# Inflation Regimes in Latin America: A Variational Bayesian Approach

Gabriel Sánchez-Hernández

December 17, 2024

# Motivation

- High and volatile inflation rates challenge price stability, a fundamental goal of central banks worldwide
- Inflation reduces purchasing power and create uncertainty about future prices, disrupting business and consumer decision-making and potentially slowing investment and economic growth (Briault, 1995; Lucas, 1972)
- Essentially a tax on those who hold money (Mishkin, 2009)

# Objective

Uncover inflation regimes, their dynamics and persistence, and analyze the determinants of inflation across nine Latin American countries: Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Guatemala, Honduras, Mexico, and Peru.
Period: January 2008 to December 2023

# Content

# Variational Bayesian Methods

- In model selection, the Bayesian approach is to calculate the posterior distribution over a set of models, given a priori knowledge and data, $\boldsymbol{y}$
- The knowledge is represented in the form of a prior over model structures $p(m)$ and their parameters $p(\boldsymbol{\theta}|m)$
- By Bayes' rule:

$$p(m|\boldsymbol{y}) = \frac{p(m)p(\boldsymbol{y}|m)}{p(\boldsymbol{y})} = \frac{p(m)\overbrace{\int p(\boldsymbol{y}|\boldsymbol{\theta}, m)p(\boldsymbol{\theta}|\boldsymbol{m})d\boldsymbol{\theta}}^{\text{marginal likelihood}}}{p(\boldsymbol{y})}$$

# Variational Bayesian Methods

There are multiple ways to approximate the marginal likelihood:

- Estimating the value at a single point estimate of $\boldsymbol{\theta}$

$$\boldsymbol{\theta}_{\mathrm{ML}} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{y}|\boldsymbol{\theta}, m)$$

$$\boldsymbol{\theta}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|m) p(\boldsymbol{y}|\boldsymbol{\theta}, m)$$

- Estimate it numerically at different $\boldsymbol{\theta}$, using Monte Carlo methods (MCMC)

The difficulty is that these methods are either computationally intensive or impossible to calculate

# Variational Bayesian Methods

A third way of estimating the marginal likelihood is using *variational* methods. The variational approach relies on

- Approximating the integral with a tractable form, forming a lower *bound*
- Optimizing the bound, making it as close as possible to the true value

# Signal Models

*Real-world processes generally produce observable outputs which can be characterized as signals. A problem of interest is characterizing such signals in terms of models (Rabiner, 1989)*

- Sources: stationary or nonstationary
- Signals: discrete or continuous, pure or corrupted
- Models: deterministic or statistical

# Statistical Signal Models

- The second class of models utilize the statistical properties of the signals
- Assumptions: the signal can be modeled as a parametric random process, and the parameters can be estimated
- Gaussian processes, Poisson processes, Markov processes,...

# Discrete Markov Models

- A system that can be, at any time instant $t \in \{1, ..., T\}$, in one state $j \in \{1, \ldots, J\}$
- Each state corresponds to an observable event
- The state transition probabilities are

$$\pi_{i,k} = p(t = k | t-1 = i)$$

$$\forall i, k \in \{1, ..., J\}$$

$$\forall t \in \{2, ..., T\}$$

$$0 \leq \pi_{i,k} \leq 1 \qquad \sum_{k=1}^{J} \pi_{i,k} = 1$$

# Discrete Markov Models

- Assume a five-period, observable Markov model
- Consider two states: sunny and rainy. A possible collection of outcomes could be

$$\boldsymbol{y_t} = \begin{bmatrix} \text{rainy} & \text{rainy} & \text{rainy} & \text{sunny} & \text{sunny} \end{bmatrix}$$

- And the state transition probabilities

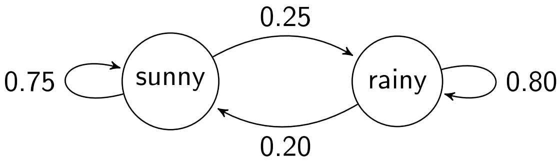



Figure 1: A Markov chain to describe the weather

# Discrete HMMs

What happens when the real-world process can't be observed?

