torchonometrics mathematical notes

This document provides a detailed mathematical exposition of the estimators implemented in the `torchonometrics` library. The library focuses on GPU-accelerated econometric estimation using PyTorch, with particular emphasis on method of moments estimators.

1. Generalized Method of Moments (GMM)

1.1 Basic Framework

The Generalized Method of Moments (GMM) framework, developed by Hansen (1982), provides a unified approach to estimation and inference in econometric models. Consider a parameter vector $\theta \in \Theta \subset \mathbb{R}^p$ that we wish to estimate, and a vector of moment conditions:

$$E[g(Z_i, \theta_0)] = 0$$

where $g : \mathcal{Z} \times \Theta \to \mathbb{R}^q$ is a vector-valued function, Z_i represents the observed data, and θ_0 is the true parameter value.

1.2 GMM Estimator

Given a sample $\{Z_i\}_{i=1}^n$, the sample moment conditions are:

$$\bar{g}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} g(Z_i, \theta)$$

The GMM estimator is defined as:

$$\hat{\theta}_{GMM} = \arg\min_{\theta \in \Theta} \bar{g}_n(\theta)' W_n \bar{g}_n(\theta)$$

where W_n is a $q \times q$ positive definite weighting matrix.

1.3 Optimal Weighting Matrix

For efficiency, the optimal choice of weighting matrix is:

$$W_{opt} = \Omega^{-1}$$

where $\Omega = E[g(Z_i, \theta_0) g(Z_i, \theta_0)']$ is the variance-covariance matrix of the moment conditions.

In practice, we estimate Ω using:

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} g(Z_i, \hat{\theta}_{first})g(Z_i, \hat{\theta}_{first})'$$

where $\hat{\theta}_{first}$ is a first-stage consistent estimator obtained using the identity weighting matrix.

1.4 Two-Step GMM

The efficient two-step GMM procedure is:

1. **First Step**: Minimize $\bar{g}_n(\theta)' I_q \bar{g}_n(\theta)$ to obtain $\hat{\theta}_{first}$
2. **Second Step**: Compute $\hat{W} = \hat{\Omega}^{-1}$ and minimize $\bar{g}_n(\theta)'\hat{W}\bar{g}_n(\theta)$

1.5 Asymptotic Distribution

Under regularity conditions, the GMM estimator has the asymptotic distribution:

$$\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{d} N(0, (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1})$$

where $G = E[\nabla_\theta g(Z_i, \theta_0)]$ is the Jacobian matrix of moment conditions.
When $W = \Omega^{-1}$ (optimal weighting), this simplifies to:

$$\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{d} N(0, (G'\Omega^{-1}G)^{-1})$$

1.6 HAC-Robust Covariance

For time series data or spatial dependence, we use the Newey-West HAC estimator:

$$\hat{\Omega}_{HAC} = \hat{\Gamma}_0 + \sum_{j=1}^{L} w_j(\hat{\Gamma}_j + \hat{\Gamma}_j')$$

where: - $\hat{\Gamma}_j = \frac{1}{n}\sum_{t=j+1}^{n} g(Z_t, \hat{\theta})g(Z_{t-j}, \hat{\theta})'$ - $w_j = 1 - \frac{j}{L+1}$ is the Bartlett kernel weight - L is the lag truncation parameter, typically chosen as $L = \lfloor 4(n/100)^{2/9} \rfloor$

1.7 J-Test for Overidentifying Restrictions

When $q > p$ (overidentified), we can test the validity of the overidentifying restrictions using Hansen's J-test:

$$J = n\bar{g}_n(\hat{\theta})'\hat{W}\bar{g}_n(\hat{\theta}) \xrightarrow{d} \chi^2_{q-p}$$

1.8 Instrumental Variables as GMM

The canonical IV regression model:

$$y_i = x_i'\beta + \epsilon_i$$
$$E[\epsilon_i|z_i] = 0$$

corresponds to the moment condition:

$$g(z_i, y_i, x_i, \beta) = z_i(y_i - x_i'\beta)$$

2. Generalized Empirical Likelihood (GEL)

2.1 Motivation

GEL methods, introduced by Smith (1997) and Newey and Smith (2004), provide an alternative to GMM that avoids the choice of weighting matrix while maintaining efficiency properties.

2.2 Empirical Likelihood Problem

The empirical likelihood approach solves:

$$\max_{p_1,\dots,p_n} \sum_{i=1}^{n} \log p_i$$

subject to: - $\sum_{i=1}^{n} p_i = 1$ - $\sum_{i=1}^{n} p_i g(Z_i, \theta) = 0$ - $p_i \geq 0$ for all i

2.3 Generalized Empirical Likelihood

GEL extends this by replacing the log-likelihood with a general discrepancy function $\rho(\cdot)$:

$$\max_{p_1,\dots,p_n} \sum_{i=1}^{n} \rho(p_i)$$

subject to the same constraints.

2.4 Saddle Point Representation

The constrained optimization problem has the saddle point representation:

$$\hat{\theta}_{GEL} = \arg\max_{\theta} \min_{\lambda} \sum_{i=1}^{n} \rho(\lambda' g(Z_i, \theta))$$

where $\lambda \in \mathbb{R}^q$ is the vector of Lagrange multipliers.

2.5 Common GEL Estimators

The library implements three main GEL estimators:

2.5.1 Empirical Likelihood (EL)

$$\rho_{EL}(v) = \log(1 - v)$$

This requires $v < 1$ for all observations.

