How to optimally macro hedge credit risk?

Johann Power
Faculty of Economics, University of Cambridge
April, 2025

Abstract

Banks inevitably face credit risk, the risk that a borrower does not pay back the lender. We specifically look at credit risk arising on derivatives trading desks. When you trade over-the-counter uncollateralised derivatives these trades have a continuous mark-to-market and at maturity one party will owe the other a payable sum which if unpaid will result in default. When entering multiple trades of different maturities, the bank will have a credit risk curve (credit risk for different tenors) with each counterparty. To manage this risk, banks buy credit default swaps (CDS) – a type of insurance which pays out to the purchaser when a specific entity defaults. Though, the question is, for holding a given level of credit risk, how much CDS should you buy? We will answer this by assuming we hedge only with a broad market CDS index like iTraxx EUR Main or colloquially 'Main' (like an equity index but for CDS) instead of buying specific single-name CDS (which can be illiquid or not exist). Our methodology utilises two machine learning techniques: weighted K-means clustering (KMC) and principal component analysis (PCA) performed on data gathered from Bloomberg. Weighted KMC is used to group counterparties of similar fundamentals like financial and geographical characteristics which we will then perform PCA on each bucket to calculate the key drivers of CDS curve movements in each group. This method not only improves on current basic models by capturing non-parallel credit curve movements, but also results in a hedging ratio reduction of approximately 20% hence reducing the bank's costs in paying CDS premiums.

1. Current Approach

Currently, a basic hedging approach would be to buy the amount of CDS whose change in value after purchasing for a 1 basis point (bp) increase ('spread DV01') would offset a 1bp increase in all the individual credit risks ('CS01') you are exposed to. Thus, with this level of CDS you could say that you are hedged against a 1bp increase in all credit exposures. However, this approach assumes a simplistic 1bp parallel increase along all tenors of the credit curve. This is unrealistic as different tenors move differently to one another from day-to-day e.g. short-term CDS tenors may move relatively less than longer-term CDS tenors. PCA is useful here as given a universe of companies, we can extract the key drivers (principal components) of CDS curve movements.

2. Data

We use historical daily business day CDS data for 138 companies between 01/01/2019-01/03/2025 from Bloomberg. The code also has data inputs for the spread DV01 of 2 possible hedging indices ('Main' & 'XOVER' – a high yield CDS index alternative), an example simulated credit risk dataset (cva_dv01_sensi.csv), example counterparty characteristics information (ricos_info.csv) and the current index cds levels (index_cds.csv). Model inputs are defined in the config.yaml file, which include parameters for performing the PCA and KMC as well as other general parameters like relative file paths.

3. Methodology

3.1. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a powerful dimensionality reduction technique that transforms a dataset into a new coordinate system where the directions of maximum variance become the principal components (Pearson, 1901). The use of PCA in financial data is most commonly applied in interest rate markets. Juneja shows how PCA can reduce US treasury yields (for various tenors) into just 3 components which together account for over 90% of the variation observed (Juneja, 2012). These components can be interpreted as level, slope and curvature movements in the interest rate term structure. This same methodology can be applied instead to CDS markets. We use 7 different CDS tenors: ['6m', '1y', '2y', '3y', '5y', '7y', '10y', '30y']. By applying PCA we can distil daily CDS rate movements to their principal components.