- The states and the signals do not belong to the same set (and could also be different in nature)
- We then define a Hidden Markov Model (HMM) as

*An underlying stochastic process that is not directly observable, but can be observed through another set of stochastic processes that produce the sequence of observations* (Baum & Petrie, 1966; Baum et al., 1970)

# MGHMM

Consider a Multivariate Gaussian Hidden Markov Model with

1. $\boldsymbol{Y}$, the sequence of observations $\boldsymbol{y}_t$, taking variables in $\mathbb{R}^p$
   The distributions of $\boldsymbol{Y}$ depend only on contemporary $\boldsymbol{z}_{t,j}$
2. $\boldsymbol{Z}$, the set of latent variables $\boldsymbol{z}_{t,j}$
3. $\boldsymbol{\Pi}$, the state transition matrix, with elements $\pi_{i,j}$
4. $\boldsymbol{\pi}$, the initial state probability vector
5. $\boldsymbol{\Theta}$, the parameter matrix

For each timestep $t \in \{1, \ldots, T\}$, and each hidden state $i, j \in \{1, \ldots, J\}$

# MGHMM

Recalling Bayes' rule:

$$\underbrace{p(\boldsymbol{\Theta}|\boldsymbol{Y})}_{\text{Posterior}} = \frac{\overbrace{p(\boldsymbol{Y}|\boldsymbol{\Theta})}^{\text{Likelihood}}\overbrace{p(\boldsymbol{\Theta})}^{\text{Prior}}}{\underbrace{p(\boldsymbol{Y})}_{\text{Evidence}}} \tag{1}$$

- The parameters are treated as unknown, as well as the hidden variables in the posterior
- VB approximates the distribution with a simpler distribution

# MGHMM

Let the log-likehood be defined as

$$\ln p(\boldsymbol{Y}|\boldsymbol{\Theta}) = \ln \int p(\boldsymbol{Y}, \boldsymbol{Z}|\boldsymbol{\Theta})\, d\boldsymbol{Z} \tag{2}$$

$$= \ln \int q(\boldsymbol{Z}) \frac{p(\boldsymbol{Y}, \boldsymbol{Z}|\boldsymbol{\Theta})}{q(\boldsymbol{Z})}\, d\boldsymbol{Z} \tag{3}$$

$$\geq \int q(\boldsymbol{Z}) \ln \left\{ \frac{p(\boldsymbol{Y}, \boldsymbol{Z}|\boldsymbol{\Theta})}{q(\boldsymbol{Z})} \right\} d\boldsymbol{Z} \tag{4}$$

by applying Jensen's inequality (Jensen, 1906). Rewriting:

$$\ln p(\boldsymbol{Y}|\boldsymbol{\Theta}) = \int q(\boldsymbol{Z}) \ln \left\{ \frac{p(\boldsymbol{Y}, \boldsymbol{Z}|\boldsymbol{\Theta})}{q(\boldsymbol{Z})} \right\} d\boldsymbol{Z} \tag{5}$$

$$+ \int q(\boldsymbol{Z}) \ln \left\{ \frac{q(\boldsymbol{Z})}{p(\boldsymbol{Z}|\boldsymbol{Y}, \boldsymbol{\Theta})} \right\} d\boldsymbol{Z}$$

$$= \mathcal{L}(q, \boldsymbol{\Theta}) + \mathrm{KL}(q||p) \tag{6}$$

# MGHMM, Bound

We have then

$$\ln p(\boldsymbol{Y}|\boldsymbol{\Theta}) = \mathcal{L}(q, \boldsymbol{\Theta}) + \text{KL}(q||p)$$

- $\mathcal{L}(q, \boldsymbol{\Theta})$, the variational lower bound
- $\text{KL}(q||p)$, the Kullback-Leibler divergence between the variational distribution and the hidden variable posterior

