our five binary Middle topic indicators and (ii) our binary election period indicator via a series of χ^2 tests, we find that two out of five Middle topics exhibit a statistically significant ($p < 0.05$) association with our presidential election indicator; whereas four out of five top-5 Denied topics exhibited a significant association in Table A.6 and only one of five Provided topics exhibited a significant association. Likewise, the average χ^2 value for our five Middle topics in this case is 5.239—whereas we obtain an average χ^2 value of 9.664 for our top-5 Denied topics and an average chi-square value of 1.233 for the top-5 Provided topics. These results clearly suggest that our Middle five topics lie between our top-5 Denied topics and our top-5 Provided topics in their strength of association with Mexican presidential election cycles—and thus also potentially in their levels of politicization overall.[8]

In sum, our Middle$_{\#1}$-Middle$_{\#5}$ topics clearly exhibit lower levels of political scrutiny than do our Denied$_{\#1}$-Denied$_{\#5}$ topics. While several of these Middle leverage topics may at times contain a small number of politically sensitive requests, such requests are the exception rather than the norm and the vast majority of each Middle topic's requests are decidedly apolitical. In these respects, the requests underlying Middle$_{\#1}$-Middle$_{\#5}$ are closer in content to those of Provided$_{\#1}$-Provided$_{\#5}$ than they are to those of Denied$_{\#1}$-Denied$_{\#5}$. This is to be expected given that the 123rd, 124th, 125th, 126th, and 127th ranked topics in our sLDA model's sorted coefficient estimates each exhibit a net decrease on the probability of a "denied request." That being said, the few distinct occurrences of potentially sensitive requests within Middle$_{\#1}$-Middle$_{\#5}$—most notably in requests pertaining to government expenses and spending breakdowns in Middle$_{\#1}$, environmental impact evaluations in Middle$_{\#4}$, and labor union disputes in Middle$_{\#5}$—do suggest that these topics may at times encompass more politically sensitive requests in comparison to Provided$_{\#1}$-Provided$_{\#5}$. The χ^2 values discussed above reaffirm this interpretation. Together this implies that our sLDA model, and its ranked coefficient estimates, are performing as expected: they are grouping requests into topics not only in terms of distinct thematic areas (as an LDA model would do) but also in terms of their perceived political sensitivity.

[8]We also re-operationalized our five Middle request indicators as a single 'Middle request' topic share of total monthly in-sample requests. As was the case for our top-5 Denied and top-5 Supplied topics, we then compared this proportion measure to our election indicator via a two-sided t-test. Here we found that our Middle request proportion measure was not not significantly larger (or smaller) during our outside of Mexican presidential election cycles. Thus, our top-5 Denied topics are the only subset among those examined here that exhibit a statistically significant association with our election month indicator at this level of aggregation.

Table A.7: Topwords for "Middle-Leverage" Topics
(English Topwords, Based on Score)

$\text{Middle}_{\#1}$	total, amount, number, year, spent, fiscal year, require, concept, itemized, expenses
$\text{Middle}_{\#2}$	cut, team, villa, provider, dental, technical personnel, code, building, fulfillment, brushes
$\text{Middle}_{\#3}$	documents, plaza, length, order, institute, code, registration, nature, container, catheter
$\text{Middle}_{\#4}$	tourists, tourism, development, promotions, Baja California Sur, La Paz, project, hotel, plan, Mexican
$\text{Middle}_{\#5}$	institute, national, sport, period, polytechnic, request, weather, provide, part, engineering

Table A.8: Topwords for "Middle-Leverage" Topics
(Destemmed Spanish Topwords, Based on Score)

$\text{Middle}_{\#1}$	total, monto, numero, año, erogado, ejercicio, requiero, concepto, desglosado, gastos
$\text{Middle}_{\#2}$	corte, equipo, villa, proveedor, dentales, tecnico, clave, construcción, cumplimiento, cepillos
$\text{Middle}_{\#3}$	documentos, pza, longitud, consigne, instituto, clave, evidencie, naturaleza, envase, cateter
$\text{Middle}_{\#4}$	turistas, turismo, desarrollo, promociones, Baja California Sur, paz, proyecto, hotel, plan, mexicano
$\text{Middle}_{\#5}$	instituto, nacional, deporte, periodo, politecnico, solicito, tiempo, proporcionar, parte, ingeniería