This PCA is performed by calculating the daily percentage changes in CDS rates, applying a standard scaler ($z_{ij} = \frac{x_{ij}-\mu_j}{\sigma_j}$) and computing the covariance matrix amongst features ($\mathbf{C} = \frac{1}{n-1}\tilde{X}^\top \tilde{X}$). Then perform an eigen decomposition on $\mathbf{C}$ by solving $\mathbf{C}\mathbf{v}_j = \lambda_j \mathbf{v}_j$ to obtain the eigenvalues λ_j and their corresponding eigenvectors $\mathbf{v}_j$, where each eigenvalue represents the variance explained by its eigenvector. The eigenvalues are then ordered in descending order, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$, and the top k eigenvectors (here $k = 2$) are selected to form the projection matrix $\mathbf{W} = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k\end{bmatrix}$. Finally, the original data is projected onto this new k-dimensional subspace by calculating $\mathbf{Z} = \tilde{X}\mathbf{W}$, and the proportion of variance explained by each principal component is given by $\frac{\lambda_j}{\sum_{i=1}^{p}\lambda_i}$, which provides insight into the importance of each component in capturing the variability of the data. The following code is used to implement this which takes as input the daily percentage changes in CDS curves for a given bucket:

```
def pca_per_bucket(df: pd.DataFrame):
    """
    Apply PCA to a bucket of CDS curves.
        Returns the transformed data,
    loading factors, and explained variance.
    """
    pca = PCA(n_components=number_components
        )

    # Standardise data
    scaler = StandardScaler()
    daily_rates_changes_unit = scaler.
        fit_transform(X=df.to_numpy())

    # Fit PCA & transform into new dataset
        with lower number of components (
        variables)
    pcs = pca.fit_transform(
        daily_rates_changes_unit)
    pca_df = pd.DataFrame(data=pcs, columns=
        pc_component_list)
    pca_coefs_dct = {}

    for v in range(0, len(pca.components_)):
       pca_coefs_dct[f'PC{v+1}'] = list(pca.
           components_[v])

    coefs = pd.DataFrame(pca_coefs_dct)
    coefs['Tenor'] = tenors

    # Want 1st PC to have a level increase
        interpretation
    for pc in list(coefs.columns)[:-1]:
       # Check tenor we are normalizing on e.
           g. 5y
       if coefs[coefs['Tenor']==
           normalize_tenor][pc].iloc[0] < 0:
          coefs[pc] = coefs[pc] * -1

    variation_explained = list(pca.
        explained_variance_ratio_)

    return pca_df, coefs,
        variation_explained
```

Listing 1. Python code used to apply PCA to CDS buckets

This process reduces the dimensionality of the dataset while retaining the most important variance. The loading factors in PCA are the coefficients that define the linear combination of the original variables for each principal component. The magnitude of a loading factor indicates the strength of the relationship between the original variable and the principal component. Over the full dataset we observe:

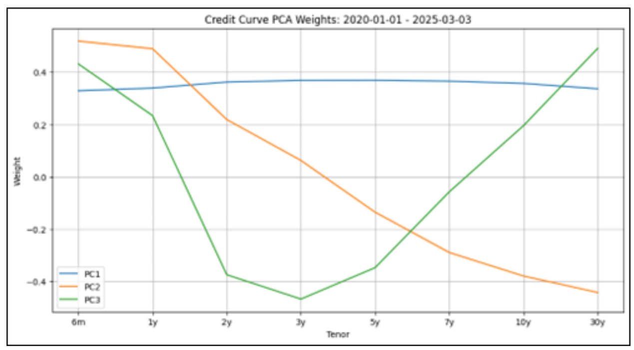

Figure 1. Loading factors for the first 3 principal components from PCA analysis on 138 companies (2020-2025).

Here we see the ‘level shift’ interpretation of the first component (PC1), the ‘slope’ effect of PC2 and the ‘curvature’ effect of PC3. PC1 can be interpreted as a broad market movement across all tenors whereas PC2 is a ‘compression trade’ with shorter term maturities moving in the opposite direction to longer term tenors. PC3 is interpreted as a butterfly effect with the ends of the curve moving in opposite directions to the middle tenors, however we don’t consider PC3 in this analysis as it only constitutes a small degree of explained variance and would otherwise result in unnecessary larger hedging implications (i.e. it’s ultimately not worth the cost). These principal component interpretations are similar to what you would expect to see in interest rate markets. PC1 and PC2 here account for over 95% of the variation observed as shown in the following elbow plot:

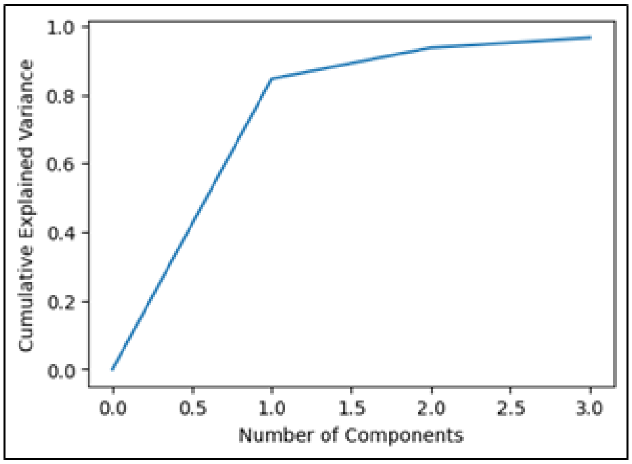

Figure 2. Percentage of the variation explained by the first 3 principal components.