The goal is obtaining a $q(\boldsymbol{Z})$ that maximizes $\mathcal{L}(q, \boldsymbol{\Theta})$, or

$$\arg\min_{q} \text{KL}(q||p) \tag{7}$$

# MGHMM, Distribution

The model joint distribution is

$$p(\mathbf{Y}, \mathbf{Z}, \mathbf{\Pi}, \mu, \mathbf{\Lambda}) = p(\mathbf{Y}|\mathbf{Z}, \mu, \mathbf{\Lambda})p(\mu|\mathbf{\Lambda})p(\mathbf{\Lambda})p(\mathbf{Z}|\mathbf{\Pi})p(\mathbf{\Pi}) \tag{8}$$

with $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Lambda})$. The factorization of the variational distribution is assumed to be

$$\begin{aligned} q(\mathbf{Z}, \mathbf{\Pi}, \mu, \mathbf{\Lambda}) &= q(\mathbf{Z})q(\mathbf{\Pi}, \mu, \mathbf{\Lambda}) && (9) \\ &= q(\mathbf{Z})q(\mathbf{\Pi})q(\mu, \mathbf{\Lambda}) && (10) \\ &= q(\mathbf{Z}) \prod_{i=1}^{J} q(\boldsymbol{\pi}_i) \prod_{j=1}^{J} q(\boldsymbol{\mu}_j, \mathbf{\Lambda}_j) && (11) \end{aligned}$$

# MGHMM, Priors

For each state $j \in \{1, \ldots, J\}$, the prior knowledge is assumed to be

- Independent univariate Gaussian prior distribution for $\boldsymbol{\mu}_j$, conditional on the precisions
- Independent Wishart prior distribution for $\boldsymbol{\Lambda}_j$
- Independent Dirichet prior distribution for $\boldsymbol{\pi}_j$

Given that the priors are assumed to be from a conjugate-exponential family, the lower bound is now tractable

# MGHMM, VBEM

The VB Expectation-Maximization (VBEM) algorithm allows to iteratively maximize the variational lower bound (LB). Given initial hyperparameters:

- Expectation step: derivate the LB over the hidden states
- Maximization step: derivate the LB over the hidden parameters

Each step is guaranteed to increase, or leave unchanged, the LB

# MGHMM, Advantages

MacKay (1997) is the first to use variational methods on hidden markov models. The are multiple advantages:

- It takes into account model complexity, against methods like ML and MAP (Beal, 1998)
- It is impossible to run into local minima (Gruhl & Sick, 2016)
- When the number of hidden states is unknown, unnecessary states are eliminated as the lower bound converges to a solution (McGrory & Titterington, 2009)

# Inflation Drivers

Kinlaw et al. (2023) derive a way to analyze inflation drivers in the United States. Later, Sánchez (2023) adapts the methodology to Latin American economies by including international drivers. Let the Mahalanobis distance be defined as

$$\delta_{t,j} = (\boldsymbol{x}_t - \bar{\boldsymbol{x}}_j)' \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{x}_t - \bar{\boldsymbol{x}}_j) \tag{12}$$

with

- $\boldsymbol{x}_t$, observations taking variables in $\mathbb{R}^{p-1}$
- $\bar{\boldsymbol{x}}_j$, the average of the drivers in a certain regime
- $\boldsymbol{\Sigma}_j$, the covariance matrix of $\boldsymbol{X} = \{\boldsymbol{x}_t\}$ in regime $j$

for $j \in \{1, \ldots, J\}$, and $t \in \{1, \ldots, T\}$

# Inflation Drivers

Modifying (12) and rescaling

$$\rho_{t,j} = \frac{L_{t,j}}{\sum_{j=1}^{J} L_{t,j}} \tag{13}$$

where $L_{t,j}$ is the likelihood of a Gaussian distribution. Taking its derivative with respect to the observations gets

$$\zeta_t = \sum_{j=1}^{J} \eta_j \left| \frac{\partial \rho_{t,j}}{\partial \boldsymbol{x}_t} \right| \tag{14}$$

with $\eta_j$ the weight of regime $j$ in the period of study. Finally, rescaling with the standard deviations of the full sample:

$$\psi_t = \frac{\zeta_t \boldsymbol{\sigma}}{\sum |\boldsymbol{\zeta_t \sigma}|} \tag{15}$$