Table A.9: Topwords for "Middle-Leverage" Topics
(Stemmed Spanish Topwords, Based on Score)

$\text{Middle}_{\#1}$	total, mont, numer, año, erog, ejerc, requier, concept, desgl, gast
$\text{Middle}_{\#2}$	cort, equip, vill, proveedor, dental, tecnic, clav, const, cumpl, cepill
$\text{Middle}_{\#3}$	document, pza, longitud, consign, institut, clav, evidenci, naturalez, env, catet
$\text{Middle}_{\#4}$	turist, turism, desarroll, promocion, bajacaliforniasur, paz, proyect, hotel, plan, mexic
$\text{Middle}_{\#5}$	institut, nacional, deport, period, politecn, solicit, tiemp, proporcion, part, ingeni

Table A.10: Topwords for "Middle-Leverage" Topics
(English Topwords, Based on Posterior Probability)

$\text{Middle}_{\#1}$	total, amount, number, year, spent, require, like this, fiscal year, concept, same
$\text{Middle}_{\#2}$	cut, team, provider, villa, technical personnel, code, fulfillment, building, dental, delivery
$\text{Middle}_{\#3}$	documents, institute, code, plaza, order, length, registration, nature, container, request
$\text{Middle}_{\#4}$	tourists, tourism, development, promotions, Baja California Sur, La Paz, project, plan, hotel, Mexican
$\text{Middle}_{\#5}$	national, institute, request, period, sport polytechnic, part, weather, provide, engineering

Table A.11: Topwords for "Middle-Leverage" Topics
(Destemmed Spanish Topwords, Based on Posterior Probability)

$\text{Middle}_{\#1}$	total, monto, numero, año, erogado, requiero, asi, ejercicio, concepto, mismo
$\text{Middle}_{\#2}$	corte, equipo, proveedor, villa, tecnico, clave, cumplimiento, construcción, dentales, entrega
$\text{Middle}_{\#3}$	documentos, instituto, clave, pza, consigne, longitud, evidencie, naturaleza, envase, solicito
$\text{Middle}_{\#4}$	turistas, turismo, desarrollo, promociones, Baja California Sur, paz, proyecto, plan, hotel, mexicano
$\text{Middle}_{\#5}$	nacional, instituto, solicito, periodo, deporte, politecnico, parte, tiempo, proporcionar, ingeniería

Table A.12: Topwords for "Middle-Leverage" Topics
(Stemmed Spanish Topwords, Based on Posterior Probability)

$\text{Middle}_{\#1}$	total, mont, numer, año, erog, requier, asi, ejerc, concept, mism
$\text{Middle}_{\#2}$	cort, equip, proveedor, vill, tecnic, clav, cumpl, const, dental, entreg
$\text{Middle}_{\#3}$	document, institut, clav, pza, consign, longitud, evidenci, naturalez, env, solicit
$\text{Middle}_{\#4}$	turist, turism, desarroll, promocion, bajacaliforniasur, paz, proyect, plan, hotel, mexic
$\text{Middle}_{\#5}$	nacional, institut, solicit, period, deport, politecn, part, tiemp, proporcion, ingeni

Classification Without Government Agency Features

All sLDA analyses employed in our main paper used a unified set of features that together encompassed (i) respondents' actual request texts and (ii) the names of each target federal agency that a request was made to. This Supplemental Appendix section evaluates the "value-added" of the latter (agency based) information for the accurate classification of government (non)responsiveness. To do so, we removed all *target agency* features from our original request text data, and re-ran our sLDA model on our (now revised) training sample of 250,000 randomly selected requests, again while using our primary "denied response" outcome variable. We then generated comparable sets of out-of-sample predictions, random classifiers, and classification statistics to those discussed in the main paper while using an updated test sample of 300,000 request texts that now also excludes target agency names. The results from this exercise appear in Table A.13 (in-sample) and Table A.15 (out-of-sample) below.