With these PCA loading factors we can project credit risk form different tenors onto our key reference tenor of the 5Y point (this can be any tenor but it makes more sense to normalize on the most 'liquid' point). PCA will also be used for across bucket projection as discussed later.

3.2. Weighted K-Means Clustering (KMC)

KMC is used as we have economic intuition that the key component drivers may vary in intensity for structurally different types of companies and so we need to group similar companies into buckets. More precisely, weighted k-means clustering is used which is an extension of the standard k-means algorithm (MacQueen, 1967) whereby each data point x_i is assigned a non-negative weight w_i with the objective to partition the data into k clusters such that the weighted sum of squared distances, given by

$$J = \sum_{i=1}^{n} w_i \min_{1 \leq j \leq k} \|x_i - \mu_j\|^2,$$

is minimized. The algorithm begins with an initialization step, where initial centroids $\mu_1, \mu_2, \ldots, \mu_k$ are selected, here we use the quantile value of the reference tenor's (5Y) CDS level. In the assignment step, each point x_i is allocated to the cluster of the nearest centroid by computing

$$c_i = \arg \min_{j \in \{1,\ldots,k\}} \|x_i - \mu_j\|^2,$$

where the weights do not affect the nearest-neighbor determination. Each data point is assigned to the nearest centroid based on Euclidean distance. Categorical features use one-hot encoding to convert them into numerical vectors. Following this, the centroids are updated using the weighted mean of the points assigned to each cluster (with 5Y CDS values weighted 5x more than all other features), with the new centroid for cluster j calculated as

$$\mu_j = \frac{\sum_{i:c_i=j} w_i \, x_i}{\sum_{i:c_i=j} w_i}.$$

These assignment and update steps are iteratively repeated until convergence is achieved either when cluster assignments stabilize, the changes in centroids fall below a predefined threshold or a maximum number of iterations is reached. This weighted approach is particularly advantageous in situations where certain features should exert a greater influence on the final clustering outcome. When grouping companies into 7 buckets using weighted KMC (with features 5Y CDS values, country of operations and sector) we observe:

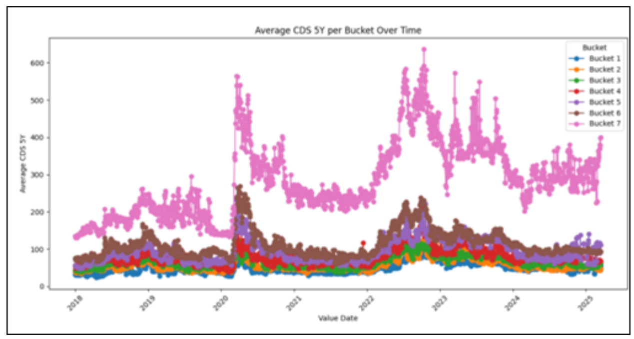

Figure 3. Clustering companies into 7 buckets weighted primarily by 5Y CDS value but also by country of operations and sector.

3.3. Optimal Hedging Solution

In each grouping, we perform PCA across tenors to project risk from non-5Y tenors onto the 5Y reference point. To do so, we first divide 1 by the 5Y CDS level to get the relative move of a 1bp shift in the 5Y. We then multiply this relative move by the difference in principal component loading factors between tenors. This gives the relative move for each tenor (which is why we perform PCA on percentage changes). Multiplying this by the CDS value at that tenor gives the absolute move. Thus, we get the absolute move across tenors for a 1bp increase in the 5Y. We do this for each principal component and calculate total PnL under each component by multiplying the absolute move by CVA sensitivity per tenor (from `cva_dv01_sensi.csv`), and then weight each PnL by the proportion of variance explained by that component.

We then hedge this total PnL move with an amount of Main CDS index. Before solving for the final quantity, we need to adjust the spread DV01 of the hedging index by the same PCA implied movement across tenors for a 1bp rise in 5Y (depending on which bucket Main is assigned to). If Main is in a different bucket (let's say bucket a) to the company we are analysing (which will often be the case with multiple buckets) then we need to project how a 5Y 1bp increase in bucket x impacts the 5Y in bucket a. To do this, we perform PCA across buckets and use the first principal component loading factors which is roughly equivalent to a linear regression across buckets. We then multiply the adjusted spread DV01 in matrix A by the absolute move of a principal component factor multiplied by $\frac{1}{\text{loading factor relative moves between bucket } x \text{ and } a}$, and finally, with as many hedging tenors as principal components we wish to hedge against (to ensure the matrix A is square and thus invertible) we solve the following system of simultaneous equations (for i principal components – the below example uses 3 factors):