# Content

# Data

Table 1: Data and Measures

| Theory | Data | Variable | Measure |
|---|---|---|---|
| | Headline CPI | $cpi_t$ | MoM (%) |
| | Core CPI | $core_t$ | MoM (%) |
| Monetary | Broad money | $broad_t$ | 36M change (%) |
| | Policy-related interest rate (%) | $mpr_t$ | YoY (p.p.) |
| International | US headline CPI | $cpi_t^*$ | MoM (%) |
| | US core CPI | $core_t^*$ | MoM (%) |
| | Nominal exchange rate to $1 | $er_t$ | YoY (%) |
| Demand-pull | Household's consumption | $priv_t$ | YoY (%) |
| | Public sector's consumption | $gov_t$ | YoY (%) |
| Expectations | Inflation expectations, 12M (%) | $exp_t$ | YoY (p.p.) |
| Cost-push | Producer price index (PPI) | $ppi_t$ | YoY (%) |

Data retrieved from: IMF, FLAR, SECMCA, FRED, central banks.
*Note*: all time series are available from January 2008 to December 2023, except for $exp_t$ and $ppi_t$, available from January 2015 to December 2023. Moreover, $ppi_t$ is unavailable for the Dominican Republic, Guatemala, and Honduras.

# Models

Table 2: Model Specifications

| Model | Variables | Period |
|---|---|---|
| 1 | $cpi_t$ | 08M1-23M12 |
| 2 | $core_t$ | 08M1-23M12 |
| 3 | $cpi_t$, $broad_t$, $mpr_t$, $cpi^*_t$, $er_t$, $priv_t$, $gov_t$ | 08M1-23M12 |
| 4 | $core_t$, $broad_t$, $mpr_t$, $core^*_t$, $er_t$, $priv_t$, $gov_t$ | 08M1-23M12 |
| 5 | $cpi_t$, $broad_t$, $mpr_t$, $cpi^*_t$, $er_t$, $priv_t$, $gov_t$, $exp_t$ | 15M1-23M12 |
| 6 | $cpi_t$, $broad_t$, $mpr_t$, $cpi^*_t$, $er_t$, $priv_t$, $gov_t$, $exp_t$, $ppi_t$ | 15M1-23M12 |

*Note*: Model 6 is unavailable for the Dominican Republic, Guatemala, and Honduras.

# Estimation

The estimation for each country, and each model specification, is as follows:

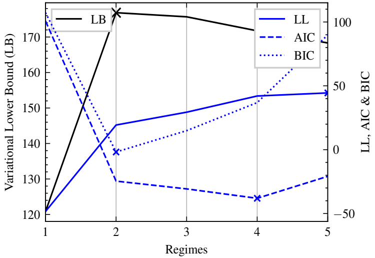
1. Define weakly informative priors, and a maximum number of possible hidden states ($J = 5$)
2. For $j \in \{1, \ldots, 5\}$, initialize the model a hundred times, and train a million iterations of VBEM
3. Chose the model complexity ($J^*$) based on the variational lower bound, $\mathcal{L}(q, \boldsymbol{\Theta})$
4. Obtain the optimal hidden regime sequence
5. Get the expected value of the parameters ($\boldsymbol{\Pi}$ and $\boldsymbol{\mu}_j$), dependent on the regimes
6. Calculate the relative importance of the inflation drivers, $\boldsymbol{\psi}_t$

# Content

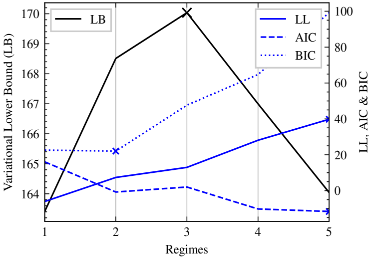
# Model Complexity

- The LL suggests a higher number of states than the LB, often reaching $J^* = 5$
- Akaike criteria tend to choose higher states than the BIC
- The LB is more conservative and does not overfit, suggesting $J^* = 2$ or $J^* = 3$ for models and countries