Table A.13: In-Sample Classification Statistics (Just Text)

	AUC-PR	**AUC**	**Precision**	**Recall**	**F1score**	**Accuracy**
sLDA	24.29	72.39	33.30	12.18	17.83	89.12
$\xi = \frac{1}{2}$	09.61	50.36	09.57	49.51	16.04	49.75
$\xi = \bar{y}$	09.69	50.00	09.70	10.05	09.87	82.21

Table A.14: Out-of-Sample Classification Statistics (Just Text)

	AUC-PR	**AUC**	**Precision**	**Recall**	**F1score**	**Accuracy**
sLDA	23.19	71.47	32.18	11.31	16.74	89.10
$\xi = \frac{1}{2}$	09.69	50.03	09.69	50.01	16.25	50.04
$\xi = \bar{y}$	09.75	50.15	09.94	10.25	10.09	82.32

In both the in-sample and out-of-sample contexts, the "requests only" sLDA model performs slightly worse than the full sLDA model discussed previously. For example, the "text only" sLDA model's in-sample and out-of-sample AUCs are 72.39 and 71.47, respectively. Each of these AUCs are noticeably smaller than that of our primary sLDA model (of 74.09 and 73.23), where the observed differences are roughly comparable to the differences obtained in moving from the in-sample to out-of-sample settings within either model. We obtain similar conclusions in comparing the overall in- or out-of-sample AUC-PR, precision, recall, and F1 scores of these two sLDA models, where in each case, the inclusion of target agency names within our text features leads to a small, but consistent and non-negligible, improvement in classification. At the same time, the "requests only" sLDA classifier continues to exhibit superior accuracy than either random classifier in terms of AUC, precision, F1 score, and overall accuracy—thereby underscoring the overall predictive benefits of applying sLDA to our raw request texts, relative to random chance.

Comparison to Alternate Approaches

Supervised Classification Comparison

Returning to our primary sample of texts, we next compare our sLDA approach to three widely used alternatives: logistic regression with a LASSO penalty, logistic regression with a ridge penalty, and standard logistic regression. For all four models, we use a document term matrix[9] that includes (i) all unique (processed) wordstems and (ii) each request's target federal government agency name, as features. We then leverage the remaining 10% of our full 2003-2015 Mexican request text corpus for these model-based comparisons, which is equivalent to roughly $100,000$ total requests. Herein, we randomly sub-divide this $100,000$ request sample into new sets of training requests ($n = 25,000$) and test requests ($n = 75,000$). We then use the $25,000$ training documents to re-estimate a new sLDA model alongside our logit, LASSO, and ridge estimators, so as to ensure that the out-of-sample predictions generated by (i) our sLDA model and (ii) our comparison models are comparable in terms of the size of the training sample used. LASSO and ridge regression have a tuning parameter (λ) whereas logit does not. For the former two models, we hence perform five-fold cross-validation within our $25,000$-document training sample, so as to identify the best performing LASSO and ridge models for out of sample prediction. We keep our final sLDA

[9]I.e., document-level counts.

model's topic number and hyperparameters fixed at the values identified earlier.

After training each model on our new $n = 25,000$ training sample, we generate predictions of "denied requests" for the remaining (held out) $75,000$ requests that comprise our test sample for this comparison. The results from this exercise appear in Table A.15. Here we can note that sLDA outperforms LASSO, ridge, and logistic regression across the most relevant metrics considered. For instance, sLDA's AUC-PR (32.23) is noticeably higher than that of our comparison metrics, which are each approximately equal to 29. The same can be said for the AUCs in Table A.15, which clearly favor sLDA (AUC= 66.15) over each alternative model (AUCs= 55.70 $\leftrightarrow$ 57.66). Turning to precision, recall, and the F1 scores reported in Table A.15, we find that Ridge and LASSO each perform better in precision than sLDA, but perform substantially worse than sLDA in terms of recall. As a result, sLDA's F1 score is roughly 10-to-20 times higher than those of LASSO and Ridge, and almost double that of logit—suggesting that sLDA is a noticeably more accurate classifier in this context. At the same time, our overall accuracy statistics *do* suggest that logit, LASSO and ridge accurately classify roughly 5-10% more test cases than does sLDA. Given the overall imbalance of our denied request outcomes, this discrepancy likely arises due to the overprediction of zeroes in the cases of logit, LASSO, and ridge, relative to sLDA.