$$\begin{aligned}
\text{SDV01}_A \cdot \Delta S_{A,\text{PC1}} \cdot Q_A + \text{SDV01}_B \cdot \Delta S_{B,\text{PC1}} \cdot Q_B + \text{SDV01}_C \cdot \Delta S_{C,\text{PC1}} \cdot Q_C &= \text{PnL}_{\text{PC1}} \\
\text{SDV01}_A \cdot \Delta S_{A,\text{PC2}} \cdot Q_A + \text{SDV01}_B \cdot \Delta S_{B,\text{PC2}} \cdot Q_B + \text{SDV01}_C \cdot \Delta S_{C,\text{PC2}} \cdot Q_C &= \text{PnL}_{\text{PC2}} \\
\text{SDV01}_A \cdot \Delta S_{A,\text{PC3}} \cdot Q_A + \text{SDV01}_B \cdot \Delta S_{B,\text{PC3}} \cdot Q_B + \text{SDV01}_C \cdot \Delta S_{C,\text{PC3}} \cdot Q_C &= \text{PnL}_{\text{PC3}}
\end{aligned}$$

expressed in matrix form as:

$$\underbrace{\begin{bmatrix} \text{SDV01}_A \cdot \Delta S_{A,\text{PC1}} & \text{SDV01}_B \cdot \Delta S_{B,\text{PC1}} & \text{SDV01}_C \cdot \Delta S_{C,\text{PC1}} \\ \text{SDV01}_A \cdot \Delta S_{A,\text{PC2}} & \text{SDV01}_B \cdot \Delta S_{B,\text{PC2}} & \text{SDV01}_C \cdot \Delta S_{C,\text{PC2}} \\ \text{SDV01}_A \cdot \Delta S_{A,\text{PC3}} & \text{SDV01}_B \cdot \Delta S_{B,\text{PC3}} & \text{SDV01}_C \cdot \Delta S_{C,\text{PC3}} \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} Q_A \\ Q_B \\ Q_C \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} \text{PnL}_{\text{PC1}} \\ \text{PnL}_{\text{PC2}} \\ \text{PnL}_{\text{PC3}} \end{bmatrix}}_{b}$$

where:

- $\text{PnL}_{\text{PC}i}$ = PnL under the i-th principal component (weighted by variance explained)
- Q_x = hedge quantity of CDS index at tenor x
- SDV01_x = adjusted spread DV01 of index at tenor x
- $\Delta S_{x,\text{PC}i}$ = change in CDS spread at tenor x under PC i

Put together, for a given set of credit exposures, each company will be clustered into a group, with the risk across tenors for that company calculated by the loading factors from the first 2 principal components for that group to get the relative movements across the credit curve normalizing on a 1bp increase in the 5Y tenor. After also adjusting the spread DV01 of the hedging index and projecting back onto the hedging index's bucket, we solve the matrix equation to get the optimal hedging solution. Summaing these solutions over all counterparties gives us our final hedging amount.

4. Results

This results in a hedge ratio of $\left(\frac{\text{total CDS hedge spread DV01}}{\text{total CS01 of all exposures}}\right) \approx 0.8$. This implies a 20% reduction in required hedging compared with the basic baseline hedging approach of assuming credit curve parallel shifts generating a hedge ratio of 1 (as you buy sufficient CDS spread DV01 to offset your CS01 risk). Hence, not only do we more accurately predict our credit risk, but it is also cheaper than the baseline approach, as we need to buy less CDS (and therefore have fewer premiums to pay).

5. Area of Improvement

An area which could be improved is in projecting risk from different buckets onto the hedging index's bucket to know how a 1bp increase in the 5Y from a given bucket correlates with the hedging index's bucket. Here the hedging index is assigned to a bucket using weighted KMC, like all other counterparties, and the first principal component from across bucket PCA is used to project risk onto the hedging index bucket. But we could also explicitly regress the movement in 5Y CDS levels in the hedging index bucket against the average movement in 5Y CDS for each bucket or explore other non-parametric relations.

6. Conclusion

Overall, by using weighted KMC to group companies and PCA to project risk across tenors and groups, this results in more accurate credit risk hedging and cost reductions and thus an improvement on current hedging methods.

References

Juneja, J. Common factors, principal components analysis, and the term structure of interest rates. *International Review of Financial Analysis*, 2012.

MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pp. 281–297. University of California Press, 1967.

Pearson, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11):559–572, 1901.