Figure 2: Regime Selection Criteria, Model 1



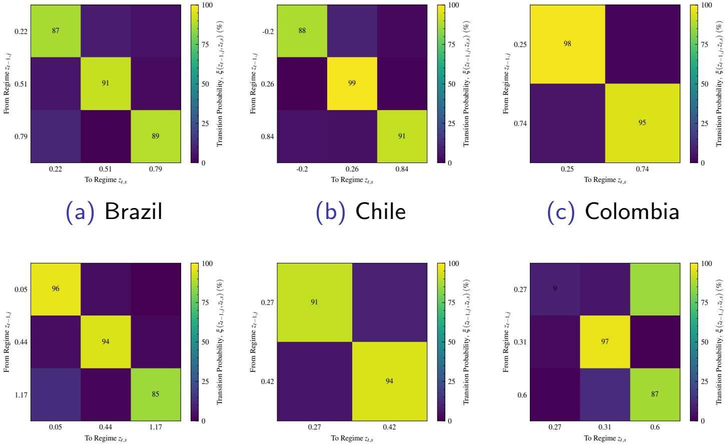

(a) Colombia




(b) Mexico

# Transition Probabilities

Figure 3: Transition Probabilities (%), Model 1




(a) Brazil (b) Chile (c) Colombia

(d) Costa Rica (e) Guatemala (f) Mexico

# Transition Probabilities

For some models, there could be hidden states with very similar dependent means. To better analyze this, there is the transition coefficient:

Table 3: Transition Coefficient (%)

| Cty. | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|---|---|---|---|---|---|---|
| BRA | 89.43 | 95.02 | 93.78 | 97.92 | 93.66 | 92.74 |
| CHL | 96.85 | 94.69 | 97.93 | 97.92 | 98.17 | 98.17 |
| COL | 97.23 | 79.0 | 97.42 | 98.45 | 95.45 | 94.62 |
| CRI | 90.03 | 97.38 | 96.87 | 97.94 | 97.24 | 97.24 |
| DOM | 92.57 | 90.16 | 96.41 | 97.17 | *nc* | - |
| GTM | 93.1 | 95.98 | 95.85 | 96.9 | 95.41 | - |
| HND | 89.84 | 90.38 | 96.89 | 95.87 | 97.25 | - |
| MEX | 92.91 | 90.91 | 94.81 | 95.84 | 93.59 | 94.52 |
| PER | 97.78 | 51.2 | 95.49 | 97.92 | 97.28 | *nc* |

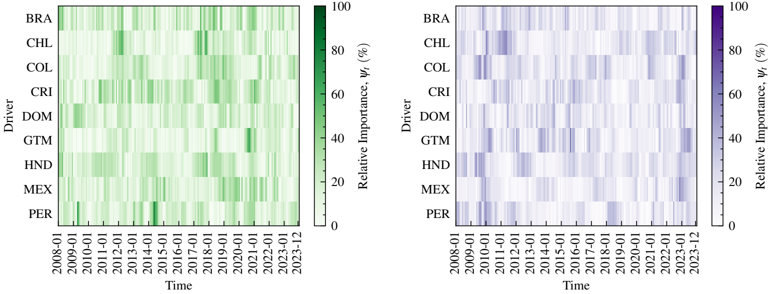
*nc* = the model does not converge.

# Dynamic and Patterns

- Monetary policy variables have a positive relationship with inflation, with a couple exceptions for the $mpr_t$ in Costa Rica (M5 and M6) and Honduras (M5)
- International inflation rates present a strong relation with domestic inflation. An exception is Brazil (M5 and M6)
- Exchange rates do not follow a certain pattern. The relation is positive in Chile, Costa Rica, the Dominican Republic, Guatemala, and Colombia
- Consumption factors present a strong positive pattern, except for Costa Rica with private expenditure (M5 and M6)
- Inflation expectations have a direct relationship with inflation, except in Costa Rica (M5 and M6)
- The relation with producer prices is positive for all countries