Table A.15: Out-of-Sample Comparisons

	AUC-PR	**AUC**	**Precision**	**Recall**	**F1score**	**Accuracy**
sLDA	35.23	66.15	34.11	50.18	40.61	66.91
Ridge	28.93	57.66	62.73	1.39	2.73	77.57
LASSO	29.08	55.70	41.78	6.48	11.21	76.87
Logit	27.64	56.47	31.67	19.89	24.43	72.25

Structural Topic Model Comparison

Our primary (sLDA) application conditioned our estimated topics on an external, document-level dependent variable ("denied request"). An alternative strategy would be to condition these topics on an external, document level *independent* variable, via the structural topic model (Roberts et al., 2014). As noted in our main paper, we believe that the modeling of "denied request" as an outcome variable—via sLDA—rather than as an explanatory variable—via the structural topic model (STM)—is most appropriate in our context, given that official responses to request texts arise *after* (as opposed to prior to) the creation of our request texts themselves. Nevertheless, it is likely that an STM that has been conditioned on our "denied request" indicator will offer similar topic insights to that of the sLDA model, given the underlying similarities between these two models. Estimating such an STM accordingly has the potential to (i) provide insights into the stability of the Denied and Provided sLDA findings discussed in the main paper, (ii) sharpen our understandings of high leverage (Denied or Supplied) topics in the Mexico context more broadly, and (iii) offer novel evidence of the STMs effectiveness for inductive "needle-in-a-haystack"-type research tasks.

In light of the above points, we estimate a 250-topic STM on the same sample of ATI requests that were used to generate our main paper's in-sample sLDA results. Recall that this sample corresponded to 10% of our total corpus (i.e., approximately 100,000 documents). We include "denied request" as the sole prevalence covariate within our STM specification. To address multi-modality concerns, we utilize different starting parameters to estimate five separate 250-topic STMs, and store the exclusivity and semantic coherence of topics' words for each model. Following past practice (e.g., Roberts et al., 2014; Bagozzi and Berliner, 2018) we choose the STM run (i.e., from this set of five estimated models) that maximizes the semantic coherence and exclusivity of our topic-word vectors. Using that selected STM's estimates, we then calculate the estimated effect of a 0-to-1 change in "denied request" upon topical prevalence for each of our 250 topics, and rank these estimated effects in size. This ranking allows us to identify the top five Denied and top five Provided topics, as identified by the STM, which we report via topwords based upon score (Tables A.16-A.18) and posterior probability (Tables A.19-A.21) below.

With these results in hand, we now discuss the similarities and differences between (i) our STM's Denied and Supplied topics and (ii) our original sLDA model's Denied and Supplied topics. To assess thematic overlap of these topics, we compared our original sLDA-Denied and sLDA-Supplied topics' topwords to those of the top 15 STM-Denied and top 15 STM-Supplied topics. In each case, we rely on the score-based topwords for our topword comparisons. Our findings are comparable if one instead relies upon topwords derived from highest posterior probability instead of score. In addition to the top five most highly associated topwords for each relevant topic (which we report for our sLDA models above and in the main paper, and for our STM models below), we also additionally consider the top 20 most highly associated (score) topwords for our topic comparisons. Doing so helps us to avoid the tendency towards false positives in our topword comparisons, given that some top five words frequently occur across multiple topics. Our discussion below begins with an assessment of the similarities and differences among our sLDA-Denied topics and the STM-Denied topic results. This is followed by an evaluation of the similarities and differences that arise between our sLDA-Provided topics and STM-Denied topics. Finally, we conclude this section with a broader discussion of our STM findings in light of our main paper's theoretical contentions.