2.5.2 Exponential Tilting (ET)

$$\rho_{ET}(v) = 1 - e^v$$

This is numerically stable and recommended by Imbens, Spady, and Johnson (1998).

2.5.3 Continuously Updated Estimator (CUE)

$$\rho_{CUE}(v) = -\frac{1}{2}v^2 - v$$

This corresponds to continuously updated GMM.

2.6 GEL Algorithm

The estimation proceeds by:

1. **Outer Loop**: For each candidate θ, solve the inner minimization
2. **Inner Loop**: Find $\hat{\lambda}(\theta) = \arg\min_\lambda \sum_{i=1}^n \rho(\lambda' g(Z_i, \theta))$
3. **Optimization**: Find $\hat{\theta} = \arg\max_\theta \left(-\min_\lambda \sum_{i=1}^n \rho(\lambda' g(Z_i, \theta))\right)$

2.7 Asymptotic Distribution

Under regularity conditions:

$$\sqrt{n}(\hat{\theta}_{GEL} - \theta_0) \xrightarrow{d} N(0, (H_{\theta\lambda} H_{\lambda\lambda}^{-1} H_{\lambda\theta})^{-1})$$

where: - $H_{\lambda\lambda} = E[\rho''(\lambda_0' g(Z_i, \theta_0)) g(Z_i, \theta_0) g(Z_i, \theta_0)']$ - $H_{\theta\lambda} = E[\nabla_\theta g(Z_i, \theta_0)]$

2.8 Derivatives of Tilt Functions

Empirical Likelihood:

- $\rho'_{EL}(v) = -\frac{1}{1-v}$
- $\rho''_{EL}(v) = -\frac{1}{(1-v)^2}$

Exponential Tilting:

- $\rho'_{ET}(v) = -e^v$
- $\rho''_{ET}(v) = -e^v$

CUE:

- $\rho'_{CUE}(v) = -v - 1$
- $\rho''_{CUE}(v) = -1$

2.9 GEL J-Test

The GEL J-statistic for testing overidentifying restrictions is:

$$J_{GEL} = 2n \sum_{i=1}^{n} \rho(\hat{\lambda}' g(Z_i, \hat{\theta})) \xrightarrow{d} \chi^2_{q-p}$$

3. Linear Regression Models

3.1 Ordinary Least Squares

The OLS estimator minimizes:

$$\hat{\beta}_{OLS} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - x_i'\beta)^2$$

This corresponds to the moment condition:

$$g(x_i, y_i, \beta) = x_i(y_i - x_i'\beta)$$

3.2 Fixed Effects Regression

For panel data with individual effects α_i and time effects γ_t:

$$y_{it} = x_{it}'\beta + \alpha_i + \gamma_t + \epsilon_{it}$$

The within transformation eliminates fixed effects:

$$\ddot{y}_{it} = \ddot{x}_{it}'\beta + \ddot{\epsilon}_{it}$$

where $\ddot{z}_{it} = z_{it} - \bar{z}_{i\cdot} - \bar{z}_{\cdot t} + \bar{z}_{\cdot\cdot}$.

3.3 Heteroskedasticity-Robust Standard Errors

For heteroskedasticity-robust inference, we use:

$$\hat{V}_{robust} = (X'X)^{-1} X' \hat{\Omega} X (X'X)^{-1}$$

where $\hat{\Omega} = \text{diag}(\hat{\epsilon}_i^2)$ for HC0, or with finite-sample corrections for HC1, HC2, HC3.

4. Maximum Likelihood Estimation

4.1 General Framework

For a parametric model with density $f(z_i|\theta)$, the likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} f(z_i|\theta)$$

The MLE maximizes the log-likelihood:

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \sum_{i=1}^{n} \log f(z_i|\theta)$$

4.2 Logistic Regression

For binary outcomes $y_i \in \{0, 1\}$:

$$P(y_i = 1|x_i) = \frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)} = \Lambda(x_i'\beta)$$

The log-likelihood is:

$$\ell(\beta) = \sum_{i=1}^{n} [y_i x_i'\beta - \log(1 + \exp(x_i'\beta))]$$

4.3 Poisson Regression

For count data with $y_i|x_i \sim \text{Poisson}(\exp(x_i'\beta))$:

$$\ell(\beta) = \sum_{i=1}^{n} [y_i x_i'\beta - \exp(x_i'\beta) - \log(y_i!)]$$

4.4 Asymptotic Properties

Under regularity conditions:

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1})$$

where $I(\theta) = -E[\nabla^2 \log f(Z_i|\theta)]$ is the Fisher information matrix.

5. Computational Implementation

5.1 GPU Acceleration

The library leverages PyTorch's automatic differentiation and GPU acceleration for: - Matrix operations in moment condition evaluation - Gradient computation for optimization - Parallel processing of large datasets

5.2 Numerical Optimization

The library uses: - **L-BFGS-B** for constrained optimization - **LBFGS** for unconstrained problems (PyTorch backend) - **Scipy optimizers** for robust fallback options

5.3 Regularization Techniques

For numerical stability: - Eigenvalue regularization for ill-conditioned matrices - Pseudo-inverse for singular Jacobian matrices - Gradient clipping for unstable optimization paths

6. References

- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4), 1029-1054.
- Newey, W. K., & Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, 72(1), 219-255.
- Imbens, G. W., Spady, R. H., & Johnson, P. (1998). Information theoretic approaches to inference in moment condition models. *Econometrica*, 66(2), 333-357.
- Smith, R. J. (1997). Alternative semi-parametric likelihood approaches to generalised method of moments estimation. *Economic Journal*, 107(441), 503-519.
- Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703-708.