# Drivers

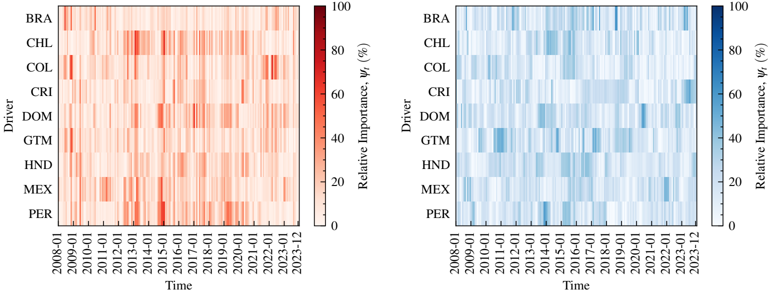
Figure 4: Monetary Policy Drivers, Model 3




(a) $broad_t$ (b) $mpr_t$

# Drivers

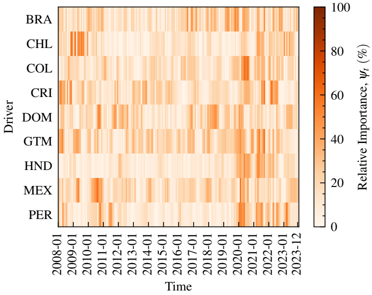
Figure 5: International Drivers, Model 3



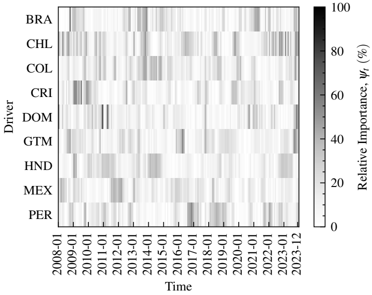

(a) $cpi_t^*$ (b) $er_t$

# Drivers

Figure 6: Demand-pull Drivers, Model 3



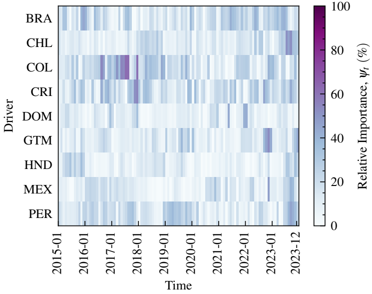

(a) $priv_t$



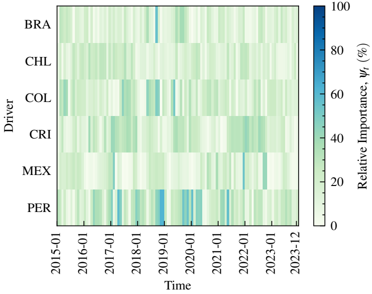

(b) $gov_t$

# Drivers

Figure 7: Expectations (M5) and Cost-push (M6) drivers




(a) $exp_t$




(b) $ppi_t$

# Content

# Conclusions

- Using a variational inference approach offers flexibility in model complexity selection compared to traditional criteria, and avoids overfitting the model
- Inflation regimes generally present high persistence across countries and models
- Monetary policy instruments play a significant role in driving inflation, with its impact peaking during periods of economic turbulence, like the financial crisis and the COVID-19 pandemic
- International inflation determinants strongly drive the inflation dynamics in Latin America
- Private expenditure stands out as the dominant force among demand-pull factors

# Content

# References I


Baum, L. E., & Petrie, T. (1966).Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, *37*(6), 1554–1563.

Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970).A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, *41*(1), 164–171.

Beal, M. (1998). *Variational Algorithms for Approximate Bayesian Inference* [Doctoral dissertation, University College London].

Briault, C. (1995).The Costs of Inflation. *Bank of England Quarterly Bulletin*, *35*(1), 33–45.

Gruhl, C., & Sick, B. (2016).Variational bayesian inference for hidden markov models with multivariate gaussian output distributions. *arXiv preprint arXiv:1605.08618*.

Jensen, J. L. W. V. (1906).Sur les fonctions convexes et les inegalites entre les valeurs moyennes. *Acta mathematica*, *30*(1), 175–193.