A comparison of our top five sLDA-Denied topics to the STM's top 15 Denied topics reveals a number of notable similarities. For instance, our sLDA-Denied$_{\#4}$ topic shares clear overlap with STM-Denied$_{\#5}$, with multiple shared topwords across each topic's score-based topword vectors. In this case, examples of overlapping words include 'solicit,' 'naturalez,' 'documental,' 'consign,' and 'corre,' suggesting that these two topics both pertain to corruption-oriented procurement requests. Interestlingly, whereas our sLDA-Denied$_{\#4}$ arrived at this conclusion based upon the topwords and associated documents, STM-Denied$_{\#5}$ further reaffirms this interpretation with its additional inclusion of the topwords 'anti-corruption,' and 'crusade' in Table A.16. Second, while sLDA-Denied$_{\#2}$ is not as clearly represented within the top five STM-Denied topics, further examination of the top 15 highest leverage "denied request" topics from our STM model suggests that sLDA-Denied$_{\#2}$ does share substantial overlap with the 12th highest ranking Denied topic in the STM model. In this case, the top twenty topwords for sLDA-Denied$_{\#2}$ and STM-Denied$_{\#12}$ each include the following financial sector-oriented topwords: 'banc,' 'financier,' 'ahorr,' 'deposit' and 'bancari,' suggesting substantial thematic overlap in this regard. Similarly, while sLDA-Denied$_{\#3}$ does not appear directly within the top five STM-Denied topics, further examination of our top 15 STM-Denied topics suggests that sLDA-Denied$_{\#3}$ shares a total of eight topwords with STM-Denied$_{\#8}$.[10] On the other hand, we find less evidence of topword overlap between sLDA-Denied$_{\#1}$ (or sLDA-Denied$_{\#5}$) and our remaining top 15 STM-Denied topics.

Turning to our Provided topics, we can note that sLDA-Provided$_{\#1}$ exhibits substantial thematic overlap with STM-Provided$_{\#1}$, including nine shared topwords among these two topics' respective top twenty word vectors: 'educ,' 'secretaría-de-educación-pública, 'sep,' 'plantel,' 'colegi,' 'cetis,' 'dgeti,' 'bachiller,' and 'docent.' Likewise, sLDA-Provided$_{\#2}$ exhibits comparably high overlap with STM-Provided$_{\#1}$[11] and also with sLDA Provided$_{\#5}$.[12] Together these results suggest that the STM and sLDA models identify very similar education-focused topics within the ATI topics identified as most likely to see a provided response. While sLDA-Provided$_{\#4}$ has less evidence of overlap with any of the top five STM-Provided topics, we find that there is noticeable conceptual overlap in the top 20 score-based topwords between this specific sLDA-Provided topic and STM-Provided$_{\#8}$, including shared words such as: 'presupuest, 'asign' and 'ejerc.' Further, while sLDA Provided$_{\#5}$ likewise fails to exhibit substantial overlap with any of our top five STM-Provided topics, it has substantial thematic overlap with STM-Provided$_{\#11}$, including a total of nine shared topwords across these two topics top 20 topword vectors.[13] Even in this case of sLDA-Provided$_{\#3}$, which does not exhibit substantial topword overlap with any of our STM-Provided topics, one can observe that the top five words for sLDA-Provided$_{\#3}$ are conceptually similar to those of STM-Provided$_{\#3}$, in that both topics largely appear to be requests for specific numeric amounts (or increases in amounts) that should be fairly easy for government agencies

[10] Specifically, 'sps,' 'procuradur,' 'republ,' 'coordin,' 'general,' 'trabaj,' 'pormenoriz,' and 'puest.'

[11] Ten overlapping words, including: 'educ,' 'secretaría-de-educación-pública,' 'sep,' 'administración-federal-de-servicios-educativos-en-el-distrito-federal-(afsedf),' 'maestr, 'bachillerat,' 'escolar,' 'basic,' 'preescol,' 'docent.'

[12] Ten overlapping words, including: 'escuel,' 'alumn,' 'docent,' 'superior,' 'secundari,' 'escol,' 'profesor,' 'primari,' 'cicl,' and 'preescol.'

[13] The overlapping words in this case are 'puest,' 'sueld,' 'tabul,' 'mensual,' 'percepcion,' 'compens,' 'brut,' 'remuner,' and 'net.'

to provide. Hence, the results discussed here and above suggest that our sLDA and STM approaches are identifying similar top Denied and top Provided topics and themes.

Taken together, the findings summarized above help to reinforce the stability and validity of our primary in-sample sLDA topics. They also suggest that the STM, like the sLDA model, is an effective method for the inductive identification of small subsets of unique documents within larger corpora of government texts. As further evidence of the latter point, we can note that several uniquely identified STM-Denied and STM-Provided topics within Table A.16 help to further sharpen the theoretical conclusions obtained within our primary analyses. For example, STM-Denied$_{\#3}$ did not exhibit substantial overlap with any of our top five sLDA-Denied topics. Yet, with top words related to corruption ('corrupt,' 'forged,' 'honesty') and conflict-oriented terms ('robbery,' 'revolutionary'), it would appear that this topic is capturing an additional distinct subset of politically sensitive requests that are even more overt and confrontational than those identified within our sLDA analysis. Likewise, STM-Denied$_{\#2}$ uniquely encompasses a range of terms related to medical device procurement[14] that appear to relate to investigations into the safety of such devices ('sanitary,' 'risks,' 'probes'). As above, this topic helps to expand our understandings of the types of potentially politically sensitive requests that arise (and face frequent denial) in Mexico, thereby further underscoring our earlier conclusion that the STM has similar levels of usefulness to the sLDA model for inductive "needle-in-the-haystack"-type tasks.

[14] Which oftentimes may encompass relatively benign requests.

Table A.16: Topwords for STM Topics Associated with “Denied Request”
(English Topwords, Based on Score)

Denied$_{\#1}$	alienation, settlement, sheets, Giralt, SAE, electronic, Abreu, goods, loan, director
Denied$_{\#2}$	sanitary, Cofepris, register, bureaucratic procedure, probes, Foley (catheter), latex, caliber, commissions, risks
Denied$_{\#3}$	honesty, corrupt, robbery, truce, bring, forged, noble, barracks, we will honor, revolutionary
Denied$_{\#4}$	copy, simple, I request, versions, constitutive, legible, Sandra, arbitration, friends, open
Denied$_{\#5}$	anti-corruption, documentary, crusade, without, contact us, order, insertions, registration, assembly, correspondent
Provided$_{\#1}$	education, SEP, scholarships, staff, MA, school, secondary, CETIS, school, DGETI
Provided$_{\#2}$	medication, pieces, purchase, request for bids, number, paragraph, hospital, differentiating, adjudication, medication
Provided$_{\#3}$	how much, which, increase, increased, question, these, why, Diconsa, they come, get
Provided$_{\#4}$	program, support, beneficiaries, SEDESOL, census, subsidies, regulation, community, PROCAMPO, opportunities
Provided$_{\#5}$	school, students, teacher, higher education, highschool, school, professors, primary, cycle, MONEX

Table A.17: Topwords for STM Topics Associated with “Denied Request”
(Destemmed Spanish Topwords, Based on Score)

Denied$_{\#1}$	enajenación, liquidación, fojas, giralt, sae, electrónico, abreu, bienes, comodato, director
Denied$_{\#2}$	sanitario, cofepris, registro, tramite, sondas, foley, latex, calibre, comisiones, riesgos
Denied$_{\#3}$	honestidad, corruptos, rapiña, tregua, traciona, forjan, noble, cuartel, honremos, revolucionaria
Denied$_{\#4}$	copia, simple, solicito, versiones, consistutiva, legible, sandra, arbitrales, amigos, abierta
Denied$_{\#5}$	anticorrupcion, documental, cruzada, únet, contactanos, consigne, inserciones, evidencie, asamblea, correspondiente
Provided$_{\#1}$	educación, sep, becas, plantel, maestría, colegio, bachillerato, cetis, escolar, dgeti
Provided$_{\#2}$	meidcamento, piezas, compra, licitación, numero, parrafo, hospital, diferencial, adjudicación, medicamento
Provided$_{\#3}$	cuanto, cual, asciende, ascendió, cuestión, estas, porque, diconsa, provienen, obtiene
Provided$_{\#4}$	programa, apoya, beneficiarios, sedesol, padron, subsidios, reglamento, comunitario, procampo, oportunidades
Provided$_{\#5}$	escuela, alumnos, docente, superior, secundaria, escolar, profesores, parimaria, ciclo, monex