# References II


Kinlaw, W., Kritzman, M., Metcalfe, M., & Turkington, D. (2023).The determinants of inflation. *Journal of Investment Management*, *21*(3), 29–41.

Lucas, R. E. (1972).Expectations and the Neutrality of Money. *Journal of Economic Theory*, *4*(2), 103–124. https://doi.org/https://doi.org/10.1016/0022-0531(72)90142-1

MacKay, D. J. (1997). *Ensemble learning for hidden markov models* (tech. rep.). Citeseer.

McGrory, C. A., & Titterington, D. M. (2009).Variational bayesian analysis for hidden markov models. *Australian & New Zealand Journal of Statistics*, *51*(2), 227–244.

Mishkin, F. S. (2009). Will Monetary Policy Become More of a Science? In *Monetary policy over fifty years* (pp. 93–119). Routledge.

Rabiner, L. R. (1989).A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286.

Sánchez, G. (2023). Inflation Regimes in Latin America, 2020-2022: Persistence, Determinants, and Dynamics.

# Content

# Measures

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, an integrable real-valued random variable $X$, and a concave function $\varphi$. Jensen's Inequality states that

$$\varphi(\mathbb{E}[X]) \geq \mathbb{E}[\varphi(X)] \tag{16}$$

Consider two probability distributions $P$ and $Q$. The Kullback-Leibler divergence is defined by

$$KL(P||Q) = \int P(x) log \left( \frac{P(x)}{Q(x)} \right) dx \tag{17}$$

# Probability Distributions

**Gaussian-Wishart distribution** (conjugate prior for $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda})$)

$$p(\boldsymbol{\mu}, \boldsymbol{\Lambda} \mid \boldsymbol{\mu}_0, \beta, \boldsymbol{W}, \nu) = \mathcal{N}(\boldsymbol{\mu} \mid \boldsymbol{\mu}_0, (\beta\boldsymbol{\Lambda})^{-1})\mathcal{W}(\boldsymbol{\Lambda} \mid \boldsymbol{W}, \nu) \quad (18)$$

where

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Lambda}^{-1}|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Lambda}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

$$\mathcal{W}(\boldsymbol{\Lambda}|\boldsymbol{W}, \nu) = B(\boldsymbol{W}, \nu)|\boldsymbol{\Lambda}| \cdot \exp\left\{-\frac{1}{2}\mathrm{Tr}(\boldsymbol{W}^{-1}\boldsymbol{\Lambda})\right\}$$

and

$$B(\boldsymbol{W}, \nu) = |\boldsymbol{W}|^{-\nu/2} \cdot \left(2^{\nu D/2}\pi^{D(D-1)/4}\prod_{i=1}^{D}\Gamma\left(\frac{\nu+1-i}{2}\right)\right)^{-1}$$

# Probability Distributions

**Dirichlet**

$$\text{Dir}(\boldsymbol{\mu} \mid \boldsymbol{\alpha}) = C(\boldsymbol{\alpha}) \prod_{k=1}^{K} \mu_k^{\alpha_k - 1} \tag{19}$$

$$\sum_{k=1}^{K} \mu_k = 1 \tag{20}$$

$$0 \leq \mu_k \leq 1 \tag{21}$$

$$|\mu| = |\alpha| = K \tag{22}$$

# MGHMM

$$q(\mathbf{Z}) \propto \prod_{t=1}^{T}\prod_{j=1}^{J}(b_{t,j})^{z_{t,j}} \prod_{t=1}^{T}\prod_{j=1}^{J}\prod_{s=1}^{J}(a_{j,s})^{z_{t,j}, z_{t+1,s}} \tag{23}$$

$$a_{j,s} = \exp\{\mathbb{E}[\ln \tilde{\pi}_{j,s}]\} \tag{24}$$

$$b_{t,j} = \exp\{\mathbb{E}[\ln p(\boldsymbol{y}_n | \boldsymbol{\mu}_j, \boldsymbol{\Lambda}_j^{-1})]\} \tag{25}$$