Table A.18: Topwords for STM Topics Associated with "Denied Request" (Stemmed Spanish Topwords, Based on Score)

Denied$_{\#1}$	enajen, liquid, foj, giralt, sae, electr, abreu, bien, comodat, director
Denied$_{\#2}$	sanitari, cofepris, registr, tramit, sond, foley, latex, calibr, comision, riesg
Denied$_{\#3}$	honest, corrupt, rapiñ, tregu, tracion, forj, nobl, cuartel, honr, revolucionari
Denied$_{\#4}$	copi, simpl, solicit, version, consistut, legibl, sandr, arbitral, amig, abiert
Denied$_{\#5}$	anticorrupcion, documental, cruz, únet, contactan, consign, insercion, evidenci, asamble, corre
Provided$_{\#1}$	educ, sep, bec, plantel, maestr, colegi, bachillerat, cetis, dgeti, bachiller
Provided$_{\#2}$	meidcament, piez, compr, licit, numer, parraf, hospital, diferencial, adjud, medic
Provided$_{\#3}$	cuant, cual, asciend, ascend, cuest, estas, porque, dicons, provien, obtien
Provided$_{\#4}$	program, apoy, beneficiari, sedesol, padron, subsidi, regl, comunitari, procamp, oportun
Provided$_{\#5}$	escuel, alumn, docent, superior, secundari, escol, profesor, parimari, cicl, monex

Table A.19: Topwords for STM Topics Associated with "Denied Request" (English Topwords, Based on Posterior Probability)

Denied$_{\#1}$	settlement, sheets, alienation, goods, director, federal, characteristics, electronic, cease, SAE
Denied$_{\#2}$	sanitary, register, Cofepris, bureaucratic procedure, commissions, risks, I request, entry, protections
Denied$_{\#3}$	public, transparent, corruption, Mexican, servers, honesty, should, corrupt, I request, information
Denied$_{\#4}$	copy, I request, simple, versions, constitutive, legible, Sandra, arbitration, friends, open
Denied$_{\#5}$	documentary, I request, crusade, order, anti-corruption, assembly, approved, information, following, insertions
Provided$_{\#1}$	education, SEP, scholarships, MA, staff, school, school, basic, secondary, secondary
Provided$_{\#2}$	information, number, medication, pieces, current, request/application, favor, purchase, paragraph, latest
Provided$_{\#3}$	how much, which, increase, increased, question, these, why, they come, get, Diconsa
Provided$_{\#4}$	program, support, beneficiaries, census, opportunities, subsides, SEDESOL, regulation, community, PROCAMPO
Provided$_{\#5}$	school, higher education, highschool, students, teacher, primary, school, professors, cycle, group

Table A.20: Topwords for STM Topics Associated with “Denied Request” (Destemmed Spanish Topwords, Based on Posterior Probability)

Denied$_{\#1}$	liquidación, fojas, enajenación, bienes, director, federal, características, electrónico, cesar, sae
Denied$_{\#2}$	sanitario, registro, cofepris, tramite, comisiones, riesgos, solicito, ingreso, protecciones
Denied$_{\#3}$	públic, transparente, corrupcion, mexicano, servidores, honestidad, debe corruptos, solicito, información
Denied$_{\#4}$	copia, solicito, simple, versiones, constitutiva, legible, sandra, arbitrales, amigos, abierta
Denied$_{\#5}$	documental, solicito, cruzada, consigne, anticorrupcion, asamblea, aprobada, información, siguiente, inserciones
Provided$_{\#1}$	educación, sep, becas, maestría, plantel, colegio, escolar, basico, bachillerato, bachillerato
Provided$_{\#2}$	información, numero, medicamento piezas, presente, solicitud, favor, compra, parrafo, ultimo
Provided$_{\#3}$	cuanto, cual, asciende, ascendió, cuestión, estas, porque, provienen, obtiene, diconsa
Provided$_{\#4}$	programa, apoyo, beneficiaros, padron, oportunidades, subsidios, sedesol, reglamento, comunitario, procampo
Provided$_{\#5}$	escuela, superior, secundaria, alumnos, docente, primaria, escolar, profesores, ciclo, grupo

Table A.21: Topwords for STM Topics Associated with “Denied Request” (Stemmed Spanish Topwords, Based on Posterior Probability)

Denied$_{\#1}$	liquid, foj, enajen, bien, director, federal, caract, electr, ces, sae
Denied$_{\#2}$	sanitari, registr, cofepris, tramit, comision, riesg, solicit, ingres, proteccion, sond
Denied$_{\#3}$	públic, transparent, corrupcion, mexic, servidor, honest, deb, corrupt, solicit, inform
Denied$_{\#4}$	copi, solicit, simpl, version, constitut, legibl, sandr, abitral, amig, abiert
Denied$_{\#5}$	documental, solicit, cruz, consign, anticorrupcion, asamble, aprob, inform, siguient, insercion
Provided$_{\#1}$	educ, sep, bec, maestr, plantel, colegi, escolar, basic, bachillerat, bachiller
Provided$_{\#2}$	inform, numer, medicament, piez, present, solicitud, favor, compr, parraf, ultim
Provided$_{\#3}$	cuant, cual, asciend, ascend, cuest, estas, proqu, provien, obtien, dicons
Provided$_{\#4}$	program, apoy, beneficiari, padron, oportun, subsidi, sedesol, regl, comunitari, procamp
Provided$_{\#5}$	escuel, superior, secundari, alumn, docent, primari, escol, profesor, cicl, grup

References

Bagozzi, Benjamin E. 2015. "The Multifaceted Nature of Global Climate Change Negotiations." *Review of International Organizations* 10(4):439–464.

Bagozzi, Benjamin E. and Daniel Berliner. 2018. "The Politics of Scrutiny in Human Rights Monitoring: Evidence from Structural Topic Models of U.S. State Department Human Rights Reports." *Political Science Research and Methods* 6(4):662–677.

Bagozzi, Benjamin E., Daniel Berliner and W. Zack Almquist. 2016. "Predicting Government (Non)Responsiveness to Freedom of Information Requests with Supervised Latent Dirichlet Allocation." *Proceedings of the International Conference on the Advances in Computational Analysis of Political Text - PolText* .

Barberá, Pablo, Richard Bonneau, John T. Jost, Jonathan Nagler and Joshua Tucker. 2014. "Is There Anybody Out There? The Effects of Legislators' Communication with their Constituents." Paper presented at the ISA Annual Convention.

Berliner, Daniel, Benjamin E. Bagozzi and Brian Palmer-Rubin. 2018. "What Information Do Citizens Want?: Evidence from One Million Information Requests in Mexico." *World Development* 109:222–235.

Berliner, Daniel, Benjamin E. Bagozzi, Brian Palmer-Rubin and Aaron Erlich. 2020. "The Political Logic of Government Disclosure: Evidence from Information Requests in Mexico." *Journal of Politics* .

Feinerer, Ingo, Kurt Hornik and David Meyer. 2008. "Text Mining Infrastructure in R." *Journal of Statistical Software* 25(5):1–54.
URL: *http://www.jstatsoft.org/v25/i05/*

Porter, Martin. 1980. "An Algorithm for Suffix Stripping." *Program* 14(3):130–137.

Puschmann, Cornelius and Tatjana Scheffler. 2016. "Topic Modeling for Media and Communication Research: A Short Primer." HIIG Discussion Paper Series No. 2016-05. URL: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2836478`.

Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin and Dragomir R Radev. 2010. "How to Analyze Political Attention with Minimal Assumptions and Costs." *American Journal of Political Science* 54(1):209–228.

Roberts, Margaret E., Brandon M. Steward, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Gadarian, Bethany Albertson and David G. Rand. 2014. "Structural Topic Models for Open Ended Responses." *American Journal of Political Science* 4(58):1064–1082–442.

Zhang, Yilin, Marie Poux-Berthe, Chris Wells, Karolina Koc-Michalska and Karl Rohe. 2018. "Discovering Political Topics in Facebook Discussion Threads with Graph Contextualization." *Annals of Applied Statistics* 12(2):1096–1123.